INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Volatility and Spillover Effects in Stock Markets - A Multivariate GARCH Approach

Bernardo Rodrigues Marques

Master in Finance

Supervisor:
PhD José Joaquim Dias Curto, Associate Professor with Aggregation, ISCTE-IUL

October, 2022

BUSINESS
SCHOOL

Department of Finance

Volatility and Spillover Effects in Stock Markets - A Multivariate GARCH Approach

Bernardo Rodrigues Marques

Master in Finance

Supervisor:
PhD José Joaquim Dias Curto, Associate Professor with Aggregation, ISCTE-IUL

October, 2022

Acknowledgements

The preparation of this dissertation was a long journey and the support of some people was essential to be successfully completed.

First, I want to thank my supervisor José Dias Curto for monitoring and driving me through the whole process. Your help and advice were very important, as was the knowledge you gave me. It was a pleasure to receive some of your expertise.

Last but not least, I want to thank my parents, Vagner and Elisabete, and my sister Catarina for the support they always give me.

Resumo

O objetivo desta dissertação é estudar o comportamento da volatilidade dos mercados financeiros. Através de análises univariadas e multivariadas, conseguimos encontrar os modelos que preveem de uma forma mais precisa a volatilidade futura dos índices de mercado, tendo em conta os efeitos de contágio existentes entre eles. Os dados utilizados neste estudo são os preços diários dos índices S&P 500, DAX e Nikkei 225, que replicam os mercados norte-americano, alemão e japonês, respetivamente. A parte prática deste estudo divide-se em duas fases. A primeira é uma análise univariada dos índices, na qual é feito o diagnóstico inicial aos dados, especificada a equação média condicional, testada a existência de heteroscedasticidade condicional no processo e estimados os modelos GARCH univariados. Na segunda parte, mais dedicada à análise conjunta dos índices, é aplicado o teste de causalidade emparelhada de Granger, estimados e aplicados os modelos VAR, estimados os modelos GARCH multivariados e tiradas as conclusões. Durante o processo, utilizam-se critérios de informação para tomar decisões. Os resultados indicam que a volatilidade dos índices tende a ser maior em períodos de crise económica e menor em períodos de estabilidade. Quando o objetivo é prever o comportamento da volatilidade destes índices de mercado de forma isolada, devemos estimar um modelo EGARCH(1,1). De entre os índices em análise, o S&P 500 e o DAX são os que apresentam maiores efeitos de contágio entre si. Se quisermos prever a volatilidade destes mercados considerando os efeitos de contágio, o modelo indicado a aplicar é o modelo DCC simétrico.

Classificação JEL:

C58, G15

Palavras-Chave:

Índices de Mercado, Volatilidade, Efeitos de Contágio, Modelos GARCH, GARCH Multivariado

Abstract

The goal of this dissertation is to study the behavior of financial market's volatility. Performing univariate and multivariate analyses, we were able to find the models that can more accurately predict future market indices' volatilities, considering the spillover effects between them. The data used in this study are the daily prices of S&P 500, DAX and Nikkei 225 indices, that replicate the North American, German and Japanese markets, respectively. The practical part can be divided into two phases. The first part is a univariate analysis of the indices, in which we make the initial diagnosis of the data, specify the conditional mean equation, test the existence of conditional heteroscedasticity in the process and estimate the univariate GARCH models. In a second part, more dedicated to the joint analysis of the indices, we start by applying the pairwise Granger causality test, estimate and diagnose the VAR models, estimate the multivariate GARCH models and take conclusions. Throughout the process, we use information criteria to make decisions. The results indicate that the volatility of the indices tends to be higher in periods of economic crisis and lower in periods of stability. When the goal is to predict the future behavior of the volatility of these market indices separately, we should use the EGARCH(1,1) model. Between the indices under analysis, S&P 500 and DAX are the pair with more presence of spillover effects between them. If we want to forecast market volatilities considering spillover effects, the most suitable model is the symmetric DCC model.

JEL Classification:

C58, G15

Keywords:

Market Indices, Volatility, Spillover Effects, GARCH Models, Multivariate GARCH

Table of Contents

List of Figures

List of Tables

Glossary

S&P 500 – Standard & Poor's 500

DAX – *Deutscher Aktienindex*

ACF – Autocorrelation Function

PACF – Partial Autocorrelation Function

ARMA – Autoregressive Moving Average

ARCH – Autoregressive Conditional Heteroscedasticity

GARCH – Generalized Autoregressive Conditional Heteroscedasticity

EGARCH – Exponential Generalized Autoregressive Conditional Heteroscedasticity

VAR – Vector Autoregressive

MGARCH – Multivariate Generalized Autoregressive Conditional Heteroscedasticity

DCC – Dynamic Conditional Correlation

A-DCC – Asymmetric Dynamic Conditional Correlation

BIC – Bayes Information Criteria

AIC – Akaike Information Criteria

1. Introduction

We live in a global world with respect to almost all branches of society, and the economy is not an exception. The global markets are real, since the biggest economies influence each other and work in harmony in most of the cases.

The market indices capture movements registered on stock exchanges, being a credible indicator for market analysis. As known, the shares of each market tend to follow group movements, appreciating when the majority of the stocks that are part of the index appreciate and losing value when they depreciate, always depending on how the index is composed. Therefore, in a global world, it is important to understand the repercussions that the movements registered in one market have on the others. Thus, would it not be useful to understand which models should be used to predict the impact that changes in the behavior of one market have on the others? How does market volatility behave? Do the events that occur in each market only affect that same market, or they have repercussions on different markets? Is the most correct model to fit the relationship between S&P 500 and DAX also the most correct model to fit the relationship between DAX and Nikkei 225? This dissertation aims to answer these questions, applying GARCH models to three of the most well-known market indices worldwide, from three different continents.

The data used in the practical application are the daily prices of S&P 500, DAX and Nikkei 225 indices, representing the North American, German and Japanese markets, respectively. It is analyzed a time horizon of 22 years, between January 2000 and December 2021. Considering a large time horizon, it is possible to understand the changes in the behavior of the markets during periods of crisis and recession and to compare them with periods of relative stability.

The empirical part of this dissertation starts by testing the stationarity of the series composed by the collected data. Then, and once the stationarity has been proven/reached, we specify the conditional mean equation using an ARMA process and test for the presence of conditional heteroscedasticity issues in the ARMA process. After being confirmed the conditional heteroscedasticity issues by the ARCH-LM test, it is necessary to estimate univariate ARCH models, namely the GARCH model, the GJR model and the EGARCH model. Once the univariate analysis is concluded, we proceed to the multivariate analysis. That starts with the application of Pairwise Granger Causality test to search for feedback-effects among the indices. Proving the existence of feedback-effects, VAR models are estimated and some diagnostic tests are applied to those models. Since heteroscedasticity problems were found in those diagnostic tests, we proceed to the estimation of multivariate GARCH models. Considering the data under analysis, we decided to estimate the standard DCC model and the A-DCC model. Some of the decisions taken throughout this study, namely the final decision about the ideal model for each pair of indices, are based on information criteria.

The application of the tests and formulation of the models led us to a considerable set of conclusions. Regarding the univariate analysis of each index, it was possible to conclude that the volatilities of the indices tend to be higher in periods of economic crisis and lower in periods of stability. Furthermore, the univariate GARCH model that best fits the returns of the indices under analysis is the EGARCH(1,1) model assuming a student's t for the conditional distribution of the errors. About the multivariate analysis, it was found that joint bad news does not have a greater impact in the correlation of indices than joint good news. Consequently, the information criteria indicate that the symmetric DCC model is the model that presents the most accurate results when we study the existence of spillover effects between the indices or when our goal is to forecast future evolutions. Through the analysis of covariances and correlations between the indices, it was also possible to verify that the spillover effects between S&P 500 and DAX are greater than the spillover effects between these two indices and Nikkei 225, and *vice versa*. These impacts, in some cases, tend to be exponential during periods of economic crisis and reduced during periods of stability. The results of this dissertation allow investors to make thoughtful decisions about the simultaneous maintenance of these assets in their portfolios, as they will perceive the interconnection between assets and their reaction to market events.

From here on, the dissertation is structured as follows: Chapter 2 includes a literature review, in which are mentioned the main authors and studies that were made on this topic or on topics directly related; Chapter 3 describes in detail the data studied throughout the dissertation; Chapter 4 illustrates the methodology used to achieve the desired results; Chapter 5 presents the entire process of testing and estimating models in an exhaustive and detailed way, so that the process can be clearly understood; Chapter 6 summarizes the findings of the empirical study.

2. Literature Review

Experts have long been concerned with understanding how financial markets work. That way, they will be able to try to predict future events as accurately as possible and extract gains from those situations. With the emergence of several innovations and new econometric models over the years, the accuracy of forecasts seems to be increasing. However, each situation is different from the other. In addition to the fact that future is unpredictable, the models that should be used vary according to the situation under analysis. Since their creation, market indices have aroused the interest of investors and, in most of the cases, are part of their portfolios. Thus, it is important for investors to apply forecasting models to these assets, as a way of predicting future events with the maximum possible accuracy. Furthermore, as important as understanding the behavior of each market separately, it is important to understand the relationship between them and the impact they have on each other. In order to answer to these needs, several models have been developed and published by several different authors over the years.

The Autoregressive Moving Average (ARMA) models were introduced by Box and Jenkins (1970). This class of models emerged from the need to predict the future behavior of markets. The ARMA models include two components: the Autoregressive (AR) part and the Moving Average (MA) part. The AR part refers to the relationship between current and past values, i.e., the autocorrelation of the observations. When a model requires the presence of autocorrelation in the observations, it assumes that past observations can be used to predict the future. On the other hand, the MA part represents a relationship between a current observation and a component of a past observation, where the component is the appropriately lagged random shock.

The ARMA models result from processes that assume the stationarity of the time-series. Consequently, they consider that volatility is constant over the time-period in analysis. However, when we analyze financial data, we verify that this assumption is not realistic. The volatility clustering concept, introduced by Mandelbrot (1963), tells us that large variations in prices tend to be followed by large variations in prices and small variations tend to be followed by small variations. Still, those variations can either be positive or negative. According to Schwert (2011), the stock volatility in North American, British and Japanese stock markets tends to be higher during financial crisis and recession periods, while it tends to be lower during periods of relative financial stability. Naturally, the periods of crisis vary depending on the market in analysis. Billio, Caporin and Gobbo (2006) also reached the conclusion that the correlation tends to be higher in periods of high volatility and the volatility has an impact on the persistence and magnitude of the autocorrelation. Thus, the identification of heteroscedasticity issues is extremely common in financial data, as the volatility of prices/returns tends not to be constant over a long time period.

Another factor that investors and researchers should consider when choosing the models to be adopted in their studies is the presence of asymmetries regarding the impact of market shocks on the volatility of returns. According to Black (1976) and Christie (1982), bad news tends to have more impact on volatility than good news, i.e., the impact of different shocks is asymmetric. There is a relationship between those asymmetries and the concept of leverage effect, which says that the weight of equity in a company's capital structure decreases if the weight of debt increases, assuming that there is no change in the remaining parameters. To consolidate the idea of asymmetry, Campbell and Hentschel (1992) concluded that news (regardless of being good or bad) increase the asset's required rate of return, due to the increase in expected volatility. This increase in the required rate of return leads to a decrease in the value of the asset, as expected. However, the impact of bad news on expected volatility tends to be magnified.

As a way of solving the non-constant volatility problem, Engle (1982a) introduced the Autoregressive Conditional Heteroscedasticity (ARCH) models. These are models that consider conditional volatility instead of constant volatility. The ARCH models were the first to contemplate this possibility and, for that reason, they are used worldwide. Bera and Higgins (1993) made a complete analysis of the ARCH models proposed by Engle, in which they discuss the possible interpretations of the model, talk about the changes made to the model during the decade following its appearance, showed how the model is implemented and presented its possible applications. In Bollerslev et al. (1992, 1994), main developments and upgrades made to the ARCH model proposed by Engle are analyzed, as well as its main characteristics and conditions. Applications of the model are also made using financial data. The authors suggest some features, such as the implementation of information transmission mechanisms or dynamic hedging strategies. Some authors who made complete and detailed analyzes of the ARCH-type models were Pagan (1996), Palm (1996) and Cox et al. (1996), among others.

Four years after the ARCH models were proposed, Bollerslev (1986) suggested the Generalized Autoregressive Conditional Heteroscedasticity (GARCH) models. As the name suggests, these are a generalization of traditional ARCH models. The GARCH models are simple models to be estimated, which makes them very famous and largely applied. Furthermore, these models consider that the main stylized facts of daily returns and the forecasts made with the models are usually identical to those made by models with a higher degree of complexity. One of the main disadvantages of GARCH model is the fact that it is a symmetrical model. The incapacity to capture asymmetries in the volatility of returns makes this a model that is often ineffective when working with financial data, as concluded by Hansen and Lunde (2005).

In response to the fact that GARCH models are not able to capture asymmetry in the volatility of returns, some authors have been presenting univariate asymmetric GARCH models that should be more effective when analyzing financial data. Some examples of these models are: the Exponential GARCH (EGARCH) model proposed by Nelson (1991); the GJR-GARCH model of Glosten et al. (1993); the Asymmetric Power ARCH (APARCH) model presented by Ding et al. (1993); and the Threshold ARCH (TARCH) model of Zakoian (1994).

In this study, we chose to apply the EGARCH model and the GJR model. The reason behind the decision of incorporating the GJR model in this study is the fact that it incorporates leverage effects in its regression, which is not the case in the original GARCH model. The EGARCH model will be applied to the data under analysis because it applies a logarithmic transformation to the conditional variances that eliminates the need to impose non-negativity restrictions. Another characteristic of the EGARCH model is the possibility of evaluating the persistence of shocks to variance.

It is important to analyze the indices individually, but it is also very important to understand the relationships they establish and the impacts they have on each other. When multivariate models are applied, the objective is to detect the existence of spillover effects, i.e., to understand whether seemingly independent events that occur in one market impact other markets in a positive or negative way. This is a topic that has been extensively addressed and studied in the literature. Dungey and Martin (2007) studied the transmission of volatility across the markets in order to detect the presence of spillovers and contagion in different equity and currency markets during the East Asian crisis. Five years later, Kenourgios and Padhi (2012) applied the Vector Error Correction Model (VECM) and the asymmetric DCC model to three emerging markets (Asian, Russian and Argentine) to study the financial contagion between the markets during the global financial crisis. Still about studies on emerging markets, Aloui et al. (2011) used the stock markets of BRICS and proved the evidence of dependence between the markets during the global financial crisis. To that propose, they implemented copula formulas and multivariate GARCH models with leptokurtic distributions. Their conclusions were confirmed by Dimitriou et al. (2013). The authors were able to reach the same conclusions using two different multivariate GARCH models: the Fractionally Integrated Asymmetric Power ARCH (FIAPARCH) model and the DCC model. Focusing now on studies applied to developed markets, we verify that Baele (2005) used data from US market and different European markets to find that shock spillover intensities increased at the end of the 20th century, as a consequence of the increase in trade integration, equity market development and low inflation values. Another topic that led to the analysis of spillover effects at the beginning of the 21st century was the introduction of the Euro as the currency of some European markets. Savva et al. (2009) and Bartram et al. (2007) concluded that the introduction of Euro increased the cross-market correlations between some European markets. The first article still found out that, among the different European markets, only UK and German markets were affected by the US market. Expanding the analysis worldwide, Singh et al. (2010) analyzed the price and volatility spillover effects between American, European and Asian stock markets. The conclusion points to the existence of regional influence between Asian and European stock markets.

To detect the presence of spillover effects, there is a wide variety of models that can be applied. If we want to study the existence of linear dependence among multiple variables, we can use the Vector Autoregressive (VAR) model proposed by Sims (1980). The VAR(p) model appeared as a special case of Vector Autoregressive Moving Average (VARMA(p, q)) model when $q = 0$. As for the case of ARMA-process, the VAR-process is also linear and stationary. The model incorporates feedback-effects, as the variables in analysis are allowed to affect each other. The feedback-effects can be found applying the Pairwise Granger Causality test (Granger, 1969).

However, as mentioned earlier, assuming that volatility is constant becomes unrealistic when we are working with financial data. In response to this problem, some authors have been creating and publishing several classes of multivariate GARCH models. The multivariate GARCH models can be constructed using three different approaches: (i) direct generalizations of the univariate GARCH model of Bollerslev (1986); (ii) linear combinations of univariate GARCH models; (iii) nonlinear combinations of univariate GARCH models. The third approach contains constant and dynamic conditional correlation models, the general dynamic covariance model and the copula-GARCH models. According to Bauwens et al. (2006), multivariate modeling frameworks lead to more relevant empirical results than working with separate univariate models.

The constant conditional correlation (CCC) model was developed by Bollerslev (1990) and became very popular since it was presented. Applying this model, we estimate a univariate GARCH for each asset separately and then build the correlation matrix. The time-varying conditional covariances are parametrized so that they are proportional to the product of corresponding conditional standard deviations, what makes the estimation of the models simpler and reduces the computational costs. However, since it assumes that conditional correlations are constant, the model can be unrealistic for many empirical applications.

Five years later, Engle and Kroner (1995) introduced the Baba, Engle, Kraft and Kroner (BEKK) model. It is a complex model, in the sense that it includes many parameters. Even if we are facing a 3-dimensional time-series, the volatility equation of a BEKK(1,1) has 24 different parameters. In practice, an unrestricted BEKK(1,1) model can only be applicable for small dimensional time series. Recently, Rast et al. (2020) presented the *positive diagonal* BEKK (pdBEKK) parametrization, that imposes a positivity constraint on the diagonal element of the parameters and simplifies the estimation of the parameters.

Over the years, it has been found that the use of constant conditional correlation models applied to financial data is unrealistic and the results of applying these models are not accurate. Thus, some recent authors have chosen to create models that assume a dynamic conditional correlation matrix.

As a solution for the practical unrealism of CCC model, Engle (2000) and Tse and Tsui (2002) proposed models with time-dependent conditional correlation matrices. Those models are called dynamic conditional correlation (DCC) models and they are parsimonious parametric models for the correlations. The DCC models “have the flexibility of univariate GARCH models coupled with parsimonious parametric models for the correlations” (Engle, 2000, p. 1). Following the same logic as for CCC models, for DCC models we estimate the univariate GARCH separately for each asset and then find the correlation matrix. The main difference is in the way the conditional correlation matrix is estimated. In the DCC model of Engle (2000), the number of parameters to be estimated is independent of the number of series to be correlated. The suitability of using dynamic models when working with financial data can be proven, for example, by the article published by Billio, Caporin and Gobbo (2006). According to the authors, the correlation tends to be higher in periods of high volatility and the volatility has an impact on the persistence and magnitude of the autocorrelation. In this sense, it can be risky to assume models that consider a constant conditional correlation matrix.

After several implementations of the DCC models, it was concluded that they did not consider the possibility of having asymmetric patterns in the cross-correlation. A study by Yu and Wu (2001) identified that the two main sources of asymmetric cross-correlation are the difference in the sensitivity of stock returns to economic factors and the difference in the quality of information that exists between small and large companies. For that reason, it was felt the need to create and present a new model that considers that asymmetry in cross-correlations.

In response to that need, Cappiello, Engle and Sheppard (2006) proposed an asymmetric dynamic conditional correlation (A-DCC) model. It appeared as an individual case of asymmetric generalized DCC (AG-DCC) model that replaces the original parameter matrices into scalars. The main differences between the A-DCC models and the original DDC models are the existence of asymmetric dynamics in the correlations and the asset specific correlation news impact curves.

The multivariate GARCH models are current and widely applied by those who pretend to study spillover effects among multiple variables, as we can see by looking at recent articles and publications in the financial markets' area. Since these models have been exposed to changes and upgrades over the years, it is important to make an appropriate choice of the models to implement, considering the data under analysis. The importance of the correct choice of these models was proven by Kartsonakis-Mademlis and Dritsakis (2020), in a study in which the performance of some asymmetric GARCH models when studying spillover effects between the global crude price and the stock market indices of G7 was examined. In this article, it was concluded that the choice of multivariate GARCH models is important, since the results obtained varied considerably depending on the estimated model. An advantage of these models is the possibility of being applied to the most diverse financial instruments. Hou et al. (2019) applied the DCC model with the objective of analyzing the existence of spillover effects between the Chinese fuel oil futures market and the Chinese stock index futures market. On the other hand, Sivrikaya et al. (2021) estimated multivariate GARCH models in order to relate the returns of American stocks and Bitcoin with the risk aversion of investors, having concluded, among other things, that those who invest in American stocks tend to be more risk averse than those who invest in Bitcoin. This year, Ampountolas (2022) published an article in which he analyzes intraday high-frequency cryptocurrencies volatility spillover effects among multiple cryptocurrency markets, through the estimation of DCC model.

Throughout this dissertation, the theories and models referred to in this literature review will be characterized and implemented. The application of these models and theories to real financial data from market indices will allow to draw conclusions about the volatility behavior of each index, as well as to identify the existence of spillover effects between the indices. Furthermore, it will be possible to understand which models should be used when the goal is to study the joint behavior of the indices under analysis.

3. Data

The data used in the empirical study is composed of the daily closing prices of three different stock indices: Standard & Poor's 500 (S&P 500); *Deutscher Aktienindex* (DAX); and Nikkei 225.

Starting with the S&P 500 index, it integrates 500 large companies traded in North American stock exchanges, chosen with respect to eight distinct criteria. Some examples of those criteria are the market capitalization of the asset, liquidity in the exchange markets, financial viability, among others. This is a capitalization-weighted index, since the weight that each asset has in the index depends on the total market value of outstanding shares.

Moving on to the DAX index, it consists in the 40 best performing companies in Frankfurt stock exchange, the biggest stock exchange in Germany and one of the most well-known in Europe. DAX is considered a total return index, since it assumes that cash distributions are reinvested and the movements in price of the components are tracked. Sometimes, the index may not perfectly reflect the actual behavior of German market, as the number of companies covered is small.

The Nikkei 225 includes 225 large publicly owned companies traded in Tokyo exchange market. The weights are defined based on the share price of each company, what makes this a price-weighted index.

The sample under analysis covers the period between January 2000 and December 2021, comprising a time horizon of 22 years. That way, we can include in our analysis periods of relative stability in the markets, as well as period of financial crisis, such as the Global Financial Crisis registered in 2007-2008 or the Coronavirus pandemic that began in 2020. It may be useful to understand the influence of those periods in our results and extract interest conclusions from there.

All the data was extracted from Yahoo Finance website and the prices are expressed in its original currencies. In the sample formulation, only the days in which there was information available for all indices under analysis were considered. We took that decision because it is necessary that individual samples of each index have the same size so that we can analyze them together and apply some multivariate models.

4. Methodology

In order to reach the conclusions that we want to obtain about the relationship between the indices under analysis, there is a whole set of steps and tests that need to be applied to our sample. These follow a logical sequence and some come as a consequence of the result of the previous ones.

4.1. Stationarity

In this empirical study, our sample is composed of financial data and we need to apply some econometric models. As several of those models imply that the time series are stationary, the first step of the study is to test the stationarity of the time series of prices. A time series is stationary if the historical observations can be generalized to the future. That generalization is possible if the mean and variance of the time series are constant and the covariances do not change over time.

4.1.1. Unit Root Tests

The visual interpretation of plotted observations allows us to have a general idea about stationarity, since we can see if mean and variance seem to be relatively constant. However, it is recommended the application of empirical tests to support our conclusions. Consequently, we decided to test the stationarity using the unit root tests. The unit root tests, as the name suggests, test the existence of unit roots in the univariate processes. If there is a unit root, it means that there is a stochastic trend and we assume that the time series has been generated through a "random walk" (RW) process. Thus, the time series is non-stationary.

The unit root tests we applied in this empirical study are the Augmented Dickey-Fuller (ADF) test, the Phillips and Perron (PP) test and Kwiatowski, Phillips, Schmidt and Shin (KPSS) test. In ADF and PP tests, the null hypothesis refers to the existence of a unit root, i.e. to the non-stationarity of the series. The main difference between those and the KPSS test is in the formulation of hypotheses. When we perform a KPSS stationarity test, the null hypothesis states that the time series is stationary.

4.1.2. Data Transformation

In most of the cases, especially when we are dealing with time series of prices, the stationarity is rejected because the mean and variance of prices are not usually constant over time. To try to solve this problem, the prices should be transformed into logarithmic returns. Firstly, we take the logarithms of the original time series to make the variance more constant. Then, we compute the first differences to make the mean more constant and remove the trend.

$$r_t = \ln(P_t) - \ln(P_{t-1}) = \ln(\frac{P_t}{P_{t-1}}) \quad (1)$$

4.2. Normality

The models applied throughout this study assume that the sample observations follow a certain empirical distribution. Often, normal distribution is assumed as the default distribution, but if this assumption is taken and the observations follow a different empirical distribution, the results may not be reliable.

The skewness and kurtosis of the distribution allow us to draw some conclusions, since there is a direct comparison of the empirical distribution with the normal distribution. When we are studying financial data, negative skewness and excess kurtosis are expected. Negative skewness exists when the left tail of the distribution is fatter than the right tail. About the kurtosis, we say that there is excess kurtosis when there are more extreme observations than in the normal distribution.

The Jarque-Bera test is a hypotheses test that assumes as null hypothesis the normality of the residuals. If the null hypothesis is rejected, we should not assume the normal distribution in the formulation of the models. The Jarque-Bera test results are reliable when we are dealing with large samples. The test statistic is computed using the skewness and kurtosis values.

$$JB = n * \left(\frac{(\hat{s})^2}{6} + \frac{\left(\hat{k} - 3\right)^2}{24} \right) \quad (2)$$

where n = sample size; $\hat{s}$ = skewness and $\hat{k}$ = kurtosis.

The test statistic follows a Qui-squared distribution with 2 degrees of freedom ($X^2_{(2)}$). For a given significance level (α), the normality is rejected if $JB > X^2_{(2)}(a)$.

4.3. Conditional Mean

The next step in our analysis is to use the past returns to model a conditional mean equation. To that propose, we will use an Autoregressive Moving Average (ARMA) process. The ARMA(p, q) model is represented by the following equation:

$$y_t = \mu + \varphi_1 y_{t-1} + \cdots + \varphi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \quad (3)$$

where:

- y_{t-i} is the return at time $t-i$.
- p denotes the number of autoregressive parameters incorporated in the model.
- q denotes the lag-order in moving-average terms.
- $\varphi_1, \dots, \varphi_p$ represent the relationship between y_t and $y_{t-1}, \dots, y_{t-p}$.
- $\theta_1, \dots, \theta_q$ represent the relationship between y_t and past random shocks.
- ε_t follows a white noise process.

In order to find the values for p and q terms to estimate the ARMA(p, q) model, we will look at the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) applied to the returns in analysis. Since we are working with returns, those should already be stationary and a pure autoregressive model should fit our returns better. Being the returns stationary, the ACF coefficients should decay exponentially and the number of autoregressive terms (p) to incorporate in our model may be suggested by the lag-order for which the autocorrelation and partial autocorrelation coefficients are higher.

After estimating the ARMA(p, q) model for each variable in analysis, we will apply the Ljung-Box test to the residuals. That way, we will be able to understand if the residuals of our model are correlated. The Ljung-Box test (Ljung & Box, 1978) is used to test the null hypothesis of no-autocorrelation.

$$\begin{cases} H_0: \rho_1 = \cdots = \rho_m = 0 \\ H_1: \exists \rho_J \neq 0 \end{cases}$$

If the series is stationary, we can admit that the residuals have been generated by a white noise process. When the null hypothesis is not rejected, it means that the time series was generated through a non-autocorrelated process and it confirms the adequacy of the estimated ARMA(p, q) model.

The test statistic of Ljung-Box test is computed according to the following expression:

$$\tilde{Q} = n(n+2) \sum_{k=1}^{m} \frac{\rho_k^2}{n-k} \tag{4}$$

where n is the number of observations, m represents the number of lags in null hypothesis and ρ_k denotes the autocorrelation at lag k.

The null hypothesis is rejected if $\tilde{Q} > X^2_{1-\alpha,m}$, assuming α as the significance level.

In order to proceed with the diagnosis of our ARMA models, we may still apply the ACF and PACF to the squared residuals of the model. That way, it is possible to check whether the correlation is high and, consequently, to find if there are still dependencies left in the data.

4.4. Conditional Heteroscedasticity

Before concluding the analysis of the ARMA models, it is necessary to understand the behavior of the errors' variance in the models. Thus, we will apply the ARCH-LM test to the residuals of the ARMA models. It tests the presence of autoregressive conditional heteroscedasticity (ARCH) effects by applying the Lagrange Multiplier (LM) test proposed by Engle (1982b).

In the ARCH-LM test, it is assumed that squared residuals are dependent according to:

$$e_t^2 = \hat{\alpha}_0 + \sum_{i=1}^{q} \hat{\alpha}_i \, e_{t-i}^2 + u_t \tag{5}$$

where q is representing the number of residuals included.

$$\begin{cases} H_0: \alpha_1 = \cdots = \alpha_q = 0 \\ H_1: \exists \alpha_j \neq 0, for\ j = 1,2, \dots, q \end{cases}$$

As we can see in the test hypotheses above, the null hypothesis of the test corresponds to a situation where there is conditional homoscedasticity in the residuals of the model. Oppositely, if the null hypothesis is rejected, it means that we have a situation in which there is conditional heteroscedasticity and, consequently, the presence of ARCH effects in the model is confirmed.

4.5. Univariate Models

Being the conditional heteroscedasticity presence confirmed in the data, it is necessary to proceed our study using Autoregressive Conditional Heteroscedasticity (ARCH) models. As the name suggests, those models are applicable to time series that present conditional heteroscedasticity issues.

4.5.1. Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model

The Generalized ARCH (GARCH) model, proposed by Bollerslev (1986), appeared as a generalization of traditional ARCH model. It has the same properties as the ARCH model without the problems related to the estimation of several parameters subject to the non-negativity restrictions (Engle & Bollerslev, 1986).

The conditional variance of a GARCH(p, q) model is estimated using the following regression:

$$y_t = x_t\beta + u_t \tag{6}$$

$$u_t = \epsilon_t\sigma_t \tag{7}$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i u_{t-i}^2 + \sum_{i=1}^{p} \delta_i \sigma_{t-i}^2 \tag{8}$$

where:

- y_t is the dependent variable.
- x_t is the vector of explanatory variables.
- β is a vector of unknown parameters.
- $E(\epsilon_t) = 0$ and $VAR(\epsilon_t) = 0$.
- $u_t|\Phi_{t-1} \sim N(0, \sigma_t^2)$ and $\Phi_{t-1} = \{u_{t-1}, u_{t-2}, \ldots\}$.

The restrictions $\alpha_0 > 0$, $\alpha_i \geq 0$ for $i = 1,2, \ldots, q$ and $\delta_j \geq 0$ for $j = 1,2, \ldots, p$ guarantee the non-negativity of conditional variance. The stationarity of the process is assured when the restriction $\sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{p} \delta_j < 1$ is respected.

4.5.2. Glosten, Jagannathan and Runkle (GJR) model

The GJR model was introduced by Glosten et al. (1993). According to the authors, the standard GARCH model has a weakness that must be amended, as it assumes that positive and negative shocks have the same impact on the volatility of returns. Thus, GJR model incorporates the leverage effect in its regression, as you can see next:

$$\alpha_i^* = \alpha_i(1 - \gamma_i)^2 \tag{9}$$

$$\gamma_i^* = 4\alpha_i\gamma_i \tag{10}$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i u_{t-i}^2 + \sum_{i=1}^{q} \gamma_i^* u_{t-i}^2 I_{t-i} + \sum_{i=1}^{p} \delta_i \sigma_{t-i}^2 \tag{11}$$

where:

- γ_i^* is the asymmetric coefficient in conditional variance.
- u_{t-i}^2 corresponds to the weight given to positive and negative shocks.
- I_t is a dummy variable that assumes the value 1 if $u_t < 0$ or the value 0 if $u_t \geq 0$.

4.5.3. Exponential Generalized Autoregressive Conditional Heteroscedasticity (EGARCH) model

The EGARCH model, introduced by Nelson (1991), is also an asymmetric volatility model. Similarly to the case of GJR, the EGARCH model incorporates an asymmetric coefficient that gives different weights for positive and negative shocks. The model applies a logarithmic transformation on conditional variances, which eliminates the need to impose restrictions to guarantee that $\sigma_t > 0$.

Comparing EGARCH model with GJR model, the difference is in the way the conditional variance is estimated:

$$\ln \sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \frac{|u_{t-i}|}{\sigma_{t-i}} + \sum_{i=1}^{q} \gamma_i \frac{u_{t-i}}{\sigma_{t-i}} + \sum_{i=1}^{p} \delta_i \ln \sigma_{t-1}^2 \tag{12}$$

The impact that positive and negative values of u_{t-i} have on volatility at time t is not always the same. Assuming that $\gamma_i < 0$, the variance increases when u_{t-i} is negative and decreases when u_{t-i} is positive.

4.6. Feedback-effects

If our goal is to analyze multiple variables simultaneously, as it is in our study, we should start by trying to understand if there are feedback-effects between the variables. Basically, we want to find the part of the actual returns of a variable that can be explained by the returns of a different variable. To that propose, we apply the Pairwise Granger Causality test, proposed by Granger (1969), in which we test the causality effect between 2 different stationary variables.

When applying the test, each regression will consider a lag-order equal to 5. We believe that it is the right choice to do, since 5 is the usual number of trading days per week. Thus, the regressions applied to assess the causality between variables X and Y are:

$$X_t = \alpha_0 + \alpha_1 X_{t-1} + \cdots + \alpha_5 X_{t-5} + \beta_1 Y_{t-1} + \cdots + \beta_5 Y_{t-5} + \varepsilon_t \tag{13}$$

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \cdots + \alpha_5 Y_{t-5} + \beta_1 X_{t-1} + \cdots + \beta_5 X_{t-5} + \varepsilon_t \tag{14}$$

where α_t and β_t are parameters to be estimated by the regression and ε_t is the error term.

Applying the Granger Causality test to the regressions, we will achieve a p-value associated to the test, which will allow us to decide whether to reject the null hypothesis or not. Our decision will depend on the significance level assumed in the analysis. Proceeding with the use of X and Y variables as example, the test hypotheses are:

$$\begin{cases} H_0: \beta_1 = \beta_2 = \cdots = \beta_5 = 0 \\ H_1: \exists \beta_j \neq 0, for\ j = 1, 2,3,4,5 \end{cases}$$

The null hypothesis states that Y does not granger causes X in the first regression and X does not granger causes Y in the second regression. Oppositely, if the null hypothesis is rejected, Y granger causes X in the first regression and X granger causes Y in the second regression.

4.7. Multivariate Models

As we saw earlier, the main objectives of this empirical study are the analysis of the relationship between different indices and the identification of the best models to carry out that analysis. Thus, now is the time to introduce the multivariate models, which will allow us to make a joint analysis and to identify interactions between variables.

4.7.1. Vector Autoregressive (VAR) models

The VAR(p) model appears as a special case of Vector Autoregressive Moving Average (VARMA) process with order p and q, where $q = 0$. The goal of VAR models is to study the existence of linear dependencies in different time series.

The expression of a k-dimensional VAR(p) model is:

$$y_t = \Phi_0 + \Phi_1 y_{t-1} + \cdots + \Phi_p y_{t-p} + \varepsilon_t \tag{15}$$

where:

- $y_t = (y_{1t}, y_{2t}, \ldots, y_{kt})'$ represents a vector of stationary variables.
- Φ_0 is a k-dimensional constant vector.
- Φ_1, Φ_2,…, Φ_p are the coefficients to be estimated.
- ε_t is a vector of innovations, with mean zero and a positive-definite time invariant covariance matrix.

The structure of the system integrates the feedback-effects that will possibly be identified in the Pairwise Granger Causality tests. However, since we are using returns instead of prices to guarantee that y_t is composed of stationary variables, we can capture only short-term dependencies among the variables.

After formulating the VAR(p) models, it is important to apply some diagnostic tests to our models. It may happen that those are not adequate models to fit our data. Firstly, we look at the modulus of the eigenvalues of the companion matrix so that we can prove the stability of the VAR-process. The normality of the process is tested using the Jarque-Bera test. The Portmanteau tests, both asymptotic and adjusted, analyze the autocorrelation in the residuals of the model, while the Breusch-Godfrey LM test is used to find autocorrelation problems in the VAR-process. Finally, the existence of conditional heteroscedasticity issues is tested through the ARCH-LM test applied to the residuals of the models.

4.7.2. Multivariate GARCH Models

The presence of conditional heteroscedasticity issues is very common when we are working with financial data, as we are in this study. Therefore, when we want to apply econometric models to our sample, we should use models that allow the errors' variance not to be constant. The Multivariate General Autoregressive Conditional Heteroscedasticity (MGARCH) models, introduced by Bauwens et al. (2006), are a good choice to deal with conditional heteroscedasticity. According to the authors, the most obvious application of MGARCH models is the study of the relations between volatilities and co-volatilities of several markets.

In this empirical study, only time-varying conditional correlation models are used. Models that assume constant conditional correlations will not be considered. The main difference between the time-varying conditional correlation models we use is in the way the correlation matrix is estimated.

4.7.2.1. Dynamic Conditional Correlation (DCC) model

The DCC model, presented by Engle (2000), appeared as a realistic alternative to the Constant Conditional Correlation (CCC) model (Bollerslev, 1990) that assumes that the conditional correlation matrix is time dependent.

To estimate the parameters of the model, we need to previously estimate a univariate GARCH model for each index. Using the residuals of univariate models, we will be able to estimate the time-varying conditional correlation matrix. Thus, a dynamic conditional correlation matrix is estimated as follows:

$$R_t = Q_t^{*-1} Q_t Q_t^{*-1} \tag{16}$$

$$Q_t = (1 - \alpha - \beta)\bar{Q} + \alpha \varepsilon_{t-1} {\varepsilon'}_{t-1} + \beta Q_{t-1} \tag{17}$$

where:

- Q_t^* is a diagonal matrix composed of the square root of the diagonal elements of Q_t.

- α and β are non-negative parameters.
- ε_t represents the errors standardized by their conditional standard deviation.
- $\bar{Q}$ is the covariance matrix of ε_t.

One of the strengths of estimating the DCC model is the possibility of estimating the associated covariance matrix:

$$H_t \equiv D_t R_t D_t \tag{18}$$

where D_t is the time-varying conditional standard deviation matrix that results from univariate GARCH models and R_t is the time-varying conditional correlation matrix.

The elements in the diagonal matrix D_t are estimated according to the expressions of chapter 4.5., depending on the univariate model that is being used. Additionally, in order to guarantee that H_t is positive definite, α and β are non-negative scalar parameters that satisfy the condition $\alpha^2 + \beta^2 < 1$.

4.7.2.2. Asymmetric Dynamic Conditional Correlation (A-DCC) model

The A-DCC model was proposed by Cappielo, Engle and Sheppard (2006). The model is similar to the symmetric DCC model proposed by Engle (2000), but there is a difference in the way of estimating the conditional correlations.

As in the DCC model, we start by estimating the univariate GARCH model and then we use its residuals to estimate the parameters of A-DCC model. The conditional correlation matrix is estimated considering asymmetry in the matrix:

$$Q_t = (\bar{Q} - \alpha^2\bar{Q} - \beta^2\bar{Q} - g^2\bar{N}) + \alpha^2\varepsilon_{t-1}\varepsilon'_{t-1} + g^2 n_{t-1}n'_{t-1}\beta^2 Q_{t-1} \tag{19}$$

$$n_t = I[\varepsilon_t < 0] \circ \varepsilon_t \tag{20}$$

$$\bar{N} = E[n_t n'_t] \tag{21}$$

For Q_t to be positive definite, it is sufficient that $(\bar{Q} - \alpha^2\bar{Q} - \beta^2\bar{Q} - g^2\bar{N})$ is also positive definite, being $\alpha^2 + \beta^2 + \delta g^2 < 1$, where δ is the maximum eigenvalue $[\bar{Q}^{-\frac{1}{2}}\bar{N}\bar{Q}^{-\frac{1}{2}}]$.

4.8. Information Criteria

During the entire empirical study, it is necessary to evaluate and compare the estimated models. In those cases, the decisions are based in one or more information criteria.

In this study, in most of the comparisons, only the Bayes Information Criteria (BIC) (Schwarz, 1978) is used, but sometimes it is complemented with the Akaike Information Criteria (AIC) (Akaike, 1973). In both cases, when contrasting 2 models, we should choose the one with the lowest value for BIC and/or AIC. According to Liew (2006), there is a direct relationship between the accuracy of each information criteria and the sample size. For samples up to 480 observations, AIC is more suitable than BIC. In turn, when the number of observations exceeds 480, BIC outperforms AIC. As our sample is much larger than 480, we will always rely more on the decision suggested by BIC, but sometimes complementing it with AIC. The information criteria can be estimated using the maximum value for the log-likelihood function:

$$BIC = k\ln(n) - 2\ln(\hat{L}) \tag{22}$$

$$AIC = 2k - 2\ln(\hat{L}) \tag{23}$$

where n represents the sample size, k measures the number of parameters of the model and $\hat{L}$ is the maximum value for the log-likelihood function.

5. Results

5.1. Visual Interpretation and Stationarity

The graphical representation of indices’ prices allows us to take some conclusions about the movements in the prices and to compare those movements with the ones registered in other indices within the same period.

Figure 1 – Daily Prices of S&P 500 (2000-2021)

Figure 2 – Daily Prices of DAX (2000-2021)

Figure 3 – Daily Prices of Nikkei 225 (2000-2021)

In the early 2000's, all the indices presented a decrease in its price due to the recession that was felt in the entire world during that period. From 2003, the indices started to recover from the recession and the graphics show us an upward trend that is common to all of them. Between 2007 and 2008, there was a global financial crisis that had a strong impact in the economies worldwide. That is why we can see a sudden decrease in the price of the indices during that period. From 2009, the indices' prices returned to the upward trend and remained in that line until 2020. The COVID-19 pandemic that stopped the world's economies in 2020 caused a fast and widespread decline in prices. From the last months of 2020, has been verified a gradual recovery in the markets.

Looking at the graphical representation of prices, we can anticipate that there is a stochastic trend. The existence of a stochastic trend points for the non-stationarity of the series. However, in order to draw more concrete and sustained conclusions about the stationarity of the series of prices, we performed unit root tests with an intercept and a trend in the test regression. Analyzing the results presented in table 1, we can confirm that the time series are non-stationary. The null hypothesis is not rejected for ADF and PP tests and rejected for KPSS test, which points for the non-stationarity of the series.

Table 1 – Unit Root Tests on Indices' Prices

	Critical Value (1% significance level)	Value of test-statistic		
		S&P500	DAX	Nikkei 225
ADF	-3.96	0.3338	-2.9042	-2.4302
PP	-3.97	0.2893	-2.7050	-2.2823
KPSS	0.119	8.6899	5.3289	6.6336

Due to this result, we transformed next the prices into returns. In most of the cases, it should solve our non-stationarity problem. The first step was to take the logarithms of the prices so that we can make the variance more constant. Then we computed the first differences of the logarithms to make the mean more constant and remove the trend. Figures 4 to 6 present the graphical representation of the daily returns for each index.

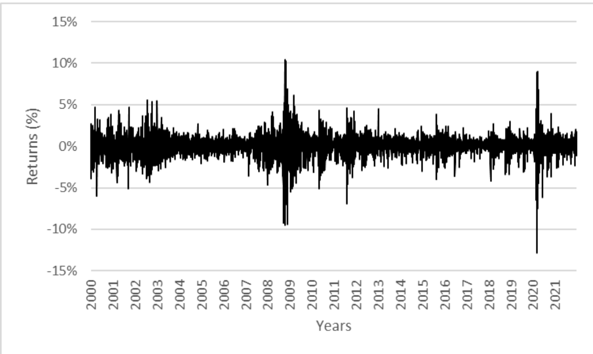


Figure 4 – Daily Returns of S&P 500 (2000-2021)

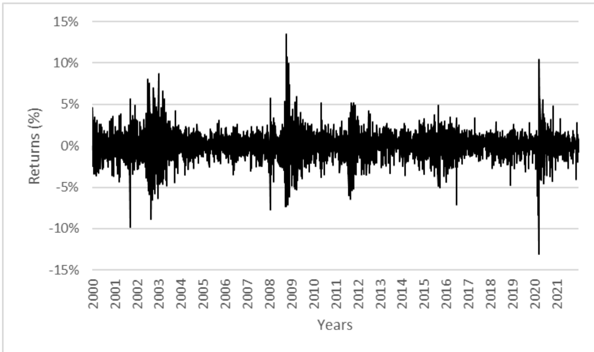


Figure 5 – Daily Returns of DAX (2000-2021)

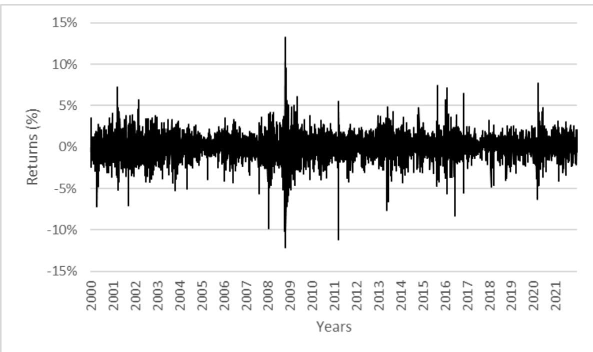


Figure 6 – Daily Returns of Nikkei 225 (2000-2021)

After the transformation, the mean and the variance became more constant and the stochastic trend is no longer present in the time series. Thus, the time series of returns seems to be stationary. Looking at the returns, we can see that volatility clustering is verified. In periods of financial crisis, there is greater volatility. In contrast, in periods of economic growth, returns are more constant and, consequently, volatility is lower.

5.2. Descriptive Statistics of Returns and Diagnostic Testing

Once our goal is to study the behavior of the returns, it is important to perform some preliminary computations and tests that allow us to have a global view about the returns. That will be the starting point for our analysis.

Table 2 – Descriptive Statistics of Returns

	S&P500	DAX	Nikkei 225
Mean	0.000232	0.000167	0.000081
Median	0.000685	0.000798	0.000393
Maximum	0.1042	0.1346	0.1323
Minimum	-0.1277	-0.1305	-0.1211
Std. Deviation	0.0128	0.0153	0.0151
Skewness	-0.4651	-0.1262	-0.3764
Kurtosis	10.5609	6.5736	6.2706

As we can see in table 2, the mean of the returns is positive but very close to zero for all indices. The positive mean shows that the returns tent to be positive within the period in analysis. Looking at each index separately, the S&P 500 was the one that presented the higher mean value of returns. Oppositely, Nikkei 225 registered the slower uptrend in its returns. The median value was also positive for the three indices under analysis, being DAX the one with the higher median value. The higher positive return and the most negative return were both registered in DAX index, which makes it the index with the higher returns' range value. The lower returns' range value was observed in S&P 500. The standard deviation of returns follows a similar pattern, being DAX the index in which the observations show, on average, the greater deviation from the mean. Contrarily, the S&P 500 index presented the lower standard deviation, which indicates that this index can be a good choice for risk-averse investors.

Changing the focus of our analysis slightly, we verify in table 2 that all indices have negative skewness estimates. Thus, we can conclude that the distributions of returns are asymmetric negative, i.e. there is a greater number of observations (or observations with higher value) located in the left tail of the distribution than in the right tail. The kurtosis estimate is also higher than 3 in all the cases, which means that the number of extreme observations registered in the empirical distribution is higher when compared to the normal distribution. Therefore, the distribution is leptokurtic and that is very common when we are dealing with financial data.

Table 3 – Unit Root Tests on Indices' Returns

	Critical Value (1% significance level)	**Value of test-statistic**		
		GSPC	**GDAXI**	**N225**
ADF	-3.96	-53.0831	-34.0611	-50.9910
PP	-3.97	-81.0802	-71.7732	-74.0784
KPSS	0.119	0.0323	0.0555	0.0544

The stationarity of the returns is a fundamental condition for the application of several financial models. Even though the graphical representation of figures 4 to 6 shows that the stationarity issues seem to have been solved by transforming the prices into returns, it is important to consolidate that idea with empirical tests. To that propose, we performed again the unit root tests that were previously applied to the prices, but this time to the returns. The values of the test statistic were calculated with an interception and a trend in the test regression. The results are presented in table 3 and reveal a contrary outcome to what happened with the time series of prices. The stationarity of the returns is proved in the three tests, for all indices, and that means we can proceed the analysis using the time series of returns.

Table 4 – Diagnostic tests to Indices' Returns

	p-value		
	GSPC	**GDAXI**	**N225**
Jarque-Bera	2,2E-16 **	2,2E-16 **	2,2E-16 **
Ljung-Box	0,6713	0,6713	0,6965
ARCH-LM	0 **	0 **	0 **

** denotes 1% significance level.

When we are estimating financial models, we need to assume distributions to represent the observations in analysis. The closer the chosen distribution approaches the empirical distribution of the observations, the more accurate will be the results. Some models assume the normal distribution by default but that may not be the right decision to make. In the null hypothesis of the Jarque-Bera test, it is assumed the normality of the returns. The results of the normality test can be seen in table 4. As the p-value of the test is lower than 0.01, the normality of the returns is rejected even considering a 1% significance level. This conclusion was already expected, since we had already seen that we are dealing with a leptokurtic empirical distribution. Consequently, from now on, our models will assume the student’s t as the distribution that better fits the returns.

The next step is to estimate the ARMA-type models that better describe our data. Those estimations should be done separately for each index.

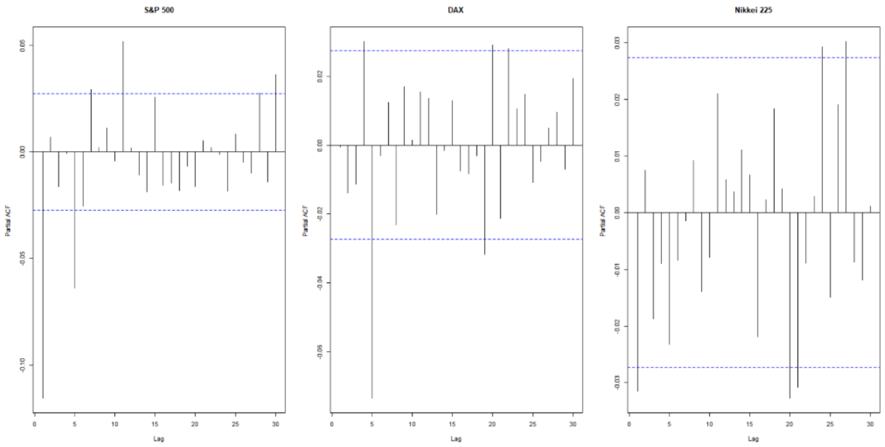


Figure 7 – PACF of S&P 500, DAX and Nikkei 225's returns

The figures above show us the partial autocorrelation functions (PACF) for each index. The partial autocorrelation coefficients are considered in the Y-axis and the lag operator in represented in X-axis. Looking at the PACF of S&P 500, we can see that there is partial autocorrelation of 5th order in the returns. The autocorrelation functions (ACF) are reported in Annex A and support the conclusions of the PACF analysis. Thus, an ARMA(5,0) model should be the right choice for S&P 500 returns. In order to verify the adequacy of the model, we applied the Ljung-Box test to the ARMA model until the 10th order serial correlation. The results are presented in table 4 and, once the null hypothesis is not rejected, we can confirm that an ARMA(5,0) model was able to capture the correlation structure observed in the S&P 500 returns. The same process was repeated for DAX and Nikkei 225. The PACF plots suggest an ARMA(5,0) model for DAX returns and an ARMA(1,0) model for Nikkei 225 returns. Looking at the Ljung-Box test's results, they also confirm the adequacy of the models to the data.

Since the visual interpretation of the returns showed signs of volatility clustering, it is important to apply the ACF to the squared residuals series. In figure 8, we can verify that the squared residuals are highly correlated and there are still some dependencies left in the data. This conclusion was expected because of the volatility clustering in returns mentioned before.

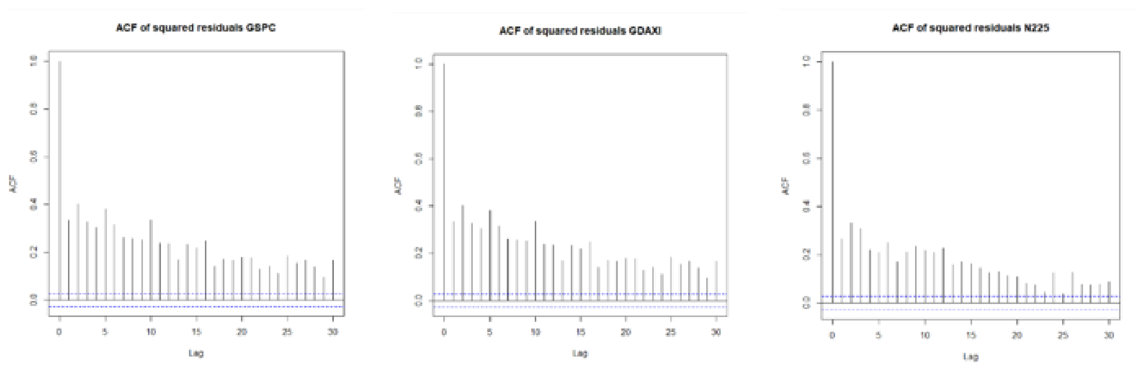


Figure 8 – ACF of squared residuals of estimated ARMA models

To conclude the analysis of ARMA models, we tested the presence of conditional heteroscedasticity issues in our sample using the ARCH LM test. The results presented in table 4 allow us to verify that the null hypothesis is rejected for all the cases. Thus, there are conditional heteroscedasticity issues in the data under analysis. In order to solve this problem, we had to estimate ARMA-GARCH models.

5.3. Univariate GARCH Models Estimation

We considered in our analysis three distinct univariate GARCH models, one of them being symmetric and the others asymmetric in terms of the weight given to positive and negative news. The standard GARCH model is the symmetric model. Oppositely, GJR and EGARCH were the asymmetric models applied in this empirical study. The estimation of the models was done by using functions from "rugarch" package available in R software. The estimated coefficients for the conditional mean equations and the GARCH models' coefficients can be found in table 5.

Table 5 – Estimates for conditional mean equation and univariate GARCH models' coefficients

	S&P 500			DAX			Nikkei 225		
	GARCH(1,1)	GJR(1,1)	EGARCH(1,1)	GARCH(1,1)	GJR(1,1)	EGARCH(1,1)	GARCH(1,1)	GJR(1,1)	EGARCH(1,1)
c	0.000564**	0.000335**	0.000324**	0.000587**	0.000294*	0.000253	0.000471**	0.000249	0.000185
θ_1	-0.078333**	-0.062096**	-0.060939**	-0.022584	-0.012251	-0.006082	-0.034708*	-0.022131	-0.023717
θ_2	-0.052979**	-0.035111*	-0.027019	-0.031532*	-0.017210	-0.017403	-	-	-
θ_3	-0.037750*	-0.014913	-0.009394	-0.027980*	-0.011432	-0.005810	-	-	-
θ_4	-0.038326*	-0.013434	-0.007868	-0.016293	-0.003829	0.003441	-	-	-
θ_5	-0.047594**	-0.027645*	-0.018283	-0.043124**	-0.031486*	-0.023544	-	-	-
ω	0.000001	0.000002	-0.214035**	0.000002	0.000002	-0.173544**	0.000003*	0.000005**	-0.294066**
α_1	0.113919	0.000000	-0.148299**	0.086697**	0.000000	-0.129098**	0.088505**	0.027616**	-0.105906**
β_1	0.881350**	0.890289**	0.977102**	0.906366**	0.910649**	0.980325**	0.897854**	0.880735**	0.966077**
γ_1	-	0.186705**	0.152928**	-	0.151965**	0.128323**	-	0.132446**	0.176338**

* and ** denote 5% and 1% significance level, respectively.

Looking at the outputs, we can observe that the estimates for conditional variance are, in general, statistically significant. That is a good indicator and supports our decision of estimating ARMA-GARCH models to deal with conditional heteroscedasticity problems.

Before analyzing table 5 in detail, it is important to understand the impact that different variables can have on the conditional variance. The variable α_1 refers to the impact that new information has in conditional variance, while β_1 measures the autoregressive persistence. Both α_1 and β_1 should assume positive values and respect the condition $\alpha_1 + \beta_1 < 1$. When the sum of the variables is close to 1, it means that the actual volatility is strongly impacted by the past volatility.

The standard GARCH estimates registered positive values for both α_1 and β_1, regardless of the index we are evaluating. However, we can confirm that there is a small positive impact of market news in the conditional variance but there is a strong persistence. Although in the case of S&P 500 the estimate of α_1 is not statistically significant, we noticed that in all indices the sum of α_1 and β_1 is close to 1. As mentioned before, it is a sign that past volatility strongly affects the actual volatility. These conclusions can be extended to GJR and EGARCH models, since the estimates for β_1 in both cases are statistically significant and assume high values. It suggests that, regardless the index and the GARCH-type model that we are applying, there is a strong persistence in the conditional variance. The estimates for α_1, which are in general statistically significant, assume small values but influence the sum of the variables to be close to 1.

The estimates for γ_1 in GJR model are positive and statistically significant for all the indices. The conclusion we can get from there is that bad news have a stronger impact in volatility than the good news and it probably makes more sense to use an asymmetric model. Looking at the EGARCH outputs, the γ_1 estimates (i.e. α_1 in the output) are negative and statistically significant. Consequently, as we already concluded in the previous model, it probably makes more sense to fit the data using asymmetric models. In order to consolidate this idea, we plotted the News Impact Curve (NIC) for each model applied to each index and it can be found in Annex B. Comparing the different graphical representations, we can conclude that the NIC for GJR and EGARCH suggest a leverage effect, confirming that negative shocks have a greater impact on volatility than positive shocks. Based on the information so far, we expected that asymmetric GARCH-type models describe better our data.

To decide about the best model for each index, we compared the Bayes Information Criteria (BIC) for all models.

Table 6 – Bayes Information Criteria Estimates for Univariate GARCH models

	S&P 500	DAX	Nikkei 225
GARCH(1,1)	-6,461802	-5,927400	-5,823703
GJR(1,1)	-6,492373	-5,956243	-5,838837
EGARCH(1,1)	-6,496457	-5,961320	-5,844328

Our expectations were confirmed by the results summarized in table 6. As predicted, the asymmetric models showed better results when compared to the symmetric model. However, analysing both asymmetric models, the EGARCH(1,1) model is the model that presents the lower BIC, regardless the market index under analysis. Before making a final decision, it is important to look at the ARCH-LM test and weighted Ljung-Box test on standard residuals of EGARCH(1,1) models. The output can be found in Annex C. The null hypothesis is not rejected for all the indices at the 1% significance level. That is a good indicator for the models. Taking into account all the information we have, we decided to proceed our analysis using the univariate EGARCH(1,1) to fit all the time series of returns.

5.4. Pairwise Granger Causality Tests

At this point, we were able to start studying the relationship between the indices. That study began with the results of the pairwise Granger causality test, whose goal is to analyze the existence of feedback-effects between the returns. However, before looking at the results of the test, it is important to mention that causation is not the same as correlation. The Granger causality test is applied to see how much of current value of a variable can be explained by past values of the other and then to see whether adding lagged values of the second can improve that explanation.

Table 7 – Pairwise Granger Causality Test

	p-value
DAX do not Granger-causes **S&P 500**	6,67E-11
S&P 500 do not Granger-causes **DAX**	2,2E-16
Nikkei 225 do not Granger-causes **S&P 500**	0,3857
S&P 500 do not Granger-causes **Nikkei 225**	2,2E-16
Nikkei 225 do not Granger-causes **DAX**	0,9095
DAX do not Granger-causes **Nikkei 225**	2,2E-16

The results of the test, using a lag order equal to 5, can be found in table 7. Starting by the relationship between S&P 500 and DAX, we can verify that p-value is lower than the significance level (5%) in both tests, which means that DAX Granger-causes S&P 500 and S&P 500 Granger-causes DAX. There is a feedback-effect among the variables. However, when we relate those two variables with Nikkei 225, the scenario is quite different. We could verify that S&P 500 and DAX Granger-cause Nikkei 225's returns, once the p-value of the tests is lower than the significance level. On the other hand, we see that Nikkei 225 do not Granger-causes S&P 500 nor DAX. In those cases, the null hypothesis is not rejected. Thus, the Granger causality runs one-way from S&P 500 to Nikkei 225 and from DAX to Nikkei 225.

Accounting the results of the test, it has been proved that there is a significant relationship between the returns of the indices. Adding that to the fact that we have already proved that the time series of returns are stationary, the next step to take is to estimate Vector Autoregressive (VAR) models and test if these are the most suitable models we can use to study and predict the joint behavior of the indices we are analyzing.

Before estimating the VAR models, we decided to test the cointegration of the variables. It is considered that the variables are cointegrated if the residuals that result from a regression formed by the variables are stationary. Thus, the Phillips-Ouliaris cointegration test was applied. The null hypothesis of the test considers the absence of unit roots in the residuals of the regression. As we can see in table 8, the p-value associated to the test is higher than the significance level of 5% and the null hypothesis of the test is not rejected for any pairs of indices. Consequently, variables are not cointegrated and VAR models can be estimated.

Table 8 – Phillips-Ouliaris Cointegration Test

	p-value		
	S&P 500 / DAX	S&P 500 / Nikkei 225	DAX / Nikkei 225
Phillips-Ouliaris Cointegration test	0,15	0,15	0,15

5.5. Vector Autoregressive (VAR) models

From now on, we are working with models that relate two variables at the same time, so we have to formulate new data series in order to be able to jointly analyze the returns of two different indices simultaneously. Then we used the "VARselect" function, that is available in the "vars" package of R software. The goal was to find which lag order should be used in the VAR models (individually for each pair of indices) according to different information criteria.

Table 9 – Results of "VARselect" Function Based on Information Criteria

	S&P 500 / DAX				**S&P 500 / Nikkei 225**				**DAX / Nikkei 225**			
	AIC(n)	**HQ(n)**	**SC(n)**	**FPE(n)**	**AIC(n)**	**HQ(n)**	**SC(n)**	**FPE(n)**	**AIC(n)**	**HQ(n)**	**SC(n)**	**FPE(n)**
Lag-order	9	5	2	9	5	2	2	5	5	5	1	5

Initially, we decided to consider the lag order proposed by the most conservative information criteria. However, after diagnosing the residuals, we concluded that, using those lag orders, the VAR models presented autocorrelation problems. That makes it not advisable to continue the empirical study assuming those lag-orders. The most common solution to this problem is to increase the lag-order. After some experiments, we reached the result that the minimal lag-orders that guarantee the elimination of those autocorrelation problems are the ones proposed by the less conservative AIC(n) and FPE(n) information criteria. In order to test the adequacy of the models, we applied a set of multivariate diagnostic tests. The different tests performed are presented below.

Table 10 – Diagnostic tests to VAR models

	p-value		
	S&P 500 / DAX **VAR(9)**	S&P 500 / Nikkei 225 **VAR(5)**	DAX / Nikkei 225 **VAR(5)**
Jarque-Bera	2,2E-16	2,2E-16	2,2E-16
Portmanteau - Asymptotic	0,01648	0,001555	0,5338
Portmanteau - Adjusted	0,01602	0,001492	0,5295
Breusch-Godfrey	0,06132	0,4354	0,6561
ARCH-LM	2,2E-16	2,2E-16	2,2E-16

Regardless the pair of indices under analysis, the modulus of the eigenvalues of the companion matrix for the proposed VAR model are lower than one, which guarantees the stability of the VAR(n)-process with a constant as deterministic regressor. The modulus of the eigenvalues of the companion matrix are reported in Annex D.

The normality of the VAR process has been tested using the Jarque-Bera normality test. The null hypothesis was rejected at any significance level, for every pair of indices, which means that VAR-process is not multivariate normal distributed. Consequently, the multivariate student's t distribution will be considered in the rest of this empirical study.

The Portmanteau tests aim to analyze the autocorrelation in the residuals. Looking at table 9, we can verify that, in two of the three estimated models, the null hypothesis is rejected for both asymptotic and adjusted tests at 5% significance level. It means that, in those cases, the residuals are autocorrelated and that is not a good sign. The exception is the VAR(5) model representing the relation between DAX and Nikkei 225. There, the Portmanteau tests do not reject the null hypothesis, which proves the no-autocorrelation in the residuals. That is an important indicator to the VAR(5) model. Although some models have verified autocorrelation issues in the residuals, that is not an impediment to proceeding with their use, nor will compromise the results that we may have from now on in this study.

The autocorrelation problems in the VAR-process, that have already been mentioned before, were checked using the Breusch-Godfrey LM test. As we can see in table 9, the null hypothesis is never rejected at 5% significance level. Thus, it has been proved that the autocorrelation problems in the VAR-process were solved with the increasing of the lag-order in the models.

To conclude the diagnosis, we applied the multivariate ARCH-LM test. The null hypothesis of the test has been rejected, regardless the pair of indices under analysis. Thus, there are conditional heteroscedasticity issues. This finding disables us to continue our analysis using Vector Autoregressive models. To solve the problem, we had to move on to the estimation of multivariate GARCH models.

5.6. Multivariate GARCH models – DCC and A-DCC

Different variations of multivariate GARCH models can be found in the literature. Although, there are bigger or smaller significant differences between them. One of the main features that allows us to differentiate them is the way the conditional correlation matrix is defined. In this empirical application, only the models that assume time-varying conditional correlations will be considered. The constant conditional correlation models have been disregarded by the existing literature over the last few years, especially because they are considered less realistic than time-varying conditional correlation models. Thus, the models that will be applied in this study are the traditional Dynamic Conditional Correlation (DCC) model and the Asymmetric Dynamic Conditional Correlation (A-DCC) model.

The conditional variance of the indices, when we are implementing multivariate GARCH models, is designed considering univariate GARCH specifications. As we have already verified, EGARCH(1,1) is the most suitable to model the returns of each index separately (according to BIC). Thus, we decided to use the univariate EGARCH(1,1) model in the estimation of time-varying variances, regardless the pair of indices under analysis. However, the mean equation underlying the model is different depending on the index that is being studied. The estimation of DCC and A-DCC parameters was made using the "rmgarch" package available in R software.

Table 11 – Multivariate GARCH models' Coefficients (S&P 500 & DAX) and Information Criteria

	DCC(1,1)	aDCC(1,1)
DCC(a)	0,028023**	0,028023**
DCC(b)	0,938742**	0,938743**
DCC(g)		0,000000
AIC	-12,935	-12,934
BIC	-12,899	-12,897

** denote 1% significance level.

The estimates for the parameters of DCC and A-DCC models applied to the returns of S&P 500 and DAX can be found in table 10. We can see that DCC(a) and DCC(b) estimates assume positive values and are statistically significant. This is a feature that helped us to confirm our decision of applying only time-varying conditional correlation models, since the non-negativity of these parameters shows that time-varying models are adequate to fit our sample. Similarly to what we saw in the univariate GARCH analysis, also in the multivariate GARCH models there is a variable that measures the impact of new information in the volatility (DCC(a)) and a variable that measures the persistency when a new market event occurs (DCC(b)). Looking at table 10, we can see that DCC(a) and DCC(b) estimates present almost equal values for both symmetric and asymmetric DCC models. It happens because the asymmetric parameter (DCC(g)) is null in A-DCC model. Thus, there is no asymmetric response in the correlations after joint bad news in the markets. Putting this in a simple way, the autocorrelation between S&P 500 and DAX returns will be similar after joint bad news or joint good news, assuming that both occur in the same magnitude. The DCC(a) estimates assume values close to zero, meaning that new information has a small impact on the volatility. Oppositely, the estimates for DCC(b) record values higher than 0.93, which shows that there is a high persistency in the volatility after a new market event.

In the bottom of table 10, we can find the AIC and BIC values. Those allow us to decide which multivariate GARCH model better fits the relationship between S&P 500 and DAX indices. Since in both cases the least value recorded was referring to symmetric DCC model, we can conclude that the symmetric DCC model is the right choice when we want to study the joint behavior of these two market indices.

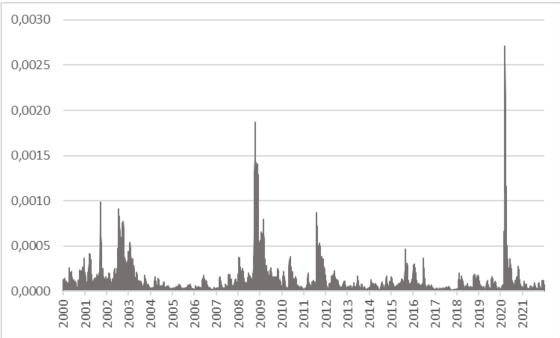


Figure 9 – Conditional Covariance Estimates - DCC model (S&P 500 & DAX)

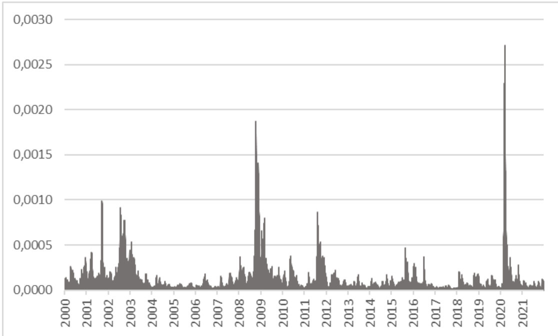


Figure 10 – Conditional Covariance Estimates - A-DCC model (S&P 500 & DAX)

Analyzing the conditional covariance estimates (for DCC and A-DCC) represented in figures 9 and 10, we may verify that in both models the covariances follow a similar pattern. The covariances tend to be very close to zero in periods of financial stability. In contrast, when the world is living periods of financial recession, the covariance of these two markets tends to be greater. This phenomenon can be evidenced by the fact that during periods of economic crisis (early 2000, 2008-2009 and 2020), have been recorded the highest covariance values between S&P 500 and DAX. That idea is verified regardless of the model we are evaluating. Since the models have different conditional correlation matrices, the absolute values are not the same for both models. However, once the asymmetric parameter (DCC(g)) is not statistically significant, the difference between the models should be almost null. For example, during the COVID-19 crisis (2020), the greater conditional covariance value registered was around 0.2713% (03/17/2020). If we look deep into the number, the difference between the models can be found if we round the number using 9 decimal places. That is a trend that persist over the entire time period.

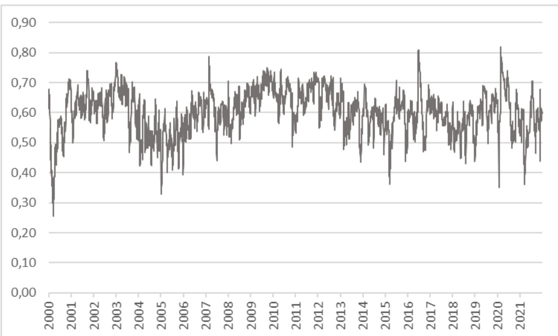


Figure 11 – Conditional Correlation Estimates - DCC model (S&P 500 & DAX)

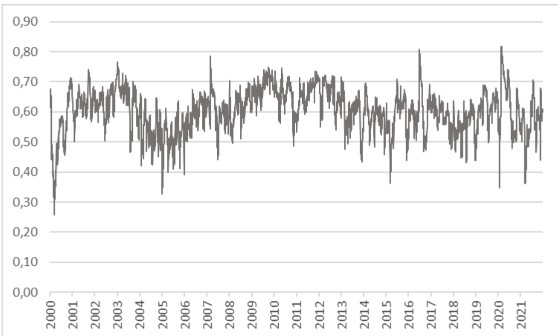


Figure 12 – Conditional Correlation Estimates - A-DCC model (S&P 500 & DAX)

As our analysis focuses on time-varying conditional correlation models, it is important to look at the conditional correlations estimated by both models. In figures 11 and 12, we can find the graphical representation of conditional correlations. Since the asymmetric parameter is not statistically different from zero, we expect the values to be almost equal for both cases. If we compare the conditional correlation value registered in 27th of March 2020, we need to leave 6 decimal places to visually see the difference between the estimates.

The correlation assumes positive estimates over the time-period in analysis, which means that both indices tend to move in the same direction: when one index records a positive return, the other index also tends to record a positive return. However, there were only 8 days, over the 22 years under analysis, in which the models estimated a strong correlation. Those were verified in 2 specific events: the BREXIT vote, in the end of June 2016; and in the adoption of measures by governments worldwide to control the growth in the number of COVID-19 cases, in the end of February 2020. Looking at the graphical representation, and based on what we saw before, we may conclude that the correlation is stronger in periods of instability in financial markets, while in periods of relative stability the correlation between the variables is positive but not strong. The lowest value recorded was about 0.2573 (03/13/2000), which makes the total range of estimated conditional correlation values over the 22 years to be approximately 0.5604.

Depending on the strategy that an investor wants to adopt, these can be assets that he wants to have simultaneously in his portfolio. Although the tendency is for index returns to move in the same direction, the correlation between them is not very strong and can allow investors to "play" with certain market scenarios. However, if the strategy is to diversify, the investor will have to combine these same market indices with other assets that behave in a different way.

Table 12 – Multivariate GARCH models' Coefficients (S&P 500 & Nikkei 225) and Information Criteria

	DCC(1,1)	**aDCC(1,1)**
DCC(a)	0,000830	0,000836
DCC(b)	0,999033**	0,999027**
DCC(g)		0,000001
AIC	-12,384	-12,383
BIC	-12,353	-12,351

** denote 1% significance level.

Moving on to the analysis of the relationship between S&P 500 and Nikkei 225's returns, we can find the estimates for DCC and A-DCC parameters in table 11. There, we can see that DCC(a) parameters are not statistically different from zero. In contrast, the DCC(b) parameters are statistically significant at 99% confidence level and their value is very close to 1. That means there is a very high persistence after the occurrence of a market event, regardless the model in analysis. Also, there is no asymmetric response in the correlations after bad news in the markets, once the asymmetric parameter (DCC(g)) is not statistically different from zero. Identically to what we saw in the joint analysis of S&P 500 and DAX, we expect the estimates for correlations and covariances to be very similar for both symmetric and asymmetric models.

The information criteria values may also be found in the bottom of table 11. Both AIC and BIC estimates assume a smaller value for DCC model, when compared with the A-DCC model. Thus, it led us to conclude that the symmetric DCC model is the right choice if we want to study the joint behavior of S&P 500 and Nikkei 225.

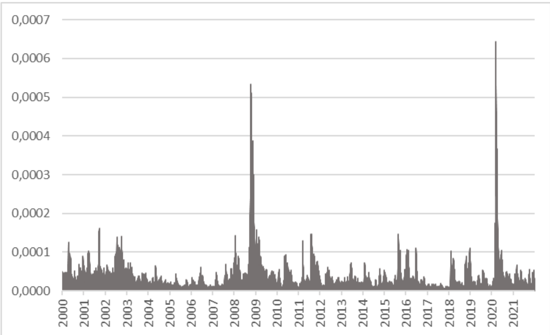


Figure 13 – Conditional Covariance Estimates - DCC model (S&P 500 & Nikkei 225)

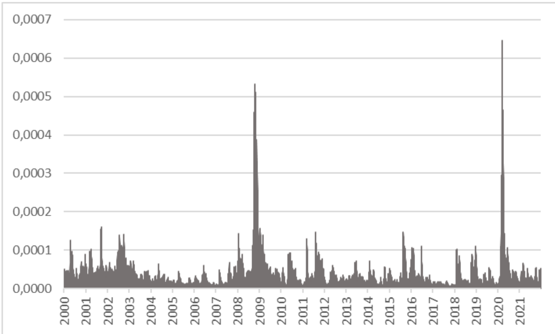


Figure 14 – Conditional Covariance Estimates - A-DCC model (S&P 500 & Nikkei 225)

Through the graphical representation of the conditional covariances (figures 13 and 14), we can confirm that both models assume very similar values for covariances. It happens because the asymmetric parameter is not statistically significant. Choosing a random day, for instance the 5th of October 2009, we need to round the covariance estimate using 8 decimal places to see that there is a slight difference in the value of estimates for both DCC and A-DCC models.

The peaks in the covariance values verified in this case are coincident with the ones that have been verified in the joint analysis of S&P 500 and DAX. The main difference is in the early 2000's, in which the values are not so high. The higher values where registered in 2009 and 2020, coincident with periods of economic recession in the worldwide economies. These results confirm that the covariances between the indices tend to be higher in periods of financial crisis and lower in periods of relative stability in the markets. Analyzing the absolute values, we see that the highest covariance value was recorded on the 17th of March 2020 (0.0644%). Looking at the period that cover the 2008-2009 financial crisis, the highest covariance value was recorded on the 21st of January 2009 (0.0157%). When compared with the covariances estimated in the joint analysis of S&P 500 and DAX, we can see that those covariance estimates are much bigger than the ones estimated in this case. The gap between the higher value recorded in each analysis is about 0.2069 percentage points, which is a notable difference.

Figure 15 – Conditional Correlation Estimates - DCC model (S&P 500 & Nikkei 225)

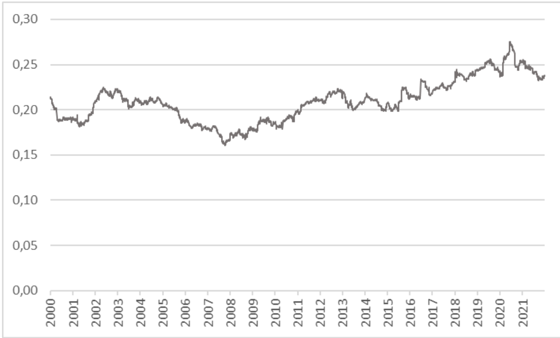


Figure 16 – Conditional Correlation Estimates - A-DCC model (S&P 500 & Nikkei 225)

The graphical representations of time-variant conditional correlations for both models show us some differences from what we concluded when we studied the relationship between S&P 500 and DAX indices. Paying attention to figures 15 and 16, we see that the correlations are constantly positive during the entire time period in analysis. It is a signal that the returns of indices tend to move in the same direction. When S&P 500 registers a positive return, DAX also tends to register a positive return. This trend had already been verified for the pair of indices analyzed previously, but here it is possible to verify that the correlation estimates assume much smaller values. This behavior drives us to believe that there is a stronger link between the European and North American markets' returns than between the North American and Japanese markets' returns.

Even though the correlation estimates do not register very dispersed values over time, it is possible to verify that the highest peak of correlation occurred during the year 2020, which coincides with the pandemic crisis that affected the entire world economies. More specifically, the highest estimate of symmetric DCC model (0.2749) was recorded on the 12th of June 2020. Comparing this value with the estimate of asymmetric DCC model for the same date (0.2755), we can observe that there is a small difference in the estimates. It is related to the fact that, although the asymmetric parameter is not statistically significant, it assumes a value slightly higher than the one observed in the joint analysis of S&P 500 and DAX. However, the difference in the correlation coefficients remains small for all days under analysis.

The smaller conditional correlation estimate for the symmetric DCC model (0.1609) was recorded on the 16th of October 2007, but the estimates for the following months assume also small values. Thus, during the financial crisis of 2008-2009, the correlation between the indices was small. That is a fact that leads us to conclude that, in the relationship between S&P 500 and Nikkei 225, periods of economic recession do not always mean periods of high correlation between indices' returns. Considering the maximum and minimum values recorded, the range is 0.1140. This is a much lower range than in the S&P 500 and DAX joint analysis.

From the investor's point of view, these two indices can simultaneously be part of the same investment portfolio if the objective of the investor is to diversify. Despite the returns of the two indices tend to move in the same direction, their magnitude is not the same (for the same period) and, therefore, they do not offer similar return to the investor.

Table 13 – Multivariate GARCH models' Coefficients (DAX & Nikkei 225) and Information Criteria

	DCC(1,1)	aDCC(1,1)
DCC(a)	0,006146*	0,006146*
DCC(b)	0,975466**	0,975465**
DCC(g)		0,000000
AIC	-11,933	-11,932
BIC	-11,902	-11,900

* and ** denote 5% and 1% significance level, respectively.

Turning now to the analysis of the relationship between DAX and Nikkei 225 indices, we can see that the results of DCC and A-DCC models' estimation have many points that coincide with those mentioned for the different sets of indices analyzed before. The estimates for DCC and A-DCC parameters are presented in table 12. The values of DCC(a) and DCC(b) are positive and statistically significant, which consolidates the idea that time-varying conditional correlation models are adequate to fit the data under analysis. As has been a trend throughout our empirical study, the DCC(a) estimates present a value close to zero, which means that the impact of new information on volatility is very small. Inversely, the value of the estimates for DCC(b) are much higher, reaching values close to 1, what indicates a high persistence in the volatility after a market event. About the asymmetric parameter (DCC(g)), it is one more time not statistically different from zero. Bad news in the market do not result in asymmetric responses in the correlations. Positive and negative shocks, when occur in the same magnitude, should have the same level of impact on the correlation of returns.

As we can see in the bottom of table 12, the information criteria indicate that the symmetric DCC is the model that better fits our data. Both AIC and BIC present lower information criteria values for the DCC model, just like in all previous situations.

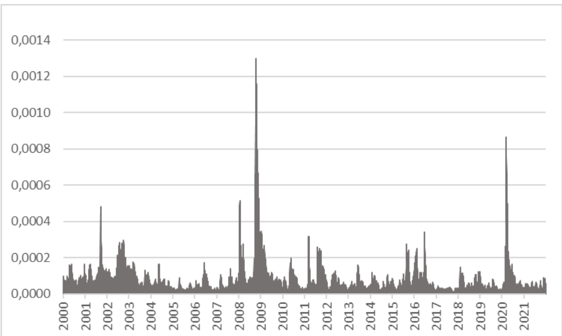


Figure 17 – Conditional Covariance Estimates - DCC model (DAX & Nikkei 225)

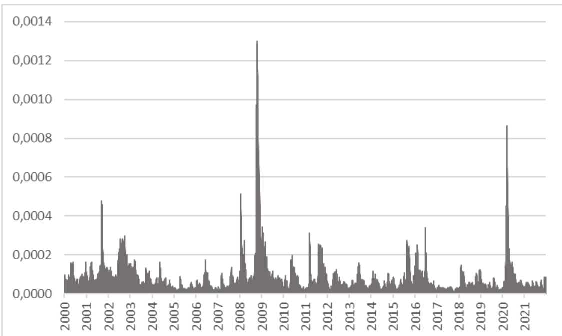


Figure 18 – Conditional Covariance Estimates - A-DCC model (DAX & Nikkei 225)

The conditional covariances presented in figures 17 and 18 show us that, as expected, the covariance estimates are very similar for both DCC and A-DCC models. Once again, it is related to the fact that the asymmetric parameter is not statistically different from zero. For instance, in the last day of our sample (12/30/2021), we need to round the estimates using, at least, 11 decimal places to find a difference between the values.

Looking at the graphical representations, we can see that the higher conditional covariances were registered in periods of financial crisis, namely the financial crisis of 2008-2009 and the financial crisis that resulted from COVID-19 pandemic. However, there is a curious fact that shows a difference in this relationship when compared to those that we analyzed before. Contrary to what happened in the previous cases, the greater covariance value occurred during the 2008-2009 financial crisis, where was recorded a maximum covariance value of 0.1297%. Comparing it with the maximum covariance value recorded during the COVID-19 pandemic (0.0865%), we can achieve that the difference is about 0.0432 percentage points. Regardless that difference, we can conclude that the covariance between DAX and Nikkei 225's returns tends to be higher in periods of economic recession and lower in periods of relative stability. If we look at the average values of covariance, we can see that they are lower than those recorded in the analysis of the first pair of indices (S&P 500 and DAX) but higher than those recorded in the analysis of the second pair (S&P 500 and Nikkei 225).

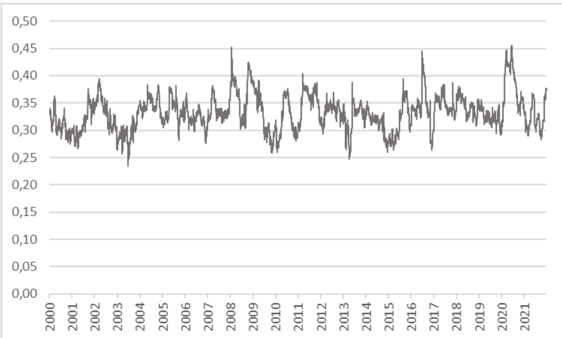


Figure 19 – Conditional Correlation Estimates - DCC model (DAX & Nikkei 225)

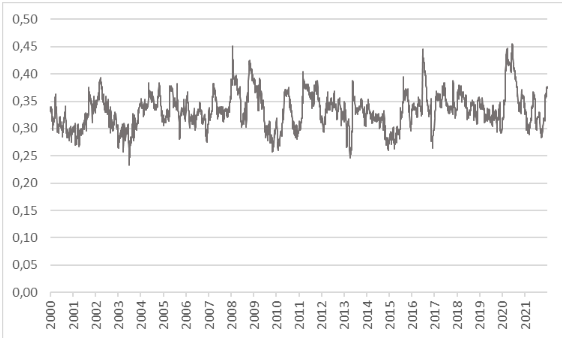


Figure 20 – Conditional Correlation Estimates - A-DCC model (DAX & Nikkei 225)

Turning our attentions to the conditional correlations (figures 19 and 20), we may verify that the estimates for both DCC and A-DCC models also present very similar estimates. In this case, the difference in the estimate for the last day in analysis is only visible if we round the numbers using at least 7 decimal places. As mentioned before, it is related to the fact that the asymmetric parameter is not statistically different from zero.

The conditional correlation estimates are positive all over the time period in analysis, meaning that the returns of both indices tend to move in the same direction. If DAX index presents a positive return, on a given day, Nikkei 225 also tends to present a positive return in the same day. The average of the conditional correlation estimates is about 0.3326, which places this pair of indices in an intermediate point between the first pair of indices (S&P 500 and DAX) and the second pair of indices (S&P 500 and Nikkei 225). Those registered an average of 0.6015 and 0.2089, respectively.

The highest correlation periods coincide with some events that have already been mentioned in this study, namely the financial crisis of 2008-2009, BREXIT vote and Coronavirus pandemic. The absolute highest value was estimated for the 17th of June 2020, during COVID-19 pandemic, where the conditional correlation was 0.4551. During 2008-2009 financial crisis and BREXIT vote period, the greater correlation estimates were 0.3922 and 0.4455, respectively. Oppositely, during the entire period in analysis, the smaller correlation value recorded was 0.2343 (07/02/2003). Thus, looking at the graphical representations, we can confirm that the correlation between DAX and Nikkei 225 indices' returns is higher in periods of financial crisis and smaller during periods of financial stability. The total maximum-minimum gap in the correlations was 0.2208, that is again in the middle between the gap values of both pair of indices mentioned before.

From the perspective of an investor that wants to diversify his portfolio, it can be useful to have assets from both indices simultaneously. Although it is expected that the returns will present the same sign on the same day, the magnitude of gains/losses is not expected to be the same, since the correlation values are never close to 1. However, they will have to be combined with other assets, as they alone do not guarantee total portfolio diversification.

6. Conclusions

For investors in financial markets, it is very important to understand the functioning of the markets so that it is possible to predict as accurately as possible their future movements and, from there, maximize gains and minimize losses. With this empirical study, our objective was to draw conclusions about the behavior of some market indices, combining the individual analysis of each index with a joint analysis of their behavior.

The main answers that we expected to obtain were related to the volatility of the markets and the impact that the variations caused by events that occur in each market have on the other markets under analysis, considering the same time period. Based on this, we expected to reach conclusions about the models that best predict the behavior, isolated and together, of the markets based on their historical data.

To carry out this study, it was considered data from three well-known market indices: S&P 500, DAX and Nikkei 225, representing the North American, German and Japanese financial markets. Daily prices recorded over 22 years were considered, between January 2000 and December 2021, to cover not only periods of relative financial stability, but also periods of major economic and financial crises.

When we looked at the returns of each index, we found that there were signs of volatility clustering. In addition, it was possible to visually verify that volatility tends to be higher in times of crisis and lower in periods of relative economic stability, which is in line with what was concluded by Billio, Caporin and Gobbo (2006). Through the estimation of univariate GARCH models, we also managed to conclude that the historical volatility values affect very significantly the current volatilities, regardless the index under analysis. This finding is important when the goal is to forecast because it enables the use of past information to predict future events.

When our objective is to use univariate models to predict future behaviors of index volatilities, it is advisable the application of EGARCH(1,1) model. In the formulation of the model, it should be considered a student's t distribution to fit the returns. Based on some information criteria, the EGARCH(1,1) is the univariate model that produces the most accurate results, regardless of the market index we are considering (within the three indices under analysis).

After drawing conclusions about the behavior of each index, we focused our attention on the spillover effects and the joint behaviors. To that propose, time series of returns were created for each pair of market indices. We started by trying to use VAR models, but we quickly concluded that they were not appropriate because they presented conditional heteroscedasticity problems that could compromise the results of the estimations and, consequently, could lead to biased conclusions.

Through the implementation of multivariate time-varying conditional correlation models, namely DCC and A-DCC models, we were able to reach the conclusions we wanted about spillover effects. The first finding was that joint bad news have no more impact in the correlation of indices than joint good news. After estimating the models, we found that the asymmetric parameter of A-DCC model was not statistically significant for any pair of indices under analysis, which allowed us to conclude that possibly a symmetric multivariate model would be the most indicated to fit our sample. In addition, the estimated parameters for the models have allowed us to discover that new events in the markets have a reduced impact on volatility. On the other hand, the persistence of the impacts is very high. This was a cross-cutting conclusion to all pairs of indices under analysis, although it was even more expressive when we applied the models to the time-series composed of S&P 500 and Nikkei 225's daily returns. In that case, the parameter that measures the immediate impact of news on volatility is not statistically significant, which explains the increase in the persistence parameter.

Regarding the conditional covariance values estimated by each model, for each time-series, it was possible to assess that the covariance values tend to be very low (almost nil) in periods of economic stability and substantially higher in periods of economic recession. Among the conditional covariance values recorded by the two models, there is no significant difference. That happens because the asymmetric parameter of A-DCC model is not statistically significant. Comparing the values recorded for each pair of indices, we found that, on average, the models applied to the time-series composed of S&P 500 and DAX's returns registered the higher covariance values, while the models applied to the S&P 500 and Nikkei 255 were the ones with the lowest values.

Focusing our attention on conditional correlations, it is possible to find some differences depending on the indices that are being considered. Starting with the application of the models to the returns of S&P 500 and DAX, we found that the correlation values are positive throughout the entire period, which indicates that the returns of the indices tend to move in the same direction. The correlation of indices is not very strong, with a tendency to be higher during periods of financial crisis and lower in periods of stability in the markets. Moving on to the values resulting from the application of the models to the S&P 500 and DAX indices, we reached that they are equally positive over the 22 years and, consequently, the tendency of the returns of both indices is to have the same signal on the same day. The correlation is equally positive and weak, but the big difference is that, in this case, it is not possible to draw conclusions about the relationship of crisis/stability periods with the conditional correlation values. Although the correlation values increase in some time periods that correspond to relevant market events, it cannot be considered a trend. In general, the correlation is low all over the period under analysis and the range of the values is reduced, which indicates that there are no large correlation peaks. Paying attention to the models applied to the DAX and Nikkei 225 indices, we verified that, once again, the correlation estimates are positive and the correlation is weak throughout the time period. In this case, it was possible to verify that the correlation is higher in periods of economic crisis and lower in periods of relative financial stability. When we compared the average values recorded for each multivariate time-series, we found that the average conditional correlation is higher in the models applied to S&P 500 and DAX indices, followed by DAX and Nikkei 225, and finally, the average is lower when applied to the S&P 500 and Nikkei 225. However, the difference between the first and the second is substantially greater than the difference between the second and the third placed.

According to Bayes Information Criteria and Akaike Information Criteria, the multivariate model that better fits the returns of the indices and, consequently, is the one that produces the most accurate results, is the symmetric DCC model. If an investor or researcher wants to understand the spillover effects between the market indices under analysis in this empirical study and predict future impacts, the symmetric DCC model is the right choice, according to BIC and AIC.

This empirical study can be used by investors who want to diversify their portfolio. From here they will be able to draw conclusions about the spillover effects between market indices that replicate the stock performance of the main companies from completely opposite geographies and check if it is a good idea to have these assets simultaneously in their portfolio. The study allows us to have not only a univariate and multivariate idea of the behavior of market volatilities, but also can be used as a basis for someone who wants to make his own tests and draw his own conclusions.

A limitation of this dissertation is linked to the fact that the sample under analysis is composed of daily data from three market indices from opposite points of the globe, so the trading days of the markets do not match. Due to this situation, only the days for which there is information available for the three indices were accounted. The ideal scenario would be to consider all the daily data available for each market, but this would make it impossible to apply the multivariate models.

The fact that the time horizon under analysis is large has allowed periods of crisis and relative stability to be covered and enabled the choice of the ideal model considering these two market situations. However, the model that better fits the returns of the indices may be different in times of crisis and relative stability. We recommend and encourage future researchers to study both periods separately, so that the comparison of models for the two distinct periods can be made, as well as the comparison of those results with the ones that were achieved in this dissertation.

Bibliographic References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *2nd International Symposium on Information Theory* (pp. 267–281). Akadémiai Kiadó.

Aloui, R., Aïssa, M. S. B., & Nguyen, D. K. (2011). Global financial crisis, extreme interdependences, and contagion effects: The role of economic structure? *Journal of Banking & Finance*, *35*(1), 130–141.

Ampountolas, A. (2022). Cryptocurrencies Intraday High-Frequency Volatility Spillover Effects Using Univariate and Multivariate GARCH Models. *International Journal of Financial Studies*, *10*(3), Art. 3.

Baele, L. (2005). Volatility Spillover Effects in European Equity Markets. *Journal of Financial and Quantitative Analysis*, *40*(2), 373–401.

Bartram, S. M., Taylor, S. J., & Wang, Y.-H. (2007). The Euro and European financial market dependence. *Journal of Banking & Finance*, *31*(5), 1461–1481.

Bauwens, L., Laurent, S., & Rombouts, J. V. K. (2006). Multivariate GARCH models: A survey. *Journal of Applied Econometrics*, *21*(1), 79–109.

Bera, A. K., & Higgins, M. L. (1993). Arch Models: Properties, Estimation and Testing. *Journal of Economic Surveys*, *7*(4), 305–366.

Billio, M., Caporin, M., & Gobbo, M. (2006). Flexible Dynamic Conditional Correlation Multivariate GARCH Models for Asset Allocation. *Applied Financial Economics Letters*, *2*, 123–130.

Black, F. (1976). Studies of stock price volatility changes. In *Proceedings of the 1976 Meeting of the Business and Economic Statistics Section* (pp. 177–181). American Statistical Association.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327.

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.

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), 5–59.

Bollerslev, T., Engle, R. F., & Nelson, D. B. (1994). Chapter 49 Arch models. In *Handbook of Econometrics* (Vol. 4, pp. 2959–3038). Elsevier.

Box, G. E. P., & Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control*. Holden-Day.

Campbell, J. Y., & Hentschel, L. (1992). No news is good news: An asymmetric model of changing volatility in stock returns. *Journal of Financial Economics*, *31*(3), 281–318.

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.

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.

Cox, D. R., Hinkley, D. V., & Barndorff-Nielsen, O. E. (1996). *Time Series Models: In econometrics, finance and other fields*. CRC Press.

Dimitriou, D., Kenourgios, D., & Simos, T. (2013). Global financial crisis and emerging stock market contagion: A multivariate FIAPARCH–DCC approach. *International Review of Financial Analysis*, *30*, 46–56.

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.

Dungey, M., & Martin, V. L. (2007). Unravelling financial market linkages during crises. *Journal of Applied Econometrics*, *22*(1), 89–119.

Engle, R. F. (1982a). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, *50*(4), 987–1007.

Engle, R. F. (1982b). A general approach to lagrange multiplier model diagnostics. *Journal of Econometrics*, *20*(1), 83–104.

Engle, R. F. (2000). Dynamic Conditional Correlation—A Simple Class of Multivariate GARCH Models. *SSRN Electronic Journal*.

Engle, R. F., & Bollerslev, T. (1986). Modelling the persistence of conditional variances. *Econometric Reviews*, *5*(1), 1–50.

Engle, R. F., & Kroner, K. F. (1995). Multivariate Simultaneous Generalized ARCH. *Econometric Theory*, *11*(1), 122–150.

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.

Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, *37*(3), 424–438.

Hansen, P. R., & Lunde, A. (2005). A forecast comparison of volatility models: Does anything beat a GARCH(1,1)? *Journal of Applied Econometrics*, *20*(7), 873–889.

Hou, Y., Li, S., & Wen, F. (2019). Time-varying volatility spillover between Chinese fuel oil and stock index futures markets based on a DCC-GARCH model with a semi-nonparametric approach. *Energy Economics*, *83*, 119–143.

Kartsonakis-Mademlis, D., & Dritsakis, N. (2020). Does the Choice of the Multivariate GARCH Model on Volatility Spillovers Matter? Evidence From Oil Prices and Stock Markets in G7 Countries. *International Journal of Energy Economics and Policy*, *10*(5), 164–182.

Kenourgios, D., & Padhi, P. (2012). Emerging markets and financial crises: Regional, global or isolated shocks? *Journal of Multinational Financial Management*, *22*(1), 24–38.

Liew, V. K.-S. (2006). *Which Lag Length Selection Criteria Should We Employ?* (SSRN Scholarly Paper N.º 885505).

Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, *65*(2), 297–303.

Mandelbrot, B. (1963). *The Variation of Certain Speculative Prices*. *36*(4), 29.

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

Pagan, A. (1996). The econometrics of financial markets. *Journal of Empirical Finance*, *3*(1), 15–102.

Palm, F. C. (1996). 7 GARCH models of volatility. In *Handbook of Statistics* (Vol. 14, pp. 209–240). Elsevier.

Rast, P., Martin, S. R., Liu, S., & Williams, D. R. (2020). A new frontier for studying within-person variability: Bayesian multivariate generalized autoregressive conditional heteroskedasticity models. *Psychological Methods*.

Savva, C. S., Osborn, D. R., & Gill, L. (2009). Spillovers and correlations between US and major European stock markets: The role of the euro. *Applied Financial Economics*, *19*(19), 1595–1604.

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, *6*(2), 461–464.

Schwert, G. W. (2011). Stock Volatility during the Recent Financial Crisis. *European Financial Management*, *17*(5), 789–805.

Sims, C. A. (1980). Macroeconomics and Reality. *Econometrica*, *48*(1), 1–48.

Singh, P., Kumar, B., & Pandey, A. (2010). Price and volatility spillovers across North American, European and Asian stock markets. *International Review of Financial Analysis*, *19*(1), 55–64.

Sivrikaya, A., İren, P., & Umay, T. (2021). The Relationship between Stock Returns, Bitcoin Returns, and Risk Aversion: Evidence from a Multivariate GARCH Model. *Sosyoekonomi*, *29*(47), Art. 47.

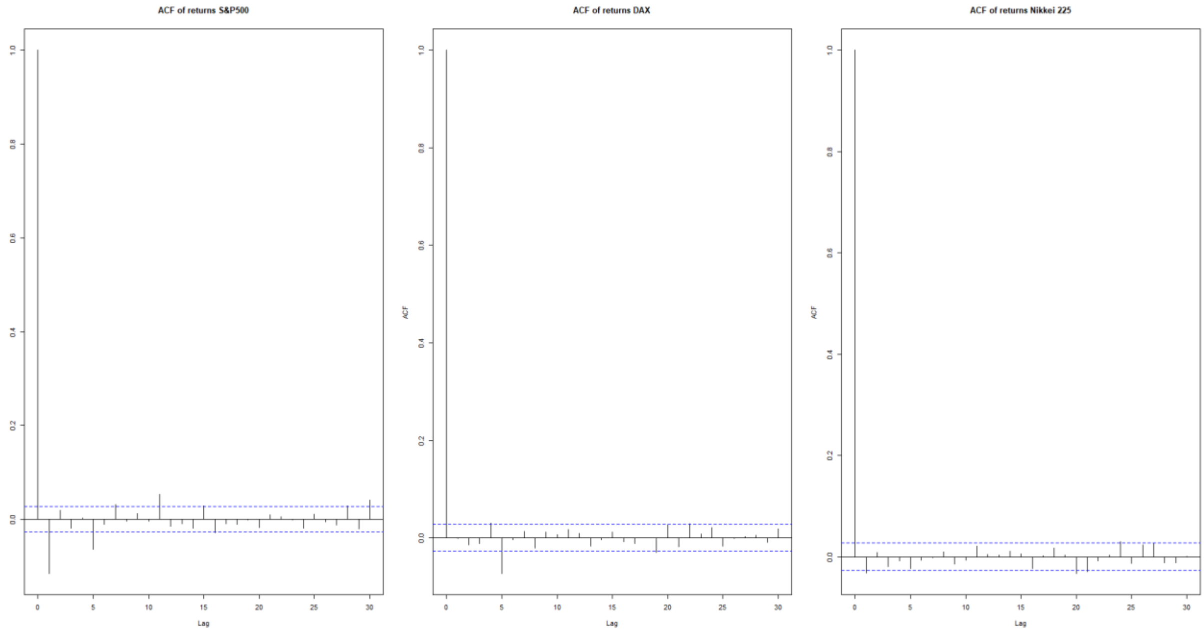
Tse, Y. K., & Tsui, A. K. C. (2002). A Multivariate Generalized Autoregressive Conditional Heteroscedasticity Model With Time-Varying Correlations. *Journal of Business & Economic Statistics*, *20*(3), 351–362.

Yu, C.-H., & Wu, C. (2001). Economic sources of asymmetric cross-correlation among stock returns. *International Review of Economics & Finance*, *10*(1), 19–40.

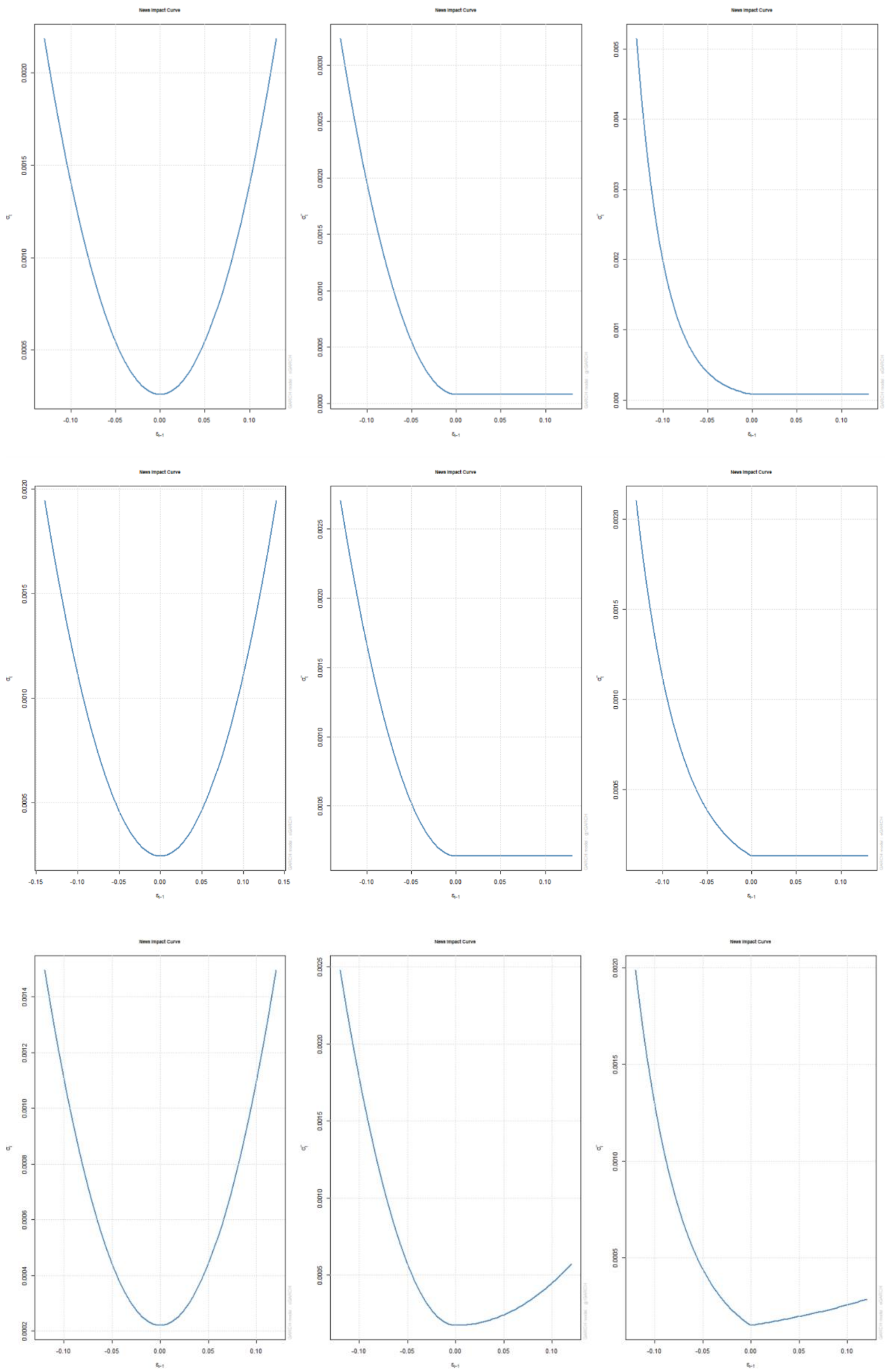
Zakoian, J.-M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and Control*, *18*(5), 931–955.

Annexes

Annex A –Autocorrelation Functions (ACF) applied to indices' returns



Annex B – News Impact Curves (NIC)



Annex C – Diagnostic of EGARCH models - weighted Ljung-Box test applied to standardized square residuals and weighted ARCH-LM test

		p-value		
		S&P 500	DAX	Nikkei 225
Ljung-Box	Lag (1)	0.7251	0.1521	0.9980
	Lag (5)	0.5180	0.1274	0.7777
	Lag (9)	0.7152	0.1810	0.7569
ARCH-LM	Lag (3)	0.6637	0.0333	0.1801
	Lag (5)	0.5547	0.0740	0.4017
	Lag (7)	0.7477	0.1212	0.4268

Annex D – Modulus of the eigenvalues of the companion matrix for the proposed VAR models

```
> roots(VAR9.1)
 [1] 0.7487595 0.7487595 0.7398798 0.7398798 0.7161693 0.7161693 0.7039871 0.7039871 0.6969494 0.6969494
 0.6939339 0.6939339 0.6486279 0.6486279 0.6203432 0.6203432 0.5807938 0.5807938
> roots(VAR5.2)
 [1] 0.5899651 0.5899651 0.5651928 0.5651928 0.5270124 0.4909266 0.4375557 0.4375557 0.3716533 0.3716533
> roots(VAR5.3)
 [1] 0.6068840 0.6068840 0.5890876 0.5890876 0.5783937 0.5322216 0.5115780 0.5115780 0.4748585 0.4748585
```