UNIVERSITÀ DI TRENTO
Dipartimento di Economia e Management

Master's Degree in

FINANCE

STUDY OF DYNAMIC CORRELATIONS FOR RISK ASSESSMENT IN FINANCE: AN ANALYSIS OF THE NASDAQ 100 WITH AND WITHOUT THE EFFECT OF COVID-19

Supervisor
Professor Giuliana Passamani

Student:
Giovanni Pedone

Academic Year 2023/2024

Acknowledgements

I would like to take this opportunity, in these few lines, to express my personal and deep gratitude to all those who enabled me to develop this thesis.

In particular, I would like to thank my supervisor, Prof. Giuliana Passamani, for her unceasing availability and constant support at every stage of the work. Even before that, I would like to thank her for providing me, during her course, with the indispensable tools for writing this dissertation. I would also like to express my gratitude to the University of Trento, whose excellent quality has contributed invaluably to my academic and personal foundation.

Finally, sincere thanks go to my family, who allowed me to reach this milestone without ever asking for anything in return.

Index

List of Figures

List of Tables

Introduction

Data analysis and the ability to extract useful information from it have always played a major role in the study of financial markets. In order to study these markets, knowledge of the main data analysis methodologies has become indispensable, both for those who are approaching the world of finance for the first time and for experienced players in the sector.

Exploratory data analysis (EDA), originally developed by the American mathematician John Tukey in the 1970s (Tukey, 1977), can be used for preliminary analysis. EDA aims to explore the data preliminarily, summarising its main characteristics through various steps, ranging from the *collection, cleaning* and *pre-processing* of the data to the identification of the relevant descriptive statistics, passing through univariate, bivariate and multivariate analyses, up to the use of techniques for visualising the results (scatter plot, box plot, heatmap, and others).

Once the exploratory analysis is completed and initial conclusions are drawn, these results can be used for advanced analysis or more sophisticated data modelling. A relevant example within EDA is *cluster analysis*, a set of methods for analysing multivariate data that allows observations to be divided into groups (*clusters*) based on similarities between variables. In other words, this technique allows the identification of hidden structures within the data, using algorithms such as *K-Means*, *hierarchical clustering*, DBSCAN, and others, thus offering a deeper insight into the characteristics of the data (Jha et al., 2023). In addition, cluster analysis can also be considered a method of data reduction, since the resulting groups can be analysed separately or further in-depth (Corter, 2023).

Such approaches also find application in the financial sector, since, after helping to understand the structure of variables, they facilitate the implementation of econometric financial models, supporting informed investment decisions and the formulation of risk management strategies. In financial econometrics, the mentioned techniques are particularly effective in identifying groups of securities within specific financial indices, selecting them based on their risk-return characteristics (Da Silva et al., 2005). These results can then be used to apply econometric models that study typical characteristics of financial time series, such as *heteroschedasticity*, *volatility clustering*, and the *presence* of *asymmetric effects* (Cont, 2001).

Analysis of these properties requires models that estimate volatility and conditional correlations. In this context, Engle was the first to propose that variance followed a conditional autoregressive process, dependent on past values and information available in the previous period. This approach is known as the *Autoregressive Conditional Heteroschedasticity* (ARCH) model (Engle, 1982) and was developed for the purpose of capturing volatility *clustering* phenomena. Later, in 1986, Bollerslev introduced a generalization of ARCH, the GARCH model, which proved more effective in predicting asset volatility (Bollerslev, 1986).

Nevertheless, these models and their extensions become very complex to apply in a multivariate context where one wants to consider the temporal interdependence among multiple time series of returns. For this reason, over the years, academics and practitioners have developed multivariate extensions of GARCH models with the aim of analysing more complex phenomena within financial time series, such as the study of conditional covariances among multiple assets.

A significant contribution in this field was made by Engle and Sheppard in 2002, with the introduction of the class of models for the study of *Dynamic Conditional Correlation* (DCC) (Engle, 2002), which gained popularity in academia because of its computational tractability and ability to study dynamic correlations between the returns of multiple assets. Nonetheless, although the DCC model is widely recognized and used in a variety of financial contexts, its specific applications to stock indices, such as the Nasdaq 100, are less frequent than, for example, more common studies of the S&P 500. In addition, even fewer analyses consider or exclude the effects of pandemic COVID-19 on correlations (Ji et al., 2022).

As a result, we decided to conduct a study of conditional second moments between a purpose-built macroeconomic-financial index and a set of stocks belonging to the Nasdaq 100, selected through cluster analysis based on risk-return characteristics and market capitalization. To provide a contribution to the existing literature, we also included an analysis excluding the initial months of the COVID-19 pandemic in order to assess its actual impact on dynamic correlations. The decision to place our focus on the Nasdaq 100 is motivated not only by the limited application of the DCC model to the U.S. index, but also by the growing importance of the technology sector, which has played a crucial role in supporting financial markets globally in recent years.

In parallel to the ability to understand and predict the time dependence of second-order moments of returns, DCC models are also relevant for risk management, especially market risk. In this regard, we decided to employ the dynamic matrix estimated through DCC to calculate the *Value at Risk* (VaR) and *Expected Shortfall* (ES) of an Equally-Weighted (EW) portfolio, adopting the variance-covariance approach. Again, we conducted a specific analysis to assess the impact of pandemic.

Therefore, the structure of the thesis is organized as follows. In the first chapter, we describe the methodology adopted for stock selection using cluster analysis, complemented by the construction and description of the macroeconomic-financial index and dataset used. In the second part, we introduce the theory behind the models applied to achieve the research objectives. The third chapter provides an analysis of the main risk measures currently provided by regulation, particularly VaR and ES. Each chapter progressively sets out the empirical results obtained during the development of the thesis, in a way that highlights the contribution of each section to the overall objective of the work.

Finally, we report the following key contents in the appendices: in Appendix A, the table and how to calculate the average returns and annualized volatility of Nasdaq 100 stocks; in Appendix B, the graphical representations of some conditional variance-covariance matrices related to the calculation of VaR and ES; in Appendix C, the most important Python and MATLAB codes to reproduce the main results, with the exception of the codes for creating the graphs, which can be easily replicated.

Cluster Analysis and Dataset Description

As defined in 1996, the field of *Knowledge Discovery in Databases* (KDD) "is the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable *patterns* in data" (Fayyad et al., 1996). It follows from this definition that *patterns*, once identified, must not be previously known, must lead to useful research actions and must be understandable in such a way as to facilitate understanding of the underlying data.

One of the main steps in the KDD process is the extraction and identification of relevant data features, commonly known as *exploratory data analysis* (EDA). EDA involves the application of analysis algorithms aimed at examining data and discovering relevant hidden or unknown patterns (Agarwal et al., 2024). The main objectives of EDA include understanding and describing the dataset by categorising and classifying the data, often using regression and classification methods. One of the key techniques in EDA is *cluster analysis*, which aims to break down sets of statistical units into homogeneous subsets called *"clusters"*.

In the first part of the elaboration, we will try to identify stocks that are representative of the American Nasdaq 100 technology index, by means of cluster analysis.

Cluster Analysis

This is a multivariate statistical model in which a number of *k* variables are observed over *n* statistical units. The *k* variables are then used to group the units in such a way that each group contains similar units with respect to the values of the variables. The term *clustering* refers to the procedure of dividing a given database into groups (Cai et al., 2016).

In cluster analysis, the definition of the measure of the degree of similarity between units becomes crucial. When working with data of a quantitative nature, the function that calculates similarity is called *"distance"* and must be minimised so that clusters are only formed by data that have similar characteristics. However, it may be preferable to refer to *similarity* when the objective is to maximise the likeness between objects within the same cluster and minimise it between objects belonging to different clusters (Mehta et al., 2020).

Suppose we have a single variable x, for which the numerator of the variance is given by:

$$D = \sum_{i=1}^{n} (x_i - \overline{x})^2$$

If, on the other hand, two variables are considered, the *total deviance* will be given by:

$$D_T = \sum_{g=1}^{K} \sum_{i=1}^{n_g} (x_{ig} - \overline{x})^2$$

This formula calculates the sum of the differences, squared, between each observation belonging to cluster g, x_{ig}, and the overall mean of the dataset, $\overline{x}$. In geometric terms, we consider the variables x_1 and x_2, as illustrated in the following figure:

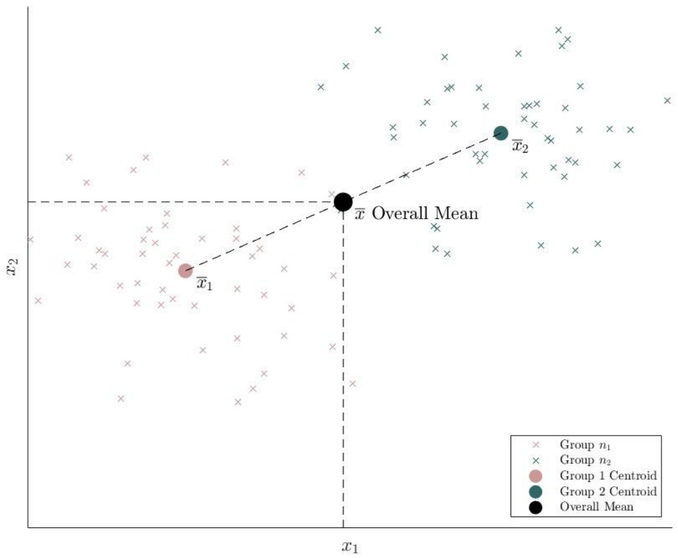


Figure 1. Representation of Cluster Analysis with identification of groups, their respective centroids, and the general centroid.

where the overall mean of the dataset – also called *"centroid"* – has the mean values of the variables x_1 and x_2 as its coordinates. If the total deviance, D_T, expresses the overall dispersion of the data with respect to the overall centroid, then, for each group of observations (n_1, n_2), we can calculate the mean to obtain the respective centroid, i.e. the averages referring to each subgroup ($\overline{x_1}$, $\overline{x_2}$).

At this point, the total deviance can be divided into two different components of variability:

- *Between* or *inter-cluster deviance*, i.e. the sum of the distances, squared, between the overall means of the groups and the overall centroid, weighted by the number of observations in each group:

$$D_B = \sum_{g=1}^{K} n_g (\overline{x_g} - \overline{x})^2$$

- *Within* or *intra-cluster deviance*, i.e. the sum of the distances, squared, between each point of the group and its overall mean:

$$D_W = \sum_{g=1}^{K} \sum_{i=1}^{n_g} (x_{ig} - \overline{x_g})^2$$

Consequently, the total deviance, D_T, is calculated as the sum of *between* and *within deviances*:

$$D_T = D_B + D_W$$

The ultimate objective of clustering is therefore to maximise the distance between different clusters (D_B) and minimise the dispersion of points within each cluster (D_W).

K-Means

Once the data to be worked on have been prepared, cluster analysis contemplates various algorithms for organising observations into homogeneous groups. These include *K-Means* (MacQueen, 1967), a *non-hierarchical* clustering technique in which the number of *K* clusters is established a priori by the user. It is a *prototype-based* or *center-based* method, where each generated cluster is represented by its own centroid.

The simple and intuitive K-Means algorithm technique (Jung et al., 2014) is described by the following execution steps:

Algorithm *K-Means*

1. random selection of *K* points as initial centroids,
2. assignment of each point to the closest centroid, according to the criterion of minimum distance,
3. recalculation of the centroids of each cluster,
4. iteration of steps 2 and 3 until the new centroids are "significantly" different from the previous ones.

In other words, the desired number of clusters *K* is arbitrarily chosen. Subsequently, each point in the dataset is assigned to the nearest centroid and, at the end of the allocation process, *K* new partitions are formed. Finally, as the initial centroids of each cluster are not representative of the final ones, new ones are recalculated based on the clusters obtained in the previous step. This procedure of assigning and recalculating the centroids is iterated until they remain unchanged, and the actual clusters are identified. Therefore, the algorithm terminates when the *intra-cluster* distance is minimal.

A crucial step in the algorithm is the assignment of each data point to the nearest centroid, such that the distance is minimal. Here, we chose to use the so-called *"elbow method"*, a graphical representation that utilises the output generated by the K-Means algorithm to determine the optimal number of clusters to work with. The method consists of displaying the explained variances (or *Within-Cluster Sum of Squares*[1], WCSS) as a function of the number of clusters. The "elbow" name is derived from the point at which the WCSS is drastically reduced, visually indicating the optimal number of clusters to be used in the model (Hartigan et al., 1979).

However, one of the limitations of the technique described so far concerns the first step of the algorithm, namely the random selection of the initial *K* centroids. Indeed, each time the algorithm is run leaving the *K* parameter unchanged, the initial centroids will be different each time since they will be selected randomly (Chang et al., 2018). Consequently, the final solution will also differ each time the algorithm is run.

In this regard, in this study, we have chosen to set the number of clusters *arbitrarily*, a practice that is widely recognised in the literature to avoid too many or, on the contrary, too few clusters. This approach is supported by numerous academic studies that demonstrate its effectiveness in various fields, including big data (Ikotun et al., 2023), data science (Kodinariya et al., 2013) and financial risk management (Zhu et al., 2021).

[1] WCSS is the equivalent of *intra-cluster* deviance (D_W) and measures the sum of squared distances between the points and centroids of the respective clusters.

Empirical Results

The purpose of this section is to analyse the initial dataset of closing prices of stocks belonging to the Nasdaq 100 index to identify, by means of cluster analysis, 4 representative stocks, selected based on average return and annualised volatilities over a 5-year period. These stocks will then be used to develop econometric models to study the evolution of correlations over time.

For the analysis of this section, Python was used in the Spyder environment, an *Integrated Development Environment* (IDE) specifically designed for scientific programming. After importing the main libraries (*NumPy, Pandas, Yfinance, Plotly Express, Matplotlib, SciPy, math, Sklearn*), so-called *"Web Scraping"* was implemented, i.e. the process of automated data extraction from web pages. In particular, *tickers* and daily *adj close* prices were downloaded from Wikipedia and Yahoo Finance, respectively, covering the time interval from 1/1/2019 to 1/1/2024.

Cluster Analysis and K-Means

Before applying the unsupervised cluster analysis methodology, it is essential to conduct an exploratory dataset analysis and proper data cleaning, which are fundamental steps to ensure the quality of the KDD process (Fayyad et al., 1996). The results are shown in the table in Appendix A. Having completed the *preprocessing*[2] phase of the financial data of our interest, we proceed to the heart of the section, namely the clustering of the data using the K-Means algorithm and the visualisation of the elbow curve (elbow method).

In detail, to determine the optimal number of clusters, we train the model by iterating over a range of values between 2 and 19. We then calculate the sum of the squared distances between the samples and the centroids of the respective clusters, known as the WCSS, to assess the compactness and coherence of the clusters formed. In formula:

$$WCSS = \sum_{g=1}^{K} \sum_{i=1}^{n_g} (x_{ig} - \overline{x_g})^2$$

where x_{ig} represents the i-th observation of cluster g, $\overline{x_g}$ and n_g are, respectively, the overall mean and the number of points of cluster g, K is the total number of clusters.

[2] Which is the one preceded by the selection of data and is followed by the stages of identifying patterns and extracting knowledge from the data.

At this point, the elbow curve is plotted, which shows the value of WCSS as a function of the number of clusters. In our case, the elbow of the curve suggests that the optimal number of clusters is $K = 4$:

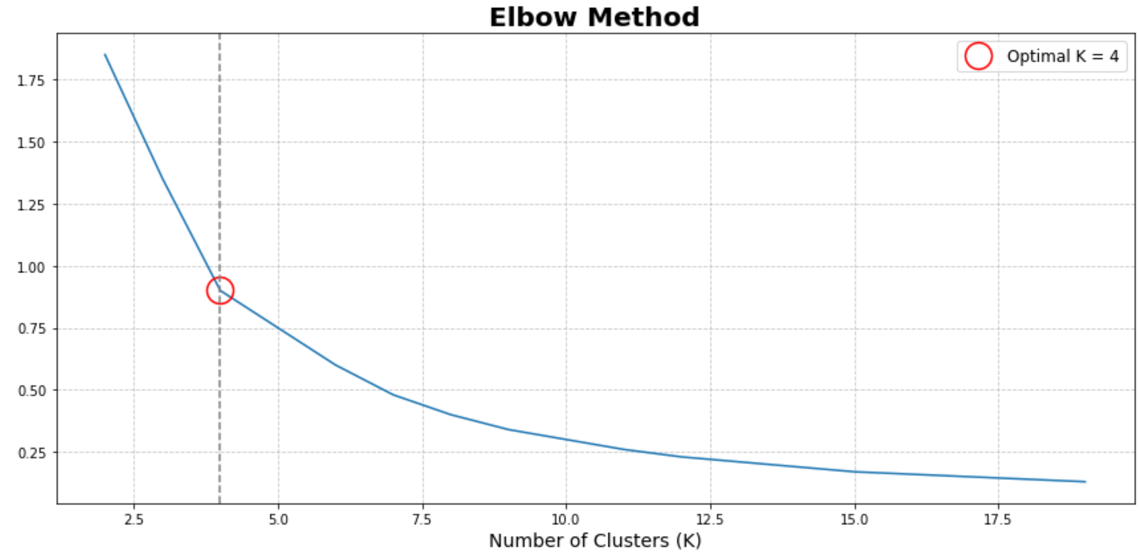


Figure 2. Representation of the elbow method and the optimal number of clusters.

Note: the reading of the elbow on the graph is from right to left.

To avoid the limitation described in the previous section, i.e. the risk of obtaining different K each time the algorithm is run, the data are *arbitrarily* divided into four clusters, regardless of the variations that might occur in subsequent runs (Zhu et al., 2021). Having done so, the stocks are then assigned to their respective clusters, and a scatter plot is created illustrating the distribution, for each *bivariate* observation, of the shares according to their risk-return characteristics (Da Silva et al., 2005):

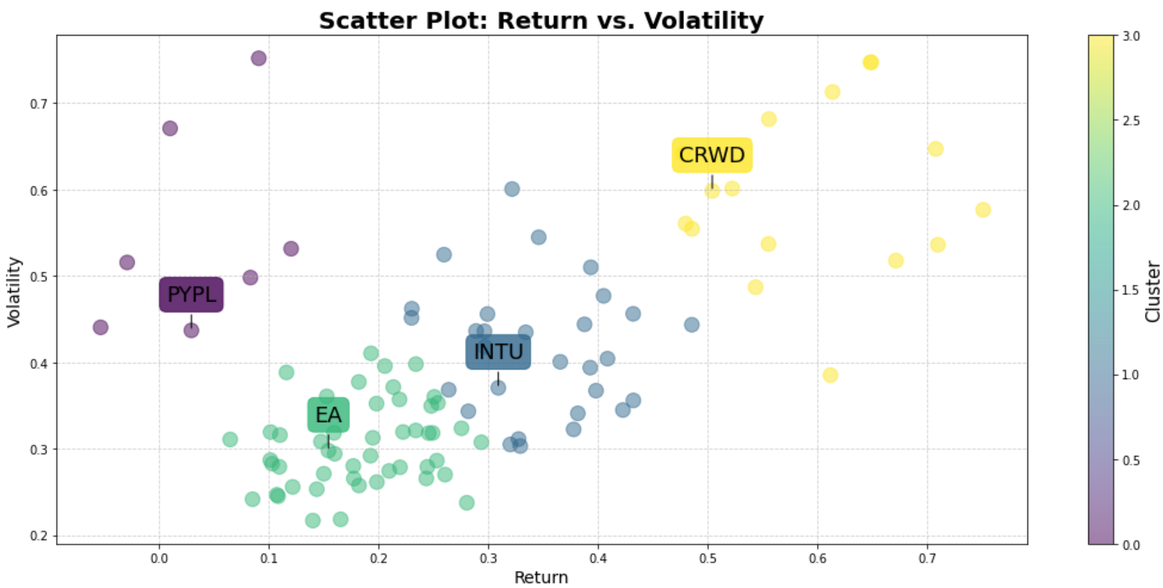


Figure 3. Stocks resulting from the identification of the 4 clusters based on risk-return characteristics.

Finally, the last part of the analysis focuses on the interpretation of the clusters obtained. Having returned the closest and most representative shares of each of the 4 clusters *(Closest Ticker)*, the results, including the months of the COVID-19 pandemic, are summarised in the following table:

Cluster	***Closest Ticker***	***Avg. Return***	***Volatility***	***Ret-Vol***
0	INTU	0.3093	0.3705	mid-mid
1	CRWD	0.5044	0.5984	high-high
2	PYPL	0.0296	0.4369	low-high
3	EA	0.1546	0.2978	low-low

Table 1. List and characteristics of equities identified by Cluster Analysis and K-Means algorithm.

The table highlights the stocks that have shown greater stability than those subject to more marked fluctuations, offering a clear and precise overview of the results. In particular, it provides a quali-quantitative description of the average return and annualised volatility of each share identified as closest to the centroid, outlining the behaviour of each over the time horizon considered.

The above is merely the starting point for a subsequent in-depth study of the dataset. The analysis to follow will focus on the description of all variables, which, together with those just mentioned, will be instrumental to the objective of the paper: the study of dynamic market correlations using multivariate econometric models.

Dataset Description

As the thesis is aimed at studying conditional second moments of financial time series, it is also necessary to identify some macroeconomic variables that allow us to consider a broader context useful for understanding financial market movements and their influence on asset returns.

The dataset chosen for the analysis is supplemented by appropriately transformed variables. In the case of financial variables, returns were calculated from *adj close* prices. Working over a 5-year time interval, we made use of daily data using different sources, such as Yahoo Finance, Bloomberg and Federal Reserve Economic Data (FRED).

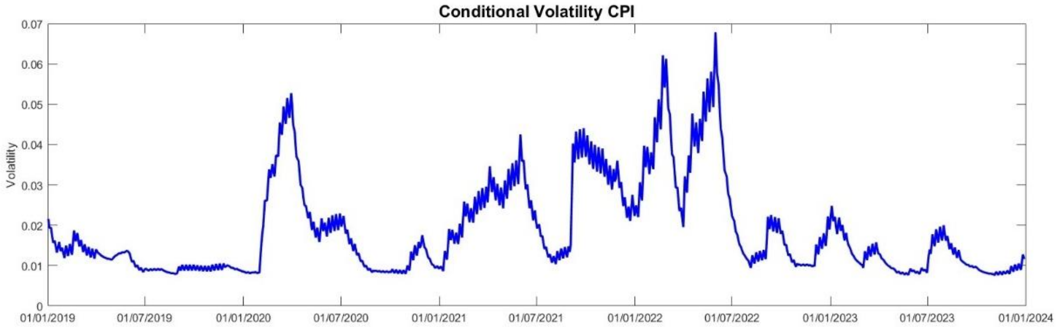
The macroeconomic and financial variables examined are those commonly suggested and considered in studies of financial models:

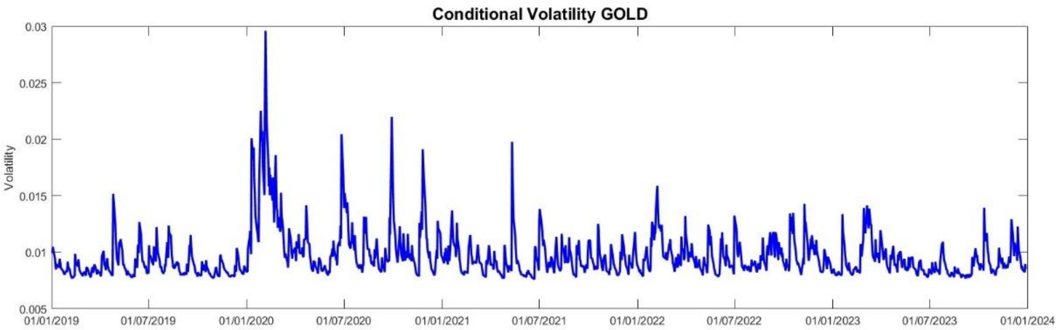
- *Consumer Price Index* (Cpi), a key indicator used to measure price inflation of goods and services purchased by consumers over a given period of time, whose changes significantly influence financial markets, the real economy, investor decisions and monetary policies (Li, 2002).

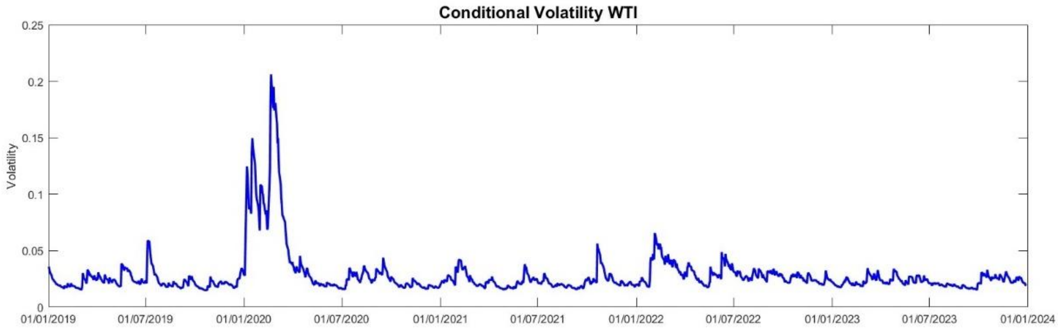
Since this data is calculated and published monthly, whereas the present study uses daily data, a linear interpolation was performed so that daily Cpi values were available. Next, the approximate percentage change in prices was calculated using the difference in the logarithm of the Cpi between consecutive periods.

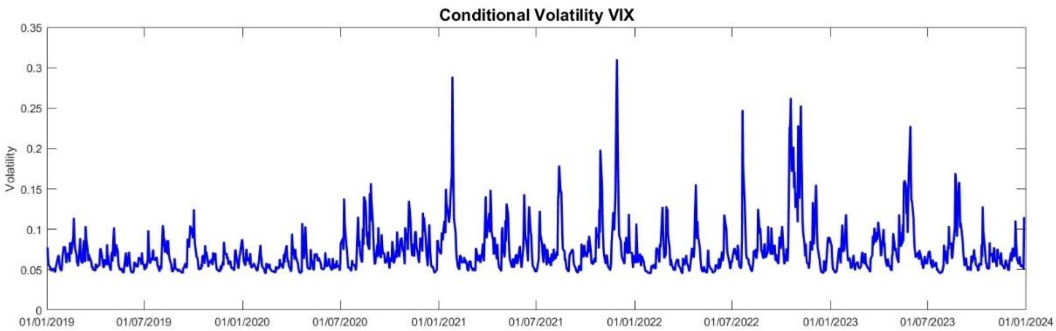
- *Gold*, historically regarded as a safe haven asset by investors, who include it in their portfolios as a form of diversification to reduce overall risk. Fluctuations in its price can influence interest rates set by central banks and/or can be seen as an indicator of inflation.
  It has been valued in terms of return.
- *West Texas Intermediate* (WTI), similar to gold, it can have a different correlation to other financial assets, and, thus, can be considered as a diversification option within portfolios to mitigate overall risk (Arouri et al., 2011). The WTI futures contract is one of the most traded, and therefore liquid, in commodity markets and its price has a significant impact on the global economy, serving as an indicator of global energy supply and demand, geopolitical events and future economic expectations.
  It has been valued in terms of return.
- *Vix*, knowns as the "fear index", measures the implied volatility of the U.S. stock market using the prices of put and call options on the S&P 500 index, with which it is negatively correlated. For this reason, it is a highly indicative variable that provides information on market sentiment and whose high reading, usually above 20-25 points, indicates increased volatility within the financial system.
  It has been valued in terms of return.
- *Fvx*, represents the yield on the 5-year Treasury and is an indicator of short-term interest rates on the bond market. This variable provides useful signals about inflation expectations, the orientation of monetary policy and market sentiment.
  It is calculated in differential form.
- *Lois*, is the spread between the 3-month Libor and the 3-month Ois and represents the risk premium of an interbank credit on a term deposit over a risk-free (*overnight*) one. It is, then, an indicator of liquidity and the degree of confidence existing in the interbank market, which can help to interpret the macroeconomic scenario and the banking market's predisposition to provide credit to companies.
  It is calculated in differential form.

The description provided thus far constitutes the preliminary basis for the implementation of the DCC model, the development of which will be undertaken in the following chapter. Since the latter requires some computational effort, it was decided to aggregate the aforementioned variables into a single index in order to obtain a concise and indicative reading of U.S. market conditions. The construction of the index went beyond that part of the econometric literature that identifies more or less sophisticated methodologies, so it was considered as a simple linear combination of variables, weighted according to normalised weights (Arrigoni et al., 2020). In order to assign the weights correctly, the conditional volatility of each of the variables introduced earlier was estimated in advance by employing the GARCH (1,1) model. The results obtained are as follows:









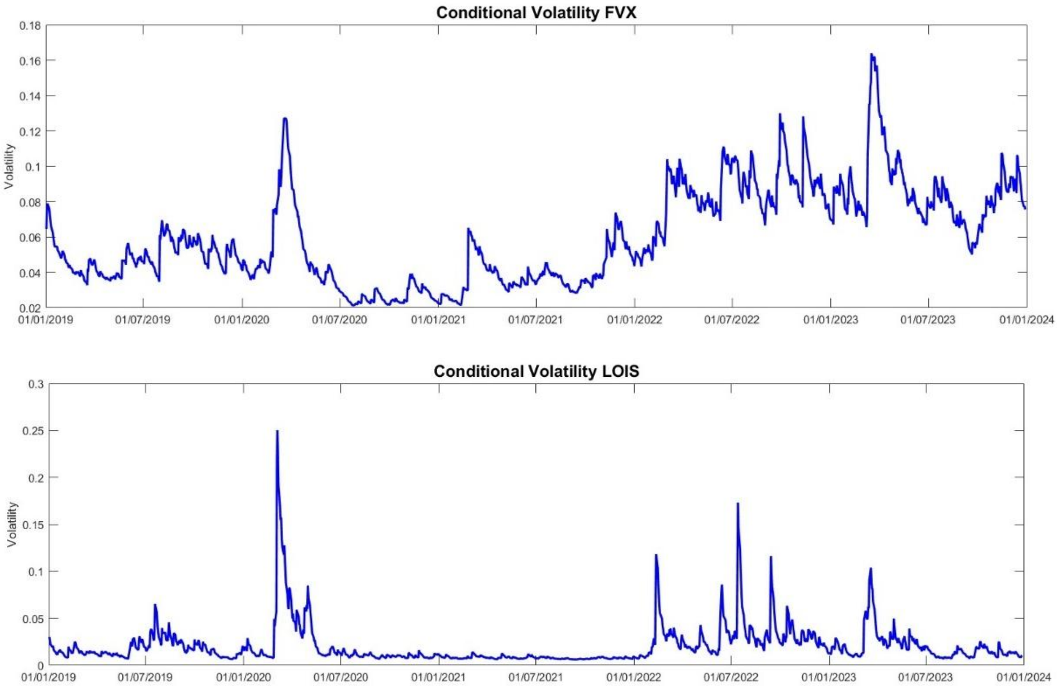


Figure 4. Conditional volatilities of macroeconomic and financial variables estimated using GARCH(1,1).

The figures depict the evolution of the volatilities of the macroeconomic and financial variables considered, over the period from January 1, 2019, to January 1, 2024. Firstly, we note how almost all variables – except for the Vix – registered significant volatility spikes during the pandemic, reflecting the turbulent period experienced in the early months of 2020.

Each plot highlights phases of higher or lower volatility, depending on the variable being analysed. For instance, Gold, Fvx, and Vix show spikes in volatility distributed throughout the period, suggesting a constant responsiveness to changes in economic conditions. In contrast, Cpi, Wti, and Lois alternate between phases of high volatility and periods of relative stability.

One particularly noticeable aspect is the reduced variability of the Lois spread between July 2020 and February 2022, which is likely due to the interventions of Central Banks during the COVID-19 pandemic, when they drastically cut interest rates to support the economy and stabilize financial markets. Such monetary policies, in fact, compressed Libor and Ois rates, and, consequently, the Lois spread. This greatly reduced the perception of risk in the interbank market, encouraging banks to lend more confidently and resulting in a prolonged period of low volatility.

	$\boldsymbol{\omega}$ ***(Omega)***	$\boldsymbol{\alpha}$ ***(Alpha)***	$\boldsymbol{\beta}$ ***(Beta)***
Gold	0.0000 *(0.0000)*	0.1413 *(0.0000)*	0.6727 *(0.0000)*
Wti	0.0000 *(0.0000)*	0.1673 *(0.0000)*	0.7925 *(0.0000)*

Vix	0.0000 *(0.0000)*	0.3650 *(0.0000)*	0.5036 *(0.0000)*
Fvx	0.0000 *(0.0000)*	0.0947 *(0.0000)*	0.9049 *(0.0000)*
Cpi	0.0000 *(0.0000)*	0.1112 *(0.0000)*	0.8888 *(0.0000)*
Lois	0.0000 *(0.0000)*	0.2993 *(0.0000)*	0.7007 *(0.0000)*

Table 2. Estimated coefficients of GARCH(1,1) models.

As the null values of the constant term, ω, clearly indicate, this parameter had no effect on the dynamics of the variables in question. As a consequence, their evolution is entirely determined by past information (α) and the persistence of volatility (β). In particular, the results suggest that Vix and Lois react more responsively to past shocks, while Fvx and Cpi show very persistent volatility over time. However, the value of β is still high for almost all variables. In addition, *p-values* of 0.000 for all parameters confirm the high statistical significance of the results, underscoring how crucial α and β are here in determining conditional volatility.

Following the extraction of conditional volatilities, these results were utilised to determine the weights to be assigned to each variable in the index, inversely proportional to their volatility. Specifically, variables characterised by greater uncertainty, and therefore higher risk, received lower weights, while those with more stable and lower volatility received higher weights (Arrigoni et al., 2020). This approach ensured that less volatile variables contributed more significantly to the index. The weights were then normalised, ensuring that the total sum was 1 and that each variable contributed proportionally to its relative risk.

Finally, we calculated the overall index log-returns as a weighted linear combination of the variables, multiplying each of them by its normalized weight, and summing the results:

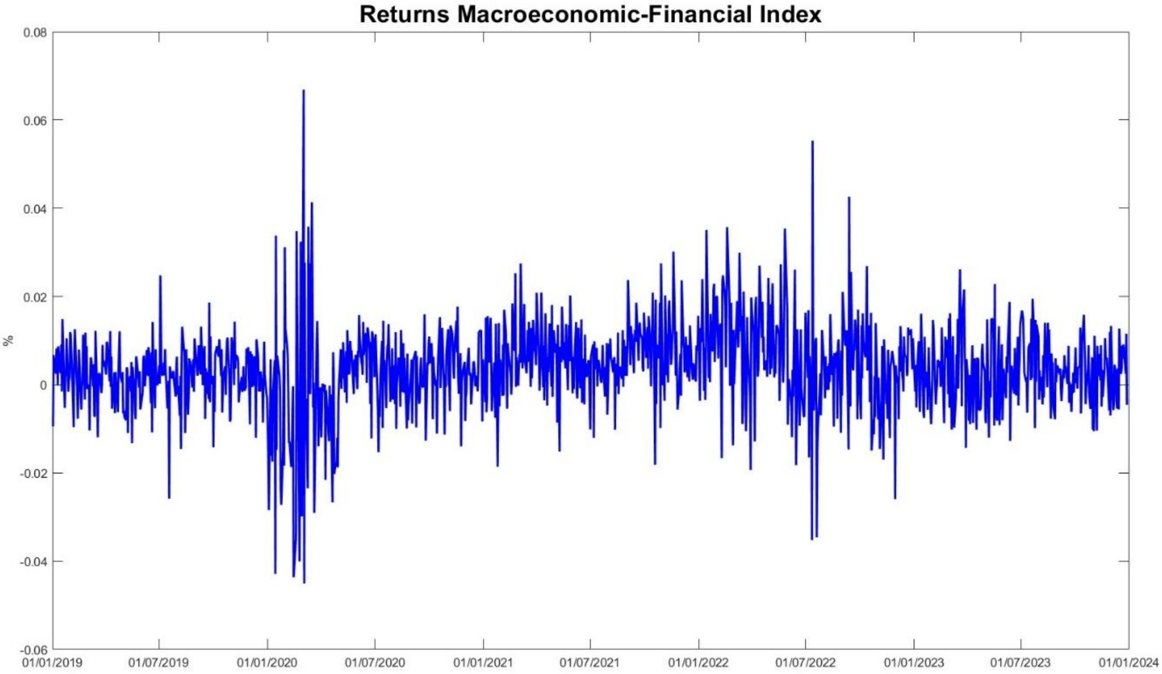


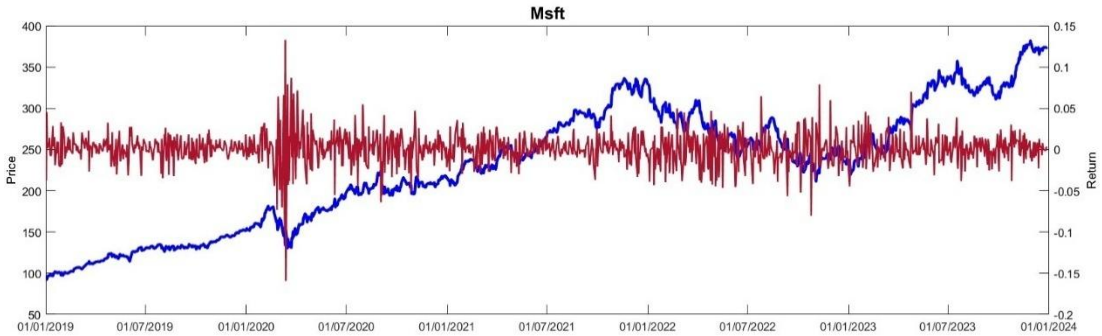
Figure 5. Representation of the returns of the purpose-built macroeconomic-financial index.

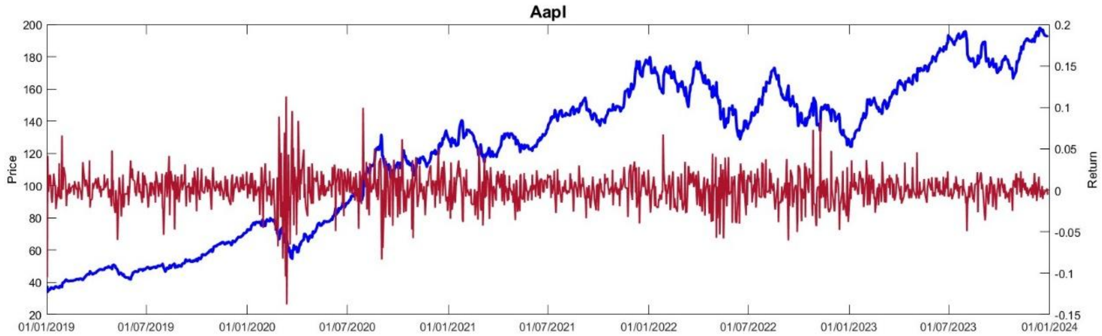
Hence, the method presented so far has enabled us to develop a dynamic index that reflects not only the variability and persistence of the fluctuations of each component, but also their weighted contribution based on expected volatility.

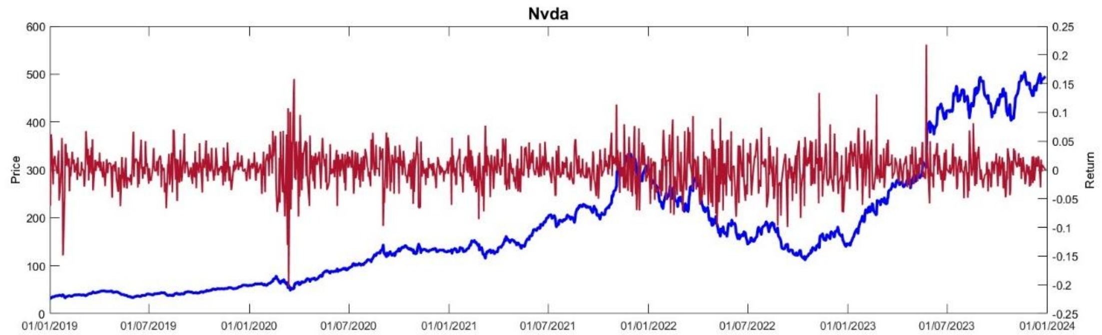
To provide a comprehensive description of the dataset, we decided to include the 4 largest capitalisation stocks of the Nasdaq 100[3], in addition to the equities selected through the cluster analysis: Microsoft, Apple, Nvidia e Amazon. The inclusion of these shares was implemented to ensure that the group of stocks utilised during the analysis was as representative as possible of the U.S. index. In contrast to the macroeconomic variables, the equities were treated on an individual basis to ensure the most accurate capture of the correlation over time of each asset with respect to the constructed index.

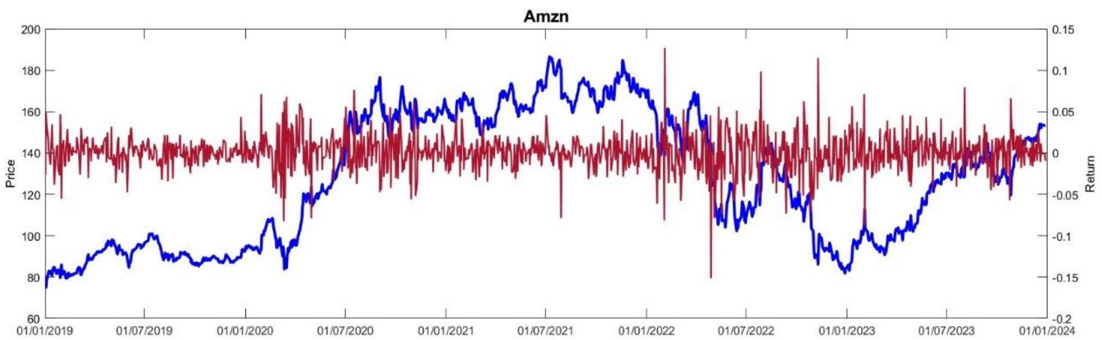
[3] The review of the companies with the largest capitalisation was conducted in the final stage of the writing of this thesis. However, due to natural market fluctuations, the names given may change in later periods.

The image below illustrates how the log-return volatilities of each stock exhibit the typical *heteroschedastic* pattern of financial time series, a phenomenon widely studied in econometrics and falling under the category of so-called *“stylized facts”* (Cont, 2001):









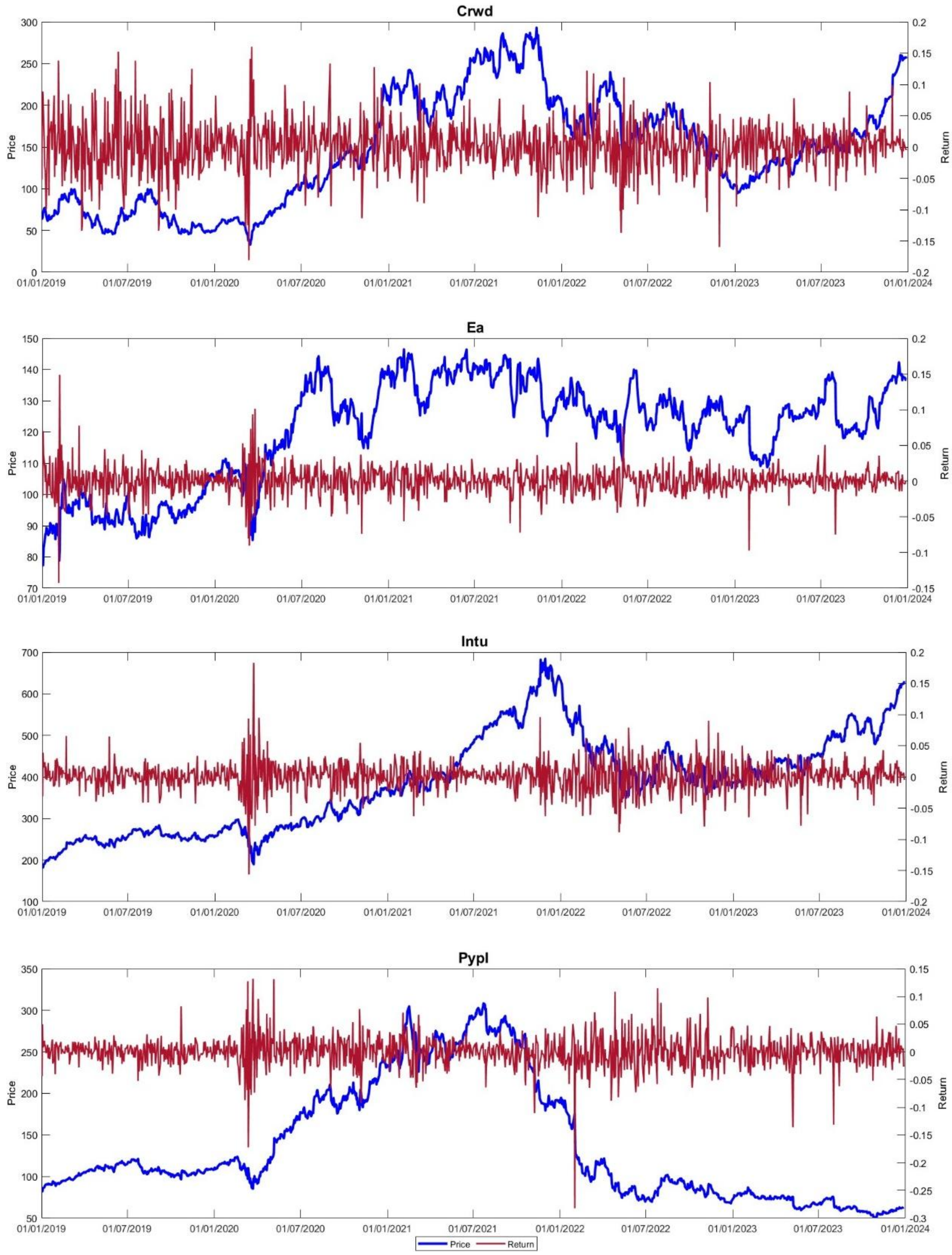


Figure 6. Combined representation of the log-returns and price trends of the stocks.

This phenomenon is also evident in the respective log-return distributions, which exhibit a pronounced concentration around the mean value, accompanied by elevated peaks and events that deviate significantly from the central tendency. This finding indicates that the data does not conform to a standard Gaussian distribution, suggesting instead that it is *leptokurtic*, with heavier tails:

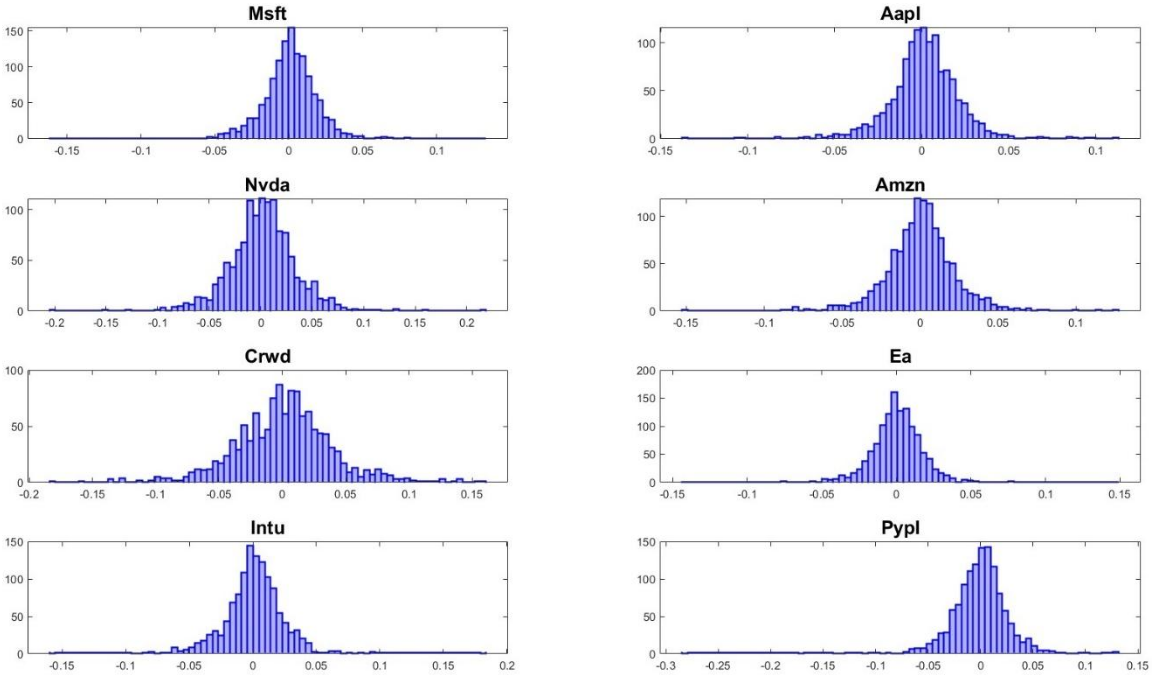


Figure 7. Log-return distributions of the financial time series.

Dynamic Conditional Correlation Model

In financial econometrics, understanding and predicting the dependence of second-order moments of asset returns plays an important role for several reasons. For example, volatilities and correlations between assets are key inputs for risk management and optimization of a financial portfolio, or for hedging strategies to derive traded option prices using the Black-Scholes formula. Furthermore, the volatilities of securities and markets tend to exhibit synchronised fluctuations over time, going to affect the correlations themselves.

Over the years, it has, thus, become necessary to develop models capable of capturing, among the patterns that characterize time series, those related to multivariate time interdependence between conditional second moments, in order to analyse the effects between different markets or returns. In other words, such models must explain how the behaviour of markets or assets affects others, or is affected by them, either directly or indirectly, through volatilities and conditional covariances (Bauwens et al., 2006). The most well-known and frequently used of these are the Diagonal VECH (DVECH) (Bollerslev et al., 1988), the Diagonal BEKK (DBEKK) (Engle et al., 1995) and the model with Conditional Constant Correlations (CCC) (Bollerslev, 1990). In 2002, Engle proposed a new model that overcomes the structural and empirical weaknesses of these frameworks by attempting to identify, in a *direct* and *parsimonious* way, the dynamic correlations between financial securities. This is known as *"Dynamic Conditional Correlation"* (DCC) model, whose evaluation and application in the context of financial risk management constitutes the basis of this thesis.

The DCC model has become popular in academia because of its computational tractability, primarily because it separates volatility and correlation modelling. The result is that the number of parameters to be estimated grows linearly rather than exponentially, as would be the case with Multivariate GARCH (MV-GARCH) models. This approach effectively addresses the issue of the number of observations needed in parameter estimation. The main steps for its implementation are as follows:

1. estimation of *time-varying* volatilities for each asset considered;
2. standardization of returns by the volatilities estimated in step 1. and estimation of the unconditional correlation matrix;
3. estimation of the correlation persistence parameters, α and β, to make the correlation between the considered assets dynamic.

In other words, in the DCC, correlation estimates are updated whenever new information on volatility-adjusted returns is recorded.

The model typically adopted to study the conditional volatility of each asset studied is the GARCH, formulated by Bollerslev (1986). If, on the other hand, one intends to check for asymmetric effects or to resolve the issue related to the violation of non-negativity constraints in time series, alternative specifications can be used, such as Threshold GARCH (TGARCH) (Zakoian, 1994), GJR-GARCH (Glosten et al., 1993) or Exponential GARCH (EGARCH) (Nelson, 1991).

Hence, the objective of this chapter is to provide an implementation of the model just introduced through the application to 9 log-return time series, one of which is the synthetically constructed macroeconomic-financial index, while the remaining series correspond to 8 representative stocks of the U.S. Nasdaq 100 index. In particular, we will go on to study the dynamic correlations between pairs of variables, one of which is the macroeconomic-financial index and the other is each of the assets, so that we can understand how the economic scenario affects the volatility, and thus the correlation, of the series considered over time. In addition, 7 types of models will be examined, of which only one will be used to later calculate *Value at Risk* (VaR) and *Expected Shortfall* (ES).

Dynamic Conditional Correlation (DCC) Model

DCC has gained popularity because of its ease of application. The models for conditional second moments mentioned in the previous paragraph all aim to estimate the evolution of correlations between returns. However, they have a significant limitation due to the large number of parameters required during the estimation process, especially when there are more than 3 or 4 time series to be analysed (Bauwens et al., 2006). This factor complicates the application of such models in contexts such as risk management and asset allocation (Engle, 2002). To clarify the concept, we refer to the main methods:

- DVECH, in addition to not guaranteeing the positivity of the covariance matrix by making predictions more volatile, it estimates a number of parameters that increases exponentially as the number of series increases;
- DBEKK, which is a special case of the previous one, almost always allows the positivity of the covariance matrix to be satisfied and has a smaller number of parameters to be estimated, which nevertheless still remains high;
- CCC, although it succeeds in reducing the number of parameters to be estimated compared with the first 2 models, empirical studies have shown that the constant correlation assumption is unrealistic, being, then, too reductive and restrictive in some applications (Chevallier, 2011).

Given the need to estimate large matrices of second moments to model co-movements over time between financial variables in a parsimonious way, Engle and Sheppard introduced the Dynamic

Conditional Correlation model in 2002. The main advantage of this model lies in the fact that the number of parameters required to estimate dynamic correlations does not depend on the number of assets considered. In fact, instead of jointly evaluating volatility and correlations, the DCC first proceeds to separately estimate the conditional volatility of each asset by univariate model, and, subsequently, integrates the multivariate component of the dynamic correlations. In this way, it takes advantage of the simplicity of univariate models to considerably reduce the computational load typical of multivariate models.

DCC belongs to the class of models that are based on decomposing the conditional covariance matrix of asset returns into the product between matrices of standard deviations and conditional correlations. It assumes that the distribution of the same returns can be approximated to multivariate normal with zero mean and variance-covariance matrix, Σ_{t+1} (Engle, 2002). Formally:

$$r_t | I_{i-1} \sim \mathbb{N}(0, \Sigma_t)$$

$$\Sigma_{t+1} = D_{t+1} \Gamma_{t+1} D_{t+1}$$

where Σ_{t+1} can be decomposed as indicated. In fact, assuming the simple case in which the number of assets is *N = 2*, we have that

- D_{t+1} is a *N x N* diagonal matrix of conditional standard deviations, $\sigma_{i,t+1}$, on the *i*-th diagonal, which, in the *2 x 2* case, is as follows:

$$D_{t+1} = \begin{bmatrix} \sigma_{1,t+1} & 0 \\ 0 & \sigma_{2,t+1} \end{bmatrix}$$

  where the volatilities of each asset are estimated separately using univariate GARCH models, assuming that the errors are normally distributed.
  Extending the model to higher-order delays, the general form of GARCH can be expressed as follows:

$$\sigma_{i,t+1}^2 = \omega_i + \sum_{p=1}^{P_i} \alpha_{ip}\, r_{i,t-p}^2 + \sum_{q=1}^{Q_i} \beta_{iq}\, \sigma^2{}_{i,t-q}$$

  where *i = 1, 2, ..., N*. Non-negativity restrictions are imposed on the GARCH to ensure that the conditional variance, $\sigma_{i,t+1}^2$, is always positive.
  At this preliminary stage of constructing the conditional covariance matrix in the DCC framework, one may also consider GARCH-type models that account for asymmetries or

leverage effects in the dynamics of volatilities, such as TGARCH, GJR-GARCH, and EGARCH.

- Γ_{t+1} is a symmetrical matrix of dynamic conditional correlations, $\rho_{ij,t+1}$, which, in the *2 x 2* case, is as follows:

$$\Gamma_{t+1} = \begin{bmatrix} 1 & \rho_{12,t+1} \\ \rho_{12,t+1} & 1 \end{bmatrix}$$

  The estimation of Γ_{t+1} facilitates the description of the evolution of asset correlations over time in response to fluctuations in returns.

By assembling the components, the variance-covariance matrix for *N = 2* can be expressed as:

$$\Sigma_{t+1} = \begin{bmatrix} \sigma_{1,t+1}^2 & \sigma_{12,t+1} \\ \sigma_{21,t+1} & \sigma_{2,t+1}^2 \end{bmatrix} = \begin{bmatrix} \sigma_{1,t+1} & 0 \\ 0 & \sigma_{2,t+1} \end{bmatrix} \begin{bmatrix} 1 & \rho_{12,t+1} \\ \rho_{12,t+1} & 1 \end{bmatrix} \begin{bmatrix} \sigma_{1,t+1} & 0 \\ 0 & \sigma_{2,t+1} \end{bmatrix}$$

Having introduced the main elements of the DCC model, it is now important to explain how one should proceed for the estimation and prediction of the mentioned matrices $\widehat{D}_{t+1}$ and $\widehat{\Gamma}_{t+1}$, respectively.

The implementation of the model can be done through two approaches, referred to as *"2-stage"* and *"3-stage"*. The main difference between the two lies in the method of estimating the dynamic correlation parameters. In this thesis, we will apply the DCC model by considering the second approach.

Going into it, first, we estimate *N* univariate GARCH models, one for each of the time series considered, to extract the conditional volatility series, $\sigma_{i,t+1}$, which we will use as input to normalize the historical returns, $R_{i,t+1}$:

$$z_{i,t+1} \equiv {}^{R_{i,t+1}}\!/_{\sigma_{i,t+1}}$$

In other words, from the estimates derived from the application of GARCH we construct the individual $\widehat{D}_{t+1}$ matrices with which we later standardize the returns. The purpose of this step is to transform these into a form such that they have zero mean and unit standard deviation, which we denote by $z_{i,t+1}$.

The second step requires using the standardized returns obtained in the first step to estimate an unconditional correlation matrix, $\bar{Q}$:

$$\bar{Q} = \frac{1}{T}\sum_{t=1}^{T} z_t z_t'$$

with T equal to the number of observations and where z_t is the vector of standardized returns. However, in the DCC framework, modelling is not done directly on the returns, but instead an auxiliary variable, $q_{ij,t+1}$, which is estimated through a GARCH(1,1) in order to calculate conditional correlations:

$$q_{ij,t+1} = \overline{\rho}_{ij} + \alpha\left(z_{i,t+1}z_{j,t+1} - \overline{\rho}_{ij}\right) + \beta\left(q_{ij,t} - \overline{\rho}_{ij}\right) \qquad \forall i,j$$

$$= \overline{\rho}_{ij} + (1 - \alpha - \beta) + \alpha\left(z_{i,t+1}z_{j,t+1} - \overline{\rho}_{ij}\right) + \beta q_{ij,t}$$

$$= \omega_{ij} + \alpha\left(z_{i,t+1}z_{j,t+1}\right) + \beta q_{ij,t}$$

where $\overline{\rho}_{ij}$ is the *i,j* element of the $\bar{Q}$ matrix, that is the unconditional correlation of the returns $z_{i,t+1}$ and $z_{j,t+1}$ resulting from the first stage.

The goal of the second step is aimed, therefore, at identifying the parameters of the matrix Q_{t+1} that will be used to recalibrate the dynamic correlations calculated in the third step in order to improve the precision of the estimates. In this way, we estimate an unconditional matrix at each point on the time axis of the dataset, going on to capture the punctual relationships between pairs of assets. The choice to estimate an unconditional matrix can be traced to two orders of reason:

1. simplification of the computational process, the calculation of a non-conditional matrix does not require the further implementation of econometric models such as GARCH;
2. reduced dependence on model parameters and simplification about the interpretation of results.

In the third stage, the dynamic correlation structure is based on the following relationship:

$$\hat{\Gamma}_{t+1} = Q_{t+1} Q_t Q_{t+1}$$

$$Q_{t+1} = (1 - \alpha - \beta)E[z_t z_t'] + \alpha z_t z_t' + \beta Q_t$$

where Q_{t+1} is the symmetric *N x N* matrix of dynamic conditional covariance, dependent on the parameters α and β that reflect the influence of innovations and past covariances. This matrix reports on the diagonal the values of the auxiliary variable, $q_{ij,t+1}$:

$$Q_{t+1} = \begin{bmatrix} q_{ii,t+1} & q_{ij,t+1} \\ q_{ji,t+1} & q_{jj,t+1} \end{bmatrix}$$

The dynamic correlation matrix, $\hat{\Gamma}_{t+1}$, is obtained by normalizing the matrix, Q_{t+1}, that is, capturing the dynamic correlations between assets on the diagonal. Specifically, the element *(i, j)* of the matrix $\hat{\Gamma}_{t+1}$ is the dynamic correlation, $\rho_{ij,t+1}$:

$$\hat{\rho}_{ij,t+1} = \frac{q_{ij,t+1}}{\sqrt{q_{ii,t+1}}\sqrt{q_{jj,t+1}}}$$

which guarantees that $\rho_{ij,t+1} \in [-1, 1]$, by construction. Furthermore, it is necessary that Σ_{t+1} be positively defined, as it is a covariance matrix. This is possible in the DCC model since Q_{t+1} is a positive semidefinite matrix (SPD) because it is a weighted average of semidefinite and positively defined matrices (Guidolin, 2018).

Consequently, both the correlation matrix, Γ_{t+1}, and the covariance matrix, Σ_{t+1}, will be semidefinite positive for $\forall_t$. The necessity for a correlation matrix with non-negative eigenvalues is driven by the requirement to guarantee the stability and reliability of the model parameter estimates that determine the dynamics of correlations over time, α and β:

$$Q_{t+1} = (1 - \alpha - \beta)E[z_t z_t'] + \alpha z_t z_t' + \beta Q_t$$

also called *"covariance targeting"*, where α and β are non-negative scalar parameters that satisfy the *stationarity* condition, $\alpha + \beta < 1$, and the non-negativity constraint, $\alpha \geq 0$ and $\beta \geq 0$. These constraints ensure that the unconditional matrix is identical to the unconditional sample correlation of the data.

Asymmetric DCC (ADCC) Model

As previously stated, this dissertation has not been exclusively limited to the consideration of symmetric models. As is the case with volatility analysis, the objective was to study how positive and negative shocks also produce different magnitudes in the field of correlations, the so-called *"leverage effect"* (Cappiello, 2006).

From a theoretical point of view, this can be achieved by rewriting the covariance targeting constraint as follows:

$$Q_{t+1} = (1 - \alpha - \beta)\hat{E}[z_t z_t'] + \alpha z_t z_t' + \beta Q_t + \gamma(g_t g_t' - \hat{E}[g_t g_t'])$$

where the expectations $\hat{E}$ will be estimated from the sample moments:

$$\hat{E}[z_t z_t'] = \frac{1}{T}\sum_{t=1}^{T}[z_t z_t']$$

while g_t vectors are defined as the negative part of z_t:

$$g_t = \begin{cases} z_{i,t} & \text{if } z_{i,t} < 0 \\ 0 & \text{if } z_{i,t} \geq 0 \end{cases}, \quad i = 1, \dots, N.$$

In summary, the correlation no longer depends on the magnitude of the standardized returns, but on their sign. Indeed, the term $\gamma(g_t g_t' - \hat{E}[g_t g_t'])$ captures the leverage effect at mixed second-order moments: when $\gamma > 0$, the correlation between asset *i* and *j* is greater when the respective standardized returns, $z_{i,t}$ and $z_{j,t}$, are negative than when they are positive. For non-negativity, we always require that $\alpha \geq 0$ and $\beta \geq 0$, further adding that $(\alpha + \gamma) \geq 0$.

Estimating the DCC Model

Adopting the specification discussed so far, DCC models can be estimated by the *Quasi-Maximum Likelihood* (QMLE) method, by construction. The reason is related to that the model is implemented in three different steps and, in each of them, only a few parameters are estimated simultaneously through numerical optimizations (Guidolin, 2018). This aspect is what makes DCC extremely tractable for risk management of large portfolios, unlike other models (Engle, 2002):

- *First stage:* parameter estimation of univariate volatility models.
- *Second stage:* standardization of returns by application of the inferences resulting from the GARCHs of the first stage, $z_{i,t+1} \equiv {}^{R_{i,t+1}}/_{\sigma_{i,t+1}}$.
- *Third stage:* estimation of the dynamic parameters of the conditional correlations, α and β, for all pairs of assets.

Importantly, in the last step, the parameters must be common to all pairs of the time series considered. Although this implies the unrealistic (but necessary) assumption that the persistence of correlations will be the same for all pairs, the level of correlation will still be different whatever the time point under consideration.

Considering the simple case where we have two sets of data, *N = 2*, the log-likelihood function to be maximized in the first step, and thus in the implementation of GARCH(1,1), is as follows:

$$\max_{\omega_i\,\alpha_i\,\beta_i}\left\{-\frac{T}{2}\,log\,2\pi - \frac{1}{2}\sum_{t=1}^{T} log\left[\omega_i + \alpha_i R_{i,t-1}^2 + \beta_i \sigma_{i,t-1}^2\right] - \frac{1}{2}\sum_{t=1}^{T}\frac{R_{i,t}^2}{\omega + \alpha_i R_{i,t-1}^2 + \beta_i \sigma_{i,t-1}^2}\right\}$$

where $\sigma_{i,0}^2$ is initialized as

$$\sigma_{i,0}^2 = \frac{\omega}{1 - \alpha_i - \beta_i}$$

Having standardized the returns in the second step by obtaining the time series pairs $z_{i,t+1} \equiv {R_{i,t+1}}/{\sigma_{i,t+1}}$ and $z_{j,t+1} \equiv {R_{j,t+1}}/{\sigma_{j,t+1}}$, we proceed to maximize the parameters of our interest, α and β:

$$\max_{\alpha,\beta} -\frac{T}{2} \log 2\pi - \frac{1}{2}\sum_{t=1}^{T} \ln\left(1 - \rho_{12,t-1}^2\right) - \frac{1}{2}\sum_{t=1}^{T} \frac{\left(z_{1,t-1}\right)^2 + \left(z_{2,t-1}\right)^2 - 2\rho_{12,t-1} z_{1,t-1} z_{2,t-1}}{\left(1 - \rho_{12,t-1}^2\right)}$$

Thus, from the maximization function we observe that the variables entering the formula are the standardized returns, $z_{i,t}$, and not the original ones. This means that the former are essentially treated as actual observations, which allows us to notice how the QMLE implies a loss of efficiency in parameter estimation in its application within the DCC model. Theoretically, we could obtain more robust parameters by using the *Maximum Likelihood* (ML) estimator to simultaneously estimate all models for volatility and correlation, but this is only possible in cases where the number of assets considered does not exceed 3 or 4 (Guidolin, 2018).

Vector Autoregressive (VAR) Model

The models presented thus far have been developed for the purpose of describing the conditional second moments of quantities of interest. In econometrics, the objective is to elucidate the dynamic behaviour of economic and financial variables. To this end, it is necessary to undertake a detailed study of the stochastic process that generates the individual observed time series, with a view to identifying its properties. However, given the involvement of multiple time series, the scope of the research extends to contemporary and lagged causal relationships between several variables.

In this regard, we introduce the *Vector Autoregressive* (VAR) model (Sims, 1980), a system of simultaneous equations that represents the natural extension of univariate Autoregressive (AR) models to multivariate time-series. The objective of the VAR is to model the conditional mean, which will subsequently be utilised in the implementation of the VAR-DCC-GARCH model.

The Vector Autoregressive model is based on the assumption that, given a vector of time series, there is a linear and dynamic interdependence between the current values, lagged ones and the error terms of each series. In formula, a VAR model of order *p* – VAR(*p*) – can be written as:

$$Z_t = c + \Phi_1 Z_{t-1} + \Phi_2 Z_{t-2} + \ldots + \Phi_p Z_{t-p} + \epsilon_t$$

where ϵ_t is a *white noise* error term, and thus i.i.d., with null expected value, $E(\epsilon_t) = 0$, and independent of lagged Z_t. In its matrix form, also called *"companion-form"*, the expression reduces to the following structure:

$$\begin{bmatrix} Z_t \\ Z_{t-1} \\ \vdots \\ Z_{t-p+1} \end{bmatrix} = \begin{bmatrix} c \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} \Phi_1 & \Phi_2 & \cdots & \Phi_{p-1} & \Phi_p \\ I & 0 & \cdots & 0 & 0 \\ 0 & I & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I & 0 \end{bmatrix} \begin{bmatrix} Z_{t-1} \\ Z_{t-2} \\ \vdots \\ Z_{t-p} \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

with I denoting the identity matrix and Φ_j for $j = 1, \ldots, p$ being matrixes *(n x n)* of coefficients measuring the influence of each variable with its own lag and that of the other variables (Nafisi-Moghadam et al., 2022).

As Sims himself observes, the advantage of the VAR model is that there is no need to specify *a priori* which variables are endogenous and which are exogenous, since all are considered endogenous. The evaluation of all components in conjunction facilitates, firstly, the development of a more parsimonious model that can incorporate a reduced number of lags and, secondly, the generation of more accurate predictions than those produced by so-called *"Simultaneous Equation Models"* (SEM)[4] (Wooldridge, 2010). This is due to the fact that the available set of information also encompasses the history of the other variables.

If we consider the simplest framework of a Vector Autoregressive model, VAR(1),

$$Z_t = c + \Phi_1 Z_{t-1} + \epsilon_t,$$

we note that *Ordinary Least Squares* (OLS) can be applied for its estimation since the set of regressors is the same for each equation. The consistency of the results derives from the independence of the error terms with respect Z_{t-1}, as specified above.

The second aspect that must be mentioned, and which is a disadvantage of the model, is the large number of parameters associated with the model. In particular, if we have *n* equations for *n* variables and *p* lags for each variable in each equation, the result will be $(n + np^2)$ parameters. Furthermore, some parameters in the Φ_1 matrix may be significantly different from zero, even if they do not explain any significant relationship in the data set. For this reason, when estimating VAR matrices,

[4] SEMs are structural models of simultaneous equations describing relationships in which, unlike VAR, the variables can be simultaneously endogenous and exogenous, allowing for two-way randomness. They therefore require identification restrictions for parameter estimation, whereas the VAR needs no further restrictions.

it is common for researchers to impose constraints and set some elements equal to zero *ex-ante* (Guidolin, 2018).

Regarding the choice of the number of lags *p*, the academically accepted practice is to prefer information criteria, such as *Akaike's information criterion* (AIC) and *Bayesian information criterion* (BIC), over autocorrelation functions (ACF) and partial autocorrelation functions (PACF). This approach has also been adopted in this thesis:

$$(M)AIC = ln\widehat{\sigma^2} + 2 \cdot \frac{p+q+1}{T}$$

$$(M)BIC = ln\widehat{\sigma^2} + 2 \cdot \frac{p+q+1}{T} \cdot lnT$$

Finally, it must be specified that the inclusion of a large number of lags may compromise, in part, the power of the test, while an inadequate number of lags may lead the test to be severely biased, resulting in the loss of validity of the asymptotic distributions.

Empirical Results

In the first chapter, we selected a set of stocks representative of the Nasdaq 100 through cluster analysis, based on risk-return and market capitalization parameters. Next, we introduced relevant macroeconomic variables and aggregated them into a single index in order to understand how financial market movements affect asset returns.

In this chapter, we will focus on analysing the dynamic conditional correlations between the chosen assets and the macroeconomic-financial index, accounting for the change in volatility over time. For this goal, we will apply the DCC model to the dataset described in Chapter 1, which consists of 9 variables, one of which is the index consisting of 6 variables weighted by their conditional volatility, and 8 are representative Nasdaq 100 equities. The multivariate time series consists of 1257 time points and covers a time interval of 5 years.

For a start, we calculated the *rolling window correlation* between each stock and the index, using it as a proxy for actual correlations. After that, we compared the results with various models, including DCC-GARCH, DCC-TARCH, DCC-GJR-GARCH, ADCC-GARCH, ADCC-TARCH and ADCC-GJR-GARCH. After selecting the most appropriate model based on the RMSE value, we examined the impact of the COVID-19 pandemic on dynamic correlations, including and excluding those that were the most volatile months for financial markets in 2020.

Finally, we compared the model chosen in the first part of the section, the DCC-GARCH, with its extended version that integrates the dynamics of the conditional mean through the VAR model, the VAR-DCC-GARCH, to determine which was the most statistically relevant and the most suitable for financial risk analysis in the third chapter.

The programming language adopted in this section of work is MATLAB R2023a, employed in conjunction with Kevin Sheppard's[5] *MFE Toolbox* to implement the DCC and ADCC models.

First Comparison: Rolling Window Correlation

As a first step in assessing the forecasting performance of dynamic correlation models, we have calculated the *rolling window correlations* between the macroeconomic-financial index and each asset, using them as a proxy for real correlations. The duration of the window has been set at 90 days, in line with the common practice of quarterly recalibration of portfolio positions. Rolling

[5] Co-author of the DCC model with Engle in 2002.

windows have been calculated on daily log-returns, allowing us to examine and evaluate fluctuations and patterns of correlations over time:

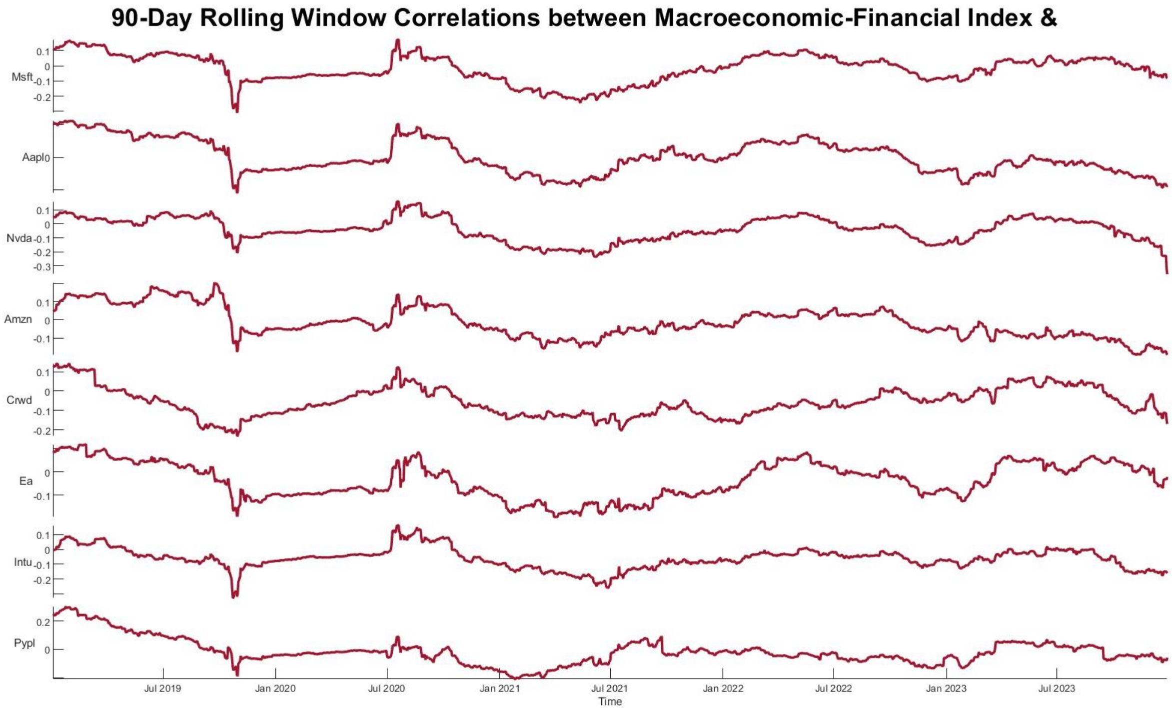


Figure 8. Representation of 90-day rolling window correlations between the macroeconomic-financial index and selected stocks, 2019-2024.

In all plots, the ordinate axes vary between -0.20 to +0.25, indicating that the interdependencies between the stocks and the index oscillate between moderate negative to positive correlations. In general, we observe that the most capitalised stocks (Microsoft, Apple, Nvidia, Amazon) show correlations around zero for longer periods, while the others resulting from the cluster analysis (CrowdStrike, Electronic Arts, Intuit, PayPal) tend to have correlations with frequent changes of sign. Despite this distinction, the set of stocks seems to register indicatively positive correlations over the entire time horizon examined. Furthermore, it can be noted that during the COVID-19 pandemic the variability of correlations is limited. Since the extraction of the rolling window is over 90 days, the correlations observed at the beginning of the crisis still reflect the influence of the pre-crisis period. As one moves forward, new data are replaced by previous one, causing potential changes in the correlations to be gradually incorporated.

Following the calculation of rolling window correlations, we proceeded to implement a series of econometric models for the analysis of dynamic correlations in our dataset. Initially, we focused on symmetric models, such as DCC-GARCH, which capture changes in volatility and correlations over time without considering possible asymmetries in market shocks. Subsequently, we extended the analysis to models that include the leverage effect, both for volatility and correlation, with the purpose of accurately identifying possible asymmetries in the time series:

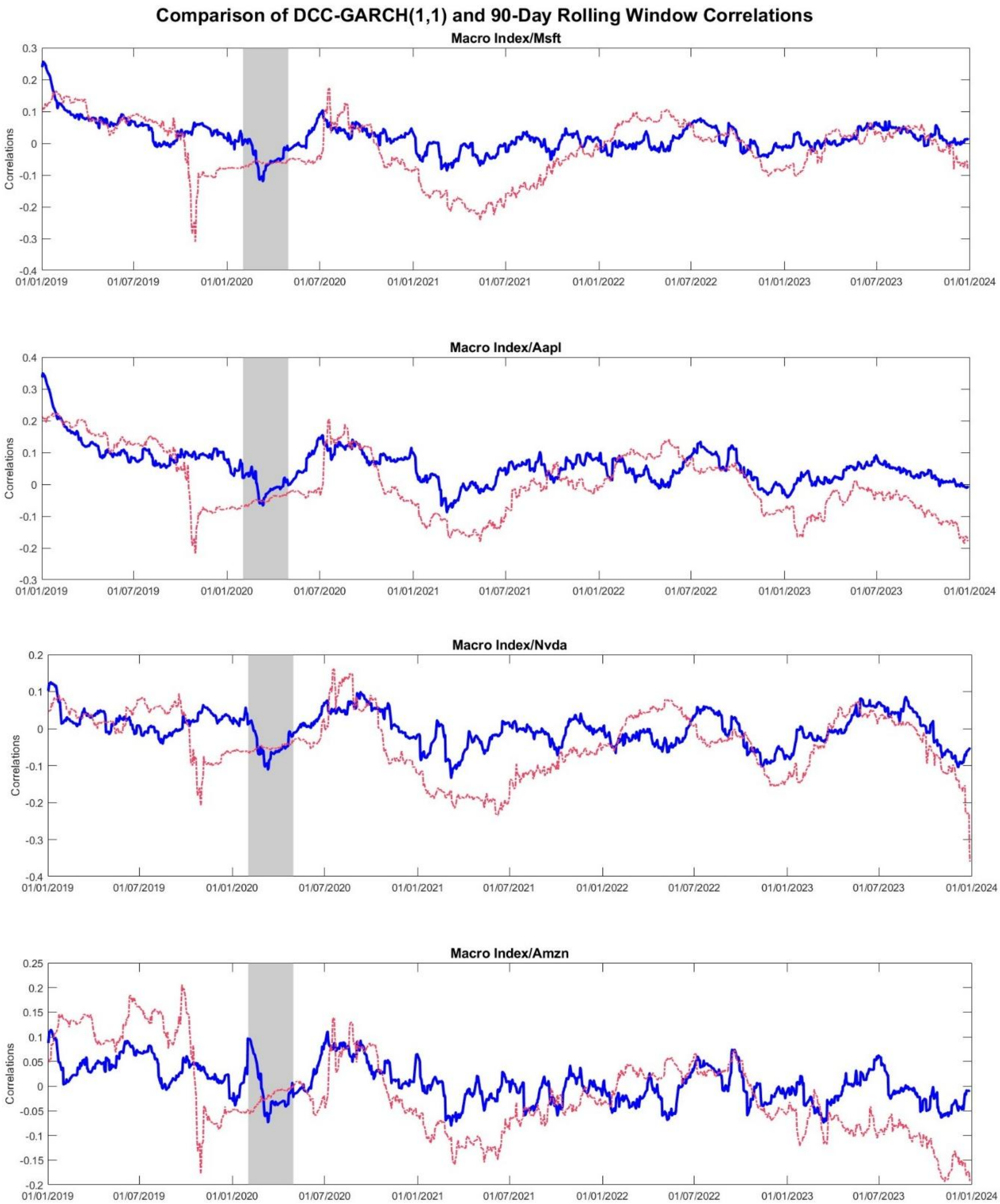
	α *(Alpha)*	β *(Beta)*	γ *(Gamma)*	*LL*	*RMSE*
DCC–GARCH (1,1)	0.0091 *(0.0001)*	0.9701 *(0.0000)*	-	30442.10	0.6074
DCC–TARCH (1,1)	0.0104 *(0.0001)*	0.9628 *(0.0000)*	-	30457.65	0.6211
DCC–GJR–GARCH (1,1)	0.0094 *(0.0001)*	0.9682 *(0.0000)*	-	30438.05	0.6141
ADCC–GARCH(1,1)	0.0012 *(0.4605)*	0.9748 *(0.0000)*	0.0224 *(0.0000)*	30478.12	0.6317
ADCC–TARCH (1,1)	0.0020 *(0.2634)*	0.9632 *(0.0000)*	0.0291 *(0.0000)*	30500.04	0.6346
ADCC–GJR–GARCH (1,1)	0.0015 *(0.8027)*	0.9696 *(0.0000)*	0.0259 *(0.0000)*	30479.19	0.6331

Table 3. Comparison of parameter estimates of the DCC and ADCC models.

The table shows the estimated parameters of the dynamic correlations for each multivariate model, together with their *p-values* (indicated in parentheses), log-likelihood (LL) and *root mean square error* (RMSE), used to compare the effectiveness and performance of each model. We can identify two main groups: the DCC and ADCC models.

In the first group, DCC-GARCH(1,1) stands out as the best performing, with an RMSE of 0.6074 and highly significant α and β, as indicated by the extremely low p-values. DCC-TARCH(1,1) and DCC-GJR-GARCH(1,1) also exhibit significant parameters, but slightly higher RMSE. In contrast, ADCC models, which incorporate asymmetry into the dynamic correlations via the γ parameter, tend to show lower performance than DCC. Although γ is always significant, the parameter α, which measures the impact of past innovations, is not significant at any level, as indicated by the high p-values.

Assessing the goodness of each model using the RMSE, we can see that the value of the statistical measure tends to increase as we refer to more and more specific models. Specifically, the former group of models tends to have lower RMSE values than the latter one, indicating that the DCC models perform better in terms of prediction accuracy. In view of the results obtained, we have chosen to use the DCC-GARCH(1,1) model, as it has the lowest RMSE of all. Below we have plotted the dynamic pairwise correlations estimated by the DCC-GARCH model for visual comparison with the previously computed rolling window correlation:


Comparison of DCC-GARCH(1,1) and 90-Day Rolling Window Correlations
Macro Index/Msft
Macro Index/Aapl
Macro Index/Nvda
Macro Index/Amzn
Correlations
01/01/2019
01/07/2019
01/01/2020
01/07/2020
01/01/2021
01/07/2021
01/01/2022
01/07/2022
01/01/2023
01/07/2023
01/01/2024

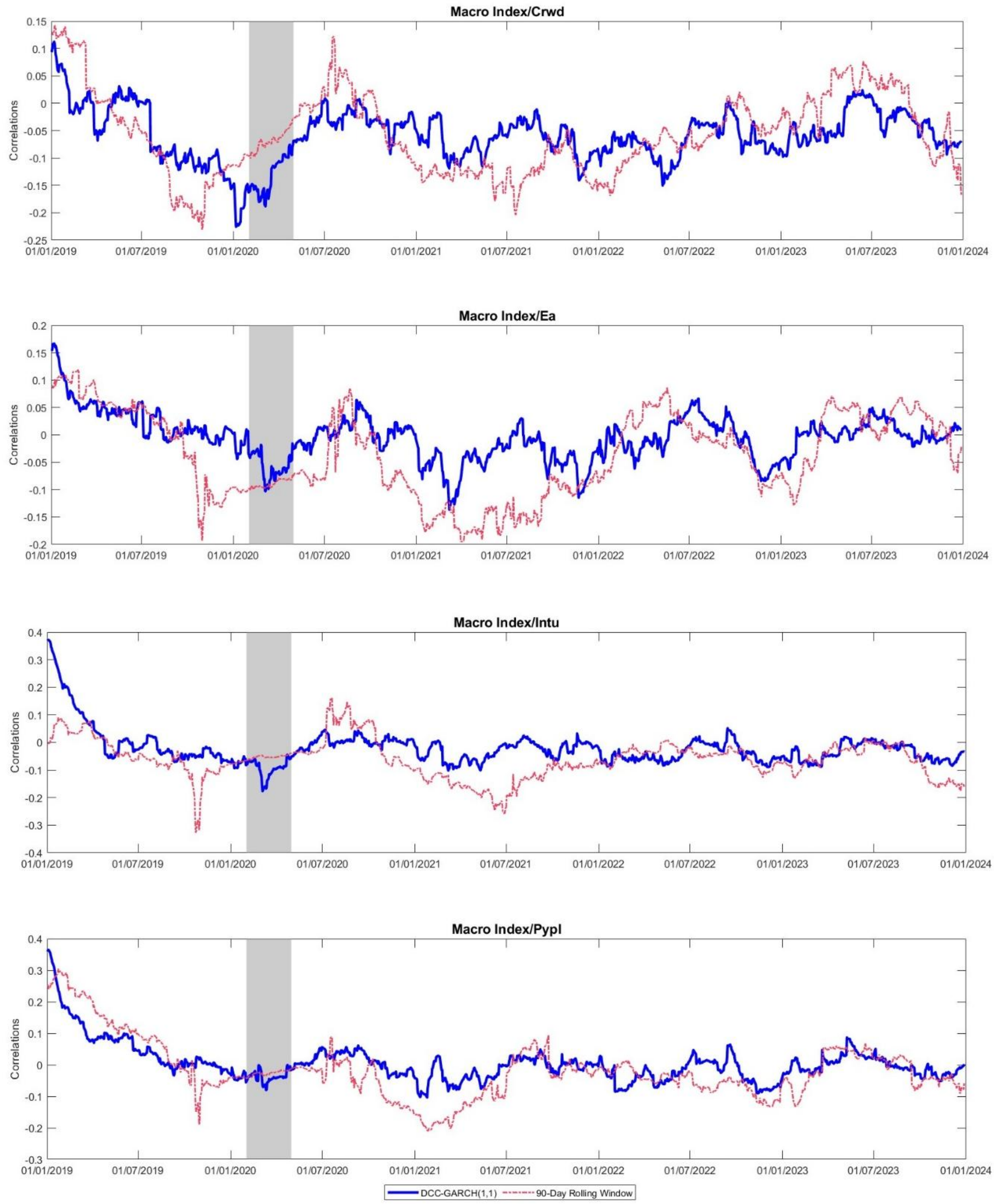
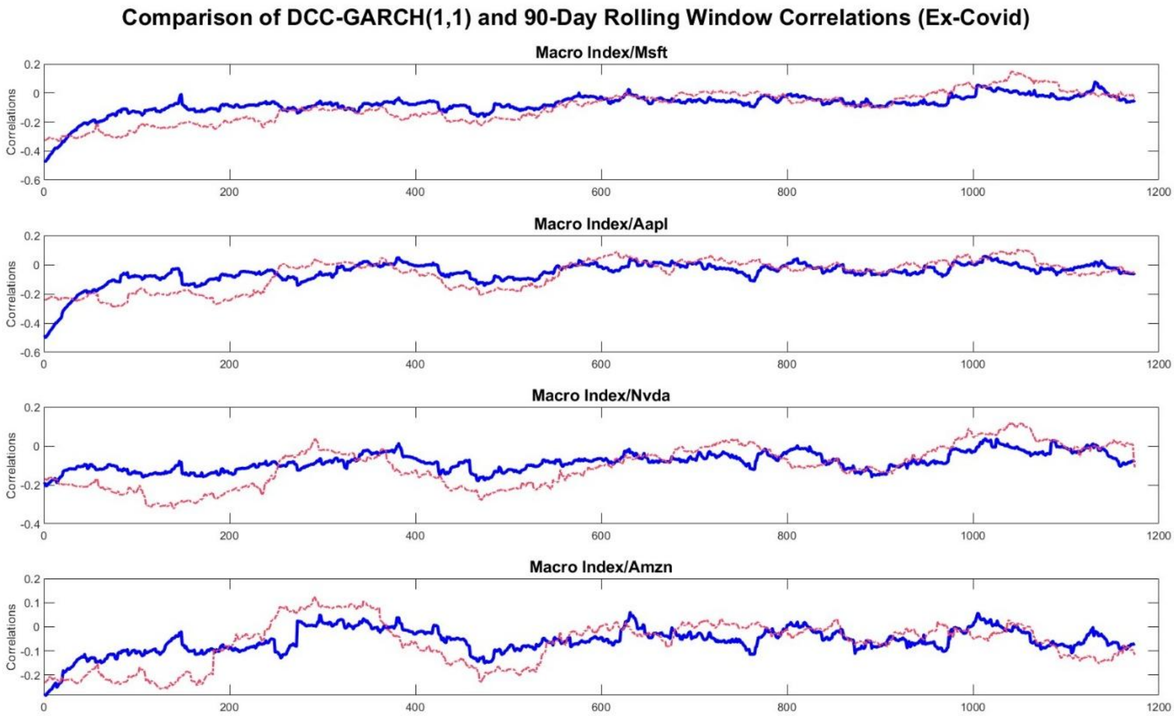


Figure 9. Comparison of estimated correlations via DCC-GARCH and rolling window between the index and selected stocks, 2019 to 2024.

In each of the 8 graphs, we can appreciate how the DCC-GARCH captures temporal variations in correlations more gradually than the rolling window method, which instead shows greater fluctuations. In fact, the blue lines present a behaviour that probably reflects the model's ability to adapt rapidly to market changes transmitted by the macroeconomic-financial index employed, with less pronounced but more regular variations. In contrast, the pink lines show more irregular and reactive changes, but less consistent over time. Furthermore, we note how the range of values of the correlations is generally between -0.20 and +0.30 in both cases, with only a few exceptions recording higher values.

Differently from the rolling window correlation, in the DCC it is not possible to immediately identify the precise impact of specific events, such as the 2020 pandemic, on the US stock market (and not only). This is because dynamic correlations show continuous variations over time. Nevertheless, if we look at the performance of the U.S. Nasdaq 100 index over the last 5 years, we can see how COVID-19 effectively caused marked volatility especially in the first 3 months of the pandemic, with the worst performance – by almost 30% – in the month between 14 February and 20 March 2020.

In view of the above, we decided to compare the dataset we have used so far with the same one, excluding, however, the months of February, March and April 2020 (Ex-Covid), obtaining the following results:



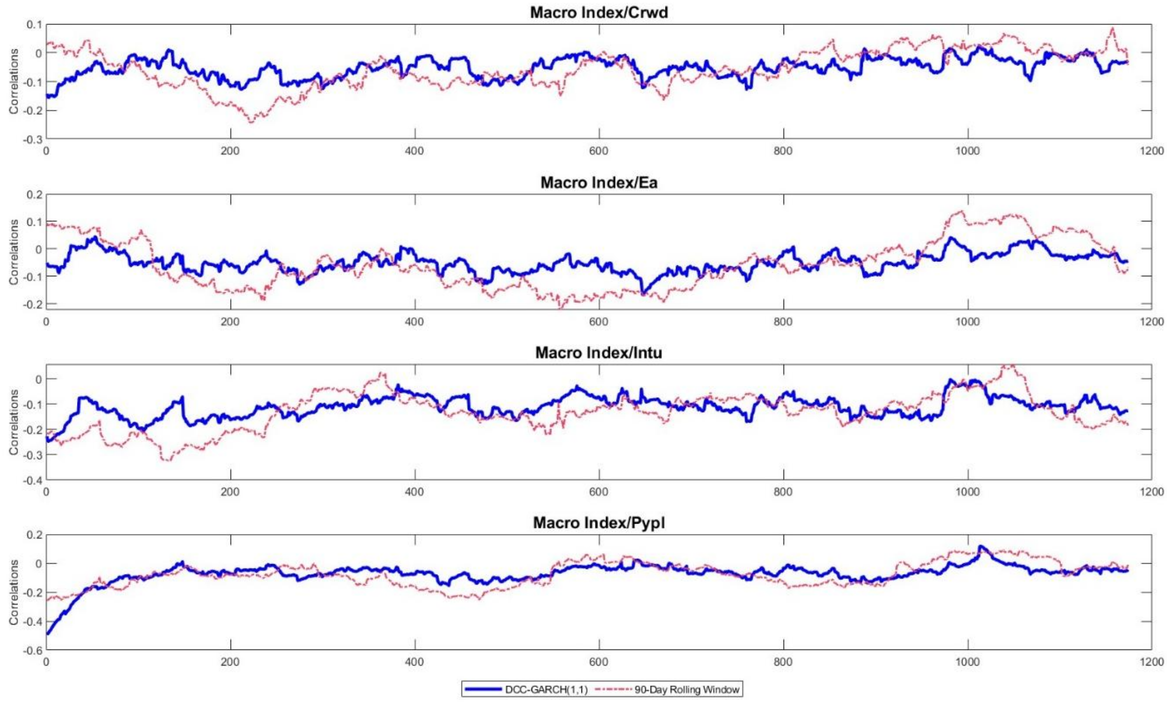


Figure 10. Comparison of estimated correlations via DCC-GARCH and rolling window between the index and selected stocks, from 2019 to 2024 and without considering the COVID-19 pandemic months.

In a scenario without considering the initial months of the COVID-19, the dynamic correlations seem to show a smoother generalised trend, in relative terms. While still maintaining an evident volatility, the DCC produces fewer regime shifts between positive and negative correlation, thus indicating a more uniform evolution precisely due to the absence of an exogenous, relevant and significant shock such as, precisely, COVID-19. This can be corroborated by the behaviour of the rolling window correlation which, in this further comparison, seems to show a more faithful alignment than the DCC curves.

Second Comparison: VAR-DCC-GARCH

Following the initial comparison with the rolling window correlation and the estimation of several models, we selected DCC-GARCH(1,1) as the most suitable, in light of its RMSE value. To ensure a more robust and in-depth analysis, we deemed it appropriate to perform an additional comparison with VAR-DCC-GARCH, working with the dataset comprising the most volatile months of the pandemic, in order to determine which of the two models is more effective and appropriate for the market risk analysis developed in Chapter Three.

To proceed with the estimation of the Vector Autoregressive (VAR) model, we first checked for the presence (or absence) of *unit root* – and thus stationarity – in our 5-year log-return dataset. By running a *for* loop on all 9 variables and applying MATLAB's *adftest* function, we verified the presence of unit root in all the time series considered:

	H_0	*p-value*	*Test Statistics*
Macro Index	1	0.0010	-31.6034
Msft	1	0.0010	-42.8367
Aapl	1	0.0010	-40.4403
Nvda	1	0.0010	-38.4510
Amzn	1	0.0010	-36.7194
Crwd	1	0.0010	-34.3158
Ea	1	0.0010	-38.1814
Intu	1	0.0010	-39.5335
Pypl	1	0.0010	-37.4011

Table 4. Results of Augmented Dickey-Fuller (ADF) statistical test for stationarity of time series.

The ADF result shows p-values below any α significance level, suggesting that the VAR can be applied without any problems to model the conditional mean of the multiple time series.

As described in the section on VAR theory and as suggested by the same MATLAB[6], we estimated the Vector Autoregressive model by choosing a number of delays equal to the number identified by the BIC, i.e. 1:

	Value	*S.E.*	*Test Statistic*	*p-value*
Constant(1)	**0.003402**	**0.000296**	**11.475245**	**0.000000**
Constant(2)	**0.001411**	**0.000561**	**2.515473**	**0.011887**
Constant(3)	**0.001748**	**0.000595**	**2.939494**	**0.003287**
Constant(4)	**0.002572**	**0.000953**	**2.697617**	**0.006984**
Constant(5)	0.000529	0.000658	0.803245	0.421833
Constant(6)	0.001543	0.001164	1.324954	0.185186
Constant(7)	0.000531	0.000555	0.956484	0.338828
Constant(8)	**0.001157**	**0.000690**	**1.676710**	**0.093599**
Constant(9)	0.000098	0.000827	0.118183	0.905923
AR{1}(1,1)	0.005448	0.028285	0.192593	0.847278
AR{1}(2,1)	-0.024538	0.053503	-0.458626	0.646503
AR{1}(3,1)	-0.064044	0.056725	-1.129019	0.258890
AR{1}(4,1)	-0.018597	0.090964	-0.204448	0.838003
AR{1}(5,1)	0.046569	0.062827	0.741228	0.458555
AR{1}(6,1)	-0.046316	0.111077	-0.416970	0.676700

[6] https://it.mathworks.com/help/econ/model-specification-structures.html.

AR{1}(7,1)	0.010446	0.052968	0.197219	0.843656
AR{1}(8,1)	-0.005957	0.065851	-0.090456	0.927925
AR{1}(9,1)	-0.028493	0.078864	-0.361293	0.717880
AR{1}(1,2)	-0.001904	0.027943	-0.068134	0.945679
AR{1}(2,2)	**-0.143909**	**0.052856**	**-2.722688**	**0.006475**
AR{1}(3,2)	**-0.122437**	**0.056039**	**-2.184844**	**0.028900**
AR{1}(4,2)	**-0.176275**	**0.089863**	**-1.961589**	**0.049810**
AR{1}(5,2)	0.054233	0.062067	0.873779	0.382239
AR{1}(6,2)	0.004243	0.109733	0.038666	0.969157
AR{1}(7,2)	-0.054841	0.052327	-1.048039	0.294621
AR{1}(8,2)	-0.060923	0.065054	-0.936490	0.349021
AR{1}(9,2)	-0.060216	0.077909	-0.772899	0.439582
AR{1}(1,3)	0.015650	0.022811	0.686092	0.492655
AR{1}(2,3)	**-0.101623**	**0.043148**	**-2.355228**	**0.018511**
AR{1}(3,3)	**-0.112878**	**0.045747**	**-2.467445**	**0.013608**
AR{1}(4,3)	**-0.152500**	**0.073359**	**-2.078820**	**0.037634**
AR{1}(5,3)	**-0.156609**	**0.050667**	**-3.090922**	**0.001995**
AR{1}(6,3)	**-0.207462**	**0.089579**	**-2.315969**	**0.020560**
AR{1}(7,3)	**-0.114144**	**0.042717**	**-2.672105**	**0.007538**
AR{1}(8,3)	**-0.137744**	**0.053106**	**-2.593755**	**0.009493**
AR{1}(9,3)	**-0.164011**	**0.063600**	**-2.578779**	**0.009915**
AR{1}(1,4)	-0.009791	0.013453	-0.727821	0.466723
AR{1}(2,4)	**0.052401**	**0.025446**	**2.059280**	**0.039467**
AR{1}(3,4)	**0.045806**	**0.026979**	**1.697819**	**0.089542**
AR{1}(4,4)	0.040760	0.043263	0.942150	0.346116
AR{1}(5,4)	0.028593	0.029881	0.956882	0.338627
AR{1}(6,4)	-0.049196	0.052829	-0.931234	0.351733
AR{1}(7,4)	0.037665	0.025192	1.495111	0.134885
AR{1}(8,4)	**0.066314**	**0.031319**	**2.117349**	**0.034230**
AR{1}(9,4)	-0.009887	0.037508	-0.263605	0.792084
AR{1}(1,5)	-0.009207	0.018646	-0.493761	0.621475
AR{1}(2,5)	0.057383	0.035269	1.626987	0.103740
AR{1}(3,5)	**0.070404**	**0.037394**	**1.882765**	**0.059732**
AR{1}(4,5)	**0.171194**	**0.059964**	**2.854949**	**0.004304**
AR{1}(5,5)	0.030089	0.041416	0.726520	0.467520
AR{1}(6,5)	**0.179753**	**0.073222**	**2.454893**	**0.014093**
AR{1}(7,5)	0.042538	0.034917	1.218250	0.223129
AR{1}(8,5)	0.046212	0.043409	1.064562	0.287074
AR{1}(9,5)	**0.102574**	**0.051987**	**1.973056**	**0.048489**
AR{1}(1,6)	-0.001573	0.008378	-0.187765	0.851061

AR{1}(2,6)	0.022831	0.015846	1.440746	0.149657
AR{1}(3,6)	0.021855	0.016801	1.300822	0.193319
AR{1}(4,6)	**0.060642**	**0.026942**	**2.250869**	**0.024394**
AR{1}(5,6)	**0.032982**	**0.018608**	**1.772467**	**0.076317**
AR{1}(6,6)	**0.054262**	**0.032899**	**1.649386**	**0.099069**
AR{1}(7,6)	**0.039654**	**0.015688**	**2.527675**	**0.011482**
AR{1}(8,6)	0.015792	0.019504	0.809674	0.418128
AR{1}(9,6)	0.024212	0.023358	1.036569	0.299937
AR{1}(1,7)	-0.014254	0.016959	-0.840490	0.400634
AR{1}(2,7)	**-0.059553**	**0.032078**	**-1.856519**	**0.063380**
AR{1}(3,7)	**-0.056641**	**0.034010**	**-1.665410**	**0.095831**
AR{1}(4,7)	-0.070852	0.054538	-1.299125	0.193901
AR{1}(5,7)	**-0.080093**	**0.037668**	**-2.126273**	**0.033481**
AR{1}(6,7)	-0.025084	0.066597	-0.376652	0.706432
AR{1}(7,7)	-0.035572	0.031757	-1.120108	0.262668
AR{1}(8,7)	-0.015621	0.039481	-0.395644	0.692368
AR{1}(9,7)	-0.020719	0.047283	-0.438180	0.661256
AR{1}(1,8)	0.023597	0.020163	1.170326	0.241870
AR{1}(2,8)	**-0.067667**	**0.038138**	**-1.774270**	**0.076019**
AR{1}(3,8)	-0.008540	0.040435	-0.211203	0.832729
AR{1}(4,8)	**-0.153136**	**0.064842**	**-2.361701**	**0.018191**
AR{1}(5,8)	-0.059970	0.044785	-1.339074	0.180546
AR{1}(6,8)	-0.057989	0.079178	-0.732380	0.463937
AR{1}(7,8)	-0.039622	0.037757	-1.049391	0.293998
AR{1}(8,8)	**-0.090525**	**0.046940**	**-1.928521**	**0.053790**
AR{1}(9,8)	-0.006307	0.056216	-0.112183	0.910678
AR{1}(1,9)	-0.015128	0.013982	-1.081943	0.279278
AR{1}(2,9)	-0.006844	0.026447	-0.258776	0.795808
AR{1}(3,9)	-0.009140	0.028040	-0.325950	0.744462
AR{1}(4,9)	-0.004128	0.044965	-0.091801	0.926856
AR{1}(5,9)	0.008468	0.031056	0.272662	0.785113
AR{1}(6,9)	0.015494	0.054907	0.282185	0.777802
AR{1}(7,9)	-0.013006	0.026183	-0.496730	0.619380
AR{1}(8,9)	0.001430	0.032551	0.043930	0.964960
AR{1}(9,9)	-0.005940	0.038983	-0.152380	0.878887

Table 5. Results of the VAR(1) model estimation for the conditional mean.

The table presents the estimated parameters of the VAR(1) model for each variable in our dataset, including constant terms and autoregressive parameters of order 1. Remember that the variables analysed are, in order, the macroeconomic-financial index, Msft, Aapl, Nvda, Amzn, Crwd, Ea, Intu and Pypl. We have also highlighted the cells that contain relevant p-value estimates.

The coefficients, *AR{1}(i,j)*, are read as follows:

- *i* represents the dependent variable, i.e. the one that is explained by the delay;
- *j* represents the explanatory variable, i.e. the one that influences *i*.

In other words, *AR{1}(i,j)* measures the influence of the lagged value of *j* on variable *i*.

Accordingly to what is described in the theoretical part, we provide below some examples to facilitate the understanding of the table and the interpretation of the results. For instance, the equation describing the influence of the lagged value of the macroeconomic index on itself is as follows:

$$Macro\_Index_t = Constant(1) + AR\{1\}(1,1)Macro\_Index_{t-1} + \epsilon_t,$$

$$Macro\_Index_t = 0.003402_{p-value=0.000000} + 0.005448_{p-value=0.847278}Macro\_Index_{t-1} + \epsilon_t$$

where:

- 0.003402 is the value of the constant for the index with a null p-value, 0.000000, suggesting that the fixed effect on the model is statistically significant and probably plays a relevant role in determining the value at time *t* of the macroeconomic index;
- on the opposite, the estimated coefficient 0.005448 does not indicate a direct relationship between the lag of the index and its value at time *t*. This is confirmed by the non-significant p-value, 0.847278.

Another example concerns the influence of the lag of Aapl on Amzn:

$$Amzn_t = Constant(5) + AR\{1\}(5,3)Aapl_{t-1} + \epsilon_t$$

$$Amzn_t = 0.000529_{p-value=0.421833} - 0.156609_{p-value=0.001995}Aapl_{t-1} + \epsilon_t$$

where:

- 0.000529 is the value of the constant for Amzn, with a very high p-value, 0.421833, indicating that the constant is not influential;
- -0.156609 is the estimated coefficient measuring the negative effect of the lagged value of Aapl on the current value of Amzn, which is associated with a very low p-value of 0.001995. Since the latter is statistically significant, the coefficient estimate suggests that, if the lagged value of Aapl increases, we expect the lagged value of Amzn to decrease in the following period.

In general, we observe that the constant terms for each variable have very small values. The associated p-values vary: some are below the commonly adopted significance levels (0.1, 0.05, 0.01), indicating that in these cases the constants are statistically different from zero and, so, contribute to the explanation of the dynamics of the model. Other p-values are, however, higher, suggesting that the constants are not statistically significant. Regarding the autoregressive coefficients, some relationships appear weak or not very influential, while others show significant interactions, both positive and negative. In some cases, the relationships are significant only around the 10% level, suggesting that there are *potentially* important effects between the variables.

In light of these findings, we can say that VAR(1) offers a detailed basis for the in-depth analysis of the dynamics between the variables. Therefore, using the residuals of the model, we implemented VAR(1)-DCC-GARCH(1,1) to test whether the dynamic correlations observed for the DCC-GARCH(1,1) model undergo relevant changes once the autoregressive relationships between the variables are integrated.

The results of both models have been represented graphically below for direct comparison:

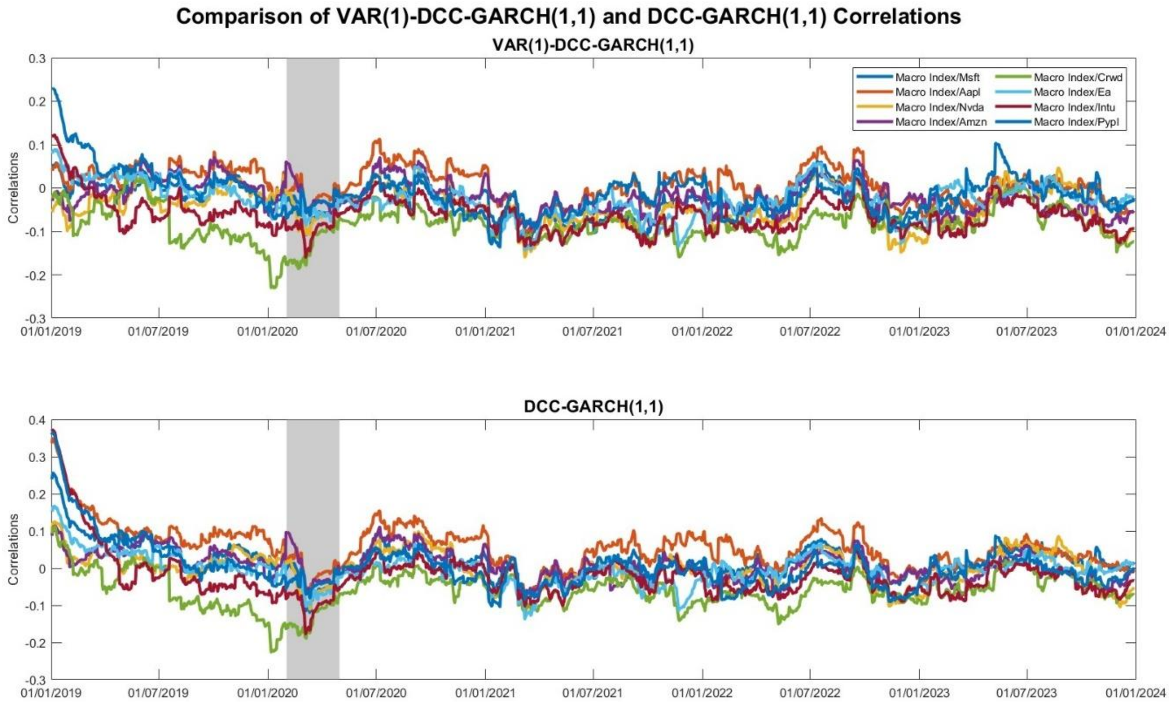


Figure 11. Comparison of estimated correlations via VAR-DCC-GARCH and DCC-GARCH between the index and selected stocks, 2019 to 2024.

We observe that the dynamic correlations obtained by the VAR(1)-DCC-GARCH(1,1) model are almost similar to those of DCC-GARCH(1,1), with slight differences limited to the starting correlations. Nevertheless, the VAR(1)-DCC-GARCH(1,1) model is preferred, as demonstrated by the *Likelihood Ratio Test* (Lrtest), which compares the likelihood estimates of the two models:

Likelihood Ratio Test

H_0	1
p-value	0

Table 6. Inference by Likelihood Ratio Test for model selection.

where we have the following results:

- *$H_0 = 1$,* indicates that the VAR-DCC-GARCH model is favoured over DCC-GARCH;
- *$\alpha = 0.05$*, default value;
- *p-value = 0,* being below the significance level, suggests strong evidence for the rejection of the null hypothesis;
- *degrees of freedom,* corresponding to the number of lags, equal to 1.

To provide a comprehensive picture, we report below the parameter estimates of the models we have just compared, together with a comparison excluding the months of the pandemic.

	***α* (Alpha)**	***β* (Beta)**
DCC–GARCH (1,1)	0.0091 *(0.0001)*	0.9701 *(0.0000)*
VAR(1)–DCC–GARCH (1,1)	0.0089 *(0.0024)*	0.9702 *(0.0000)*

Table 7. Comparison of parameter estimates of the DCC-GARCH and VAR-DCC-GARCH models.

The table demonstrates that the estimated α and β parameters are very similar between the two models. Moreover, as already noted in Table 3, both conditions of stationarity and non-negativity are fulfilled, as the parameters of the VAR-DCC-GARCH are positive and their sum is less than 1. However, although the VAR-DCC-GARCH model shows similar estimates to the DCC-GARCH, it proves to be statistically better at representing the temporal dynamics of the correlations between variables, as confirmed above by the Likelihood Ratio Test.

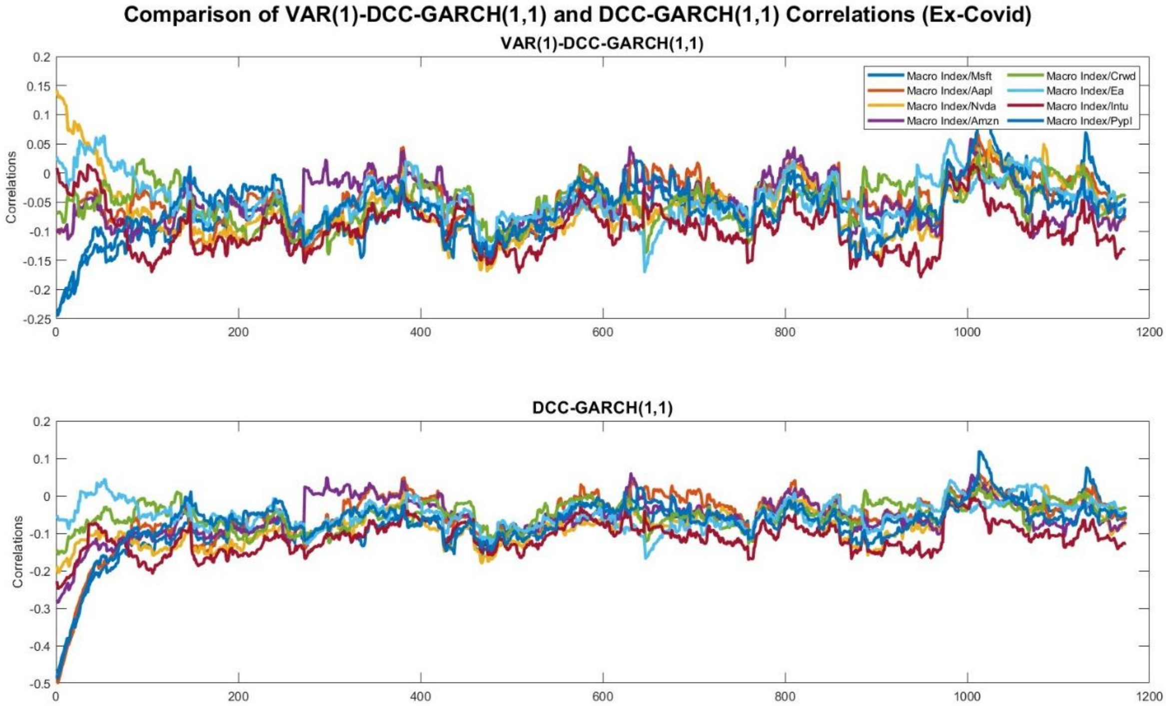


Figure 12. Comparison of estimated correlations via VAR-DCC-GARCH and DCC-GARCH between the index and selected stocks, from 2019 to 2024 and without considering the COVID-19 pandemic months.

As observed in the previous section, also in this comparison, excluding the initial months of the pandemic, the dynamic correlations between the VAR-DCC-GARCH and DCC-GARCH models show a relatively more regular pattern. Differences in the initial correlations are again present but are less pronounced than in the comparison that included the pandemic period. The absence of a relevant exogenous shock seems to reduce the instability in the correlations, favouring a more stable and homogeneous evolution.

Considering these results, the VAR-DCC-GARCH model will be used in the next chapter for an advanced financial risk analysis. In detail, by exploiting the dynamic conditional correlation matrix, key indicators such as *Value at Risk* (VaR) and *Expected Shortfall* (ES) will be calculated, through an approach that will allow for a more dynamic and precise assessment of market risk.

Value at Risk and Expected Shortfall

There are areas in finance where correlations are fundamental inputs. By analysing them through models such as MV-GARCH, it is possible to apply the results in sectors like asset allocation and risk management.

In the context of *Modern Risk Management*, risk measurement plays a crucial role. For instance, the set of regulations governing banks' and investment institutions' exposure to market risk (BCBS, 2013) requires these entities to hold a buffer amount of capital to ensure solvency in case of unexpected losses. Similarly, risk measures are frequently adopted by management as tools to identify the amount of uncertainty associated with an investment, determining whether to accept such exposure and aiming to limit its negative effects. The indicators we refer to are mainly *Value at Risk* (VaR) and *Expected Shortfall* (ES).

In finance, there are several methodologies for estimating the above measures and constructing the relevant probability distribution. These include the variance-covariance method, the historical simulation method and Monte Carlo simulation-based methods. Here, we will apply the former approach to employ the dynamic conditional correlation matrix derived from the VAR-DCC-GARCH model. The chosen criterion framework assumes that the market factors underlying the model all follow a normal distribution. Consequently, the probability distribution of gains and losses will be a linear combination of the distributions of the underlying factors.

The normality verification is also of fundamental importance also because of its close connection to the model implemented by Markowitz in 1952, *Mean-Variance Optimization* (MVO). According to this model, a portfolio is considered efficient if it provides the highest possible return for a given level of risk or, alternatively, if it provides the least risk for a given level of profitability. The overall risk is determined by the covariances among all assets in the portfolio.

When parametric distributions are adopted, there is a significant distinction regarding how the data with which one works are considered and treated:

i. *unconditional,* returns in each period are independent, identically distributed (i.i.d.) causal variables. In this approach, the distribution *adapts* to the behaviour of the variables in question without regard to past information;
ii. *conditional,* time series of financial returns depend on the set of past information, I_{t-1}, and include the current one, I_t. The data are thus adjusted by considering, for example, changes in volatility or regressing them against variables that affect them.

Operating in the field of risk management, it is common to focus on the distribution of losses in a portfolio, rather than on overall returns. The reasons are essentially related to the fact that losses are the object of interest in risk management. It is natural, therefore, to base a measure of risk on their distribution. Moreover, the concept of loss distribution makes sense at all levels of aggregation, that is, both for a portfolio consisting of a single instrument and for the trading book of a financial institution. If estimated appropriately, loss distribution adequately reflects the principle of portfolio diversification (McNeil et al., 2005).

In view of this, we can say that unconditional techniques are often referred to as *static* risk management and are of particular interest if the time horizon over which we want to measure losses is relatively large, as is the case for credit risk and insurance. In contrast, conditional techniques are known as *dynamic* risk management and are particularly relevant for market risk.

Clearly, based on the models and stylized facts being considered so far – heteroscedasticity, volatility clustering, presence of asymmetric effects in volatility and correlation – the parametric conditional approach will be applied in this thesis.

Measures of Risk

Value at Risk (VaR)

One of the most widely utilised instruments by financial institutions, which was incorporated into the Basel II (2008) and III (2013) capital adequacy requirements, is Value at Risk. VaR answers the question *'How much is it possible to lose with a probability of X% over a given time horizon?'*, thus quantifying the maximum potential fluctuation in the value of a financial portfolio, depending on a given probability and a fixed time horizon (J.P. Morgan, 1996). In general, it corresponds to the following definition:

Definition *Given a certain confidence level* $\alpha \in (0, 1)$ *and a certain time horizon, the VaR of a portfolio at the* α *level is given by the smallest value l such that the probability that the loss L exceeds l is not greater than (1 –* α*):*

$$VaR_\alpha = \inf\{l \in \mathbb{R} : P(L > l) \leq 1 - \alpha\} = \inf\{l \in \mathbb{R} :\ F_L(l) \geq \alpha\}$$

In probabilistic terms, the VaR is simply a quantile F_L^{-1} of the normal distribution of loss L, with mean μ_L and standard deviation σ_L:

$$VaR_\alpha = F_L^{-1}(\alpha) = q_\alpha(L) \sim N(\mu_L\,, \sigma_L^2),$$

so:

$$VaR_\alpha = q_\alpha(L) = \mu_L + \sigma_L \cdot q_\alpha(z)$$

Regarding α, the Basel Committee standardizes the use of VaR at the 99% confidence level for market risk, at 95% for banks' trading floor, daily (BCBS, 2013). Typically, as with all risk measures, Value at Risk grows with the confidence interval chosen, as well as depending on the time horizon under consideration. Therefore, if we have a 95% confidence level, we define z_α as that value of the standard normal that will lie on the left-hand side of the probability density function and have a value of -1.645. For an interval equal to 99%, z_α will be -2.326.

Coherence and convexity of risk measures

Below, we report the properties that a risk measure, ρ, should have for being considered coherent (Artzner et al., 1999), and thus mathematically correct:

(i) *Monotonicity (M):* $X \geq Y \longrightarrow \rho(X) \geq \rho(Y)$
If the loss of security *X* is always greater than that of security *Y*, then the risk of *X* must always exceed that of *Y*.

(ii) *Subadditivity (S):* $\rho(X+Y) \leq \rho(X) + \rho(Y)$
The overall risk of a portfolio consisting of two securities cannot be greater than the sum of the risk of each individual security, thus ensuring the diversification principle of Markowitz Theory.

(iii) *Positive Homogeneity (PH):* for any $a > 0,\ \rho(aX) = a\rho(X)$
If the size of the security directly affects the risk, then the risk increases proportionally to the exposure.

(iv) *Translation Invariance (T):* $\rho(X+k) = \rho(X) + k, \forall\, k$
Adding a certain loss, *k*, requires increasing the capital by the same amount.

The popularity of VaR essentially results from its easy implementation and understanding because it allows market risk to be quantified by calculating a single numerical value over a given period of time and for a certain level of confidence (Barone Adesi et al., 1999).

However, there is now consensus in the academic field that VaR cannot be considered a coherent, convex measure of risk because it does not satisfy all the necessary properties, particularly axiom *(ii)* (Danìelsson et al., 2005). The main problem is that VaR represents a quantile of the loss distribution and not an expectation. In other words, it does not provide information about the severity of losses beyond the chosen confidence level, thus failing to provide a complete view of the overall risk of a portfolio.

To fully understand what has just been described, let us introduce the concept of *risk surface*, which we can define as a mapping that associates the weights of a portfolio with its risk. Suppose we have a portfolio with weights equal to w_i and we denote its profit and loss as

$$\pi(\vec{w}) = \sum_i w_i X_i$$

where $\vec{w}$ is the vector of portfolio weights, X_i represents the returns on the portfolio assets.

Having chosen a risk measure $\rho(X)$, we define the following function as the risk surface of the portfolio:

$$\vec{w} \to \rho\big(\pi(\vec{w})\big)$$

which maps the vector of portfolio weights onto the portfolio's risk, $\rho\big(\pi(\vec{w})\big)$. In other words, the risk surface is a representation of the relationship between the weights and the overall risk of the portfolio, as assessed by a specific risk measure, ρ.

A natural result of this definition is that if ρ is convex, then the respective risk surface will also be convex. Therefore, if we have a portfolio with two securities, $W = X + Y$, the overall risk of the portfolio will be less than (or equal to) the sum of the risks of the two individual securities that comprise it:

$$\rho(\pi(\alpha w_i(1-\alpha)w_2 \leq \alpha\rho\big(\pi(w_1)\big) + (1-\alpha)\rho(\pi(w_2))$$

In this regard, VaR is unable to ensure this result, producing a total portfolio measure greater than that resulting from the sum of individual securities, violating the axiom of subadditivity:

$$\rho(X+Y) > \rho(X) + \rho(Y)$$

In conclusion, VaR produces *multiple fake local minima*, which prevents it from being considered a convex (Rockafellar, 1970) – and thus coherent – measure because it fails to adequately model risk. The problem persists even with large sample sizes, making the Central Limit Theorem (TLC) ineffective. In fact, *local minima* do not tend to decrease as sample size *t* increases (Acerbi, 2007).

All this can have serious consequences for risk models. First, a financial institution may not be adequately covered in terms of regulatory capital requirements. Second, there is a risk of making sub-optimal investment decisions (Danìelsson et al., 2005). For this reason, there was a need to identify a measure that could satisfy the properties defined by (Artzner et al., 1999) and overcome the limitations of VaR.

Expected Shortfall (ES)

Expected Shortfall is a coherent alternative to VaR and, like the latter, is a measure that is based on the distribution of losses and answers the question *'What is the expected (average) loss incurred in the $\alpha\%$ worst cases of our portfolio?'*. In other words, ES not only considers losses that exceed the $\alpha\%$ threshold, but also estimates the average of these extreme losses, providing a more comprehensive information setup of the associated risk.

Several authors have contributed to the definition of this risk measure, configuring it under different names, such as *"Conditional Value at Risk"* (CVaR), *"Worst Conditional Expectation"* (WCE), *"Beyond VaR"*, and others. For instance, Rockafellar and Uryasev (2002) demonstrated the consistency of CVaR, as well as its significant advantages over traditional VaR. Kibzun and Kuznetsov (2006) argued that, under normal conditions, ES is a function that satisfies convexity. Peña (2002), on the other hand, highlighted how VaR and CVaR are two measures of risk that complement each other since, while the objective of the former is to control market risk under normal conditions, that of the latter is to check risks under extreme conditions.

Regarding the ES, its theory is in fact closely related to that of the predecessor *Tail Conditional Expectation* (TCE) (Artzner et al., 1997), which corresponds to the expected value of losses, *L*, in the event of a loss greater than the VaR. Mathematically, it is defined as

$$TCE_\alpha(L) = -E[L \mid L \leq q_\alpha(L)]$$

However, subadditivity is guaranteed only in the case where the loss distribution *L* is continuous in a neighbourhood of the quantile q_α, whereas it may violate the property in the case of general (or discrete) distributions. The problem is that the probability that *L* is less than (or equal to) the quantile does not correspond to α itself: it is usually higher in the case of continuous distributions, thus not reflecting the actual risk accurately (Acerbi et al., 2002):

$$P[L < q_\alpha(L)] \neq \alpha$$

Going beyond the demonstration carried out by the authors, let us introduce the first ES formula that Acerbi and Tasche provided in their 2002 paper:

$$ES_\alpha(L) = -\frac{1}{\alpha}(\mathbb{E}\left[L \cdot 1_{\{L \leq q_\alpha(L)\}}\right] - q_\alpha(L)(\mathbb{P}[L \leq q_\alpha(L)] - \alpha$$

Although it may appear complex from a mathematical point of view, the ES equation solves the problem faced for TCE: in other words, it guarantees subadditivity even for non-continuous distributions, where the probability that *L* is less than (or equal to) the quantile α may not correspond exactly to α itself. Moving on in the formula, we transform the second part of the equation and simplify with some algebraic steps:

$$ES_{\alpha}(L) = -q_{\alpha}(L) - \frac{1}{\alpha}\mathbb{E}\big[\big(L - q_{\alpha}(L)\big) \cdot 1_{\{L \leq q_{\alpha}(L)\}}\big]$$
$$= -q_{\alpha}(L) + \tfrac{1}{\alpha}\mathbb{E}\left[\big(L - q_{\alpha}(L)\big)_{-}\right]$$

where $(L)_{-} = -\min(l, 0)$ represents the negative part of L.

The important aspect of the formulas presented so far is that they are, unlike what happens in VaR and TCE, invariant if one substitutes $<$ to $\leq$ or, equally, whatever quantile is considered, because there is only one definition of the ES (Acerbi et al., 2002).

By integrating the probability of the first ES equation, i.e., the quantile $p \to q_{p}(L)$, the authors capture exactly the average worst-case loss and provide a conclusive and mathematically rigorous definition of Expected Shortfall:

$$ES_{\alpha}(L) = -\frac{1}{\alpha}\int_{0}^{\alpha} q_{p}(L)\, dp = -\frac{1}{\alpha}\int_{0}^{\alpha} F\overleftarrow{l}(p) dp$$

where $F\overleftarrow{l}(p)$ is the so-called *"generalized inverse function"* to the left of $F(l)$, and corresponds to the lower quantile, q_{p}. In conclusion, given a confidence level α, the ES corresponds to the weighted average of all VaRs:

$$ES_{\alpha}(L) = -\frac{1}{\alpha}\int_{0}^{\alpha} VaR(L) dp$$

This means that ES not only captures the maximum expected loss for a given confidence level but also provides insight into losses that exceed VaR. Hence it is logical that $ES \geq VaR$ always, offering so a more conservative and robust measure of total risk.

Empirical Results

In the second chapter, we studied the dynamic correlations between selected stocks from the Nasdaq 100 and the developed macroeconomic-financial index, comparing different econometric models and evaluating the impact of the COVID-19 pandemic. After identifying the most suitable model, the VAR-DCC-GARCH, in this section we focus on the application of the results obtained for the analysis of financial risk, with particular attention to the Value at Risk and Expected Shortfall measures.

From this point, we move from multivariate analysis on individual assets to a univariate study of the overall portfolio. For risk assessment, in fact, we constructed an Equally-Weighted (EW) portfolio by assigning equivalent weights to each asset on an initial investment of $10,000 and calculating the daily value of the portfolio by multiplying the weighted daily returns of each asset by the amount invested. Next, to obtain the daily portfolio changes (P&L), we calculated the difference between consecutive daily values. Finally, using the dynamic conditional covariance matrix estimated through the VAR-DCC-GARCH model, we calculated the dynamic VaR and ES on the portfolio loss distribution at two significance levels, 5% and 1%.

As in Chapter 2, the analysis was conducted on two separate scenarios: one that includes the pandemic months and another that excludes them. Lastly, we examined VaR breaches, identifying times when portfolio losses exceeded risk limits, and plotted the estimated potential losses using ES.

Construction of the Equally-Weighted portfolio and calculation of risk measures

One of the main objectives of risk management is portfolio optimization, which has a solid foundation in the Markowitz Theory. This theory is based on the Mean-Variance Optimization (MVO) approach, which shows how efficient portfolios optimize the combination of expected return and risk, as measured by variance (or standard deviation) (Markowitz, 1952). In our study, to construct an EW portfolio consisting of the stocks and the index considered so far, we adopted Markowitz's classical formula for calculating the portfolio's expected return:

$$E(r_P) = \sum_{i=1}^{n} w_i E(r_i) = w^T r$$

where:

- $E(r_P)$ is the expected return of the portfolio;
- n is the number of securities i within the portfolio;
- w_i are the weights of the individual assets;

- $E(r_i)$ is the expected return of the i-th security;
- w^T is the transposed vector of the portfolio weights.

Further, we have assumed that we are operating from the perspective of a U.S. investor, a choice that justifies not considering any kind of exchange rate during the entire analysis. The initial investment is \$10,000 and there are no transaction costs involved.

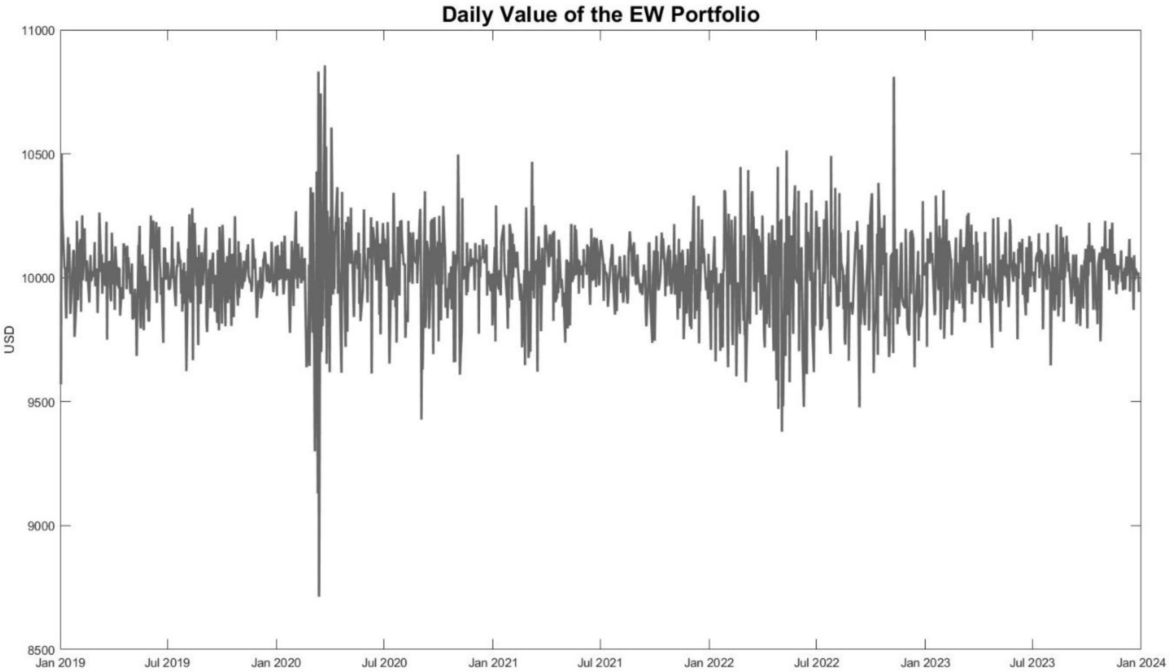


Figure 13. Representation of the daily value of the EW portfolio, assuming an initial investment of \$10,000 and no transaction costs.

Another essential aspect of MVO is the definition of risk in terms of portfolio standard deviation, which is calculated according to the following formula:

$$\sigma_P = \sum_{i=1}^{n} w_i^2 \sigma_i^2 \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j Cov_{ij} = (w\Sigma w^T)^{-1}$$

where:

- σ is the standard deviation of the portfolio;
- $Cov_{i,j}$ is the covariance between asset i and j in the portfolio;
- Σ is the variance-covariance matrix of the n-assets within the portfolio.

Markowitz considers variance as the main measure of risk, and his goal is to identify the combination of securities that maximizes return while minimizing risk (or vice versa), through diversification, to construct the so-called *"efficient frontier"*. However, in our case, the purpose is to calculate VaR and ES, two advanced measures of risk that focus on potential losses.

To do this, we adopted one of three recognized approaches for estimating these measures: the parametric (or variance-covariance) approach. This method is based on the assumption that the underlying market factors in the model all follow a normal distribution, allowing us to estimate expected losses through VaR and ES, rather than just variance (McNeil et al., 2005). In this way, the parametric approach incorporates the fundamental concepts of Markowitz Theory but uses more advanced tools to calculate the covariance matrix. In fact, in our case, we made use of the dynamic conditional covariance matrix estimated through VAR-DCC-GARCH.

The formula for expected portfolio return remains unchanged, while risk is redefined as follows:

- for VaR we have

$$VaR_\alpha(P) = \mu - z_\alpha \cdot \sqrt{w\Sigma_{VAR/DCC/GARCH} w^T}$$

  where:

  - $VaR_\alpha(P)$ is the portfolio VaR at the α level considered;
  - μ is the expected return of the portfolio;
  - z_α is the quantile of the standard normal distribution corresponding to α;
  - $\Sigma_{VAR/DCC/GARCH}$ is the conditional dynamic covariance matrix applied to the portfolio, obtained through VAR-DCC, which involves the GARCH model for volatility[7];

- for ES we have

$$ES_\alpha(P) = \mu - \frac{1}{\alpha}\int_0^\alpha z_p \cdot \sqrt{w\Sigma_{VAR/DCC/GARCH} w^T} dp$$

  where:

  - $ES_\alpha(P)$ is the portfolio ES at the α level considered;
  - z_p is the quantile of the standard normal distribution for $p \in [0, \alpha]$.

[7] Appendix B shows the graphical results of some key moments of the covariance matrix.

Having defined the main inputs required by the risk measures, i.e., time horizon over which to calculate VaR and ES (5 years in our case, 2019 to 2024) and confidence level (95% and 99%), we obtained the following results on the entire portfolio:

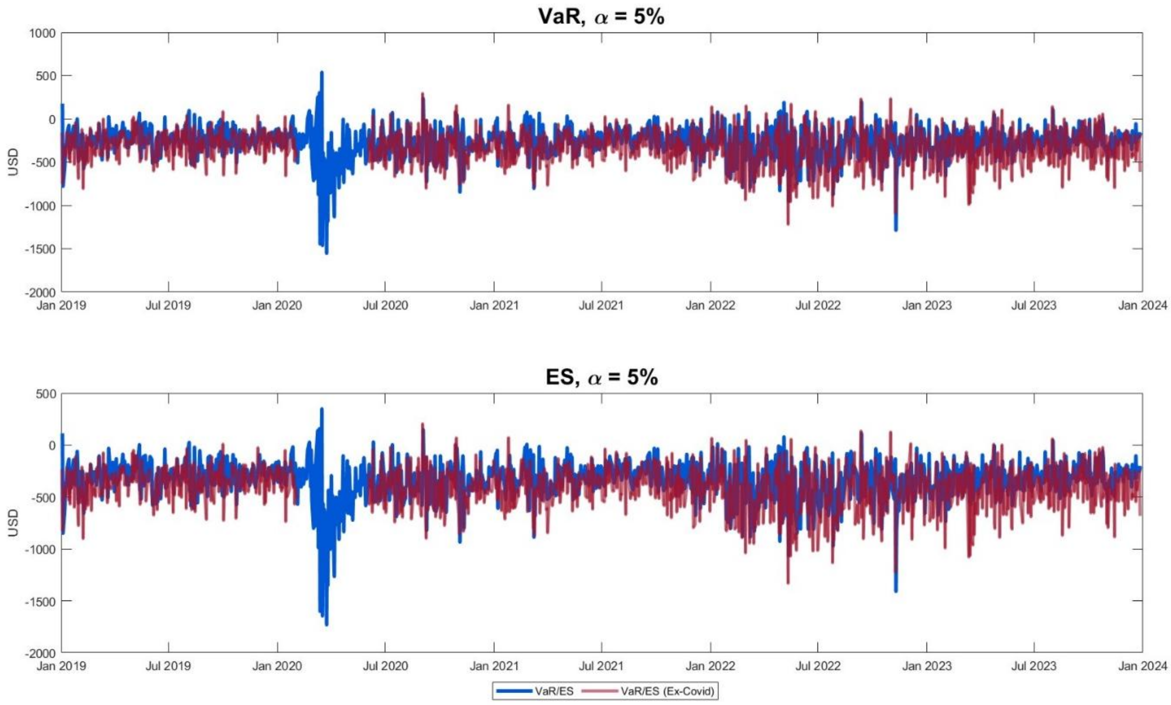


Figure 14. Comparison of Value at Risk (VaR) and Expected Shortfall (ES) at 5%, with and without the COVID-19 pandemic months.

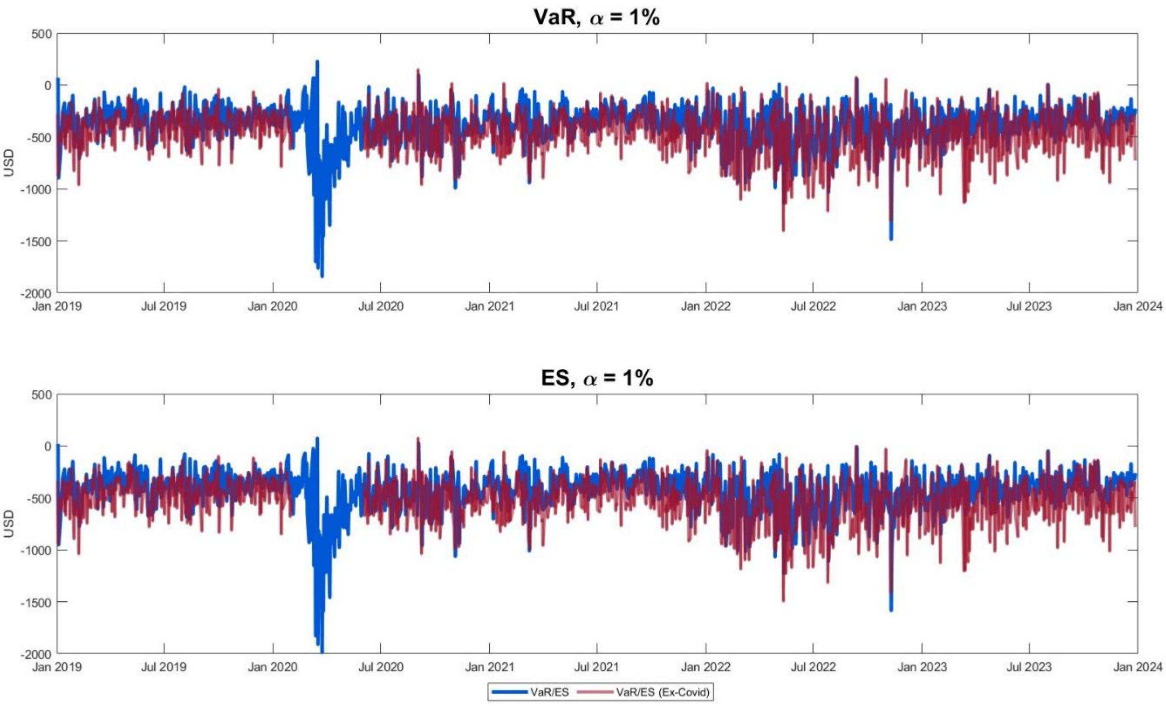


Figure 15. Comparison of Value at Risk (VaR) and Expected Shortfall (ES) at 1%, with and without the COVID-19 pandemic months.

The proposed graphs illustrate the expected daily losses of our portfolio, calculated through VaR and ES, including and excluding the effect of the COVID-19 crisis. It is evident that the pandemic had the greatest impact, with a peak in risk seen in March 2020 due to the resulting financial crisis. Prior to this period, until February 2020, expected losses were relatively small. After the pandemic peak, however, markets remained susceptible to further perturbations, reaching new instability in 2022, probably due to the Russian-Ukrainian conflict. This dynamic is reflected in both VaR and ES, showing how daily expected losses are affected by relevant market events.

VaR and ES provide different measures of the risk of loss, as the latter also considers losses that exceed VaR. Specifically:

- the quantile of $VaR_{95\%}$ is -1.645, while that of $VaR_{99\%}$ is -2.326;
- the integral of the inverse function of the standard normal distribution of $ES_{95\%}$ is -2.063, while that of $ES_{99\%}$ is -2.665.

Therefore, it is natural that the ES is always greater than the VaR (Acerbi et al., 2002). Moreover, a higher confidence level implies a smaller tail, a leftmost cut-off point, and, therefore, higher values for both magnitudes. Apart from that, the procedure for calculating the risk measures remains the same in all scenarios.

Finally, we emphasize that in our analysis VaR does not assume a constant value. The variability obtained is due to the fact that we were able to compute it on a loss distribution that varies over time in response to market changes and past information. Using the VAR-DCC-GARCH model, which estimates a dynamic conditional covariance matrix, we were able to capture the temporal dynamics of volatilities and correlations between the index and portfolio assets. In this way, the dynamic matrix contributed to a more timely and continuously updated estimate of portfolio risk.

Measures of Risk	Average	
	Covid	*Ex-Covid*
$VaR_{95\%}$	-271.34	-338.05
$VaR_{99\%}$	-378.82	-466.59
$ES_{95\%}$	-337.24	-416.86
$ES_{99\%}$	-423.27	-530.50

Table 8. Comparison of average potential VaR and ES losses, with and without COVID-19 pandemic months and for both confidence levels.

Regarding the comparison between the group considering or excluding the COVID-19 months (February, March, April 2020), the expected losses are, on average, slightly higher in the group not including the pandemic, despite the fact that the COVID-19 period was highly volatile.

The result is partly surprising, and the reason may be that excluding the pandemic left room for other market dynamics that contributed to the risk. For instance, we can mention the commodity shortage (particularly critical for the sector and technology stocks considered in our study) (Liu et al., 2022), tensions between the United States and China (Riaz et al., 2024), the conflict between Russia and Ukraine in 2022, or high inflation in recent years. These events have had a significant impact on markets, keeping risk measures high even in the absence of the pandemic period.

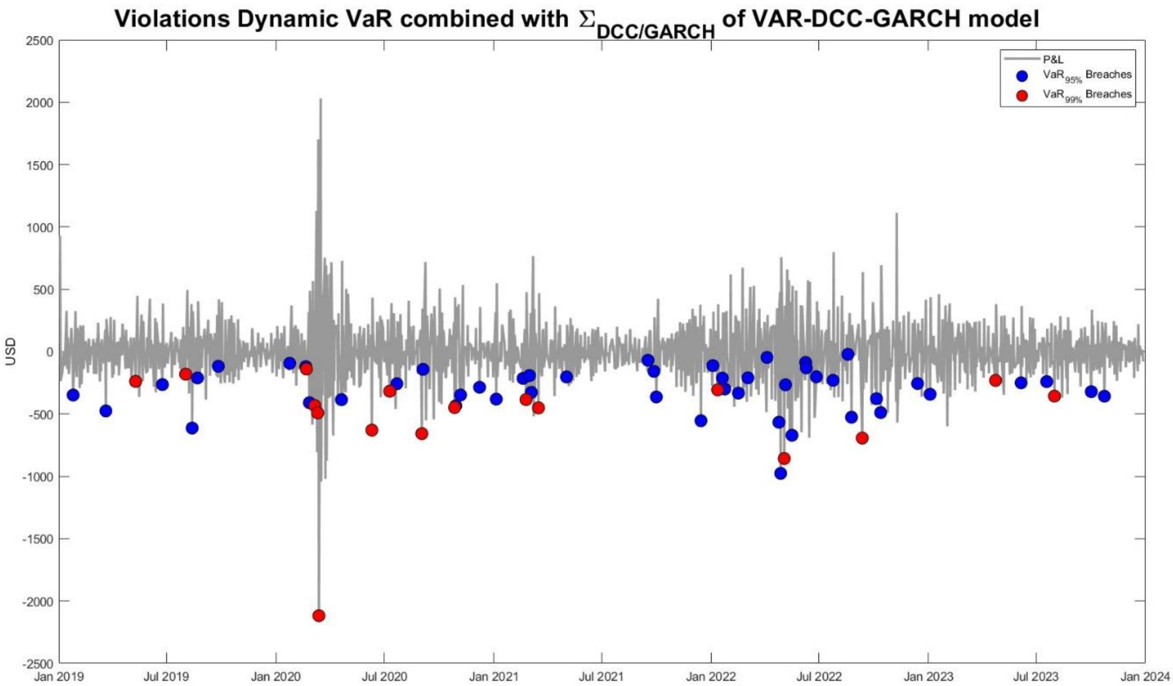


Figure 16. VaR breaches occurred from 1/1/2019 to 1/1/2024 for both confidence levels.

The graph highlights when the actual daily portfolio losses exceed the expected VaR value, for both levels of significance α. Examining the trend over time, we see that the violations are distributed throughout the observed period, with a higher concentration in periods of higher volatility. Most of them are moderate in magnitude relative to the overall portfolio value, with a few exceptions that exceed the average expected losses observed previously for both measures.

Although the parametric approach allows obtaining VaR measures that can adapt quickly to market conditions, the various violations suggest, in our case, that the model tends to underestimate real risk, especially during periods of high uncertainty. This behaviour can be attributed to the nature of VaR, which is typically based on a normal distribution, and which often does not adequately capture the large tails of the distribution of actual losses.

The recurrent deviations of losses from expected ones recorded during crisis periods indicate the need for more robust risk management strategies. These include greater portfolio diversification to reduce risk concentration, the use of derivatives for hedging purposes, or the implementation of more sophisticated risk measures. In relation to the latter point, we highlight the results obtained after implementing ES for both α significance levels:

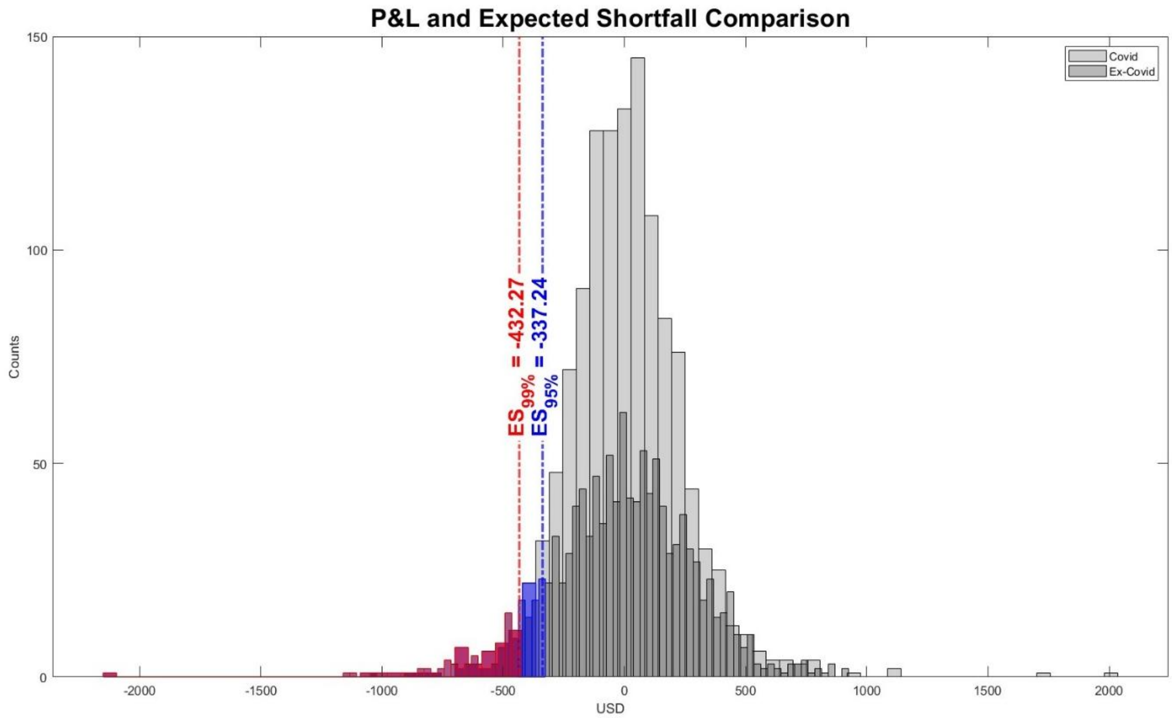


Figure 17. Comparison of potential portfolio losses by ES for both confidence levels.

ES provides a more in-depth view of potential losses in extreme scenarios than VaR, as it also considers cases that exceed the threshold set by VaR itself. The graph shows how ES values capture more severe tail events, emphasizing the role of ES as a complementary and more reliable measure than VaR in extreme risk cases (Peña, 2002).

The histogram illustrates that most observations are concentrated around the mean, with the distribution decreasing toward the tails, following an approximately normal shape. However, closer analysis reveals thicker tails than those of a normal distribution, thus suggesting a leptokurtic distribution, which is typical of financial time series (Cont, 2001). The highlighted values for ES in the graph support this reading: for a 95% confidence interval (blue line), the average loss in the worst 5% of cases is \$337.24; for a 99% interval (red line), the average loss in the worst 1% of cases is \$432.27.

Conclusions

In this thesis, we emphasized the importance of *exploratory data analysis* (EDA) as a tool for preliminary data analysis in financial markets. Using techniques such as *cluster analysis*, we were able to group data into homogeneous sets based on common characteristics. Specifically, we employed the *K-Means* algorithm, one of the most widely used *non-hierarchical clustering* methodologies, to identify representative stocks in the U.S. Nasdaq 100 index based on risk and return characteristics.

Following a detailed description of the dataset and a concise presentation of its primary characteristics, the second part of the thesis employs the results obtained to further analyse the dynamics of correlations and financial risk. Among the various models implemented, we selected the one that demonstrated the most effective RMSE indicator, thus ensuring the most accurate capture of temporal variations. In the subsequent phase of the study, a comparison was made between the selected model and a version that excluded the months of February, March and April 2020, which were identified as the most volatile months for financial markets during the period of the pandemic. It was noted that the absence of a relevant exogenous *shock* in these months resulted in a reduction in the frequency of regime shifts between positive and negative correlations, thereby improving the stability of the correlation dynamics themselves. Furthermore, a comparison was made with a version of the model that incorporated conditional mean modelling. This additional integration was found to enhance the accuracy of the results obtained in the study of conditional dynamic correlations.

In the third chapter, we applied the dynamic matrix estimated through the VAR-DCC-GARCH model to calculate VaR and ES through the variance-covariance approach. The methodology was implemented on an EW portfolio, and the impact on the latter was analysed considering the main confidence levels (95% and 99%), both including and excluding the pandemic period. The findings indicated that, on average, the exclusion of the initial months of the pandemic resulted in amplified potential losses, suggesting the presence of other significant risk dynamics during the period considered, such as the shortage of raw materials (particularly critical for the sector and technology stocks considered in our study), geopolitical tensions between the United States and China and high inflation.

Finally, analysis of VaR violations showed that 2022 also experienced high volatility – probably due to the Russia-Ukraine conflict – making it difficult for the parametric approach to accurately estimate real financial risk. The incorporation of ES has enabled us to enhance our ability to estimate losses in worst-case scenarios (5% and 1% of cases), providing a more accurate assessment of extreme risks and underscoring the necessity of considering ES as a complementary measure to VaR for effective risk management.

In conclusion, the use of more sophisticated econometric techniques is crucial to achieve more accurate and comprehensive modelling of financial correlations and risk dynamics. The adoption of ES as a complementary measure to VaR provides a more accurate view of the potential risk of an investment, especially in contexts characterized by high volatility and exogenous shocks.

Future Research

Considering how important Machine Learning models have become in financial markets in recent years, future developments related to the first part of the elaboration could explore alternative algorithms to *K-Means*, such as *hierarchical clustering* techniques, DBSCAN, *spectral clustering*, etc. In addition, *unsupervised* alternative techniques to cluster analysis could be considered, such as, for example, the well-known *principal component analysis* (PCA). Regarding further applications of the models studied, we could employ the results of dynamic correlations in the context of derivative *hedging* strategies to understand how to maintain, against financial risks, an effective and real-time hedge. In details, *dynamic hedge ratios* based on DCC models have numerous applications in the financial field, including futures markets, currency risk, multi-asset portfolios, and cryptocurrencies.

Bibliography

Cluster Analysis

Agarwal, A., Prabha, S., Yadav, R., 2024. *Exploratory Data Analysis for Banking and Finance: Unveiling Insights and Patterns*. arXiv preprint arXiv: 2407.11976.

Arouri, M., Jouini, J., Nguyen, D.K., 2011. *Volatility spillovers between oil prices and stock sector returns: Implications for portfolio management*. Journal of International Money and Finance, Elsevier, 30(7), 1387–1405.

Arrigoni, S., Bobasu, A., Venditti, F., 2020. *The simpler the better: measuring financial conditions for monetary policy and financial stability*. ECB Working Paper, No. 2451, August.

Cai, F., Le–Khac, N.–A., Kechadi, T., 2016. *Clustering Approaches for Financial Data Analysis: a Survey*. Papers 1609.08520, arXiv.org.

Chang, S., Zhenzhong, X., Xuan, G., 2018. *Improvement of K Mean Clustering Algorithm Based on Density*. CoRR, abs/1810.0 (1).

Conrad, C., Loch, K., Ritter, D., 2012. *On the Macroeconomic Determinants of the Long–Term Oil-Stock Correlation*. Working Papers 0525, Department of Economics, University of Heidelberg.

Cont, R., 2001. *Empirical Properties of Asset Returns: Stylized Facts and Statistical Issues*. Quantitative Finance, 1, 223–236, http://dx.doi.org/10.1080/713665670.

Corter, J.E., 2023. *Quantitative Research and Educational Measurement*. In: International Encyclopedia of Education (Fourth Edition).

Da Silva, S., Cunha, J., Da Costa, N. Jr., 2005. *Stock selection based on cluster analysis*. Economics Bulletin, AccessEcon, 13(1), 1–9.

Fayyad, U., Piatetsky–Shapiro, G., Smyth, P., 1996. *From Data Mining to Knowledge Discovery: An Overview*. In Advances in Knowledge Discovery and Data Mining, AAAI Press/The MIT Press.

Hartigan, J.A., Wong, M.A., 1979. *A K–Means Clustering Algorithm*. Journal of the Royal Statistical Society. Series C, 28(1), 100–108.

Ikotun, A.M., et al., 2023. *K–Means Clustering Algorithms: A Comprehensive Review, Variants Analysis, and Advances in the Era of Big Data*. Information Sciences, (622), 178–210. https://doi.org/10.1016/j.ins.2022.11.139.

Jha, A., Prajapati, B., 2023. *Machine learning and exploratory data analysis in cross–sell insurance*. Encyclopedia of Data Science and Machine Learning, 651–685. IGI Global.

Jung, Y.G., Kang, M.S., Heo, J., 2014. *Clustering performance comparison using k–means and expectation maximization algorithms*. Biotechnology & Biotechnological Equipment, 28(sup1): S44S48.

Kodinariya, T.M., Makwana, P.R., 2013. *Review on Determining Number of Cluster in K–Means Clustering*. International Journal, (1), 90–95.

Li, L., 2002. *Macroeconomic Factors and the Correlation of Stock and Bond Returns*. Working paper. Yale International Center for Finance.

MacQueen, J.B., 1967. *Some Methods for Classification and Analysis of Multivariate Observations*. Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, University of California Press, 281–297, http://projecteuclid.org/euclid.bsmsp/1200512992.

Mehta, V., Bawa, S., Singh, J., 2020. *Analytical review of clustering techniques and proximity measures*. Artif Intell Rev 53, 5995–602, https://doi.org/10.1007/s10462-020-09840-7.

The MathWorks Inc., 2023. *Datafeed Toolbox™ version: 6.4*. (R2023a), Natick, MA: The MathWorks, Inc.

The MathWorks Inc., 2023. *Econometrics Toolbox™ version: 6.2*. (R2023a), Natick, MA: The MathWorks, Inc.

The MathWorks Inc., 2023. *Financial Toolbox™ version: 6.5*. (R2023a), Natick, MA: The MathWorks, Inc.

The MathWorks Inc., 2023. *Statistics and Machine Learning Toolbox™ version: 12.5*. (R2023a), Natick, MA: The MathWorks, Inc.

Tukey, J.W., 1977. *Exploratory Data Analysis*. Reading, Massachusetts: Addison–Wesley.

Zhu, Z., Liu, N., Wang, W., 2021. *Early Warning of Financial Risk Based on K–Means Clustering Algorithm*. Complexity, Hindawi, (24), 1–12, March.

Dynamic Conditional Correlation Model

Bauwens, L., Laurent, S., Rombouts, J., 2006. *Multivariate GARCH Models: A Survey.* Journal of Applied Econometrics (21), 79–109.

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

Bollerslev, T., Engle, R.F., Wooldridge, J.M., 1988. *A Capital–Asset Pricing Model with Time–Varying Covariances.* Journal of Political Economy 96(1), 116–31.

Bollerslev, T., 1990. *Modelling the Coherence in Short–Run Nominal Exchange Rates: A Multivariate Generalized ARCH Model.* The Review of Economics and Statistics (72), 498–505.

Cappiello, L., Engle, R.F., Sheppard, K., 2006. *Asymmetric dynamics in the correlations of global equity and bond returns*. Journal of Financial econometrics (4), 537–572.

Chevallier, J. 2012. *Time–varying correlations in oil, gas and CO2 prices: an application using BEKK, CCC, and DCC–MGARCH models.* Applied Economics 44(32), 4257–4274.

Engle, R.F., 1982. *Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation*. Econometrica: Journal of the Econometric Society, 987–1007.

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

Engle, R.F., 2001. *GARCH 101: The use of ARCH/GARCH models in applied econometrics*. Journal of economic perspectives, 15(4), 157–168.

Engle, R.F., 2002. *Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models*. Journal of Business & Economic Statistics, 20(3), 339–350.

Fabozzi, F.J., Kolm, P. N., Pachamanova, D. A., Focardi, S. M., 2007. *Robust portfolio optimization and management*. John Wiley & Sons.

Glosten, L.R., Jagannathan, R., Runkle, D.E., 1993. *On The Relation between The Expected Value and The Volatility of Nominal Excess Return on stocks.* Journal of Finance 48(5), 1779–1801.

Gourieroux, C., 1997. *ARCH Models and Financial Applications.* Springer–Verlag: New York.

Guidolin, M., Pedio, M., 2018. *Essentials of Time Series for Financial Applications.* Academic Press.

He, X.–Z., Li, K., Wang, C., 2016. *Volatility clustering: A nonlinear theoretical approach*. Journal of Economic Behavior & Organization 130, 274–297.

Ji, X., Wang, S., Xiao, H., Bu, N., Lin, X., 2022. *Contagion Effect of Financial Markets in Crisis: An Analysis Based on the DCC–MGARCH Model*. Mathematics, MDPI, 10(11), 1–14, May.

Nafisi–Moghadam, M., Fattahi, S., 2022. *A Hybrid Model of VAR–DCC–GARCH and Wavelet Analysis for Forecasting Volatility*. Engineering Proceedings, 18(1): 6, https://doi.org/10.3390/engproc2022018006.

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

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

Sheppard, K., 2021. *MFE Toolbox™ for MATLAB*. Available at: https://www.kevinsheppard.com/code/matlab/mfe-toolbox/.

Wooldridge, J.M., 2010. *Econometric Analysis of Cross Section and Panel Data.* 2nd Edition, MIT Press, Cambridge.

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

Value at Risk and Expected Shortfall

Acerbi, C., Tasche, D., 2002. *Expected Shortfall: a natural coherent alternative to Value at Risk*. Economic Notes, 31(2), 379–388.

Acerbi, C., 2007. *Coherent measures of risk in everyday market practice*. Quantitative Finance, Taylor & Francis Journals, 7(4), 359–364.

Artzner, P., Delbaen, F., Eber, J.–M., Heath, D., 1997. *Thinking Coherently*. Risk, 10, 68–71.

Artzner, P., Delbaen, F., Eber, J.–M., Heath, D., 1999. *Coherent measures of risk*. Math. Fin., 9(3), 203–228.

Barone Adesi, G., Giannopoulos, K., Vosper, L., 1999. *VaR without Correlations for Portfolios of Derivative Securities*. Journal of Futures Markets.

Basel Committee on Banking Supervision (BCBS), 2008. *Consultative Document, Proposed Revisions to the Basel II Market Risk Framework*. BIS, Basel, Switzerland, http://www.bis.org/publ/bcbs140.pdf.

Basel Committee on Banking Supervision (BCBS), 2013. *Consultative Document, Fundamental Review of the Trading Book: A revised Market Risk framework*. BIS, Basel, Switzerland, http://www.bis.org/publ/bcbs265.pdf.

Danìelsson, J., de Vries, C.G., 1997c. *Value–at–Risk and Extreme Returns*. In Extremes and Integrated Risk Management (ed. P. Embrechts), London: RiskWaters Group, 85–106.

Danìelsson, J, de Vries, C.G., Samorodnitsky, G., Jorgensen, B.N., Mandira, S., 2005. *Subadditivity Re–Examined: The Case for Value–at–Risk*. FMG Discussion Papers dp. 549, Financial Markets Group.

Kibzun, A.I., Kuznetsov, E.A., 2006. *Analysis of criteria VaR and CVaR*. Journal of Banking & Finance, Elsevier, 30(2), 779–796.

Liu, X., Shehzad, K., Kocak, E., Zaman, U., 2022. *Dynamic correlations and portfolio implications across stock and commodity markets before and during the COVID–19 era: A key role of gold*. Resources Policy, Elsevier, 79(C).

Markowitz, H., 1952. *Portfolio Selection*. The Journal of Finance, 7, 77–91.

McNeil, A.J., Frey, R., Embrechts, P., 2005. *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press, Princeton, NJ.

Morgan, J.P., 1996. *Risk Metrics – Technical Document*. J.P. Morgan/Reuters, New York.

Peña, I., 2002. *La gestión de riesgos financieros de mercado y de crédito*. Madrid: Prentice Hall.

Riaz, A., Ullah, A., Xingong, L., 2024. *Does trade policy uncertainty in China and USA matter for key financial markets?*. Economic Change and Restructuring, Springer, 57(2), 1–30, April.

Rockafellar, R.T., 1970. *Convex Analysis*. Princeton University Press, Princeton, NJ.

Rockafellar, R.T., Uryasev, S.P., 2002. *Conditional Value–at–Risk for General Loss Distributions*. Journal of Banking & Finance, 26, 1443–1471.

The MathWorks Inc., 2023. *Optimization Toolbox™ version: 9.5*. (R2023a), Natick, MA: The MathWorks, Inc.

The MathWorks Inc., 2023. *Risk Management Toolbox™ version: 2.2*. (R2023a), Natick, MA: The MathWorks, Inc.

Appendix A: Average returns and annualized volatilities of Nasdaq 100 stocks

Here are the results of the *preprocessing* phase performed on the daily *adj close* prices of the Nasdaq 100. These prices were collected during the period from 1 January 2019 to 1 January 2024. The table below shows the changes in prices to get the average returns and volatilities.

Both measures were annualized: the average daily return, where return is defined as $\frac{P_t - P_{t-1}}{P_{t-1}}$ (where P_t is the closing price at day t), was multiplied by 252 (the number of stock market open days in a year) to obtain the average annual return; volatility was annualized by multiplying the daily standard deviation by the square root of 252.

Tickers	***Avg. Return***		***Δ Avg. Return***	***Volatility***		***Δ Vol.***
	Covid	*Ex-Covid*		*Covid*	*Ex-Covid*	
ADBE	0.2642	0.2625	*-0.0017*	0.3682	0.3342	*-0.0340*
ADP	0.1772	0.1755	*-0.0017*	0.2800	0.2451	*-0.0349*
ABNB	0.1205	0.1205	*0.0000*	0.5314	0.5315	*0.0001*
GOOGL	0.2460	0.2494	*0.0034*	0.3181	0.2963	*-0.0218*
GOOG	0.2495	0.2533	*0.0038*	0.3182	0.2969	*-0.0213*
AMZN	0.1984	0.2036	*0.0052*	0.3522	0.3476	*-0.0046*
AMD	0.5556	0.5660	*0.0104*	0.5369	0.5164	*-0.0205*
AEP	0.0854	0.0840	*-0.0014*	0.2415	0.2194	*-0.0221*
AMGN	0.1439	0.1426	*-0.0013*	0.2531	0.2229	*-0.0302*
ADI	0.2483	0.2421	*-0.0062*	0.3495	0.3015	*-0.0480*
ANSS	0.2510	0.2493	*-0.0017*	0.3600	0.3263	*-0.0337*
AAPL	0.3780	0.3854	*0.0074*	0.3223	0.2912	*-0.0311*
AMAT	0.4323	0.4325	*0.0002*	0.4561	0.4179	*-0.0382*
ASML	0.4089	0.4132	*0.0043*	0.4042	0.3724	*-0.0318*
AZN	0.1777	0.1805	*0.0028*	0.2655	0.2480	*-0.0175*
TEAM	0.3462	0.3581	*0.0119*	0.5447	0.5476	*0.0029*
ADSK	0.2060	0.1999	*-0.0061*	0.3957	0.3617	*-0.0340*
BKR	0.2303	0.2304	*0.0001*	0.4514	0.4265	*-0.0249*
BIIB	0.0835	0.0749	*-0.0086*	0.4980	0.4843	*-0.0137*
BKNG	0.2137	0.2159	*0.0022*	0.3714	0.3551	*-0.0163*

AVGO	0.3984	0.3997	*0.0013*	0.3672	0.3224	*-0.0448*
CDNS	0.4325	0.4402	*0.0077*	0.3558	0.3218	*-0.0340*
CDW	0.2758	0.2771	*0.0013*	0.3236	0.2893	*-0.0343*
CHTR	0.1103	0.1077	*-0.0026*	0.3158	0.2974	*-0.0184*
CTAS	0.3152	0.3192	*0.0040*	0.3102	0.2742	*-0.0360*
CSCO	0.1032	0.0975	*-0.0057*	0.2827	0.2484	*-0.0343*
CCEP	0.1596	0.1589	*-0.0007*	0.3183	0.2947	*-0.0236*
CTSH	0.1018	0.0899	*-0.0119*	0.3192	0.2700	*-0.0492*
CMCSA	0.1098	0.1076	*-0.0022*	0.2790	0.2560	*-0.0230*
CEG	0.6121	0.6121	*0.0000*	0.3851	0.3851	*0.0000*
CPRT	0.3294	0.3376	*0.0082*	0.3031	0.2772	*-0.0259*
CSGP	0.2543	0.2559	*0.0016*	0.3529	0.3281	*-0.0248*
COST	0.2807	0.2889	*0.0082*	0.2374	0.2152	*-0.0222*
CRWD	0.5044	0.5138	*0.0094*	0.5984	0.5831	*-0.0153*
CSX	0.1602	0.1548	*-0.0054*	0.2941	0.2515	*-0.0426*
DDOG	0.4777	0.4899	*0.0122*	0.6415	0.6351	*-0.0064*
DXCM	0.4053	0.4245	*0.0192*	0.4769	0.4935	*0.0166*
FANG	0.3220	0.2955	*-0.0265*	0.6006	0.5402	*-0.0604*
DLTR	0.1586	0.1528	*-0.0058*	0.3735	0.3451	*-0.0284*
DASH	0.0102	0.0102	*0.0000*	0.6707	0.6707	*0.0000*
EA	0.1546	0.1547	*0.0001*	0.2978	0.2783	*-0.0195*
EXC	0.1016	0.0952	*-0.0064*	0.2868	0.2461	*-0.0407*
FAST	0.2534	0.2557	*0.0023*	0.2862	0.2537	*-0.0325*
FTNT	0.3879	0.3933	*0.0054*	0.4439	0.4215	*-0.0224*
GEHC	0.2939	0.2940	*0.0001*	0.3074	0.3074	*0.0000*
GILD	0.1219	0.1248	*0.0029*	0.2558	0.2530	*-0.0028*
GFS	0.2599	0.2599	*0.0000*	0.5246	0.5246	*0.0000*
HON	0.1503	0.1496	*-0.0007*	0.2710	0.2423	*-0.0287*
IDXX	0.2821	0.2871	*0.0050*	0.3433	0.3237	*-0.0196*
ILMN	-0.0533	-0.0667	*-0.0134*	0.4404	0.4272	*-0.0132*
INTC	0.1168	0.1037	*-0.0131*	0.3883	0.3480	*-0.0403*
INTU	0.3093	0.3092	*-0.0001*	0.3705	0.3336	*-0.0369*
ISRG	0.2193	0.2181	*-0.0012*	0.3570	0.3296	*-0.0274*
KDP	0.1075	0.1022	*-0.0053*	0.2467	0.2041	*-0.0426*
KLAC	0.4858	0.4855	*-0.0003*	0.4435	0.3936	*-0.0499*

KHC	0.0650	0.0545	*-0.0105*	0.3107	0.2708	*-0.0399*
LRCX	0.4763	0.4769	*0.0006*	0.4778	0.4367	*-0.0411*
LIN	0.2437	0.2473	*0.0036*	0.2655	0.2363	*-0.0292*
LULU	0.3658	0.3681	*0.0023*	0.4005	0.3698	*-0.0307*
MAR	0.2343	0.2429	*0.0086*	0.3979	0.3836	*-0.0143*
MRVL	0.3937	0.3973	*0.0036*	0.5099	0.4900	*-0.0199*
MELI	0.4860	0.4954	*0.0094*	0.5543	0.5404	*-0.0139*
META	0.2891	0.2929	*0.0038*	0.4363	0.4220	*-0.0143*
MCHP	0.2967	0.2879	*-0.0088*	0.4359	0.3895	*-0.0464*
MU	0.2995	0.2960	*-0.0035*	0.4559	0.4253	*-0.0306*
MSFT	0.3203	0.3231	*0.0028*	0.3049	0.2648	*-0.0401*
MRNA	0.6494	0.7398	*0.0904*	0.7472	0.9302	*0.1830*
MDLZ	0.1656	0.1667	*0.0011*	0.2182	0.1863	*-0.0319*
MDB	0.5561	0.5659	*0.0098*	0.6815	0.6713	*-0.0102*
MNST	0.2099	0.2130	*0.0031*	0.2743	0.2517	*-0.0226*
NFLX	0.2305	0.2369	*0.0064*	0.4620	0.4632	*0.0012*
NVDA	0.6718	0.6879	*0.0161*	0.5177	0.4955	*-0.0222*
NXPI	0.3344	0.3325	*-0.0019*	0.4347	0.3982	*-0.0365*
ORLY	0.2447	0.2428	*-0.0019*	0.2790	0.2305	*-0.0485*
ODFL	0.3819	0.3902	*0.0083*	0.3409	0.3150	*-0.0259*
ON	0.4801	0.4754	*-0.0047*	0.5606	0.5149	*-0.0457*
PCAR	0.2608	0.2664	*0.0056*	0.2699	0.2456	*-0.0243*
PANW	0.3931	0.4047	*0.0116*	0.3938	0.3818	*-0.0120*
PAYX	0.1929	0.1899	*-0.0030*	0.2919	0.2495	*-0.0424*
PYPL	0.0296	0.0153	*-0.0143*	0.4369	0.4112	*-0.0257*
PDD	0.6485	0.6805	*0.0320*	0.7473	0.7660	*0.0187*
PEP	0.1404	0.1364	*-0.0040*	0.2168	0.1645	*-0.0523*
QCOM	0.2973	0.2982	*0.0009*	0.4178	0.3935	*-0.0243*
REGN	0.2226	0.2486	*0.0260*	0.3193	0.3913	*0.0720*
ROP	0.1824	0.1804	*-0.0020*	0.2574	0.2164	*-0.0410*
ROST	0.1822	0.1744	*-0.0078*	0.3773	0.3383	*-0.0390*
SIRI	0.0822	0.0842	*0.0020*	0.4013	0.3913	*-0.0100*
SBUX	0.1477	0.1415	*-0.0062*	0.3082	0.2683	*-0.0399*
SNPS	0.4229	0.4332	*0.0103*	0.3448	0.3191	*-0.0257*
TTWO	0.1530	0.1510	*-0.0020*	0.3606	0.3420	*-0.0186*

TMUS	0.2198	0.2207	*0.0009*	0.2785	0.2466	*-0.0319*
TSLA	0.7081	0.7035	*-0.0046*	0.6470	0.5982	*-0.0488*
TXN	0.1950	0.1919	*-0.0031*	0.3126	0.2760	*-0.0366*
TTD	0.6140	0.6113	*-0.0027*	0.7130	0.6826	*-0.0304*
VRSK	0.1987	0.1978	*-0.0009*	0.2614	0.2219	*-0.0395*
VRTX	0.2342	0.2374	*0.0032*	0.3211	0.3009	*-0.0202*
WBA	-0.0875	-0.0830	*0.0045*	0.3430	0.3402	*-0.0028*
WBD	-0.0290	-0.0357	*-0.0067*	0.5156	0.5163	*0.0007*
WDAY	0.1934	0.1941	*0.0007*	0.4104	0.3965	*-0.0139*
XEL	0.1086	0.1027	*-0.0059*	0.2450	0.2001	*-0.0449*
ZS	0.5226	0.5375	*0.0149*	0.6011	0.5975	*-0.0036*

Table 9. Average returns and annualized volatilities of Nasdaq 100 tickers from 1/1/2019 to 1/1/2024, with and without COVID-19 pandemic months.

Appendix B: Variance-Covariance Matrices

The conditional variance-covariance matrix illustrates the covariances among the 9 assets analysed: the macroeconomic-financial index, Msft, Aapl, Nvda, Amzn, Crwd, Ea, Intu, Pypl. This matrix shows how stock returns move together on the specific days chosen. Specifically:

- elements on the diagonal of the matrix represent the conditional variances of each asset;
- elements outside the diagonal indicate conditional covariances between asset pairs on the specific day considered.

As it is not possible to graphically report all the conditional covariance matrices of the VAR-DCC-GARCH model because they are estimated daily over a 5-year time frame, we decided to present only a few significant dates to highlight the evolution of risk at key moments. The selected dates include the beginning of the analysis period (January 1, 2019), the pre-pandemic phase (January 31, 2020), the period immediately following the most acute phase of the pandemic (May 1, 2020), and the end of the analysis period (January 1, 2024):

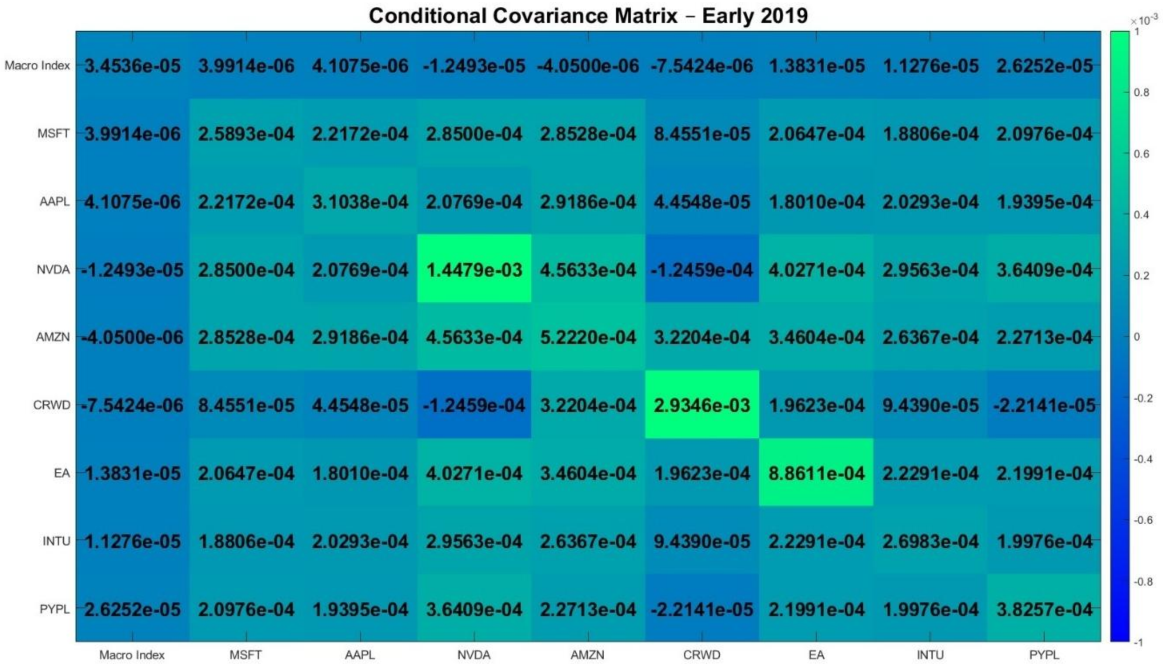


Figure 18. Conditional Dynamic Variance-Covariance Matrix for the calculation of Var and ES, as of January 1, 2019.

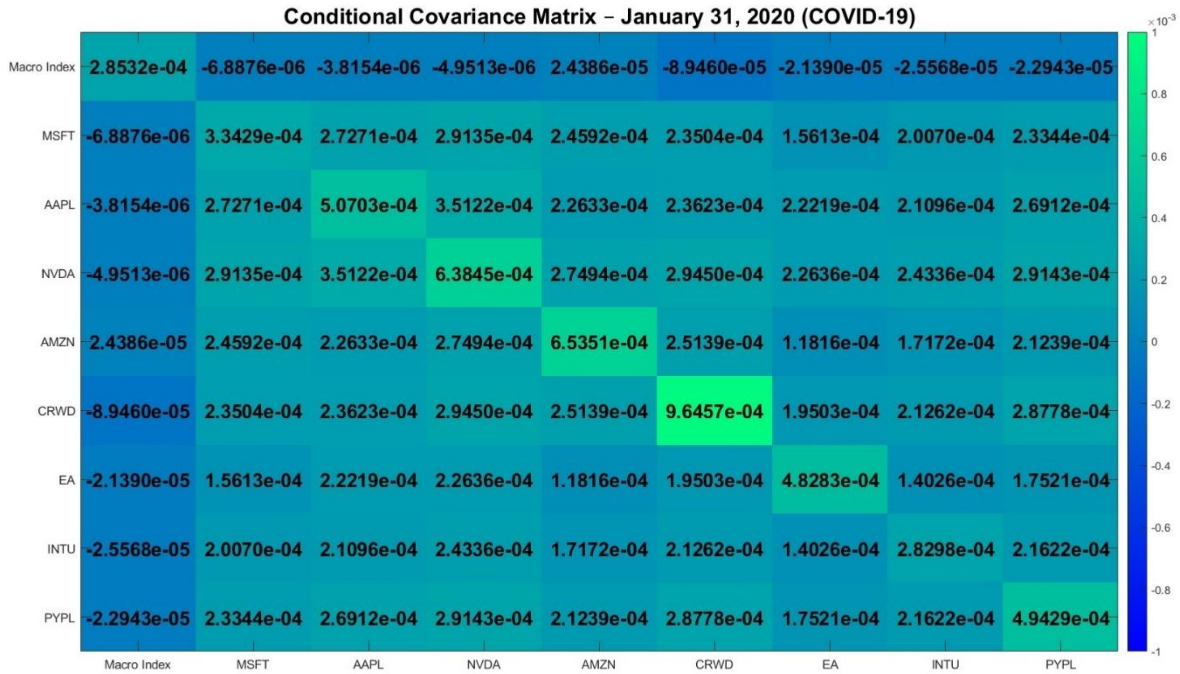


Figure 19. Conditional Dynamic Variance-Covariance Matrix for the calculation of Var and ES, as of January 31, 2020 (Pre-Covid).

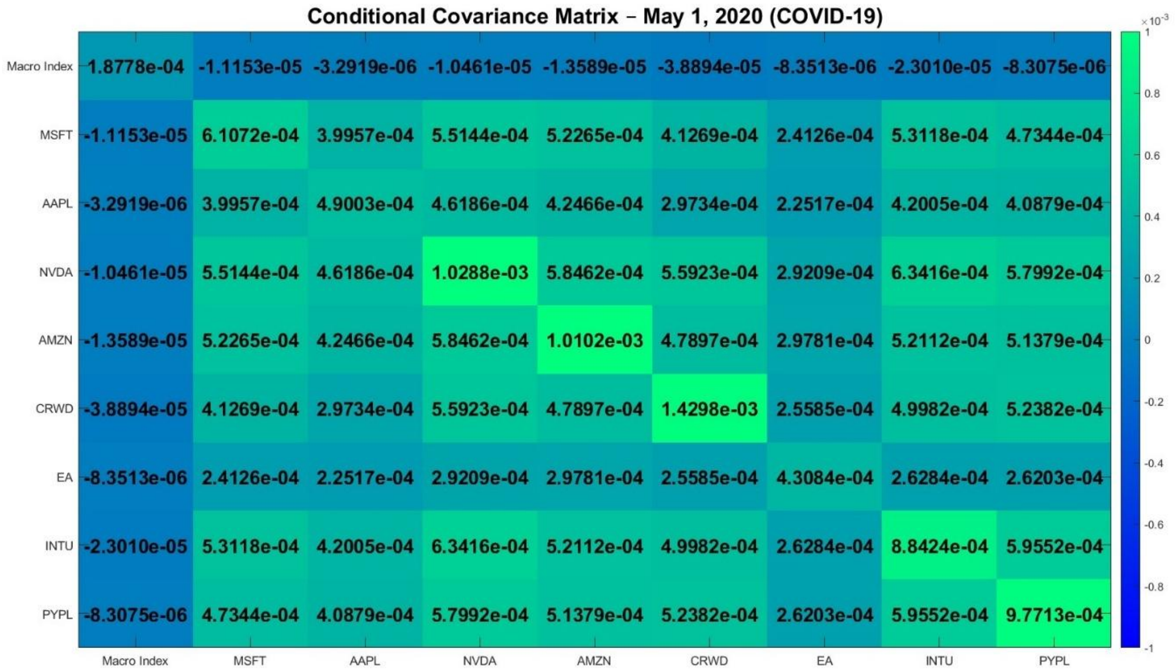


Figure 20. Conditional Dynamic Variance-Covariance Matrix for the calculation of Var and ES, as of May 1, 2020 (Post-Covid).

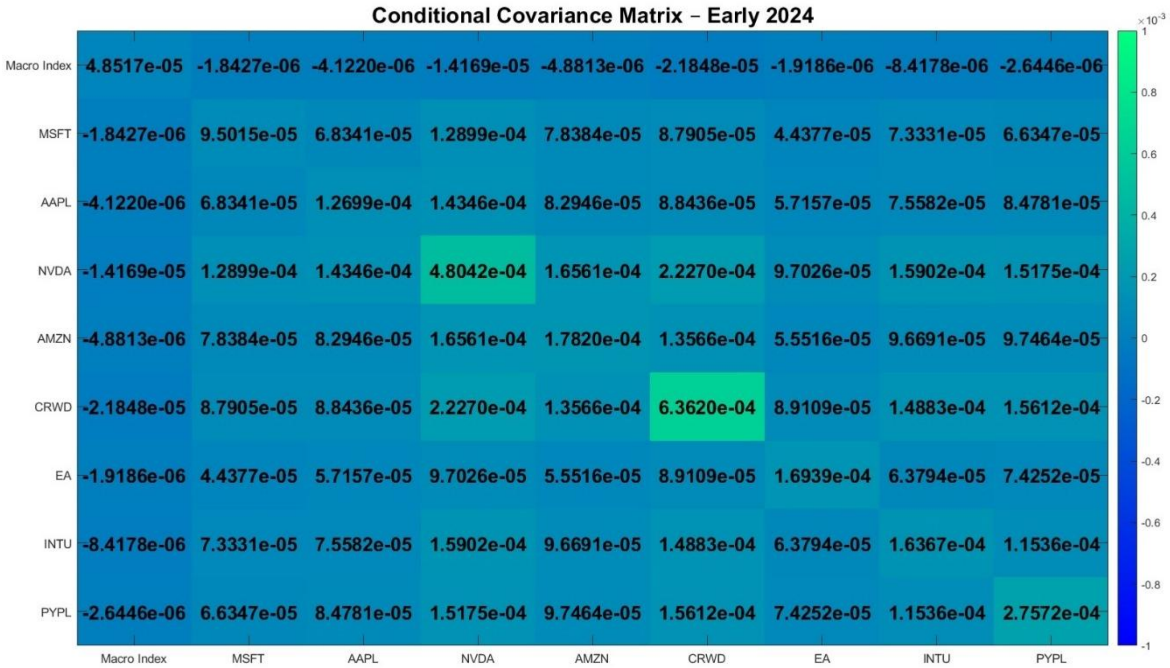


Figure 21. Conditional Dynamic Variance-Covariance Matrix for the calculation of Var and ES, as of January 1, 2024.

Appendix C: Python and MATLAB Codes

In this Appendix we report Python codes used for *Web Scraping* of Nasdaq 100 data and cluster analysis, and MATLAB codes for constructing the macroeconomic-financial index, implementing the DCC and ADCC models, and calculating VaR and ES on an EW portfolio. For estimation of multivariate models, it is recommended to download the *MFE Toolbox* developed by Kevin Sheppard, available on his official website and indicated in the *Dynamic Conditional Correlation Model* section of the Bibliography of this thesis.

The operations shown were replicated on an alternative dataset, in which the months of February, March, and April 2020 were not considered, to assess the effect of the pandemic period on the results achieved.

Python Code for Web Scraping, Cluster Analysis and K-Means

```
## Importing libraries
from datetime import datetime
from math import sqrt
import numpy as np
import pandas as pd,
from pandas_datareader import data as pdr
import yfinance as yf
from scipy.cluster.vq import kmeans,vq
from sklearn.cluster import KMeans
import plotly.express as px
from matplotlib import pyplot as plt
px.default='svg'
yf.pdr_override()

## Web Scraping of Nasdaq 100 tickers
nasdaq100_url = 'https://en.wikipedia.org/wiki/Nasdaq-100'
tickers = (
pd.read_html(nasdaq100_url)[4]['Ticker']
.str.replace('.', '-', regex=False)
.tolist()
)

## Downloading historical data from Yahoo Finance
startdate = datetime(2019, 1, 1)
enddate = datetime(2024, 1, 1)
prices_df = pdr.get_data_yahoo(
tickers,
start=startdate,
end=enddate
)['Adj Close']

# Calculation of average return and annualised volatility
returns = prices_df.pct_change().mean() * 252
```

```
volatility = prices_df.pct_change().std() * sqrt(252)
data = pd.DataFrame(
{'Returns': returns,
'Volatility': volatility
}).dropna()

## Cluster Analysis with K-Means
k = 4
kmeans = KMeans(
n_clusters=k
).fit(data)
data['Cluster'] = kmeans.labels_

## Cluster interpretation
cluster_names = {0:'mid-mid',
                    1:'high-high',
                    2:'low-low',
                    3:'low-high'}
data['Clusters_Name'] = data['Cluster'].map(cluster_names)
```

MATLAB Code for Multivariate Econometric Models

For the econometric analysis, we first imported the necessary macroeconomic variables and stock data, performing the usual data management operations, including calculating log-returns, handling missing values, and interpolating time series where necessary. After properly preparing and transforming the data, we constructed the macroeconomic-financial index by employing the conditional volatilities of the macroeconomic variables, as described in the code below.

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%  DYNAMIC CONDITIONAL CORRELATION MODEL   %%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

clear; clc;

%% Estimation of conditional volatilities of macroeconomic variables to
%% construct the macroeconomic-financial index
% Initialisation of models
models = cell(1, numel(macro_variables));
fits   = cell(1, numel(macro_variables));
vol_forecast      = cell(1, numel(macro_variables));
parameters_fitted = cell(3, numel(macro_variables))

% For loop to train GARCH models and obtain predicted volatilities
for i = 1:numel(macro_variables)

    model = garch('GARCHLags', 1, 'ARCHLags', 1, 'Offset', NaN);
    fit   = estimate(model, macro_variables{i});
    models{i} = model;
    fits{i}   = fit;

    % Extraction of parameters and predicted volatilities
    parameters_fitted(:, i) = {fit.Constant; fit.ARCH{1}; fit.GARCH{1}};
    vol_forecast{i} = sqrt(infer(fit, macro_variables{i}));
```

```
end

% Creation of inverse weights for each variable
weight_gold = 1./mean(vol_forecast{1});
weight_wti  = 1./mean(vol_forecast{2});
weight_vix  = 1./mean(vol_forecast{3});
weight_fvx  = 1./mean(vol_forecast{4});
weight_cpi  = 1./mean(vol_forecast{5});
weight_lois = 1./mean(vol_forecast{6});

% Calculation and normalisation of variable weights
weights_norm = [weight_gold, weight_wti, weight_vix, weight_fvx, weight_cpi,
weight_lois];
weights_norm = weights_norm ./ sum(weights_norm);

% Computation of the macroeconomic-financial index
index_macro = weights_norm(:, 1) .* gold + ...
    weights_norm(:, 2) .* wti + ...
    weights_norm(:, 3) .* vix + ...
    weights_norm(:, 4) .* fvx + ...
    weights_norm(:, 5) .* cpi + ...
    weights_norm(:, 6) .* lois;


%% Rolling Window Correlation
% Grouping variables into a single matrix
ndx_synt = [index_macro, msft, aapl, nvda, amzn, crwd, ea, intu, pypl];

% Defining the length of the rolling window
window_size  = 90;
total_length = length(ndx_synt);

% Matrix initialisation to contain rolling window correlations
rolling_corr = NaN(total_length, 8);

% Calculation of correlations for each rolling window
for i = 1:total_length

    % Calculation of start and end indices of the rolling window
    start_index = max(i - window_size + 1, 1);
    end_index  = min(i + window_size - 1, total_length);

    % Data selection in the mobile window
    window_data = ndx_synt(start_index:end_index, :);

    % Calculation of correlations between the index and assets
    rolling_corr(i, :) = corr(window_data(:, 1), window_data(:, 2:end));

end

% Timetable creation with rolling window correlations between index and asset
Fulldates_datetime = datetime(Fulldates);
rolling_corr_timetable = array2timetable(rolling_corr, 'RowTimes', ...
    Fulldates_datetime, 'VariableNames', stocks_names);
```

```
%% Estimation of the Dynamic Conditional Correlation Model (DCC)
% Calculation of DCC-GARCH model parameters
[parameters_garch, ll_garch, Ht_garch, vcv_garch] = dcc(ndx_synt, [], 1, 0, 1,
1, 0, 1, 2);

% Calculation of Dynamic Correlations
for t = 1:length(ndx_synt)
    for j = 2:size(ndx_synt, 2)
        if j == 1
            DCC_GARCH(t, j) = 1;
        else
            DCC_GARCH(t, j) = Ht_garch(1, j, t) / (sqrt(Ht_garch(1, 1, t)) ...
                * sqrt(Ht_garch(j, j, t)));
        end
    end
end

% Alpha and Beta parameter extraction of the DCC-GARCH Model
alpha_dccgarch = parameters_garch(end-1);
beta_dccgarch  = parameters_garch(end);

% Calculation of standard errors (SE)
se_alpha_dccgarch = sqrt(vcv_garch(end-1, end-1));
se_beta_dccgarch  = sqrt(vcv_garch(end, end));

% Calculation of t-statistics
t_alpha_dccgarch = alpha_dccgarch / se_alpha_dccgarch;
t_beta_dccgarch  = beta_dccgarch / se_beta_dccgarch;

% Calculation of p-values with nommal distribution
p_value_alpha_dccgarch = 2 * (1 - normcdf(abs(t_alpha_dccgarch)));
p_value_beta_dccgarch  = 2 * (1 - normcdf(abs(t_beta_dccgarch)));
```

In Chapter 2, we implemented and compared several extended models of the DCC-GARCH in order to identify, by RMSE, the most suitable one to capture the dynamics of the correlations between the variables considered. In the following, we only report the main inputs used for the development of the various models, as the relevant dynamic correlations, parameters and significance tests were calculated following the same procedure as just described for the DCC-GARCH:

```
%% Estimation of the DCC-TARCH Model
[parameters_tarch, ll_tarch, Ht_tarch, vcv_tarch] = dcc(ndx_synt, [], 1, 0, 1,
1, 1, 1, 1);

%% Estimation of the DCC-GJR-GARCH Model
[parameters_gjr, ll_gjr, Ht_gjr, vcv_gjr] = dcc(ndx_synt, [], 1, 0, 1, 1, 1, 1,
2);

%% Estimation of the A(symmetric)DCC-GARCH Model
[parameters_agarch, ll_agarch, Ht_agarch, vcv_agarch] = dcc(ndx_synt, [], 1, 1,
1, 1, 0, 1, 2);

%% Estimation of the ADCC-TARCH Model
[parameters_atarch, ll_atarch, Ht_atarch, vcv_atarch] = dcc(ndx_synt, [], 1, 1,
1, 1, 1, 1, 1);
```

```
%% Estimation of the ADCC-GJR-GARCH Model
[parameters_agjr, ll_agjr, Ht_agjr, vcv_agjr] = dcc(ndx_synt, [], 1, 1, 1, 1,
1, 1, 2);
```

The implementation of the models for the study of conditional second moments was inspired by Kevin Sheppard's MFE Toolbox, available on GitHub.

```
%% RMSE comparison of multivariate models
% RMSE calculation for each model
RMSE_GARCH  = rmse(DCC_GARCH(:,  2:end), rolling_corr, 'omitnan');
RMSE_TARCH  = rmse(DCC_TARCH(:, 2:end), rolling_corr, 'omitnan');
RMSE_GJR    = rmse(DCC_GJR(:, 2:end), rolling_corr, 'omitnan');
RMSE_AGARCH = rmse(ADCC_GARCH(:, 2:end), rolling_corr, 'omitnan');
RMSE_ATARCH = rmse(ADCC_TARCH(:, 2:end), rolling_corr, 'omitnan');
RMSE_AGJR   = rmse(ADCC_GJR(:, 2:end), rolling_corr, 'omitnan');

% Identification of optimal model
model_names = {'DCC_GARCH', 'DCC_TARCH', 'DCC_GJR', 'ADCC_GARCH', 'ADCC_TARCH',
'ADCC_GJR'};
metric_models = array2table(sum([RMSE_GARCH; RMSE_TARCH; RMSE_GJR; RMSE_AGARCH;
RMSE_ATARCH; RMSE_AGJR], 2), 'VariableNames', {'RMSE'});
metric_models.Properties.RowNames = model_names;
[~, best_model_index] = min(metric_models.RMSE);
```

VAR-DCC-GARCH

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%      VECTOR AUTOREGRESSIVE MODEL      %%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%% Calculation of the ADF Test for each time series to check for the
%% presence of unit root
% Table initialisation to store results
stocks_names_ADF = {'Indice Macro', 'Msft', 'Aapl', 'Nvda', 'Amzn', 'Crwd', ...
    'Ea', 'Intu', 'Pypl'};
num_assets  = length(stocks_names_ADF);
results_ADF = table('Size', [num_assets, 4], 'VariableTypes', {'double', ...
    'double', 'double', 'double'}, 'VariableNames', {'H0', 'p-Value', ...
    'Test Statistic', 'Critical Value'}, 'RowNames', stocks_names_ADF);

% Running ADF Tests on each historical series
for i = 1:num_assets

    [h, pValue, stat, cValue] = adftest(ndx_synt(:, i));
    results_ADF.('H_{0}')(i) = h;
    results_ADF.('p-value')(i) = pValue;
    results_ADF.('Test Statistic')(i) = stat;
    results_ADF.('Critical Value')(i) = cValue;

end
```

```
% Verification of presence of unit root for each time series
for i = 1:size(ndx_synt, 2)

    [h, pValue, stat, cValue] = adftest(ndx_synt(:,i));

end

% Definition of lag range to be considered for VAR implementation
maxLags = 10;

% Definition of vectors for AIC and BIC
aic  = zeros(maxLags, 1);
bic  = zeros(maxLags, 1);

% For loop execution on different lags
for p = 1:maxLags

    % Estimation of the VAR(p) Model
    Mdl = varm(size(ndx_synt, 2), p);
    EstMdl = estimate(Mdl, ndx_synt);

    % Calculation of information criteria
    results = summarize(EstMdl);
    aic(p)  = results.AIC;
    bic(p)  = results.BIC;

end

% Finding the optimal number of lags for each information criterion
[~, bestLagAIC] = min(aic);
[~, bestLagBIC] = min(bic);
optimalLags = bestLagBIC;

% VAR Model estimation with optimal lag and calculation of residuals
Mdl_VAR = varm(size(ndx_synt, 2), optimalLags);
EstMdl_VAR = estimate(Mdl_VAR, ndx_synt);
residuals_VAR = infer(EstMdl_VAR, ndx_synt);

% DCC-GARCH Model estimation on VAR residuals
[parameters_garch_VAR, ll_garch_VAR, Ht_garch_VAR, vcv_garch_VAR] =
dcc(residuals_VAR, [], 1, 0, 1, 1, 0, 1, 2);

% Calculation of dynamic correlations of the VAR-DCC-GARCH Model
for t = 1:length(ndx_synt) - optimalLags
    for j = 2:size(ndx_synt, 2)
        if j == 1
            DCC_GARCH_VAR(t, j) = 1;
        else
            DCC_GARCH_VAR(t, j) = Ht_garch_VAR(1, j, t) ./ (sqrt ...
                (Ht_garch_VAR(1, 1, t)) * sqrt(Ht_garch_VAR(j, j, t)));
        end
    end
end

% Alpha and Beta parameter extraction of the VAR-DCC-GARCH Model
alpha_vardccgarch = parameters_garch_VAR(end-1);
beta_vardccgarch  = parameters_garch_VAR(end);
```

```matlab
% Calculation of standard errors (SE)
se_alpha_vardccgarch = sqrt(vcv_garch_VAR(end-1, end-1));
se_beta_vardccgarch  = sqrt(vcv_garch_VAR(end, end));

% Calculation of t-statistics
t_alpha_vardccgarch = alpha_vardccgarch / se_alpha_vardccgarch;
t_beta_vardccgarch  = beta_vardccgarch / se_beta_vardccgarch;

% Calculation of p-values with normal distribution
p_value_alpha_vardccgarch = 2 * (1 - normcdf(abs(t_alpha_vardccgarch)));
p_value_beta_vardccgarch  = 2 * (1 - normcdf(abs(t_beta_vardccgarch)));


%% Comparison of DCC-GARCH and VAR-DCC-GARCH by Likelihood Ratio Test
dof = optimalLags; % must coincide with lags
[h_lrtest, pValue_lrtest] = lratiotest(ll_garch_VAR, ll_garch, dof);
```

MATLAB Code for Calculating Value at Risk (VaR) and Expected Shortfall (ES)

The implementation below was developed following the guidelines provided by MathWorks for estimating VaR and ES using a parametric approach. For further details, official documentation is available at Estimate VaR for Equity Portfolio Using Parametric Methods.

```matlab
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%                VaR and ES               %%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%% Input
% Equally-Weighted portfolio definition without transaction costs, with
% total investment of $10,000
Weights    = ones(1, 9) / 9;
Investment = 1e4;
portfolioReturns = sum(ndx_synt .* Weights, 2);
portfolioValue   = Investment * (1 + portfolioReturns);
portfolioPnL     = diff(portfolioValue);

% Definition of significance levels
alpha_5 = 0.05;
alpha_1 = 0.01;

% Function to calculate VaR
VaR_formula = @(mu, sigma, alpha) - (mu + sigma * norminv(1 - alpha));

% VaR integration function to calculate ES
ES_formula = @(mu, sigma, alpha) + (1 / alpha) * integral(@(p) VaR_formula(mu,
sigma, p), 0, alpha);


%% Calculation of Dynamic VaR and ES for the selected model
% For loop to calculate VaR and ES for each time point
for i = 1:length(Ht_garch_VAR)

    % Extraction of the dynamic conditional covariance matrix
    covarianceMatrix_dyn = Ht_garch_VAR(:, :, i);
```

```
    % Calculation of portfolio standard deviation
    portfolioSigma_dyn = sqrt(Weights * covarianceMatrix_dyn * Weights');
    portfolioMu_dyn = portfolioReturns(i);

    % Calculation of VaR and ES for a 1% significance level
    var_mov_1_dyn(i) = VaR_formula(portfolioMu_dyn, portfolioSigma_dyn,
alpha_1) * Investment;

    es_mov_1_dyn(i)  = ES_formula(portfolioMu_dyn, portfolioSigma_dyn, alpha_1)
* Investment;

    % Calculation of VaR and ES for a 5% significance level
    var_mov_5_dyn(i) = VaR_formula(portfolioMu_dyn, portfolioSigma_dyn,
alpha_5) * Investment;

    es_mov_5_dyn(i)  = ES_formula(portfolioMu_dyn, portfolioSigma_dyn, alpha_5)
* Investment;

end

% Calculation of dynamic VaR and ES averages for each significance level
mean_var_5 = mean(var_mov_5_dyn);
mean_var_1 = mean(var_mov_1_dyn);
mean_es_5  = mean(es_mov_5_dyn);
mean_es_1  = mean(es_mov_1_dyn);
```