Dipartimento di Economia e Finanza

Cattedra di Empirical Finance

Multivariate GARCH and Portfolio Optimization

RELATORE

Prof. Santucci De Magistris Paolo

CANDIDATO

Edoardo Brardinoni
Matricola N. 724101

CORRELATORE

Prof. Proietti Tommaso

Anno Accademico 2021/22

ABSTRACT

Multivariate GARCH models (MGARCH) have been a widely used and effective method in the field of finance for modelling conditional volatility since their introduction by Bollerslev et al. (1988). This thesis is divided into two main parts. The first part provides an extensive theoretical overview of MGARCH models, with a focus on VECH models, DCC and CCC, as well as the BEKK specification of the Diagonal VECH class. In addition, we introduce the Markowitz Efficient Portfolio theory and the Value at Risk (VaR). The second part of the thesis presents an empirical analysis where we apply the DVECH, SBEKK, DBEKK, DCC and CCC models to estimate and forecast the time conditional covariance matrix of seven ETFs. The estimation sample spans from 2012-2019, while the forecasting sample covers January 2020 to December 2022, with a total of 756 daily forecasts conducted using a one-step rolling window. We use covariance forecasts from the five different GARCH models as input in two different Portfolio optimization problems within Markowitz's framework: a minimum variance optimization and an expected return maximization under VaR constraint. In the second case, six different combinations of risk aversions are compared in a two-dimensional risk aversion parameter space, which represents investors' aversion to dispersion of returns and extreme observation respectively.

TABLE OF CONTENTS

1.GARCH MODEL

1.1Introduction

The Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model is an extension of the Autoregressive Conditional Heteroscedasticity (ARCH) model. Both models were developed to model and forecast the volatility of financial time series, but GARCH models allow for more flexibility and complexity in modeling the volatility process.

ARCH models were first proposed by Engle (1982) as a way to model and forecast the volatility of financial time series that exhibit conditional heteroscedasticity, which means that the volatility changes over time.

ARCH models assume that the conditional variance of a time series is a function of its past values. The simplest form of an ARCH model is the ARCH(1) model, which assumes that the conditional variance at time t is a function of the squared residuals (of the returns from a mean process) from the previous time period. GARCH models were first proposed by Bollerslev (1986) as an extension of ARCH models. They add an additional term to the ARCH model that allows the conditional variance to depend on past conditional variances. This allows for more flexibility in modeling the volatility process and for the volatility to persist over time. The simplest form of a GARCH model is the GARCH(1,1) model, which assumes that the conditional variance at time t is a function of the squared residuals from the previous time period and the conditional variance from the previous time period.

Understanding and predicting the temporal dependence in the second-order moments of asset returns, such as variances and covariances, is important in financial econometrics. It is widely accepted that financial volatilities move together over time across assets and markets. Recognizing this feature through a multivariate modeling framework leads to more relevant empirical models than working with separate univariate models. This opens the door to better decision tools in various areas, such as asset pricing, portfolio selection, option pricing, hedging, and risk management. These models are widely used to analyze univariate financial returns. However, when dealing with multiple assets or markets, it is important to consider the dynamic relationships between the volatilities and covariances of these different series. This led tothe development of the Multivariate GARCH (MGARCH) model, which was first proposed by Bollerslev et al. (1988).

MGARCH models are particularly useful for applications that require estimates of conditional variances, covariances, and correlations of multivariate time series. These models have been widely used in the context of systems of financial returns, with many studies applying them to areas such as asset pricing, portfolio selection, risk management, and future hedging. For example, studies have used MGARCH models to estimate the conditional variances and covariances of financial time series to compute Value-at-Risk (VaR) measures for a portfolio of assets.The largest number of implementations of MGARCH models appear in the context of systems of financial returns; for example, Bollerslev et al. (1988), Attanasio (1991), De Santis and Gerard (1997), Hansson and Hordahl (1998), Lien and Tse (2002), Engle and Colacito (2006), Andersen et al. (2007), McNeil et al. (2010), Santos et al. (2012) and Santos et al. (2013), apply these models to selected asset pricing, portfolio selection, risk management, and future hedging applications However, the original MGARCH models were rather flexible, allowing all volatilities and conditional covariances in the model to be related to each other. This made their empirical implementation difficult, as it required the estimation of a large number of parameters. Additionally, their parameters needed to be restricted to guarantee covariance stationarity and positive definiteness of conditional covariance matrices. As a result, many popular MGARCH models implemented to represent the dynamic evolution of volatilities, covariances, and correlations of real systems are restricted in a such way that parameter estimation is feasible and it is easy to guarantee covariance stationarity and positive definiteness of the conditional covariance matrices.

Multivariate GARCH (MGARCH) models are an extension of univariate GARCH models, while univariate GARCH models only consider the volatility of a single time series, MGARCH models take into account the volatility of multiple time series and the relationship between them. This allows for a more comprehensive understanding of the co-movements of the random variables, which is crucial for practical applications such as asset pricing, risk management, and asset allocation.In a multivariate GARCH model, the volatility of each variable is modeled as a function of the past values and past volatility of all the variables. This allows for the modeling of the interdependencies and interactions between the different variables, and the prediction of the volatility of each variable given the past values and past volatility of all the variables.MGARCH models are more complex than univariate GARCH models as they require more data and more parameters to estimate. However, this complexity also brings increased power, as MGARCH models can capture the relationship between the volatilities of different time series and the correlation structure among them. This is essential

for understanding the dependencies and interdependence among multiple time series and making more informed decisions. MGARCH models are a powerful tool for modeling the volatility of multiple time series and understanding the relationship between them. By carefully designing the model and choosing appropriate parameterizations, MGARCH models can provide valuable insights into the dynamics of multiple time series and inform decision-making in various fields. Additionally, model selection techniques such as AIC, BIC, and HQIC can be used to compare different MGARCH models, to select the model that best fits the data while avoiding overfitting.

GARCH model can be covariance stationary even though it is time-varying. A model is considered to be covariance stationary if the mean and the covariance of the process do not change over time. In the case of a GARCH model, the conditional mean of the process is assumed to be constant, while the conditional variance is time-varying. However, if the model is well-specified and the parameters are estimated correctly, the time-varying conditional variance will converge to a stationary process, meaning the mean and the covariance of the process do not change over time. This is achieved by ensuring that the model satisfies the necessary conditions for covariance stationarity, such as having the moduli of the eigenvalues of the model less than one.

Convergence in stationarity in GARCH models refers to the behavior of the model as the number of observations in the dataset increases. A GARCH model is considered to be stationary if the mean and variance are constant over time. In other words, the model's parameters do not change over time and the model's predictions do not depend on the specific time period for which it is being applied.

When a GARCH model is estimated using a finite sample of data, the estimated parameters will be based on the sample and will be subject to some degree of estimation error. However, as the sample size increases, the estimation error is expected to decrease, and the estimated parameters will converge to the true population values. This is known as asymptotic consistency.In addition, when a GARCH model is estimated using a rolling window of data, the estimation process is repeated for different subsets of the data, and the estimated parameters will change over time. However, as the window size increases, the estimated parameters will converge towards a stable value, this is known as asymptotic stationarity

One of the main challenges in developing MGARCH models is finding a balance between flexibility and parsimony. Flexibility is crucial for being able to represent the dynamics of the conditional variances and covariances, but having too many parameters can make the model difficult to estimate, interpret, and generalize. Overfitting is a common issue that can

arise when the number of parameters is too high. On the other hand, having too few parameters may not capture the relevant dynamics in the covariance structure, which can lead to underfitting. Additionally, ensuring the positive definiteness of the conditional covariance matrix is also important, as covariance matrices need to be positive definite by definition. One way to achieve this is to derive conditions under which the conditional covariance matrices implied by the model are positive definite, but this can be infeasible in practice. An alternative is to formulate the model in a way that positive definiteness is implied by the model structure, such as using certain types of parametric distributions or applying constraints on the parameters.

Parsimony is a key consideration in the development of multivariate GARCH (MGARCH) models. Parsimony refers to the principle of using the simplest possible model that can explain the data. In the context of MGARCH models, this means using a model with the smallest number of parameters that can still accurately capture the dynamics of the conditional variances and covariances.

Parsimony, or the principle of using the simplest explanation or solution, is an important concept in MGARCH models for several reasons. Firstly, parsimony can greatly aid in the estimation process by reducing the number of parameters that need to be estimated. This not only speeds up the estimation process but also reduces the risk of overfitting, which occurs when a model is so complex that it fits the data too well but fails to generalize to new data. Additionally, a parsimonious model is also less prone to estimation problems such as multicollinearity, and it reduces the risk of estimation bias.

Secondly, a parsimonious model is also simpler and more transparent, making it easier to understand the relationships between the different variables in the model and the impact of each parameter on the model. This is especially important in MGARCH models, where the number of parameters can be quite large, making it difficult to interpret the model. With a parsimonious model, it is easier to identify the key drivers of the model and the impact of each parameter on the model's output. Thirdly, parsimony can aid in model selection by allowing for the comparison of different models based on the number of parameters they have, with models having fewer parameters being generally preferred. This is because models with fewer parameters are considered to be more parsimonious, and are therefore less likely to be overfitting the data. Additionally, model selection can be done by comparing different models based on their goodness-of-fit statistics such as AIC, BIC, HQIC, which are penalized by the number of parameters. Finally, parsimony also helps to make the model more robust by reducing the number of parameters that need to be estimated, making the

model less sensitive to small changes in the data. This is particularly important in MGARCH models, where the number of parameters can be quite large, making the model more sensitive to small changes in the data. By reducing the number of parameters, parsimony can help to make the model more robust, ensuring that the model's output remains stable even when the data changes.This feature is important in MGARCH models for several reasons, including easier estimation, better interpretability, better model selection, and increased robustness. By using the principle of parsimony, MGARCH models can be made more efficient and accurate, providing better insights into the underlying relationships in the data.

However, parsimony should not be the only consideration in model selection. A model that is too simple may not be able to capture the relevant dynamics in the covariance structure, so it's important to find a balance between parsimony and flexibility.

In MGARCH models, parsimony is usually achieved by using a simple structure for the model, such as assuming that the conditional covariance matrix is diagonal or by assuming that the conditional covariances follow a specific structure, such as a factor model. Additionally, some MGARCH models use shrinkage techniques to reduce the number of parameters and make the model more parsimonious.

Flexibility is a crucial aspect in the design and implementation of Multivariate GARCH (MGARCH) models. It pertains to the capability of the model to adapt to the dynamic nature of conditional variances and covariances. A flexible MGARCH model can account for a vast range of potential dynamics and grasp the intricacies of the underlying relationships between variables. Another vital aspect is “Flexibility”, in MGARCH models it enables the representation of complex dynamics. By allowing for a substantial number of parameters or by enabling a more general structure for the model, a flexible model can represent a wide range of possible dynamics. This can be accomplished through models that allow for time-varying parameters, such as the VECH (Vector Error Correction Heteroskedasticity) model which enables the conditional variances and covariances to change over time, providing a more comprehensive understanding of the underlying dynamics.

Flexibility also allows for the capture of nonlinearities in the data, which is crucial for accurately representing the underlying relationships between variables covariances. Flexibility is also important for handling outliers in the data. Outliers can have a significant impact on the model's output, and a flexible model can handle them more effectively. Lastly, flexibility is also important in MGARCH models for handling asymmetries in the data. Asymmetries can arise due to factors such as seasonality, and a flexible model can handle them more effectively. A model that is too flexible may be too complex and difficult to

estimate and interpret, so it's important to find a balance between flexibility and parsimony.In MGARCH models, flexibility can be achieved by using a more general structure for the model, such as allowing for non-diagonal elements in the conditional covariance matrix, or by allowing for more general structures such as factor models. The last major problem we face in MGARCH Models is the positive definitiveness of the Variance-Covariance Matrix. Positive definiteness is a crucial aspect of MGARCH models, which pertains to the property of the conditional covariance matrix, a key component of the model, that must be met. A matrix (A) is considered positive definite if, for any non-zero vector w, the product w'Aw (A is a matrix) is a positive scalar. This means that all its eigenvalues must be positive. In particular, a matrix A is positive definite if it satisfies one of the following conditions:

- A is congruent with a diagonal matrix with positive real entries.
- A is symmetric or Hermitian, and all its eigenvalues are real and positive.
- A is symmetric or Hermitian, and all its leading principal minors are positive.
- There exists an invertible matrix Z with conjugate transpose Z^* such that $A = W^* W$

In the case of MGARCH models, the conditional covariance matrix is a vital quantity that must be estimated and must be positive definite.

There are several reasons why positive definiteness is crucial in MGARCH models. It is a fundamental requirement for covariance matrices by definition. Therefore, MGARCH models must ensure that the conditional covariance matrix is positive definite to ensure that the model is well-defined and not affected by any mathematical or computational errors. Positive definite covariance matrices are also necessary for the implementation of various statistical techniques such as maximum likelihood estimation and the Kalman filter.Positive definiteness is important for interpreting the model and its parameters. A positive definite conditional covariance matrix ensures that the variances and covariances of the variables in the model are well-defined and positive, which is crucial for understanding the underlying relationships between the variables. Additionally, positive definiteness ensures that the model can be used for prediction and forecasting, as the forecast variances and covariances are well-defined and positive.

It is also vital for estimation, as it is one of the prerequisites for the model to be identifiable and estimable, which is essential for the accuracy and reliability of the model's output. Positive definiteness also ensures that the estimation algorithm converges to the global

optimum, rather than getting stuck in a local optimum, leading to better estimation of the model's parameters.Positive definiteness also enables the use of statistical techniques such as Value-at-Risk (VaR) and Expected Shortfall (ES) for risk management.

There are several ways to ensure positive definiteness in MGARCH models. One way is to derive conditions under which the conditional covariance matrices implied by the model are positive definite, but this is often infeasible in practice. An alternative is to formulate the model in a way that positive definiteness is implied by the model structure. For example, using a factor model structure for the MGARCH model can ensure positive definiteness. Some MGARCH models use constraints on the parameters of the model to ensure positive definiteness, such as the Dynamic Conditional Correlation (DCC) model, which uses a correlation matrix as the parameter of the model, ensuring that the correlation matrix is always positive definite.Finding a good balance between Parsimony, Flexibility and the requirement of Positive Definiteness is the main challenge practitioners face when building their own MGARCH Model.

1.2 General Framework

Considering an Nx1 random vector y_t, Conditioned on the Information Set J_{t-1} ; we define Θ as a finite set of parameters and write our random process as:

$$\boldsymbol{y_t} = \mu_{\mathbf{t}}(\Theta) + \boldsymbol{\varepsilon_t}$$

Such that $\mu_t(\Theta)$ is the conditional mean of y_t, and ε_t is an error term with Expected value equal zero and conditional variance given by a Positive definite Symmetric Matrix $H_{t.}$. The randomness of the process is given by the random component ε_t, which contains the estimation error of $\mu_t(\Theta)$ as well (which is composed by a forecasting error as well if the conditional mean is obtained through an autoregressive process).

If the random vector y_t $is\ mean\ stationary$ then:

$E[\boldsymbol{y_t}] = E[\boldsymbol{y_t}|J_{t-1}] = \mu(\Theta)$

i.e. The conditional Expectation is equal to the unconditional expectation, otherwise:

$E[y_t|J_{t-1}] = \mu_t(\Theta)$

In both cases we have that the Expectation of the random component is equal to the zero vector:

$$E[\boldsymbol{\varepsilon_t}] = \mathbf{0}$$

Since this is the component that brings randomness in the process, MGARCH models are interested in modeling it. In fact, regardless of the particular model of interest, all the MGARCH propose different equations to define the variance carried by ε_t.

In general, we define ε_t as:

$$\boldsymbol{\varepsilon_t} = \boldsymbol{H_t^{1/2}}(\Theta)\boldsymbol{\eta_t}$$

Where $H_t^{1/2}(\Theta)$ is a NxN Positive definite symmetric matrix, which is the matrix counterpart of the standard deviation. In fact $H_t(\Theta)$ is the conditional variance of y_t, and we can obtain $H_t^{1/2}(\Theta)$ through its Choleski decomposition.

η_t is a Nx1 white noise process:

$$E[\boldsymbol{\eta_t}] = \mathbf{0}$$
$$\text{Var}[\boldsymbol{\eta_t}] = \boldsymbol{I_N}$$

With I_N being the NxN Identity matrix.

It follows that:

$$\text{Var}[\boldsymbol{y_t}|I_{t-1}] = \text{Var}[\boldsymbol{\varepsilon_t}|I_{t-1}]$$
$$= \boldsymbol{H_t^{1/2}}(\Theta)Var[\boldsymbol{\eta_t}|I_{t-1}]\left(\boldsymbol{H_t^{1/2}}(\Theta)\right)^T$$
$$= \boldsymbol{H_t^{1/2}}(\Theta)\boldsymbol{I_N}\left(\boldsymbol{H_t^{1/2}}(\Theta)\right)^{\boldsymbol{T}}$$
$$= \mathbf{H_t}(\Theta)$$

This shows that $H_t(\Theta)$ is the conditional variance of the random process.

It's worth noting that both $\mu_t(\theta)$ and $H_t(\theta)$ depend on the parameter vector θ, however such a vector can be split into two components θ_1 and θ_2, such that we have $\mu_t(\theta 1)$ and $H_t(\theta 2)$; they can also have some (or all) parameters in common. A particular case are the GARCH-in-mean Models in which the conditional mean is functionally dependent on the conditional variance.

The aim of the MGARCH models is to define $Ht(\theta)$, they take different approaches, Each of these models has its own strengths and weaknesses and is suited to different types of data and research questions.

We can split them into four main categories:

1) The ones that model the conditional covariance matrix directly , such as VEC and BEKK; this class is a direct generalizations of the univariate GARCH model of Bollerslev (1986). These models are among the first parametric MGARCH models developed and have been widely used in financial and economic applications such as portfolio optimization, risk management and asset pricing.

The VECH (Vector Error Correction Higher Order) is a dynamic factor model that allows for the modeling of the co-movements between multiple time series. It uses a vector error correction mechanism, which is a statistical technique that captures the long-term relationships between variables and the dynamic interactions among them. This makes the VECH model particularly useful for understanding the long-term dynamics of the system being studied.The BEKK (Baba, Engle, Kraft, and Kroner) is a multivariate GARCH model that allows for the modeling of the conditional covariances between multiple time series. This model uses a bilinear structure to model the conditional covariances, which allows for the modeling of the dynamic interactions among the variables. This makes the BEKK model particularly useful for understanding the short-term dynamics of the system being studied. Both the VEC and BEKK models have been widely used in financial and economic applications, such as portfolio optimization, risk management, and asset pricing. These models provide a powerful tool for understanding the underlying relationships between multiple time series and for making more informed decisions based on those relationships.

2) Factor models are motivated by parsimony, meaning that the random process ε_t is assumed to be generated by a limited number of unobserved heteroskedastic factors. This is in contrast to models that estimate the conditional variances and covariances directly, which can result in a large number of parameters and a complex model structure.

The use of factor models in MGARCH aims to simplify the model by reducing the number of parameters, making it more parsimonious. This not only improves the interpretability of the model but also helps to reduce the risk of overfitting. Additionally, it can also aid in the estimation process by reducing the number of parameters that need to be estimated, making it more computationally efficient.The factor model structure can be used to capture the underlying dynamics of the co-movements between multiple time series by assuming that the observed time series are driven by a smaller number of unobserved factors

3) The third class of models are the ones aiming to model the conditional variances and correlations instead of modeling the conditional covariance matrix directly. One example of this class of models is the Constant Conditional Correlation (CCC) model and its extensions such as the Dynamic Conditional Correlation (DCC) . This models that may be viewed as nonlinear combinations of univariate GARCH model. This type of model aims to simplify the estimation process by only modeling the conditional variances and correlations, rather than the entire covariance matrix.The appeal of this class of models lies in the intuitive interpretation of the correlations. The CCC model, for example, assumes that the correlations between the variables are constant over time, making it easy to interpret and understand. Additionally, models belonging to this class have received significant attention in the recent literature. The CCC model and its extensions have been widely used in various applications in Finance.

4) Semiparametric and Nonparametric models, which do not rely on a specific functional form for the conditional covariance matrix. These models are designed to address the limitations of parametric models in terms of misspecification of the structure of the conditional covariance matrices and in cases where the underlying dynamics of the conditional variances and covariances are not clear. In particular, Semiparametric models combine the strengths of both parametric and non-parametric models, such as consistency and interpretability of the parametric model and robustness against distributional misspecification of the non-parametric model. This allows for more efficient estimation of the conditional variances and covariances while being less sensitive to misspecification of the model structure.

1.3 Estimation

The most commonly used method for estimating the parameters of a GARCH models is maximum likelihood estimation (MLE), Another method is Quasi Maximum Likelihood Estimation (QMLE) which is a variant of MLE and it is robust against some distributions assumptions and is usually used in absence of Normality assumptions on the distributions of the innovations. QMLE is an estimation method that is used when the assumption of i.i.d. observations is not met or under possible misspecification of the probability density function. QMLE uses a simplified likelihood function that is based on the conditional densities of the observations given the parameters and the variances at each time point. By assuming the conditional normality of the observations, the QMLE can be applied to MGARCH models, even though the observations are not strictly i.i.d.The QMLE is consistent estimator under certain conditions and is asymptotically equivalent to the MLE.In particular, in MGARCH models, the assumption of independent and identically distributed (i.i.d.) observations is not met as the variance is time dependent, however the observation can be considered to be i.i.d in order to use MLE; since, despite being time-varying as long as the specific model is covariance-stationary the conditional covariance will converge to its population unconditional value.In order to apply MLE estimation some regulatory conditions must be satisfied, such conditions are satisfied assuming an i.i.d Gaussian distribution of the random process.. The MLE provides a tool to obtain an asymptotically efficient estimations of the parameters, in particular an estimator is defined Asymptotically efficient if it is consistent, asymptotically normally distributed and has asymptotic variance covariance matrix that is no larger than the one obtained through another consistent and asymptotically normally distributed estimator :

$$plim\ \hat{\boldsymbol{\theta}} = \boldsymbol{\theta_0}$$

$$\hat{\boldsymbol{\theta}} \overset{a}{\sim} N\left[\boldsymbol{\theta_0}, \left\{-E\left[\frac{\partial^2 \ln\left[L(\boldsymbol{\theta_0} \mid \boldsymbol{\varepsilon})\right]}{\boldsymbol{\partial\theta_0 \partial\theta_0'}}\right]\right\}^{-1}\right]$$

Where $\boldsymbol{\theta_0}$ are our true, unobservable parameters. We start by assuming ε_t is distributed as a normal with mean μ and variance σ^2 (both scalars), whose probability density function is given by:

$$f(\varepsilon_t) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{(\varepsilon_t - \mu)}{\sigma}\right)^2\right)$$

Defining $f(\varepsilon_t|\theta_0)$ the probability density function of ε_t under the true parameters, the joint density function of T i.i.d. random variables is given by:

$$L(\boldsymbol{\theta_0} \mid \varepsilon) = \prod_{t=1}^{T} f\,(\varepsilon_t \mid \boldsymbol{\theta_0})$$

This is the likelihood function. MLE, however, works with the natural logarithm of the likelihood function:

$$\ln\,[L(\boldsymbol{\theta_0} \mid \varepsilon)] = \sum_{t=1}^{T} \ln\,[f(\varepsilon_t \mid \boldsymbol{\theta_0})]$$

Assuming our random variables are normally distributed the log-likelihood function is given by:

$$\ln\,[L(\boldsymbol{\theta} \mid \boldsymbol{\varepsilon})] = \sum_{t=1}^{T} \ln\,[f(\varepsilon_t \mid \boldsymbol{\theta})] = -\frac{1}{2}\sum_{t=1}^{T}\left[\ln\,(\sigma^2) + \ln\,(2\pi) + \frac{(\varepsilon_t - \mu)}{\sigma^2}\right]$$

This is the likelihood function, a function of the sample data ε and of the unknown parameter vector θ. The problem we want to solve is:

$$\hat{\theta} = argmax_{\Phi} L(\boldsymbol{\varepsilon}, \boldsymbol{\theta})$$

Where Φ is a compact set. This is the function we want to maximize respect to the parameter vector, the values of the parameters that maximize the log-likelihood function are our MLE estimates $\hat{\theta}$. Since the logarithm is a monotonic function the values that maximize the likelihood function are the same that maximize the log-likelihood. The first order condition is:

$$\frac{\partial \ln\left[L(\boldsymbol{\theta_0} \mid \boldsymbol{\varepsilon})\right]}{\boldsymbol{\partial \theta_0}} = \mathbf{0}$$

Since θ_0 is the true parameter which maximizes the likelihood function such that its gradient is equal to zero.

By defining $A(\theta_0)$ and $B(\theta_0)$ the two extreme values that our ε can assume in the sample, by definition of CDF:

$$\int_{A(\theta_0)}^{B(\theta_0)} f(\varepsilon_t \mid \boldsymbol{\theta_0}) d\varepsilon_t = 1$$

By differentiating the equation left and right by θ_0, and applying Leibnitz's theorem we get on the LHS:

$$\frac{\partial \int_{A(\theta_0)}^{B(\theta_0)} f(\varepsilon_t \mid \boldsymbol{\theta_0})}{\boldsymbol{\partial \theta_0}} d\varepsilon_t = \int_{A(\theta_0)}^{B(\theta_0)} \frac{\partial f(\varepsilon_t \mid \boldsymbol{\theta_0})}{\boldsymbol{\partial \theta_0}} d\varepsilon_t + f(B(\boldsymbol{\theta_0}), \boldsymbol{\theta_0}) \frac{\partial B(\boldsymbol{\theta_0})}{\boldsymbol{\partial \theta_0}} - f(A(\boldsymbol{\theta_0}), \boldsymbol{\theta_0}) \frac{\partial A(\boldsymbol{\theta_0})}{\partial \boldsymbol{\theta_0}}$$

And on the RHS:

$$\frac{\partial 1}{\boldsymbol{\partial \theta_0}} = \mathbf{0}$$

Which leads to:

$$\frac{\partial \int_{A(\theta_0)}^{B(\theta_0)} f(\varepsilon_t | \boldsymbol{\theta_0})}{\boldsymbol{\partial \theta_0}} d\varepsilon_t = \mathbf{0}$$

This implies interchangeability of integration and differentiation as long as $\partial \mathrm{A}(\theta_0)/\partial\theta_0 = \partial \mathrm{B}(\theta_0)/\partial\theta_0 = 0$, i.e., as long as the range of the sample data does not depend on the parameters. If such condition holds, we can rewrite the equation as:

$$\frac{\partial \int f(\varepsilon_t \mid \boldsymbol{\theta_0})}{\boldsymbol{\partial \theta_0}} d\varepsilon_t = \frac{\int \partial f(\varepsilon_t \mid \boldsymbol{\theta_0})}{\boldsymbol{\partial \theta_0}} d\varepsilon_t$$

$$= \int \frac{1}{f(\varepsilon_t \mid \boldsymbol{\theta_0})} \cdot \frac{\partial f(\varepsilon_t \mid \boldsymbol{\theta_0})}{\partial \boldsymbol{\theta_0}} f(\varepsilon_t \mid \boldsymbol{\theta_0}) d\varepsilon_t$$

$$= \int \frac{\partial \ln [f(\varepsilon_t \mid \boldsymbol{\theta_0})]}{\partial \boldsymbol{\theta_0}} f(\varepsilon_t \mid \boldsymbol{\theta_0}) d\varepsilon_t$$

$$= E\left[\frac{\partial \ln [f(\varepsilon_t \mid \boldsymbol{\theta_0})]}{\partial \boldsymbol{\theta_0}}\right] = \mathbf{0}$$

This shows that if evaluated in $\boldsymbol{\theta_0}$, the expected value of the score is equal to zero:

$$E\left[\boldsymbol{g_{0,t}}\right] = \mathbf{0}$$

Where $\boldsymbol{g_{0,t}}$ is the mx1 gradient vector of the log-likelihood function of the t observation of our sample with respect to the parameter vector, evaluated in $\boldsymbol{\theta_0}$. In order to obtain the variability of the score with respect to the parameters, we differentiate the former result obtaining the Hessian matrix:

$$\partial(\int \frac{\partial \ln [f(\varepsilon_t \mid \boldsymbol{\theta_0})]}{\partial \boldsymbol{\theta_0}} f(\varepsilon_t \mid \boldsymbol{\theta_0}) d\varepsilon_t) / \partial \boldsymbol{\theta_0'}$$

$$= \int \left[\frac{\partial^2 \ln [f(\varepsilon_t \mid \boldsymbol{\theta_0})]}{\partial \boldsymbol{\theta_0} \partial \boldsymbol{\theta_0'}} \cdot f(\varepsilon_t \mid \boldsymbol{\theta_0}) + \frac{\partial \ln [f(\varepsilon_t \mid \boldsymbol{\theta_0})]}{\partial \boldsymbol{\theta_0}} \cdot \frac{\partial f(\varepsilon_t \mid \boldsymbol{\theta_0})}{\partial \boldsymbol{\theta_0'}}\right] \cdot d\varepsilon_t$$

$$= \int \left[\frac{\partial^2 \ln [f(\varepsilon_t \mid \boldsymbol{\theta_0})]}{\partial \boldsymbol{\theta_0} \partial \boldsymbol{\theta_0'}} \cdot f(\varepsilon_t \mid \boldsymbol{\theta_0}) + \frac{\partial \ln [f(\varepsilon_t \mid \boldsymbol{\theta_0})]}{\partial \boldsymbol{\theta_0}} \frac{\partial \ln [f(\varepsilon_t \mid \boldsymbol{\theta_0})]}{\partial \boldsymbol{\theta_0'}} \cdot f(\varepsilon_t \mid \boldsymbol{\theta_0})\right] \cdot d\varepsilon_t = \mathbf{0}$$

Since the integral of a sum is the sum of the integrals, we split the integral and take the first part on the RHS of the equation:

$$\int \frac{\partial \ln [f(\varepsilon_t \mid \boldsymbol{\theta_0})]}{\partial \boldsymbol{\theta_0}} \cdot \frac{\partial \ln [f(\varepsilon_t \mid \boldsymbol{\theta_0})]}{\partial \boldsymbol{\theta_0'}} \cdot f(\varepsilon_t \mid \boldsymbol{\theta_0}) \cdot d\varepsilon_t$$

$$= -\int \frac{\partial^2 \ln [f(\varepsilon_t \mid \boldsymbol{\theta_0})]}{\partial \boldsymbol{\theta_0} \partial \boldsymbol{\theta_0'}} \cdot f(\varepsilon_t \mid \boldsymbol{\theta_0}) \cdot d\varepsilon_t$$

Looking at the LHS of the equation above and remembering that:

$$E\left[\frac{\partial \ln\left[f(\varepsilon_t \mid \boldsymbol{\theta_0})\right]}{\partial \boldsymbol{\theta_0}}\right] = \mathbf{0}$$

We get that the variance of the gradient of the log-likelihood is given by:

$$\text{Var}\left[\frac{\partial \ln\left[f(\varepsilon_t \mid \boldsymbol{\theta_0})\right]}{\partial \boldsymbol{\theta_0}}\right] = \int \frac{\partial \ln\left[f(\varepsilon_t \mid \boldsymbol{\theta_0})\right]}{\partial \boldsymbol{\theta_0}} \cdot \frac{\partial \ln\left[f(\varepsilon_t \mid \boldsymbol{\theta_0})\right]}{\partial \boldsymbol{\theta_0'}} \cdot f(\varepsilon_t \mid \boldsymbol{\theta_0}) \cdot d\varepsilon_t$$

While:

$$\int \frac{\partial^2 \ln\left[f(\varepsilon_t \mid \boldsymbol{\theta_0})\right]}{\partial \boldsymbol{\theta_0} \partial \boldsymbol{\theta_0'}} \cdot f(\varepsilon_t \mid \boldsymbol{\theta_0}) \cdot d\varepsilon_t = E\left[\frac{\partial^2 \ln\left[f(\varepsilon_t \mid \boldsymbol{\theta_0})\right]}{\partial \boldsymbol{\theta_0} \partial \boldsymbol{\theta_0'}}\right]$$

Which leads to the general expression for the Variance:

$$\text{Var}\left[\frac{\partial \ln\left[f(\varepsilon_t \mid \boldsymbol{\theta_0})\right]}{\partial \boldsymbol{\theta_0}}\right] = -E\left[\frac{\boldsymbol{\partial^2 ln}\left[\boldsymbol{f}(\boldsymbol{\varepsilon_t} \mid \boldsymbol{\theta_0})\right]}{\partial \boldsymbol{\theta_0} \partial \boldsymbol{\theta_0'}}\right] = -E\left[\boldsymbol{H_{0,t}}\right]$$

Where $\mathbf{H_{0,t}}$ indicates the mxm Hessian matrix of the log-likelihood for the t observation, evaluated in $\boldsymbol{\theta_0}$.This shows that the expected value of the Hessian in t is the variance of the log-likelihood in t. Now we generalize the result to obtain the score and Hessian for our whole sample t=1,…,T:

$$\ln\left[L(\boldsymbol{\theta} \mid \boldsymbol{\varepsilon})\right] = \sum_{t=1}^{T} \ln\left[f(\varepsilon_t \mid \boldsymbol{\theta})\right]$$

$$\boldsymbol{g} = \frac{\partial \ln\left[L(\boldsymbol{\theta} \mid \boldsymbol{\varepsilon})\right]}{\partial \boldsymbol{\theta}} = \frac{\sum_{i=1}^{T} \partial \ln\left[f(\varepsilon_t \mid \boldsymbol{\theta})\right]}{\partial \boldsymbol{\theta}} = \sum_{t=1}^{T} \boldsymbol{g_t}$$

Since under $\boldsymbol{\theta_0}$ each $\boldsymbol{g_{0,t}}$ has expected value equal to zero, their sum is equal to zero:

$\boldsymbol{E}[\boldsymbol{g_0}] = E\left[\sum_{t=1}^{T} \boldsymbol{g_{0,t}}\right] = \mathbf{0}$

As for the variance:

$$H = \frac{\partial^2 \ln\left[L(\boldsymbol{\theta} \mid \boldsymbol{\varepsilon})\right]}{\boldsymbol{\partial\theta\partial\theta'}} = \frac{\sum_{i=1}^{T} \partial^2 \ln\left[f(\varepsilon_t \mid \boldsymbol{\theta})\right]}{\boldsymbol{\partial\theta\partial\theta'}} = \sum_{t=1}^{T} \boldsymbol{H_t}$$

Which under θ_0 becomes:

$$E[\boldsymbol{H_0}] = \mathrm{E}\left[\sum_{t=1}^{T} \boldsymbol{H_{0,t}}\right]$$

$$-E[\boldsymbol{H_0}] = \mathrm{Var}\left[\frac{\partial \ln\left[L(\boldsymbol{\varepsilon} \mid \boldsymbol{\theta_0})\right]}{\boldsymbol{\partial\theta_0}}\right] = -E\left[\frac{\partial^2 \ln\left[L(\boldsymbol{\varepsilon} \mid \boldsymbol{\theta_0})\right]}{\boldsymbol{\partial\theta_0\partial\theta_0'}}\right]$$

These two results are the score and variance of our loglikelihood function under the true parameter $\boldsymbol{\theta_0}$, in particular the last equation is known as the information matrix equality, which is the inverse of the asymptotic variance of our estimator $\widehat{\boldsymbol{\theta}}$.

These are the key quantitites to obtain the fundamental result of MLE estimation, the asymptotic Normality of the estimator:

$$\sqrt{T}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta_0}\right) \xrightarrow{d} N\left(0, \left\{-E\left[\frac{1}{T} H(\boldsymbol{\theta_0})\right]\right\}^{-1}\right)$$

$$\widehat{\boldsymbol{\theta}} \xrightarrow{a} N\left(0, \left\{-E\left[\frac{\partial^2 \ln\left[L(\boldsymbol{\theta_0} \mid \boldsymbol{\varepsilon})\right]}{\boldsymbol{\partial\theta_0\partial\theta_0'}}\right]\right\}^{-1}\right)$$

If we have N i.i.d. Gaussian vectors of the same size (Tx1) that follow a multivariate Gaussian distribution:

$$\boldsymbol{\varepsilon_t} \sim MN(\boldsymbol{\mu}, \boldsymbol{H})$$

Where μ is a Nx1 vector and H a NxN covariance matrix, the multivariate density of the Nx1 random vector becomes:

$$f(\boldsymbol{\varepsilon_t}) = (2\pi)^{-\mathrm{N}/2} . |\mathbf{H}|^{-1/2} . \exp\left(-\frac{1}{2}(\boldsymbol{\varepsilon_t} - \boldsymbol{\mu})' \boldsymbol{H^{-1}} (\boldsymbol{\varepsilon_t} - \boldsymbol{\mu})\right)$$

The likelihood function becomes:

$$L(\boldsymbol{\theta_0} \mid \boldsymbol{\varepsilon}) = \prod_{t=1}^{T} f\left(\boldsymbol{\varepsilon_t} \mid \boldsymbol{\theta_0}\right)$$

While the log-likelihood:

$$\ln\left[L(\boldsymbol{\theta_0} \mid \boldsymbol{\varepsilon})\right] = \ln \prod_{t=1}^{T} f\left(\boldsymbol{\varepsilon_t} \mid \boldsymbol{\theta_0}\right)$$

$$= \ln \prod_{t=1}^{T} \frac{1}{(2\pi)^{N/2}} |\mathbf{H}|^{-1/2} \exp\left(-\frac{1}{2}(\boldsymbol{\varepsilon_t} - \boldsymbol{\mu})' \mathbf{H^{-1}} (\boldsymbol{\varepsilon_t} - \boldsymbol{\mu})\right)$$

$$= \sum_{t=1}^{T} \left(-\frac{N}{2}\ln(2\pi) - \frac{1}{2}\ln|\mathbf{H}| - \frac{1}{2}(\boldsymbol{\varepsilon_t} - \boldsymbol{\mu})' \mathbf{H^{-1}} (\boldsymbol{\varepsilon_t} - \boldsymbol{\mu})\right)$$

$$= -\frac{NT}{2}\ln(2\pi) - \frac{T}{2}\ln|\mathbf{H}| - \frac{1}{2}\sum_{t=1}^{T} (\boldsymbol{\varepsilon_t} - \boldsymbol{\mu})' \mathbf{H^{-1}} (\boldsymbol{\varepsilon_t} - \boldsymbol{\mu})$$

Where **H** is positive semidefinite.

In the MGARCH model the random process is given by ε_t where we want to maximize the log-likelihood with respect to the parameter vector θ which is defined in the specification of the conditional covariance matrix and depends on the particular model we're referring to.

The likelihood function of MGARCH is non-linear and non-convex, so the optimization algorithm may converge to a local maximum. To avoid this, sample estimates should be used as starting values since under stationarity such estimates converge to the true parameters.

1.4 VECH Model

The VECH GARCH is the very first MGARCH presented in the literature, it was first proposed by Bollerslev, Engle, and Wooldridge (1988) in their paper "A Capital Asset Pricing Model with Time Varying Volatilities". It's one of the simplest MGARCH models available as it's basically an extension of the Univariate Garch.(p,q).

Defining:

$$\boldsymbol{\varepsilon_t} \mid I_{t-1} \sim MN(\mathbf{0}, \boldsymbol{H_t})$$

As a Nx1 random vector, where H_t is our NxN time varying conditional covariance matrix and 0 is a Nx1 vector; we can structure the Multivariate Conditional Variance Covariance Matrix as:

$$\text{Vech}\,(\mathbf{H_t}) = \mathbf{C} + \sum_{i=1}^{q} \mathbf{A_i}\text{Vech}(\boldsymbol{\varepsilon_{t-i}}\boldsymbol{\varepsilon_{t-i}}') + \sum_{j=1}^{p} \mathbf{B_j}\text{Vech}\left(\mathbf{H_{t-j}}\right)$$

Where the vech(.) operator is applied to a symmetric matrix and it is a vectorization of the lower triangular elements of the matrix starting from the main diagonal. As an example, taking a 3x3 Symmetric Matrix A:

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

$$vech(\boldsymbol{A}) = [a_{11}, a_{22}, a_{33}, a_{21}, a_{32}, a_{31}]'$$

The VECH GARCH equation is obviously an extension of the univariate GARCH(p,q) frame work:

$$h_{i,t} = c_i + \sum_{j=1}^{q} \alpha_{ij}(\varepsilon_{i,t-1}^2) + \sum_{j=1}^{p} \beta_{ij} h_{i,t-1}$$

In particular C is a 1/2 N(N+1) x 1 vector of Constants, which is the representatives of c in the univariate case, A_i and B_j are 1/2 N(N+1) x 1/2 N(N+1) matrices of coefficients (for i=1,2,...p; j=1,2,…,q), represented by α and β in the univariate framework.
(ε_{t-i} ε_{t-i}') are the cross product of the past observations with i lags (i=1,2,…,p) and H_{t-j} is the past conditional Variance Covariance matrix with j lags (j=1,2,…,q).
The estimation of the parameters is executed through Maximum Likelihood Estimation, considering the innovations distributed as a Multivariate Normal, the parameter vector is:

$$\boldsymbol{\theta} = [C', vec(\boldsymbol{A_1})', \ldots, vec(\boldsymbol{A_q})', vec(\boldsymbol{B_1})', \ldots, vec(\boldsymbol{B_p})']'$$

Where the operator vec(.) is is a linear transformation which converts the matrix into a column vector.

Referring to the the 3x3 matrix A from above we get:

$$vec(\boldsymbol{A}) = [\, a_{11}, a_{21}, a_{31}, a_{12}, a_{22}, a_{32}, a_{13,} a_{23}, a_{33}]'$$

And the log-likelihood function at time t is distributed as a multivariate Normal:

$$\ln\left[L(\boldsymbol{\theta} \mid \boldsymbol{\varepsilon_t})\right] = -\frac{N}{2}\ln(2\pi) - \frac{1}{2}\ln|\mathbf{H_t}| - \frac{1}{2}\boldsymbol{\epsilon_t}'\mathbf{H_t^{-1}}\boldsymbol{\epsilon_t}$$

Where ε_t is a Nx1 random vector. Considering our estimation sample t=1,…,T, let the values of the pre sample values of vech(ε_{t-i} ε_{t-i}^{T}) and vech($\mathbf{H_{t-j}}$) are set equal to unconditional Expected values, or if unknown through the sample estimate as the two are asymptotically equivalent under covariance stationarity assumptions. Let ε_t (t=1,…;T) be realized values of the random process ε, we can compute H_t through an iterative method.

The overall function to maximize through iterative methods is:

$$\ln\left[L(\boldsymbol{\theta} \mid \boldsymbol{\varepsilon})\right] = \sum_{t=1}^{T} \ln\left[L(\boldsymbol{\theta} \mid \varepsilon_t)\right]$$

The standard procedure to find the maximum point of the function is through numerical approximation of the first derivative $\partial L_t(\Theta)/\partial\Theta$ such as Newton-Raphson Method to obtain first order necessary conditions. Given standard regularity conditions (Crowder 1976, Wooldridge 1986) the ML estimate of Θ will be Asymptotically normal and Unbiased.

The VECH model is covariance stationary if the magnitudes of the eigenvalues of the matrix (A+B) are less than one. This concept is detailed by Engle and Kroner in 1995. Although there is currently no known general necessary conditions for the positive definitiveness of Ht, Gourieroux (1997) provides sufficient conditionsThe model is a powerful tool for modeling the dynamics of conditional variances and covariances in a multivariate GARCH setting. It is particularly useful for representing symmetric responses of these quantities to past observations and cross-products of observations. Conditional variances depend on their past values and on conditional covariances, the same apply to conditional covariances. However, this flexibility comes at a cost, as the VECH model can be challenging to estimate

in relatively large systems due to the large number of parameters involved. Specifically, the number of parameters in the VECH model equals (p + q)(N(N + 1)/2)2 + N(N + 1)/2, where N is the number of variables in the system.. This can lead to computational challenges during estimation, particularly when N is large, making the model less practical for large-scale applications.In order to overcome the computational limitations brought forwards by such an high number of parameters, Bollerslev, Engle, and Wooldridge (1988) introduced a simplification to their model; considering the matrices A and B to be Diagonal matrices. This method however eliminates the cross-sectional dependencies between variances and covariances, although it significantly reduces the number of parameters to estimate to (p+q +1)N(N +1)/2.

Here each equation can be computed separately:

$$h_{ijt} = c_{ij} + \alpha_{ij}\varepsilon_{it-1}\varepsilon_{jt-1} + \beta_{ij}h_{ij,t-1}$$

With $i, j = 1, \dots, N$.

From these relations we can see that only the lagged values of the conditional covariance and the lagged cross products of the observation impact the covariance terms hijt.

In order to ensure positive the definitiveness of the conditional variance covariance matrix, Engle and Ding (2001) proposed a different specification of the model for the case p=q=1, The DVECH equation is in fact equivalent to:

$$\boldsymbol{H_t} = \mathbf{C} + \mathbf{A} \circ \boldsymbol{\varepsilon_{t-1}\varepsilon_{t-1}}' + \mathbf{B} \circ \mathbf{H_{t-1}}$$

Where $\circ$ indicates the Hadamard product, i.e the element by element product of matrices of the same size.

Given two matrices U and V of the same size we have:

$(\mathbf{U} \circ \mathbf{V})_{ij} = (\mathbf{U})_{ij}\,(\mathbf{V})_{ij}$

In particular, since $\mathbf{H}_t$ is symmetric and $\boldsymbol{\varepsilon}_{t-1}\,\boldsymbol{\varepsilon}_{t-1}$**',**$\mathbf{H}_{t-1}$ are as well; **C, A** and be **B** have to be symmetric thanks to the element by element product characteristic of the Hadamard operator. In fact given a sum of matrices, the resulting matrix is symmetric if and only if the all the summed matrices are symmetric themselves, in our particular case this is possible if and only if **A, B** and **C** are symmetric.

This is an important result as only the lower portion of the parameters matrices have to be estimated.

In order to ensure positive definitiveness Engle and Ding make use the Shur Product Theorem and its lemmas.Formal proof of these theorems can be found in their paper.

Remembering that the sum of positive semi-definite matrix is a positive semi-definite we can obtain rewrite the DVECH Model as:

$$\mathbf{H}_t = \mathbf{CC}' + \mathbf{AA}' \circ \boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}' + \mathbf{BB}' \circ \mathbf{H}_{t-1}$$

This is the Matrix-Diagonal Model in which **CC'**, **AA',BB'** are positive semi-definite by construction. Since, **AA'.BB'** and **CC'** are positive semidefinite by construction, in order to ensure that $\mathbf{H_t}$ is positive semi-definite it's enough to consider its starting value H_0 being the sample estimate. By expressing the matrices of parameters in the form of the product of a matrix and its transpose (**CC′, AA′**, and **BB′**), rather than just the matrices themselves (**C, A**, and **B**), the positive semi-definiteness property is automatically satisfied during estimation without the need for any additional constraints. A further simplification is the Vector-Diagonal Model in which **A** and **B** are constrained to be rank one matrices, this formulation is equivalent to a diagonal BEKK:

$$\mathbf{H_t} = \mathbf{CC}' + \boldsymbol{\alpha\alpha}' \circ \boldsymbol{\varepsilon}_{\mathbf{t-1}}\boldsymbol{\varepsilon}_{\mathbf{t-1}}' + \boldsymbol{\beta\beta}' \circ \mathbf{H_{t-1}}$$

Here α and β are Nx1 vectors, since $\alpha\alpha'$ and $\beta\beta'$ are linear combination of the two vectors with themselves, the resulting matrix has rank one. In the Scalar-Diagonal Model instead, α and β are positive scalars. It implies that any linear combination of the original series will have the same ARCH and GARCH parameters in its GARCH representation. This is a significant constraint, and is equivalent to a scalar BEKK model:

$$\mathbf{H}_t = \mathbf{CC}' + \alpha\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}' + \beta\mathbf{H}_{t-1}$$

Each conditional covariance along the matrix will present the same parameters:

$$h_{ijt} = c_{ij} + \alpha\varepsilon_{it-1}\varepsilon_{jt-1} + \beta h_{ij,t-1}$$

1.5 BEKK MODEL

The BEKK model was introduced by Engle and Kroner (1995) as special case of the VECH model, it guarantees the positive definiteness of the conditional covariance matrix. This is achieved by embedding the covariance matrix in a specific structure that imposes restrictions on the parameters across equations. The BEKK model includes all positive definite diagonal VECH models and nearly all positive definite VECH representations. However, like the VECH model, it still suffers from the disadvantage of having a large number of parameters, as even for the simplest form with K = p = q = 1 their magnitude is of the same order of a generic Diagonal model. This can make the estimation process computationally heavy.
The BEKK (1,1,K) is defined as:

$$\mathbf{H}_t = \mathbf{C} + \sum_{k=1}^{K} \mathbf{A}_k' \boldsymbol{\varepsilon}_{t-1} \boldsymbol{\varepsilon}_{t-1}' \mathbf{A}_k + \sum_{k=1}^{K} \mathbf{B}_k' \mathbf{H}_{t-1} \mathbf{B}_k$$

Where K determines the generality of the process, and the matrices A,B and C are NxN matrices with C being lower triangular.The positivity of H_t is guaranteed by the specific way the model is parameterized. By using the Kronecker product of two matrices, it can be determined that the BEKK model is covariance stationary if and only if the absolute values of the eigenvalues of the sum $\sum_{k=1}^{K} A_k \otimes A_k + \sum_{k=1}^{K} B_k \otimes B_k$ are less than 1.
The number of parameters in the BEKK(1,1,1) model is N(5N +1)/2. To reduce this number, we can impose diagonality to the parameter matrices A and B, obtaining a DVECH model equivalent, in particular a vector diagonal VECH representation. Since BEKK(1,1,1) is defined as:

$$\mathbf{H}_t = \mathbf{C} + \mathbf{A}' \boldsymbol{\varepsilon}_{t-1} \boldsymbol{\varepsilon}_{t-1}' \mathbf{A} + \mathbf{B}' \mathbf{H}_{t-1} \mathbf{B}$$

And being A and B diagonal matrices, we have that:

$$\mathbf{A}' \boldsymbol{\varepsilon}_{t-1} \boldsymbol{\varepsilon}_{t-1}' \mathbf{A} = \mathbf{A} \circ \boldsymbol{\varepsilon}_{t-1} \boldsymbol{\varepsilon}_{t-1}{}'$$
$$\mathbf{B}' \mathbf{H}_{t-1} \mathbf{B} = \mathbf{B} \circ \mathbf{H}_{t-1}$$

The Diagonal BEKK(1,1,1) is equivalent to the vector diagonal VECH, where A and B are constrained to be rank one matrices.
However, unlike DVECH, BEKK model assures positive definitiveness of the conditional Variance Covariance Matrix. We can impose further restrictions on the parameters considering them to be scalars:

$$\mathbf{H_t} = \mathbf{C} + \alpha . \boldsymbol{\varepsilon_{t-1}\varepsilon'_{t-1}} + \beta \mathbf{H_{t-1}}$$

Obviously this significantly simplify parameters estimation, easing computational heaviness at the cost of increasing model misspecification.

1.6 CCC Model

The Constant Conditional Correlation (CCC) GARCH model, also known as the Engle-Bollerslev-Nelson (EBN) model, was introduced by Robert Engle, Tim Bollerslev and Mark Nelson in a paper published in 1990.

The CCC-GARCH model allows for the modeling of the conditional correlation between multiple time series, in addition to modeling the conditional variances of the individual time series. This is achieved by introducing a new parameter matrix, the correlation matrix, that captures the interactions between the different time series. The appeal of the CCC-GARCH model lies in its ability to capture the dynamic interactions between multiple time series, as well as its interpretability. The correlation matrix allows for easy interpretation of the relationships between the variables. The CCC model assumes a constant structure of the correlation matrix, such that R_t=R is a positive definite symmetric matrix. Positive definiteness of H_t is assured if and only if R is positive definite and alle the conditional variances h_{iit} are positive. The single conditional variances can be defined through any univariate GARCH (1,1):

$$h_{i,t} = \omega_i + \alpha_i(\varepsilon^2_{i,t-1}) + \beta_i h_{i,t-1}$$

The novelty introduced by Bollerslev is that he doesn't specify directly the conditional covariances, but he does it indirectly through the Correlation matrix. Thus, the focus of the model is the Estimation of the constant matrix R. Since R is symmetric and h_t a Nx1 vector,

the number of parameters to estimate is reduced to N(N+5)/2.It is possible to simplify the estimation process by assuming that conditional variances and correlations are independent, by first estimating the conditional variances, and then separately estimating the conditional correlations. This two-step procedure is relatively straightforward and can be used with high dimensional systems. The model specification is:

$$\boldsymbol{H_t = D_t R D_t}$$
$$\boldsymbol{D_t} = \text{diag}\left(\sqrt{h_{i,t}}\right)$$
$$\boldsymbol{R = D_t^{-1} H_t D_t^{-1}}$$

Where D_t is the time dependent NxN diagonal matrix containing the N conditional variances and R is the NxN symmetric positive definite matrix containing the constant correlations coefficients:

$$R = \begin{pmatrix} \rho_{1,1} & \rho_{1,2} & \rho_{1,3} \\ \rho_{2,1} & \rho_{2,2} & \rho_{2,3} \\ \rho_{3,1} & \rho_{3,2} & \rho_{3,3} \end{pmatrix}$$

$$D_t = \begin{pmatrix} \sqrt{h_{1,t}} & 0 & 0 \\ 0 & \sqrt{h_{2,t}} & 0 \\ 0 & 0 & \sqrt{h_{3,t}} \end{pmatrix}$$

With $\rho_{i,i} = 1$ ∀i, and $\rho_{i,j} = \rho_{j,i}$ ∀i, j.

In the original paper h_{ijt} is defined through a standard GARCH(1,1) model, however its specification isn't binding and is left to the practitioner as long as it results in a positive value.A simple estimate of the conditional correlation matrix is the unconditional correlation matrix of the residuals. Since here the Conditional Correlation is a time-invariant matrix, we can easily compute it through its sample estimate. Given two zero-mean random variables, we define the constant conditional correlation between them as:

$$\rho_{ij} = \frac{E_{t-1}\left(\varepsilon_{i,t}\varepsilon_{j,t}\right)}{\sqrt{E_{t-1}\left(\varepsilon_{i,t}^2\right)E_{t-1}\left(\varepsilon_{j,t}^2\right)}}$$

Whose sample estimate is:

$$\widehat{\rho_{ij}} = \frac{\sum_{t=1}^{T} \varepsilon_{i,t}\varepsilon_{j,t}}{\sqrt{\left(\sum_{t=1}^{T} \varepsilon_{i,t}^2\right)\left(\sum_{t=1}^{T} \varepsilon_{j,t}^2\right)}}$$

In CCC-GARCH the conditional correlation matrix gets estimated through multivariate MLE estimation, however under covariance stationarity it should converge to its sampe estimate which should itself converge to its unconditional value.

1.7 Time Varying Conditional Correlation

We can define the conditional time varying correlation between two random variables with Expected value equal zero as:

$$\rho_{ij,t} = \frac{E_{t-1}\left(\varepsilon_{i,t}\varepsilon_{j,t}\right)}{\sqrt{E_{t-1}\left(\varepsilon_{i,t}^2\right)E_{t-1}\left(\varepsilon_{j,t}^2\right)}}$$

In this setting, the correlation is based on information known from the previous period. Correlations for future periods can also be defined in a similar way. According to probability laws, all correlations defined in this manner must fall within the range of [-1,1]. The conditional correlation holds true for all possible past information and for all linear combinations of the variables. Our stochastic process is defined as:

$$\varepsilon_{i,t} = \sqrt{h_{i,t}}\eta_{i,t}$$
$$h_{i,t} = E_{t-1}\left(\varepsilon_{i,t}^2\right)$$
$$\varepsilon_{i,t}\left|I_{t-1} \sim N(0, h_{i,t})\right.$$

With η_{it} being a white noise, with variance one and mean zero. And $h_{i,t}$ the conditional variance at time t. And considering N normally distributed random vector of size T, i=1,…,N the stochastic process is given by:

$$\boldsymbol{\varepsilon_t = D_t \eta_t}$$

$$\boldsymbol{H_t} = E_{t-1}(\boldsymbol{\varepsilon_t}\boldsymbol{\varepsilon_t'})$$
$$\boldsymbol{\varepsilon_t}|\boldsymbol{I_{t-1}} \sim MN(\boldsymbol{0}, \boldsymbol{H_t})$$

Where 0 is a Nx1 vector and H_t the NxN conditional covariance matrix. With:

$$\boldsymbol{D_t} = \text{diag}\left(\sqrt{h_{i,t}}\right)$$

Being the NxN conditional diagonal matrix containing the N conditional standard deviations of each $\varepsilon_{i,t}$.Thus, we can rewrite the correlation as:

$$\rho_{ij,t} = \frac{E_{t-1}(\varepsilon_{i,t}\varepsilon_{j,t})}{\sqrt{E_{t-1}(\varepsilon_{i,t}^2)E_{t-1}\left(\varepsilon_{j,t}^2\right)}} = \frac{E_{t-1}(\sqrt{h_{i,t}}\eta_{i,t}.\sqrt{h_{j,t}}\eta_{j,t})}{\sqrt{E_{t-1}(h_{i,t}.\eta_{i,t}^2)E_{t-1}\left(h_{j,t}.\eta_{j,t}^2\right)}}$$
$$= \frac{E_{t-1}(\sqrt{h_{i,t}})E_{t-1}(\sqrt{h_{j,t}})E_{t-1}(\eta_{i,t}\eta_{j,t})}{E_{t-1}(\sqrt{h_{i,t}})E_{t-1}(\sqrt{h_{j,t}}).\sqrt{E_{t-1}(\eta_{i,t}^2)E_{t-1}\left(\eta_{j,t}^2\right)}}$$
$$= \frac{E_{t-1}(\eta_{i,t}\eta_{j,t})}{\sqrt{E_{t-1}(\eta_{i,t}^2)E_{t-1}\left(\eta_{j,t}^2\right)}} = E_{t-1}\left(\eta_{i,t}\eta_{j,t}\right)$$

Since $E_{t-1}\left(\eta_{i,t}^2\right) = 1$, which implies that the conditional correlation of our random variable ε_{it} is equivalent to the conditional covariance of the white noise η_{it}:

$$\boldsymbol{\eta_t}|\boldsymbol{I_{t-1}} \sim MN(0, \boldsymbol{R_t})$$

Where R_t is the NxN conditional correlation matrix of our random vector ε_t.
Many different approaches have been proposed in the literature to approximate ρ_{ijt}, with one of the most popular being the one proposed by RiskMetrics, it uses exponentially decaying weights with parameter λ, with $0 < \lambda < 1$:

$$\hat{\rho}_{ij,t} = \frac{\sum_{s=1}^{t-1} \lambda^{t-s-1}\varepsilon_{i,s}\varepsilon_{j,s}}{\sqrt{\left(\sum_{s=1}^{t-1} \lambda^{t-s-1}\varepsilon_{i,s}^2\right)\left(\sum_{s=1}^{t-1} \lambda^{t-s-1}\varepsilon_{j,s}^2\right)}}$$

The estimator clearly gives more relevance to recent observation, while the weight of past ones decay exponentially towards 0. Since $\lambda<1$ the first observations will have weight λ

with higher exponent which will make them decay towards zero; in particular, the numerator decays faster than the denominator. It's not clear at which point in time past observations become irrelevant and so the value of λ is ambiguous and depends on the dataset, Riskmetrics propose a value of 0,94. This estimator also lie in the interval [-1,1], moreover in a multivariate setting all variables must have the same λ in order to ensure a positive definite correlation matrix. The estimator can be expressed in matrix notation as:

$$\boldsymbol{H_t} = \lambda(\boldsymbol{\eta_{t-1}\eta'_{t-1}}) + (1-\lambda)\boldsymbol{H_{t-1}}$$

Defining q_{ijt} as the conditional covariance matrix of the white noise, we can rewrite the conditional correlation as :

$$q_{i,j,t} = E_{t-1}(\eta_{i,t}\eta_{j,t}) = \frac{E_{t-1}(\eta_{i,t}\eta_{j,t})}{\sqrt{E_{t-1}(\eta_{i,t}^2)E_{t-1}(\eta_{j,t}^2)}}$$

$$= \frac{E_{t-1}(\varepsilon_{i,t}\varepsilon_{j,t})}{\sqrt{E_{t-1}(\varepsilon_{i,t}^2)E_{t-1}(\varepsilon_{j,t}^2)}} = \frac{q_{ij,t}}{\sqrt{q_{ii,t}q_{jj,t}}} = \rho_{i,j,t}$$

Which means that the covariance of the white noise is equal to the correlation of our random process and to its own correlation as well. This holds as long as $E_{t-1}(\eta_{i,t}^2) = 1$. Since we have:

$$\boldsymbol{\eta_t} = \boldsymbol{D_t^{-1}\varepsilon_t}$$

In matrix form we obtain the conditional correlation matrix becomes:

$$E_{t-1}(\boldsymbol{\eta_t\eta'_t}) = \boldsymbol{D_t^{-1}H_tD_t^{-1}} = \boldsymbol{R_t} = \boldsymbol{Q_t}$$

Though exponential smoothing we obtain the following equality:

$$q_{i,j,t} = (1-\lambda)(\eta_{i,t-1}\eta_{j,t-1}) + \lambda(q_{i,j,t-1})$$

Which in matrix form is given by:

$$\boldsymbol{Q_t} = (1-\lambda)(\boldsymbol{\eta_{t-1}\eta'_{t-1}}) + \lambda \boldsymbol{Q_{t-1}}$$

Where $\boldsymbol{Q_t}$ is a NxN matrix. A different approach is given by a cross-sectional GARCH(1,1):

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

Where $\overline{\rho_{ij}}$ is the unconditional correlation between η_{it} and η_{jt}, which is not time varying and is also equal to the unconditional expectation of q_{ijt}; the variance of q_{ijt} is 1.
In matrix form we can express the estimator as:

$$\boldsymbol{Q_t} = S(1-\alpha-\beta) + \alpha(\boldsymbol{\eta_{t-1}\eta'_{t-1}}) + \beta \boldsymbol{Q_{t-1}}$$

Where S represent the unconditional correlation Matrix of η_t. In both the estimators the conditional correlation matrix is positive definite, as the conditional covariance matrix is the weighted average of a Positive definite matrix ($\mathbf{H_{t-1}}$) and positive semidefinite matrix ($\boldsymbol{\eta_{t-1}\,\eta_{t-1}}$'). In particular, the second estimator is mean reverting if $\alpha+\beta<1$ with both being non-negative; while it simplifies to the first one if $\alpha+\beta=1$.
Ding and Engle (2001) proposed a generalization of such estimators:

$$\boldsymbol{Q}_t = S \circ (\boldsymbol{\iota\iota'} - A - B) + \boldsymbol{A} \circ \boldsymbol{\eta_{t-1}\eta'_{t-1}} + \boldsymbol{B} \circ \boldsymbol{Q_{t-1}}$$

Where ι is Nx1vector of ones, ans A and B are NxN matrices It can easily be seen that by setting $\mathbf{A}=(1-\lambda)\,\boldsymbol{\iota\iota'}$ and $\mathbf{B}=\lambda\,\boldsymbol{\iota\iota'}$; we obtain the first estimator:

$$\begin{aligned}\boldsymbol{S} \circ (\boldsymbol{\iota\iota'} - (1-\lambda)\boldsymbol{\iota\iota'} - \lambda\boldsymbol{\iota\iota'}) + (1-\lambda)\boldsymbol{\iota\iota'} \circ \boldsymbol{\eta_{t-1}\eta'_{t-1}} + \lambda\boldsymbol{\iota\iota'} \circ \boldsymbol{Q_{t-1}} \\ = \boldsymbol{S} \circ (\boldsymbol{\iota\iota'} - \boldsymbol{\iota\iota'}) + (1-\lambda)(\boldsymbol{\eta_{t-1}\eta'_{t-1}}) + \lambda \boldsymbol{Q_{t-1}} \\ = (1-\lambda)(\boldsymbol{\eta_{t-1}\eta'_{t-1}}) + \lambda \boldsymbol{Q_{t-1}}\end{aligned}$$

While setting $\mathbf{A}=\alpha\,\boldsymbol{\iota\iota'}$ and $\mathbf{B}=\beta\,\boldsymbol{\iota\iota'}$ we obtain the second one:

$$\begin{aligned}\boldsymbol{S} \circ (\boldsymbol{\iota\iota'} - \alpha\boldsymbol{\iota\iota'} - \beta\boldsymbol{\iota\iota'}) + \alpha\boldsymbol{\iota\iota'} \circ \boldsymbol{\eta_{t-1}\eta'_{t-1}} + \beta\boldsymbol{\iota\iota'} \circ \boldsymbol{Q_{t-1}} \\ = \boldsymbol{S}(1-\alpha-\beta) + \alpha(\boldsymbol{\eta_{t-1}\eta'_{t-1}}) + \beta \boldsymbol{Q_{t-1}}\end{aligned}$$

$\mathbf{Q_t}$ will be positive semidefinite if the matrices **A**,**B** and (**ɯ**' – **A** – **B**) are positive semidefinite, if any of them is positive definite then $\mathbf{Q}_t$ will be positive definite.

1.8 DCC Model

The DCC-GARCH (Dynamic Conditional Correlation) model was first introduced by Engle (2002) as a way to capture the dynamic changes in the correlation structure of financial time series, it is an extension of the CCC model.

In the DCC-GARCH model, the constant conditional correlation matrix is replaced with a time-varying conditional correlation matrix. This allows the model to capture changes in the correlation structure over time.

It is used to model the correlation between multiple time series, relaxing the constant correlations assumption of the CCC model.

We can express the model as:

$$\boldsymbol{H_t = D_t R_t D_t}$$

$$\boldsymbol{\varepsilon_t \mid I_{t-1} \sim MN(0, D_t R_t D_t)}$$

$$\boldsymbol{D_t^2} = \text{diag}\,\{\omega_i\} + \text{diag}\,\{\boldsymbol{\alpha_i}\} \circ \boldsymbol{\varepsilon_{t-1}\varepsilon'_{t-1}} + \text{diag}\,\{\beta_i\} \circ \boldsymbol{D_{t-1}^2}$$

$$\boldsymbol{\eta_t = D_t^{-1} \varepsilon_t}$$

$$\boldsymbol{Q_t = S \circ (\iota\iota' - A - B) + A \circ \eta_{t-1}\eta'_{t-1} + B \circ Q_{t-1}}$$

$$\boldsymbol{R_t} = \text{diag}\,\{\boldsymbol{Q_t}\}^{-1}\boldsymbol{Q_t}\text{diag}\,\{\boldsymbol{Q_t}\}^{-1}$$

In which the third equation simply states that the conditional variances are obtained through a GARCH(1,1) specification:

$$D_{i,t}^2 = h_{i,t} = \omega_i + \alpha_i(\varepsilon_{i,t-1}^2) + \beta_i h_{i,t-1}$$

While the last equation expresses the form of the conditional correlation matrix as a function of the conditional covariance matrix of the white noise, $\boldsymbol{Q_t}$ converges asymptotically to its true value in which its main diagonal is given by the variances of $\eta_{i,t}$. Which means that diag $\{\boldsymbol{Q_t}\}^{-1}$converges to the NxN identity matrix ($\mathbf{I_N}$), such that:

$$\text{dia g}\{\boldsymbol{Q_t}\}^{-1}\boldsymbol{Q_t}\,\text{dia g}\{\boldsymbol{Q_t}\}^{-1} \overset{a}{\sim} \boldsymbol{I_N^{-1}Q_t I_N^{-1}} = \boldsymbol{Q_t}$$

The parameters of the model can then be estimated through a MLE, this is possible thanks to the first equation which expresses the gaussianity of the process, in absence of Normality we would have to estimate the parameters through Quasi-Maximum Likelihood.Assuming gaussianity of the process, the target loglikelihood function is the following application of the multivariate normal loglikelihood, seen before in its general form:

$$\ln\left[L(\boldsymbol{\theta} \mid \boldsymbol{\varepsilon})\right] = -\frac{1}{2}\sum_{t=1}^{T}\left(N\log\left(2\pi\right) + \ln|\boldsymbol{H_t}| + \boldsymbol{\varepsilon_t'H_t^{-1}\varepsilon_t}\right)$$
$$= -\tfrac{1}{2}\textstyle\sum_{t=1}^{T}\left(N\ln\left(2\pi\right) + \ln|\boldsymbol{D_tR_tD_t}|\ \boldsymbol{\varepsilon_t'D_t}^{-1}\boldsymbol{R_t}^{-1}\boldsymbol{D_t}^{-1}\boldsymbol{\varepsilon_t}\right)$$

Which remembering that:

$$\boldsymbol{\varepsilon_t} = \boldsymbol{D_t\eta_t}$$
$$\boldsymbol{\varepsilon_t'} = \boldsymbol{\eta_t'D_t}$$

Where $\boldsymbol{D_t} = \boldsymbol{D_t}'$ since it's a diagonal matrix:

$$= -\frac{1}{2}\sum_{i=1}^{T}\left(N\ln\left(2\pi\right) + 2\ln|\boldsymbol{D_t}| + +\ln|\boldsymbol{R_t}| + \boldsymbol{\eta_t'D_tD_t}^{-1}\boldsymbol{R_t^{-1}D_t}^{-1}\boldsymbol{D_t\eta_t}\right)$$
$$= -\frac{1}{2}\sum_{i=1}^{T}\left(N\ln\left(2\pi\right) + 2\ln|\boldsymbol{D_t}| + +\ln|\boldsymbol{R}_t| + \boldsymbol{\eta_t'I_NR_t^{-1}I_N\eta_t}\right)$$
$$= -\frac{1}{2}\sum_{i=1}^{T}\left(N\ln\left(2\pi\right) + 2\ln|\boldsymbol{D_t}| + +\ln|\boldsymbol{R_t}| + \boldsymbol{\eta_t'R_t^{-1}\eta_t}\right)$$
$$= -\tfrac{1}{2}\textstyle\sum_{i=1}^{T}\left(N\ln\left(2\pi\right) + 2\ln|\boldsymbol{D_t}| + +\ln|\boldsymbol{R_t}| + \boldsymbol{\eta_t'R_t^{-1}\eta_t} +\right.$$
$$\left.\boldsymbol{\varepsilon_t'D_t}^{-1}\boldsymbol{D_t}^{-1}\boldsymbol{\varepsilon_t} - \boldsymbol{\eta_t'\eta_t}\right)$$

With:

$$\boldsymbol{\varepsilon_t'D_t}^{-1}\boldsymbol{D_t}^{-1}\boldsymbol{\varepsilon_t} - \boldsymbol{\eta_t'\eta_t} = \boldsymbol{\eta_t'D_t}'\boldsymbol{D_t}^{-1}\boldsymbol{D_t}^{-1}\boldsymbol{D_t\eta_t} - \boldsymbol{\eta_t'\eta_t} = \boldsymbol{\eta_t'I_NI_N\eta_t} - \boldsymbol{\eta_t'\eta_t}$$

$$= \boldsymbol{\eta_t'\eta_t - \eta_t'\eta_t = 0}$$

In order to simplify the estimation of the parameters Engle introduced a two step procedure to maximize the log-likelihood function. We split the parameters into two subsets, one for the variances diagonal matrix ($\mathbf{D_t}$) and another one for the correlations ($\mathbf{R_t}$), respectively θ and Φ.We can now consider the total log-likelihood as the sum of the log-likelihood of the two components, one dependent on **θ** (volatility component, Lv) and the other dependent on **Φ** (correlations component, Lc):

$$ln[L(\boldsymbol{\theta}, \boldsymbol{\phi} \mid \boldsymbol{\varepsilon}))] = ln[L_V(\boldsymbol{\theta})] + ln[L_C(\boldsymbol{\theta}, \boldsymbol{\phi})]$$
$$ln[L_C(\boldsymbol{\theta}, \boldsymbol{\phi} \mid \boldsymbol{\varepsilon}))] = -\tfrac{1}{2}\textstyle\sum_{t=1}^{T} \left(\ln |\boldsymbol{R_t}| + \boldsymbol{\eta_t' R_t^{-1} \eta_t} - \boldsymbol{\eta_t'\eta_t}\right)$$
$$ln\,[L_V(\boldsymbol{\theta} \mid \boldsymbol{\varepsilon}))] = -\tfrac{1}{2}\textstyle\sum_{t=1}^{T} \left(N\ln\,(2\pi) + \ln |\boldsymbol{D_t}|^2 + \boldsymbol{\varepsilon_t' D_t^{-2} \varepsilon_t}\right)$$

It's important to note that while Lv(**θ**) only depends on the parameter vector **θ**, Lc(**Φ,θ**) depends on both **Φ** and **θ**; the term $\boldsymbol{\eta_t'\eta_t}$belongs to the correlation component as the covariance matrix of η_t is the correlation matrix itself, so the white noises are the drivers of the correlation. We can maximize the total log-likelihood by separately maximizing the two components. The volatility part of the log-likelihood is equal to the sum of its N GARCH Likelihoods which is jointly maximized by separately maximizing the single Log-likelihoods of its GARCH components. Remembering that the ML estimate of a single GARCH is defined as:

$$\ln\,[L(\theta \mid \varepsilon)] = -\frac{1}{2}\sum_{t=1}^{T}\left(\ln\,(2\pi) + \ln\,(h_t) + \frac{\varepsilon_t^2}{h_t}\right)$$

We can rewrite the Volatility component as:

$$ln[L_V(\theta \mid \varepsilon))] = -\frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{N}\left(\ln\,(2\pi) + \ln\left(h_{i,t}\right) + \frac{\varepsilon_{i,t}^2}{h_{i,t}}\right)$$

The square residuals (**$\eta_t\eta_t$'**) do not enter the first order conditions ($\partial L_C(\boldsymbol{\Phi}, \Theta)/\partial(\boldsymbol{\Phi}, \Theta)$=0) as they're not dependent on the parameters, they're only dependent on the parameters of the

univariate Garch estimates [**ω',α',β'**] through $\boldsymbol{D_t^{-1}}$, however such parameters are estimated separately through N univariate GARCH estimations. The two-step estimation approach consists in first estimating **θ**, by maximizing the volatility component Log-likelihood:

$$\widehat{\boldsymbol{\theta}} = \arg max \{ln[L_V(\boldsymbol{\theta} \mid \boldsymbol{\varepsilon}))]\}$$

Then take the $\widehat{\boldsymbol{\theta}}$ estimate as given and plug it in the correlation log-likelihood component to estimate **Φ**:

$$\widehat{\boldsymbol{\phi}} = \arg\max \left\{\ln[L_C(\boldsymbol{\phi} \mid \widehat{\boldsymbol{\theta}}, \boldsymbol{\varepsilon}))]\right\}$$

Consistency in the second step is assured if certain regularity conditions are met and the first step estimates are consistent. In fact, the result of the second step will depend on the estimates obtained in the first step, therefore, as long as the function that relates the two steps is continuous near the true parameters, consistency will be maintained.

2. PORTFOLIO SELECTION

2.1 Introduction

Modern Portfolio theory was first introduced by Harry Markowitz, an American economist, and is considered to be one of the most influential and revolutionary concepts in the field of finance. This theory was introduced in 1952 through Markowitz's paper "Portfolio Selection". The theory proposes that a portfolio of assets should not be viewed as an isolated set of assets but rather as a collection of assets that interact with each other. The risk and return of a portfolio are not simply a function of the individual assets in the portfolio, but rather are influenced by the relationship between those assets.Markowitz's work was motivated by the observation that investors often make investment decisions based solely on the expected returns of individual assets, without considering the risk associated with those assets. However, he believed that investors are risk-averse and therefore prefer portfolios with the highest expected return for a given level of risk. To create such a portfolio, an

investor would typically invest in a mix of different assets, such as stocks, bonds, and other securities, to create a diversified portfolio that minimizes risk.

Markowitz's approach to portfolio theory involves mathematical analysis to construct an optimal portfolio that considers both the expected return and risk of the individual assets, and how they combine with each other. The risk is typically measured in terms of the standard deviation of returns, which is used as a proxy for the uncertainty related to the portfolio. By considering the correlation between assets, Markowitz showed that it is possible to create a diversified portfolio that reduces overall risk without sacrificing returns. The Markowitz portfolio theory proposes a series of fundamental concepts that will eventually become the foundation of advanced structured analysis. The first is that portfolio returns are a combination of the returns of individual assets. The second is that the risk of a portfolio is not simply the sum of the risks of the individual assets, but is also influenced by the correlations between them. lastly that the expected return of a portfolio is the weighted average of the expected returns of the individual assets.

Markowitz's theory is based on several assumptions about investor behaviour. The first assumption is that investors view each investment alternative as being represented by a probability distribution of expected returns over some holding period. In other words, they consider the potential range of returns for each investment and the likelihood of each return. The second assumption is that investors maximize one-period expected utility, and their utility curves demonstrate diminishing marginal utility of wealth. This means that investors seek to maximize their returns while also considering the value of their wealth and the potential risks associated with their investments. The third assumption is that investors estimate the risk of the portfolio based on the variability of expected returns. In other words, they consider the range of potential outcomes and the likelihood of each outcome when evaluating the risk of their portfolio. The fourth assumption is that investors base their decisions solely on expected return and risk, so their utility curves are a function of expected return and the expected variance or standard deviation of returns only. This means that investors evaluate investments solely based on the expected returns and potential risks associated with each investment, rather than considering other factors like social or ethical considerations.The final assumption is that for a given risk level, investors prefer higher returns to lower returns. Similarly, for a given level of expected return, investors prefer less risk to more risk, i.e they are investors are risk-averse and prefer to minimize risk, while also seeking to maximize their returns. Markowitz's approach can be summarized in a simple two-step process. The first step involves the identification of the optimal mix of assets that

maximizes returns for a given level of risk. This can be done by plotting the expected returns and standard deviations of various asset combinations on a graph, known as the efficient frontier. The optimal portfolio lies on the efficient frontier, and the investor can choose the portfolio that best matches their risk tolerance. The second step involves the determination of the actual portfolio composition. This involves identifying the actual allocation of funds among the selected assets, taking into account any constraints, such as minimum investment levels or restrictions on short selling. The investor would then periodically rebalance the portfolio to maintain the desired asset allocation, the original framework was in fact a static multi-objective optimization problem.

One of the main benefits of the Markowitz portfolio theory is that it allows investors to achieve a higher expected return for a given level of risk, or a lower level of risk for a given expected return. This is achieved through diversification, which reduces the overall risk of the portfolio. By spreading investments across multiple assets, the investor can reduce the impact of any one asset's poor performance.

However, Markowitz's theory has many limitations. It assumes that investors are rational and have perfect information about the assets in the portfolio and that information is spread equally and instantaneously amongst all agents. In reality, investors often make irrational decisions and may not have complete information about the assets they are investing in. Another limitation is the assumption of normality of the log-returns, however most of these problems can be mitigated by manipulating the original model to fit the data and the particular scenarios of reference. Markowitz portfolio is in fact the building block upon which a much more complex structure has been built along the decades.

2.2Efficient Portfolio

We will define the return of each unit of capital invest in a given portfolio as:

$$r_p = \mathbf{w}'\mathbf{r}$$
$$\boldsymbol{w}'\boldsymbol{i} = 1$$

Where r_p is the return of the portfolio, w the Nx1 vector of weights of the portfolio, i.e the vector containing the percentage amount of the unitary capital invested in each asset *i*, and **r** is a Nx1 vector containing the returns of the single assets.

By definition $\mathbf{w'i}$ =1, where *i* is a Nx1 vector of ones, this is the constrained imposing that the sum of the percentage of the investment in each asset is equal to the unitary capital meaning that all our capital has to be invested in assets belonging to the portfolio of reference. It's worth noting that since the set A={$\mathbf{w'i}$ =1}is open and unbounded, thus Weistrass theorem can't be applied to find global extremal point over the set A, which is an affine hyperplane of dimension N-1. We can add additional constraints to the wights, such as the non negative constraints that forbids short sellling such that $w_i>0 \; \forall \; i \in N$, in such case $\boldsymbol{w} \in \boldsymbol{\Sigma_N}$, where $\boldsymbol{\Sigma_N}$ is simplex of N-1 dimensions in the real vector space (although a simplex can be defined even in the complex space), the smallest convex set that contains a given set of points the in Euclidean space.

The expected rate of return and standard deviation of returns are two crucial characteristics used to describe a portfolio of assets. In particular the correlation between assets, has a significant impact on the portfolio's standard deviation. If assets in a portfolio are positively correlated, they tend to move in the same direction, leading to an increase in the portfolio's standard deviation and risk. In contrast, when assets are negatively correlated, they move in opposite directions, which reduces the portfolio's standard deviation and risk. A low correlation between assets can reduce the portfolio's risk without affecting its expected return. This is due to the diversification benefits of combining assets with low correlation, which can lead to an overall reduction in portfolio risk. The individual assets in a portfolio can differ in their expected rates of return and standard deviations, affecting the portfolio's risk and return. Negative correlation between assets is especially effective in reducing portfolio risk. Combining two assets with a correlation of -1.0 would reduce the portfolio's standard deviation to zero only if the individual standard deviations of the assets are equal, which is a rare occurrence in practice. We define the Expected return and variance of a given portfolio as:

$$E[r_p] = \boldsymbol{w}' E[\boldsymbol{r}]$$
$$\sigma_p^2 = \boldsymbol{w}' \boldsymbol{H} \boldsymbol{w}$$

With **H** being the NxN covariance matrix of our assets. We can identify the efficient portfolio as that particular portfolio represents the optimal combination of assets is based on the concept of Pareto optimality.

Pareto optimality is a state in which it is not possible to improve the welfare of any one individual without reducing the welfare of another. In the context of portfolio management,

Pareto optimality means that the efficient portfolio represents the best possible allocation of assets, given a set of constraints such as risk tolerance and return objectives.
The efficient portfolio is structured from Pareto optimality by identifying the set of portfolios that are Pareto efficient. These portfolios are those that provide the highest expected return for a given level of risk, or the lowest level of risk for a given expected return. The set of Pareto efficient portfolios is known as the efficient frontier.
Estimating the set of all the efficient portfolio weights results in a parametric constrained optimization problem, which is usually divided into three main branches in the general framework.

1. Minimizing the risk of the portfolio, constraining the weights to ensure a minimum expected return: *min* ($\mathbf{H_p(w)}$) s.t.: $E[r_p\mathbf{(w)}] \geq r$, $\mathbf{w} \in \mathbb{R}^n$, $\mathbf{w'}\boldsymbol{i} = 1$ for a given *r*.
2. Maximize the expected return for a maximum level of risk: max ($E[r_p\mathbf{(w)}]$) s.t.: $\mathbf{H_p(w)} \leq \mathbf{H}$, $\mathbf{w} \in \mathbb{R}^n$, $\mathbf{w'}\boldsymbol{i} = 1$ for a given **H**.
3. max ($E[r_p\mathbf{(w)}] - \alpha\mathbf{H_p(w)}$) s.t: $\alpha \in [0,+\infty]$, $\mathbf{w} \in \mathbb{R}^n$, $\mathbf{w'}\boldsymbol{i} = 1$. In this case for the two limit cases $\alpha=0$ and $\alpha=\infty$ the objective function to optimize become $E[r_p\mathbf{(w)}]$ and - $\mathbf{H_p(w)}$ respectively, we can see α as a risk aversion coefficient as the maximization of - $\mathbf{H_p(w)}$ consists in minimizing the risk without regards for the expected return and viceversa for $\alpha=0$.

It's important to note that in these three class of problems $\mathbf{H_p(w)}$ doesn't necessarily refers to the variance of the portfolio, it's usually a function of it but more in general is a variable expressing the risk associated to the portfolio.

2.3 Value at Risk (VaR)

Defining the concept of risk is not straightforward, in Markowitz framework portfolio variance is used as a proxy. While the standard deviation is a widely used measure of risk, there are other approaches to assessing the potential for losses in a portfolio. One alternative is Value at Risk (VaR), which calculates the potential loss a portfolio may experience over specific time horizon at a given level of confidence. Unlike standard deviation, VaR estimates the worst-case scenario at a given confidence level, rather than the typical deviation from the mean. One advantage of VaR is that it provides a simple and

straightforward measure of risk that can be used to compare the risk of different portfolios. VaR can also be used to set risk limits and to monitor the risk of a portfolio over time.
Given $\mathbf{x} \in \mathbb{R}^n$ and y a random vector in $\mathbb{R}^m$ that represents a stochastic process, such as market prices, we denote by f(**x, y**) the loss associated to x, which can be interpreted as a portfolio with **X** representing the set of available portfolios subject to certain constraints. However, there could be other interpretations as well. It is important to note that the loss could be negative, and therefore represent a gain. For each x, f(x, y) is a random variable that has a distribution in $\mathbb{R}$, driven by the probability distribution of y in $\mathbb{R}^m$ with pdf f(**y**). The probability of *f*(x, y) not exceeding a threshold value ζ is Ψ(**x**, ζ):

$$\Psi(\mathbf{x}, \zeta) = \int_{-\infty}^{f(\mathbf{x},\mathbf{y})\leq\zeta} f(\mathbf{y})\, d\mathbf{y}$$

For a given **x**, Ψ(**x**, ζ) is the cumulative distribution function of the loss associated with **x**, Ψ(**x**, ζ) is non-decreasing in ζ. For any specified probability level α within the range of 0 and 1, the value of α-VaR associated with the loss random variable for x is denoted by ζα(x) which is defined as the minimum value of ζ in the set of real numbers such that Ψ(x, ζ) is greater than or equal to α. The value of ζα(**x**) is obtained as the left endpoint of the interval that consists of all values of ζ where Ψ(**x**, ζ) is equal to α:

$$\zeta\alpha(\mathbf{x}) = \min\{\, \zeta \in \mathbb{R} : \Psi(\mathbf{x}, \zeta) \geq \alpha \,\}$$

In the setting of portfolio optimization, with returns of the assets are assumed being realization of a random process distribute as a Multivariate Normal, the var of the Portfolio is given by:

$$\zeta_t\alpha(\mathbf{x_t}) = \mu_t(\mathbf{w_t}, \mathbf{x_t}) + Za(\mathbf{w_t}'H_t(\mathbf{x_t})\mathbf{w_t})^{1/2}$$

Where Z_a is the α quantile of the standard normal, $\mathbf{w_t}$ the vector of weights of the assets within the portfolio at time t, **x** the vector of returns of the single assets ,$H_t(\mathbf{x})$ the time conditional covariance matrix of the returns, μ_t **(w,x)** the expected value of the portfolio at time t, function of expected returns and weights. It's worth noting that in such a setting the VaR becomes a time series, displaying typical characteristics of random processes such as

variance and expected value. This framework will be used in the empirical application to apply a time varying Var constraint in the portfolio optimization algorithm.

EMPIRICAL APPLICATION

3.1 GENERAL FRAMEWORK

In this Chapter an empirical application of the M-GARCH to financial time series will be presented, for the study 7 ETF have been selected in order to have some level of diversification within the single assets to reduce the size of our portfolio thus the parameters to be estimated. The following ETF have been selected represent different sectors, geographical areas, market sizes and company policies:

1. Invesco QQQ Trust Fund (QQQ): the technology-heavy NASDAQ-100 Index managed by Invesco. The NASDAQ-100 Index is comprised of 100 of the largest and most actively traded non-financial companies listed on the NASDAQ stock exchange.

2. First Trust Financials AlphaDEX Fund (FXO): tracks the performance of the StrataQuant Financials Index. The index uses the AlphaDEX stock selection methodology, it selects stocks from the financial sector with potential for outperformance relative to traditional market-cap weighted indexes.

3. iShares MSCI Emerging Markets Fund (EEM): provides exposure to the emerging markets sector. The fund tracks the performance of the MSCI Emerging Markets Index, which includes large- and mid-cap stocks from emerging market countries such as China, Brazil, South Africa, and Russia.

4. Utilities Select Sector SPDR Fund (XLU): tracks the performance of the Utilities Select Sector Index. The fund invests in a diversified portfolio of companies in the utilities sector,

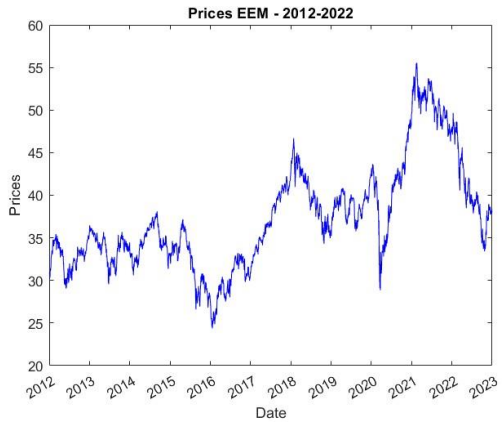
including electric utilities, multi-utilities, and independent power producers and energy traders.

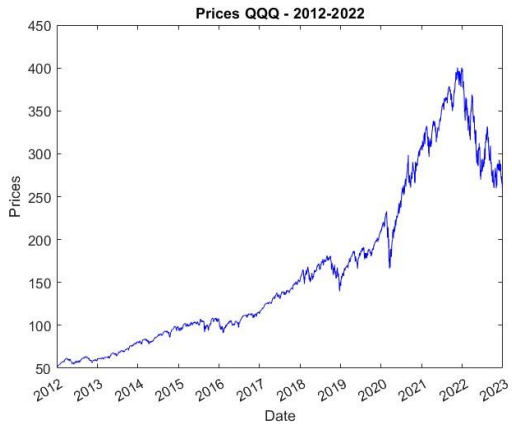
5. Invesco DB Agriculture Fund (DBA): tracks the performance of a basket of agricultural commodity futures contracts. The fund is designed to provide exposure to the agriculture sector, which includes commodities such as corn, wheat, soybeans, and sugar.

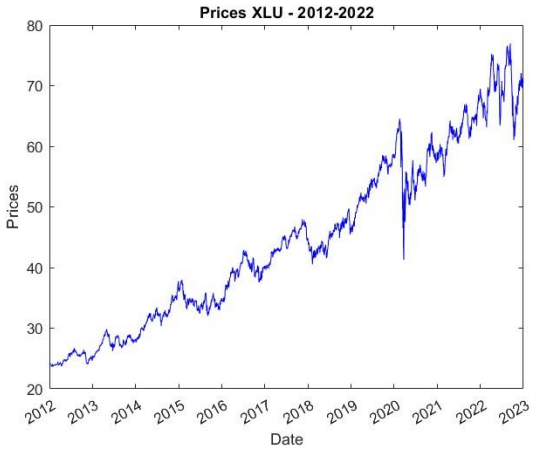
6. Vanguard Real Estate Fund (VNQ): tracks the performance of the MSCI US Investable Market Real Estate 25/50 Index. The fund invests in a diversified portfolio of publicly traded real estate investment trusts (REITs) and other real estate-related companies in the United States.

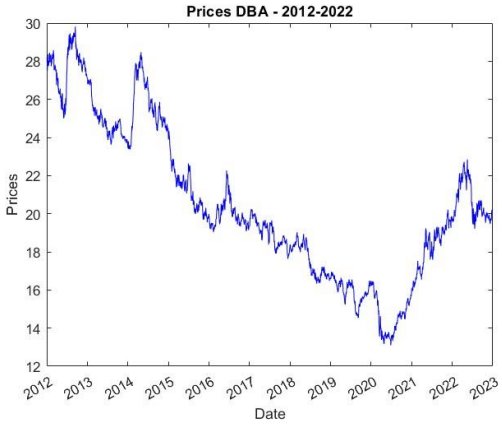
7. Health Care Select Sector SPDR Fund (XLV): tracks the performance of the Health Care Select Sector Index. The fund invests in a diversified portfolio of companies in the health care sector, including companies involved in pharmaceuticals, biotechnology, medical devices, and health care services.

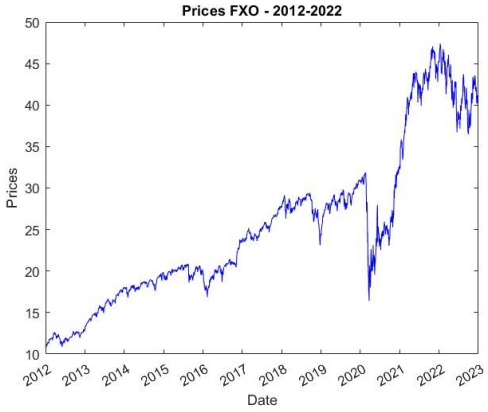
For the analysis we considered a 10 year sample ranging from January 2012 to December 2022, accounting for 2768 daily observations. In which the first seven years (for a total of 2011 observations) compose the in sample dataset on which estimates were drawn; while the last three years (2020-2022) are the forecasting sample amounting to 756 observations. The prices of reference are the adjusted closing prices. The graphs along the whoòle timeframe of reference are reported below:











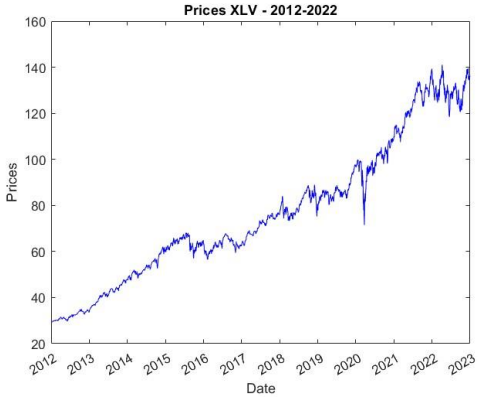
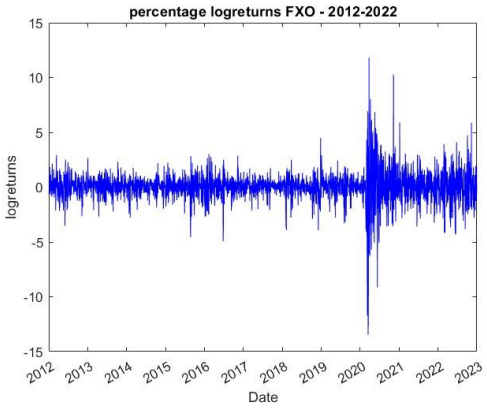


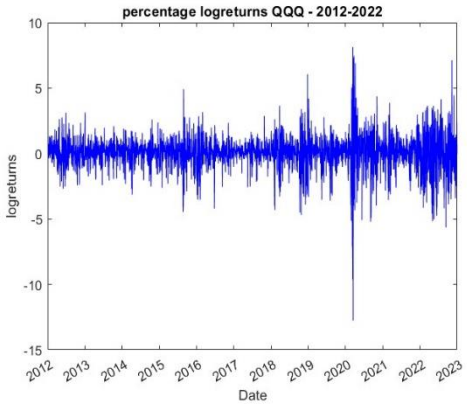
Figure 1: historical prices of the seven ETFs

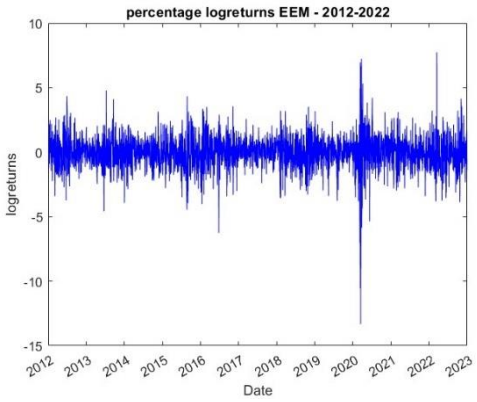
Figure 1 displays the graphs of the prices of the 7 assets we picked for our analysis. graphs there is a clear downwards peak in 2020, caused by the outbreak of COVID-19.

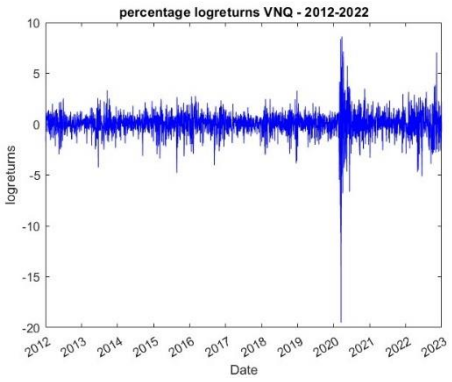
We now compute the percentage log-returns, which are plotted below in **Figure 2**, as:

$$r_{i,t} = (\log(P_{i,t}) - \log(P_{i,t-1})) * 100$$









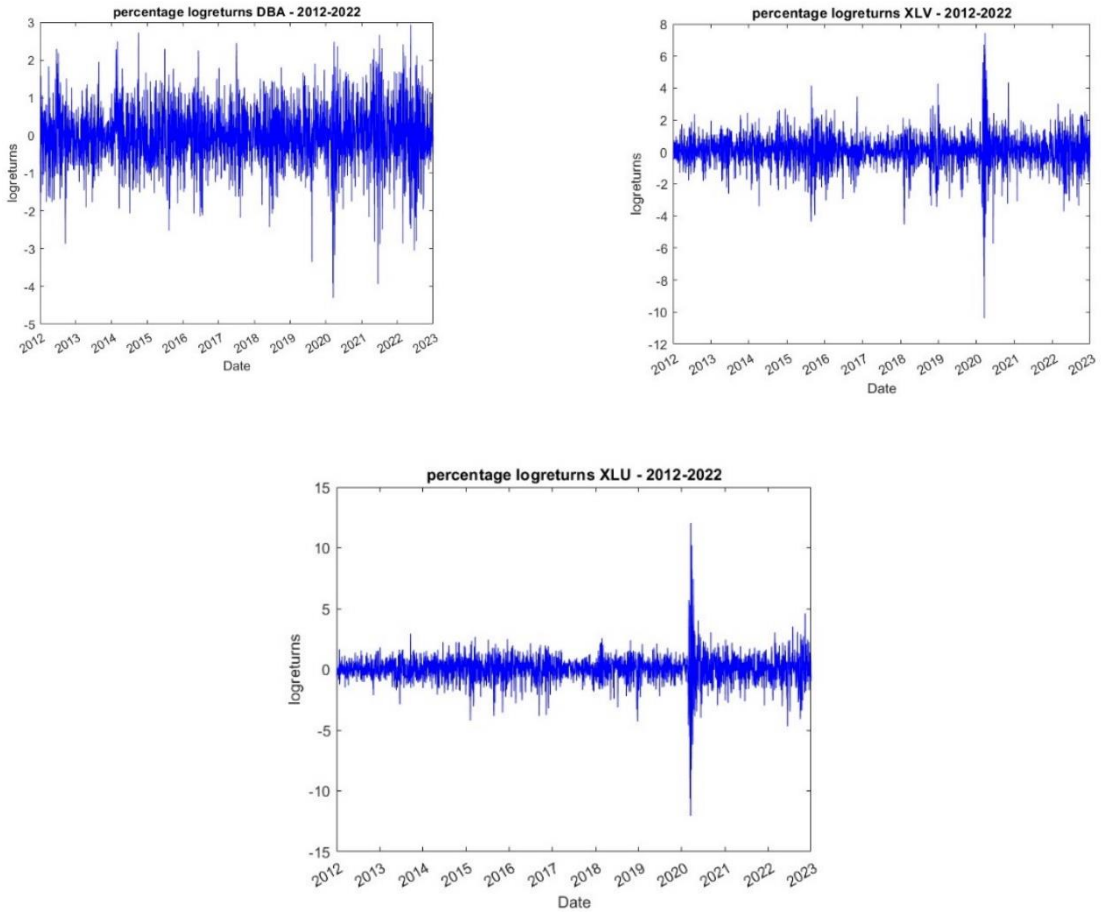


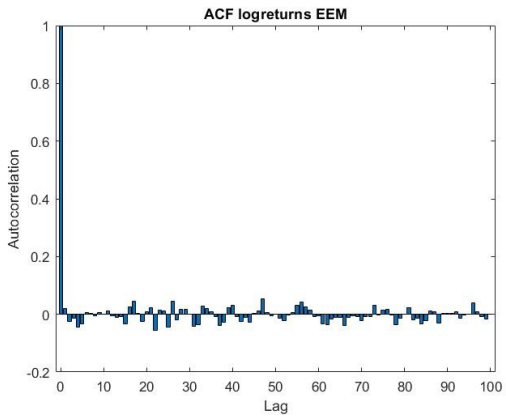
Figure 2: Log-returns of the seven ETFs

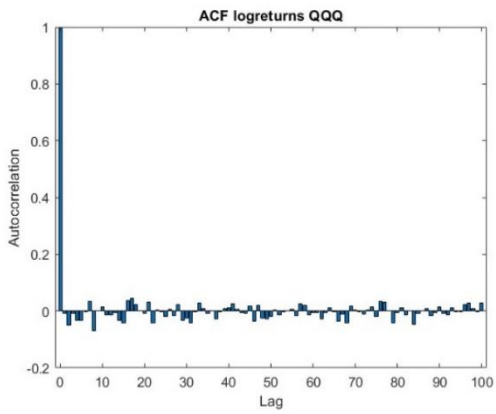
The year 2020 was a critical year for the stock market, particularly in the first half, due to the Covid-19 pandemic. This posed significant challenges for the estimation and forecasting of future data points. The returns data for 2020 highlights the significance of the year, with the first 4-6 months being particularly noisy, and the variance trajectory not following the same mapping as before. This leads to potential biases in the estimation process and can make parameters estimated from previous periods less precise in explaining future projections.To address these challenges, it is necessary to implement a model that can effectively capture extreme events like the Covid-19 pandemic. One potential approach is to use a fundamental model specifically designed for such events. However, for this particular application, we have opted to use a standard Generalized Autoregressive Conditional Heteroscedasticity (GARCH) approach to define the conditional variance structure. This approach can still provide valuable insights into the stock market and is a commonly used method for forecasting in finance. In this case, since 2020 is the first year in the "Forecasting sample," it will not affect the in-sample estimates. The in-sample parameters will be used to

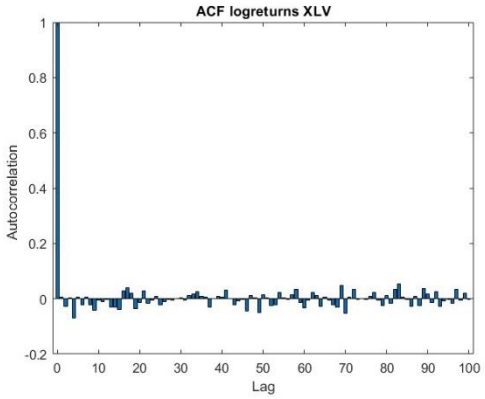
generate normal-scenario predictions in the forecasting process, assuming the GARCH assumptions hold. The variance forecasts generated by this model should be lower than the variances of the forecasting sample.It is important to note that the GARCH model is not perfect, and there may be limitations in its ability to capture the complexities of the stock market, especially during extreme events like the Covid-19 pandemic. However, it provides a starting point for forecasting and can still provide valuable insights when used with caution. Regular reassessment of the model's accuracy and necessary adjustments should be made to ensure its relevance in reflecting the current market conditions.In the absence of a time-dependent conditional mean, we can define log-returns:

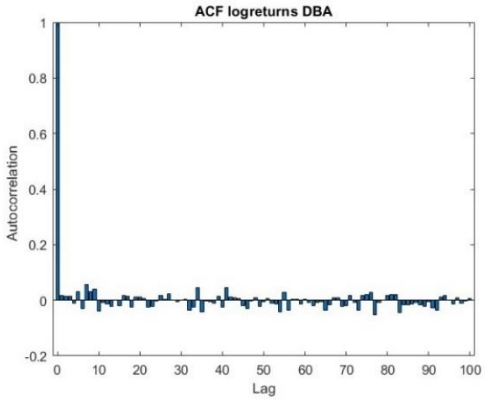
$$r_{i,t} = \mu_i + \varepsilon_{i,t}$$

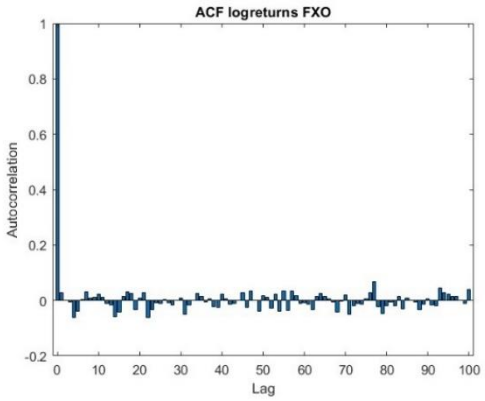
With εi,t indicating the innovations, drivers of the variance and following a GARCH process. To assess the presence of autocorrelation in the returns, the autocorrelation function (ACF) was calculated. The ACF measures the linear correlation between the returns at different time lags and helps to determine whether there is a pattern in the returns over time. The ACF was computed using the estimation sample, as the actual price dynamics of the forecasting period were unknown. To address missing values in the sample, the log returns were "smoothed" by computing the average between the previous and next observation for any missing values. The Autocorrelation Function (ACF) for returns up to 100 lags was then calculated, providing an estimate of the linear correlation along the time-series for different lags:

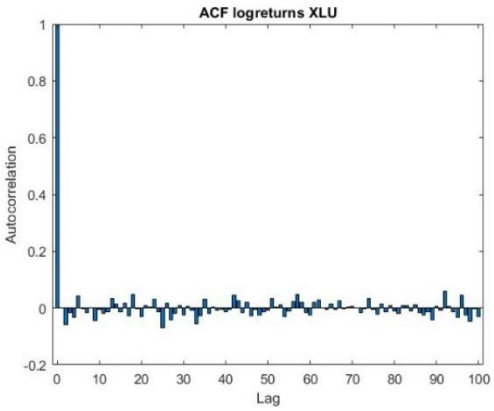












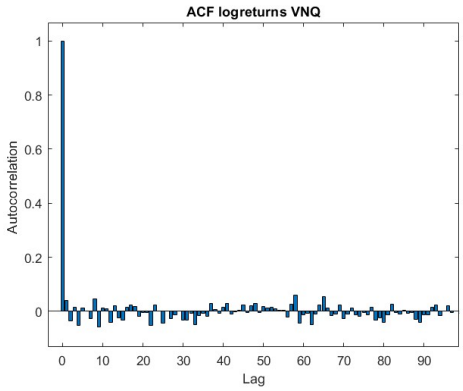


Figure 3: AFC of the ETFs' log-returns

Apparently, there is no substantial correlation present within the exchange-traded funds (ETFs). This is consistent with the obtained results of the AR(1) of the in-sample timeseris, that did not yield any significant autocorrelation even at a 10% level of significance. . However, this dynamic changes over the forecast period. By estimating the AR(1) through a one-step rolling window, it was observed that the significance of the autocorrelation

coefficients slowly increases and reaches its peak around May 2020. This increase in autocorrelation is not surprising as the high volatility seen in 2020, as depicted in the figure above, was due to a structural change in the price dynamics.To account for this change in the data, the residuals were calculated as the deviation from the sample mean in the in-sample estimation. In the forecasting application, an AR(1) was applied to define the mean-generating process from the 50th observation onwards. This approach allowed for a more accurate representation of the price dynamics and helped to mitigate the impact of the structural change.The presence of significant autocorrelation in the forecast period may indicate a persistent pattern in the data, suggesting that past values have an influence on future values. In this case, applying an AR(1) model in the forecasting process proved to be an appropriate approach for capturing this pattern and generating more accurate forecasts. Presence of autocorrelation in daily log-returns of financial assets was already discussed widely in the literature, see "Large Scale Conditional Covariance Matrix Modeling, Estimation and Testing" Zhuanxin Ding, Rober F. Engle (1996). **Table 3.1** presents the statistical summary of the returns, including the mean, standard deviation, kurtosis, and skewness.

TABLE 3.1: Descriptive Statistics Percentage Log-Returns

	Mean	Median	st.deviation	Skewness	Kurtosis	Max	Min
QQQ	0.069	0.101	1.009	-0.401	6.039	6.057	-4.685
FXO	0.058	0.132	0.874	-0.523	5.562	4.488	-4.924
EEM	0.014	0.080	1.171	-0.205	4.144	4.788	-6.262
XLU	0.045	0.113	0.841	-0.551	4.881	2.941	-4.271
DBA	-0.028	-0.069	0.708	0.050	3.698	2.722	-3.357
VNQ	0.041	0.090	0.899	-0.509	4.782	3.331	-4.795
XLV	0.061	0.098	0.893	-0.379	5.160	4.269	-4.532

QQQ had the highest average return at 0.069, followed by XLV at 0.061, FXO at 0.058, and VNQ at 0.041. EEM had an average return of 0.014, while DBA had a negative average return of -0.028. This suggests that QQQ may have been a more profitable investment compared to the other ETFs, while DBA may have resulted in losses over the observed period. The median provides a measure of the middle value of the returns data. FXO had the highest median return at 0.132, followed by XLU at 0.113, and then QQQ at 0.101. This suggests that FXO had more consistent positive returns compared to the other ETFs, which may be attractive to investors who seek a reliable return. EEM had the highest standard deviation at 1.171, followed by QQQ at 1.009. XLU had the lowest standard deviation at

0.841. Making EEM and QQQ be riskier investments due to their higher risk associated to the uncertainty.The skewness measures the asymmetry of the returns distribution. A negative skewness indicates that the distribution is skewed to the left, while a positive skewness indicates that the distribution is skewed to the right. QQQ, FXO, EEM, VNQ, and XLV all had negative skewness values, indicating a left-skewed distribution. Such that the majority of the returns were concentrated on the left side of the distribution. XLU and DBA had more symmetrical distributions. This implies that XLU and DBA may be more predictable and stable investments.The kurtosis measures the "peakiness" of the returns distribution. A higher kurtosis indicates more extreme values in the tails of the distribution. All ETFs had positive kurtosis values, suggesting that they had fat tails and were more prone to extreme events. However, EEM had the highest kurtosis value, suggesting that it may be more susceptible to extreme events compared to the other ETFs.As a normal has both third and fourth moment equal to zero, these returns are leptokurtic and platykurtic (beside DBA as it ha positive skewness).Finally, the maximum and minimum returns provide insights into the potential gains and losses of each ETF. EEM had the highest return at 4.144, while DBA had the lowest return at -3.357. This implies that EEM may have the potential for higher returns, but also a higher risk of losses compared to DBA.In conclusion, QQQ may have been a more profitable investment with higher average returns, while FXO had a more consistent positive performance. EEM had the highest volatility and kurtosis, making it a riskier investment. XLU and DBA had lower volatility and more symmetrical distributions, making them more predictable and stable investments. In **Table3.2** a set of test statistic will be presented, namely: the Jarque-Bera (JB) normality test statistic, the Ljung_Box test statistic (LB) with 10 lags for residuals, absolute and square residuals for autocorrelation, the ARCH test for heteroskedasticity with 10 lags and the Augmented Dickey–Fuller for the existence of unit roots and stationarity also with ten lags.These tests are computed over the sampling period 2012-2019, and since this time framework didn’t show significance of AR(1) autocorrelation, the residuals were computed as simple deviation from the sample mean.

TABLE 3.2: Test Statistics Percentage Log-Returns and residuals

	JB(r)	$LB_{10}(\varepsilon)$	$LB_{10}(\lvert\varepsilon\rvert)$	$LB_{10}(\varepsilon^2)$	$ARCH_{10}$ test(ε)	ADF(r)
QQQ	827.447	22.495	612.578	552.785	242.733	-14.163
FXO	641.587	15.977	392.451	292.443	163.832	-12.817
EEM	123.775	8.500	304.258	268.017	143.604	-13.785
XLU	398.282	17.983	113.584	113.167	80.613	-14.388
DBA	41.646	20.673	88.625	70.334	53.421	-12.689
VNQ	352.791	25.143	208.253	218.887	126.566	-13.586
XLV	439.124	17.430	486.084	420.083	204.948	-14.492

Starting with the Jarque test, we observe that the test statistic is extremely high for all the assets, with QQQ having the highest value at 827.45. The critical value for the test is 5.991, indicating that the null hypothesis of normality can be rejected for all the assets at a 5% significance level. This suggests that the distribution of the returns and residuals is not normal, as the Jarque-Bera test is distributed as chi-square with 2 degrees of freedom and under the null, the random variable is normally distributed. This implies that when forecasting and estimating GARCH parameters under the assumption of normality, we may incur bias driven by distribution misspecification. Moving on to the Ljung-Box test, we see that the test statistics are below the critical values for some of the ETFS, including XLU, FXO, EEM, and XLV. These levels are significantly smaller compared to the Jarque test, with the largest test statistic being 25.14 for VNQ. The table implies that only QQQ, DBA and VNQ present autocorrelation with ten lags, as the statistic is distributed as a chi-squared with 10 degrees of freedom (number of lags) and its critical value is 18.31. The null hypothesis of the absence of autocorrelation and i.i.d. of the data holds for one lag in all assets. We can see that the test statistics increase significantly for the squared and absolute residuals, which is an obvious consequence of the higher correlation present in the magnitude of the returns time series. The null is rejected for all ETFs, in particular, the test statistic is higher for absolute returns as squared returns move towards zero when $|r|<1$ while diverge faster for $|r|>1$. In both the absolute residuals and squared residuals subsets, QQQ presents the highest value of 612.57 and 552.78, respectively. The ARCH test, which tests the null hypothesis of autocorrelation in the squared residuals and the null hypothesis that a series of residuals r(t) exhibits no conditional heteroscedasticity, has the same distribution as the Ljung-Box test with 10 degrees of freedom. It is significant for all the assets, with the largest test statistic being 242.73 for QQQ. For the other assets, the value ranges from 53.42

to 163.83, leading us to reject the null of the absence of heteroskedasticity. We can conclude that the variance of the residuals is not constant over time and changes depending on the values of the previous residuals. The presence of conditional heteroscedasticity requires models that account for this feature, and GARCH models are a good fit for this data structure.Finally, the ADF test, which is distributed as a t-student with N-10 degrees of freedom and critical value=-1.941 at a 5% significance level, presents low values for all ETFs, far below the critical floor, with the lowest being XLV at -14.49 and the highest being DBA at -12.89. Thus, the null hypothesis of a unit root can be rejected for all the assets at a 5% significance level. This implies that the series is stationary, which simplifies the modeling process and improves the accuracy of the predictions within our GARCH framework. Overall, while all the assets exhibit violations of normality, conditional heteroscedasticity, and some degree of autocorrelation, their magnitudes and significance levels vary. For example, QQQ has the highest test statistic for the Jarque test and ARCH test, indicating that its distribution of returns and residuals is the most non-normal and heteroscedastic, respectively. VNQ has the highest test statistic for the Ljung-Box test, indicating that its residuals exhibit the highest degree of autocorrelation. Additionally, DBA has the largest ADF test statistic, suggesting that its series is the least stationary among all the assets, while XLV is the most stationary one.It is worth noting that the significance levels and critical values for the different tests vary depending on the number of lags used in the analysis, excluding the Jarque-Bera test. Therefore, it is crucial to choose appropriate significance levels and lag orders while interpreting the results. The use of the correct significance levels and lag orders helps to avoid spurious results and ensure that the models are well-specified.

3.2 MGARCH: ESTIMATION AND FORECASTING

Since our assets present stationarity and heteroskedasticity, it is appropriate to model variance forecasts with GARCH models. Starting from the 50th observation in the forecasting window, the data-generating process is assumed to be an AR(1) with GARCH innovations. This allows the variance of the returns to be decomposed into two distinct components: a purely GARCH component and an autocorrelation component. The GARCH component captures the conditional volatility of the returns, while the autocorrelation component links previous GARCH estimates to the variance when estimating the variance of r(t+k) given the information set I(t) for k>=2.

The GARCH component is modelled using a MGARCH process, where the autoregressive component models the persistence of volatility over time, while the moving average component captures sudden changes in volatility. By assuming that the data-generating process is an AR(1) with GARCH innovations, the model can capture both the autocorrelation and conditional volatility structure of the returns, providing a better framework for forecasting future returns. In the forecasting of the volatility for the forecasting sample (of length T=756) a one step ahead rolling window was used, this lead the volatility forecast to reduce to the GARCH component as the one step ahead forecast of AR(1) is defined only by the variance component of the residuals one-step ahead future residuals. In particular in the analysis we computed model residual within the forecasting sample for each observations as time went forward one step at a time, such residuals were used at each iteration to forecast the next period conditional variance through the various model. In particular, due to the high dimensionality of the data and the high dimensionality of the data and high numbers of parameters to be estimated, the following procedure was applied:

- Define a one step ahead rolling window in which at each itineration a new value gets added and the first of the old ones gets dropped to maintain the size dataset of interest constant at 2011 points.
- In the one step ahead rolling window compute the residuals at each itineration with the new data that becomes gradually available, one data point at a time.
- For the first 50 observation there was no significant presence of autocorrelation and so the expected value was simply compute with a moving average, residuals were computed by simple subtraction form realized log-returns
- From the 51th observation onwards, at each itineration the new parameters for the AR(1) were estimated, as they become statistically relevant, and compute the residuals
- Due to the high dimensionality of the data it was infeasible to estimate the parameters each time for all 756 itineration's, so the following simplification was employed: Divide the forecasting sample in 11 sub samples of 70 observations each, with the last one containing 56 observations. Then estimate the parameters a the end of each subsample for the next forecasting period. With the values for the first 70 forecast being the in-sample estimates of the parameters.
- All estimations along the rolling window (for AR(1) and MGARCH) were made by subtracting from the start sample the number of new entering observations in order

to keep a fixed number of observation on which running esrimates of 2011 values. Which for AR(1) implied losing one observation at the beginning in each itineration, as a new one was added; while for the MGARCH it was done in groups of 70 observations(entering new 70 ones from the forecasting sample, which are now realized values, and dropping the first 70 observations).

Table 3.3 below displays the estimates for the autoregressive and ARCH component of five different MGARCH models. The autoregressive component of the conditional variance is denoted by alpha, while beta represents its dependence on past square residuals. It is important to note that alpha_q and beta_q refer to the parameters used in estimating the correlation matrix of the innovations **Q**. In particular, beta_q represents the autoregressive component of $\mathbf{Q}_t$, while alpha represents the coefficient linking it to past square innovations ($\boldsymbol{\eta_{t-1}\eta'_{t-1}}$) These two parameters are constrained such that beta_q+alpha_q < =1.

The parameters alpha and beta are important in understanding the dynamics of the conditional variance in MGARCH models. The autoregressive component alpha captures the persistence of the conditional variance, while the ARCH component beta captures the impact of past squared residuals on the current conditional variance. By constraining beta_q + alpha_q to be less than or equal to 1, we ensure that the conditional variance remains positive and well-behaved. While in **Table 3.4**, standard deviation and t-statistics of the parameters are reported, since for large sample size the t-distribution converges to a Normal we can simply compare them with the z-score to deduce the significance of the parameters. It has to be pointed out that all the statistics reported besides alpha_q, beta_q and alpha and beta sBEKK are mean values of the parameter matrices of reference. Not all parameters were posted due to their numerosity, thus mean value are taken into consideration. The same applies to standard deviation (square root of average variance) and t-statistics .The averages were taken considering the absolute values of the single parameters/statistics in order to study the average magnitude. The DVECH estimation method involves estimating the elements of the lower triangular section of the parameters matrices Since the matrix is symmetric, we only need to estimate the lower triangular part, which has N*(N+1)/2 elements, repeated for the three parameter matrices as introduced by Bollerslev, Engle, and Wooldridge (1988). However, to ensure positive semi-definiteness of the matrix, we need to compute the parameter matrices **A** and **B** as **A=A*A'*** and **B=B*B'***, where **A*** and **B*** are NxN full matrices. This increases the number of parameters to be estimated to is N*(N+1)/2+2*N^2, which is computationally intensive and leads to incorrect and shaky

result for larger systems unless you have significant computation power and a perfected algorithm.

TABLE 3.3: Average Estimated parameters of MGARCH models

	N.of parameters	alpha	beta	alpha_q	beta_q
SBEKK	N*(N+1)/2+2 For N=7➔30	0.180	0.967		
DBEKK	N*(N+1)/2+2*N For N=7➔42	0.090	0.964		
DVECH	N*(N+1)/2+2*N^2 For N=7➔126	0.0397	0.139		
CCC	3*N+N*(N-1)/2 For N=7➔42	0.109	0.846		
DCC	3*N+N*(N-1)/2+2 For N=7➔44	0.109	0.846	0.022	0.942

TABLE 3.4: Average standard error and t-statistic of parameters MGARCH models

	sigma alpha	sigma beta	t alpha	t beta	Sigma alpha_q	sigma beta_q	t alpha_q	t beta_q
SBEKK	0.011	0.006	16.092	78.433				
DBEKK	0.019	0.011	22.689	121.89				
DVECH	0.013	0.008	5.637	49.807				
CCC	0.025	0.017	4.501	67.876				
DCC	0.025	0.017	4.501	67.876	0.019	0.005	5.179	49.445

The SBEKK model estimates the value of alpha to be 0.1802, with a standard error of 0.0112 and a t-statistic of 16.09, while the estimated value of beta is 0.9669, with a standard error of 0.0061 and a t-statistic of 157.43. These results suggest that past squared residuals have a smaller impact on the estimation of the conditional covariance matrix compared to past conditional covariance matrices. Furthermore, both parameters' t-statistics are highly significant, surpassing the critical value of 1.96 at a 5% significance level, indicating that the estimates are unlikely to be noise. The DBEKK GARCH model estimates alpha at 0.0902, with a standard error of 0.019 and a t-statistic of 22.689. The estimated value of beta is 0.9642, with a standard error of 0.0115 and a t-statistic of 121.883, similar in nature to the SBEKK GARCH. The CCC GARCH and DCC GARCH models both have the same estimated values for alpha and beta by structure, with the conditional variances estimated

through N separated univariate GARCH(1,1) models. The estimates for alpha and beta are 0.1086 and 0.8457, respectively, and both are significant. In particular, the beta_q of the DCC is much higher and closer to the SBEKK and DBEKK estimates of the conditional variance beta than to the estimates of the CCC and DCC for the univariate processes. In the DVECH GARCH model, the estimates for alpha and beta are 0.044 and 0.139, respectively, with the standard errors for these parameters at 0.013 and 0.008, respectively. The t-statistics are 5.637 and 49.807. Beta assumes the lowest values across all models, this isn't an unexpected result due to the high dimensionality of the DVECH estimator, being the largest parameterized model in the framework. Therefore, beta being much lower than the others is necessary to keep the conditional variance divergence in check in the matrix multiplication, the same logic is applied to alpha. The data clearly indicate that there is a significant persistence of conditional variance over time, which is highly significant, as evidenced by the beta parameters consistently having the highest t-statistic across all the models and highest magnitude, moreover both alpha and betas are all significant at 5%. Therefore, it is essential to choose an MGARCH specification that best fits the data. Although the differences in results among different models may appear small, they can have significant impacts on our forecasts. This highlights the importance of carefully comparing the results of different models and selecting the one that best suits the data to achieve accurate forecasts. It's crucial to consider potential biases and flaws in the estimates, and in case of substantial differences between estimates of different models, to study the roots of the issue and address them in order to mitigate possible model misspecifications or structural issues.

Parameters estimation within DVECH has been a major issue due to the dimensionality problem. In fact, the conditional variance estimates and forecasts of the DVECH are explosive moving further into the sample for N=7 in the order of 10^15, using the Sheppard Algorithm in the MLE toolbox and a function evaluation set at 3000 in the "optimset" function It is worth noting that while the estimates converged when checking the efficacy of the algorithm with five assets, they did not hold for seven, as shown by the explosive trajectories. The algorithm I developed had even worse performance in this regard, as it only produced accurate results convergent to Sheppard's when given very specific starting values. Although it ran much more quickly, using it for a large number of variables would produce results that were significantly biased, making it an unfeasible option for practical applications. The issue was eventually solved by setting the function evaluation parameter to a lower value of 1000 thus reducing the accuracy of the estimation but gaining in

parsimony. Moreover the “interior point estimation” of the “fmincon” function in MATLAB seemed to yield more consistent results in a 2011x7 dataset than the "trust-region-reflective" method, leading to an increase of computational resources to employ it in the procedure, as the “interior point method” is more computationally intensive but yield better results when there are a high number of constraints to be satisfied or when they are significantly binding.The “trust-region-reflect” method employed by default in MATLAB is a gradient-based optimization method that can handle nonlinear constraints, however, if the constraints are particularly difficult to satisfy, or if the problem has a large number of constraints it loses efficacy. Nevertheless trajectory still ended up exploding in the forecasting sample for thee forecasting subsets (namely T=3,4,8), as the increase of volatility caused by structural effect of the markets and by the Covid-19 aftereffects lead to the estimation of biased parameters in those timeframes, which caused explosive forecasts in the following periods. In order to lighten the problem I replaced the parameters matrices estimated in those windows with the ones estimated the previous iteration. In other words $\mathbf{A_3, B_3, C_3}$, and $\mathbf{A_4, B_4, C_4}$ were replaced by their counterparts estimated in forecasting period T=2 $\mathbf{A_2, B_2, C_2}$ and the same logic was applied on the matrices of the 8^{th} period which were replaced with the ones from the $7^{th.}$ Amongst these diverging parameters matrices there was hardly any difference in the matrix of constants C, while B which scales the autoregressive component was significatively larger than it was in previous and next estimations. The issue of dimensionality estimation in DVECH and other large sized parametric estimators has been addressed multiple times in the literature, such as in "A Large-Dimensional Factor Model for High-Dimensional Time Series" by Fan et al. (2008). In this paper, the authors propose a new method for estimating a large-dimensional factor model for high-dimensional time series data. They discuss the challenges of estimating the DVECH matrix when the number of assets is large, and propose a new method called the smoothed threshold factor model (STFM) that is designed to handle such situations. In particular, the authors note that the DVECH matrix can have a large number of parameters when the number of assets is large, which can make estimation difficult. They also note that the DVECH matrix can be ill-conditioned, which can lead to numerical instability in estimation procedures. The STFM method proposed in the paper is designed to address these issues by introducing a thresholding procedure that reduces the number of parameters in the DVECH matrix and improves its conditioning properties.

This was even worst for the algorithm I coded, which converged to Sheppard algorithms results only for precise starting values. Although it ran much faster application was unfeasible as results were quite biased when applied to a high number of variables

A different kind of issue ensued in computing DCC forecasts, in this case as well explosiveness of the variance estimated was detected, in this case the issue lied in the code that was used for forecasting future parameters as there was no issue in the estimation. The problem lied in the definition of the conditional correlation matrix, as we've already seen in chapter 1.8 $\boldsymbol{R_t} = \text{diag}\{\boldsymbol{Q_t}\}^{-1}\boldsymbol{Q_t}\text{diag}\{\boldsymbol{Q_t}\}^{-1}$, with $\mathbf{Q_t}$ defined as a multivariate ARMA(1,1) process, since $\mathbf{Q_t} = E_{t-1}(\boldsymbol{\eta_t}\boldsymbol{\eta_t'})$, it is equal to $\boldsymbol{R_t}$ in its population representation while It should converge to it asymptotically in its sample counterpart. Thus, the algorithm was initialised considering $\boldsymbol{Q_{t-1}} = \boldsymbol{R_{t-1}}$ this value was used to estimate $\mathbf{Q_t}$ through its ARMA representation and then the estimate was used as input to compute $\boldsymbol{R_t}$ through the formula reported above. However, while theoretically correct this approach eventually lead to explosive forecasts, the two possible reasons were either that 2011 observations weren't a significantly large enough sample for **Q** to converge to **R** or that the model itself was misspecified for the true generating process of the returns. As its explosiveness could hardly be bounded by a larger data set do to its autoregressive component a larger dataset would only have pushed the trajectories of the estimator further away from its true theoretical value. So, issue may have laid in the starting point of initialization but that didn't seem to be the case as it wasn't greatly different from other models conditional correlations or from sample corelation itself. It should be reminded that the estimation process of the parameters was carried iteratively every 70 periods, to overcome the issue I simply considered $\mathbf{Q_t} = \boldsymbol{R_t}$ i.e. I assumed its asymptotic properties to hold within our sample, and directly computed $\boldsymbol{R_t}$through $\mathbf{Q_t}$ ARMA process:

$$\boldsymbol{R_t} = \boldsymbol{S} \circ (\boldsymbol{\iota}\boldsymbol{\iota}' - \boldsymbol{A} - \boldsymbol{B}) + \boldsymbol{A} \circ \boldsymbol{\eta_{t-1}}\boldsymbol{\eta_{t-1}'} + \boldsymbol{B} \circ \boldsymbol{R_{t-1}}$$

This lead to realistic forecasting results and was used in practice to carry on with the analysis, however the issue of misspecification within our sample remains. When checking for the RMSE_Adj (RMSE for volatility forecasting power, adjusted for scale) for the different models in the portfolio optimization it resulted that under this metrics DCC fared worse than any other model in forecasting power over volatility in a portfolio setting which puts the emphasis mostly conditional correlation itself. Summing up these considerations it should be reasonable to conclude that DCC model has proved itself quite unable to explain the time varying dynamics between the data structure of the 2020-2022 forecasting sample.

Tables containing different descriptive statistics for the portfolio optimization are listed further ahead, amongst them RMSE_Adj for DCC was the highest in all the different settings, indicating "bad fit" of the model.

Tables 3.5A and **3.5B** display the in-sample correlation matrices of the seven ETFs, where the first is the sample correlation matrix, while the second the correlation matrix estimated with CCC model:

Table 3.5A: Sample correlation Matrix (2012-2019)

	QQQ	FXO	EEM	XLU	DBA	VNQ	XLV
QQQ	1.000	0.766	0.701	0.274	0.144	0.485	0.768
FXO	0.766	1.000	0.672	0.298	0.154	0.573	0.713
EEM	0.701	0.672	1.000	0.302	0.265	0.480	0.574
XLU	0.274	0.298	0.302	1.000	0.046	0.608	0.344
DBA	0.144	0.154	0.265	0.046	1.000	0.118	0.110
VNQ	0.485	0.573	0.480	0.608	0.118	1.000	0.499
XLV	0.768	0.713	0.574	0.344	0.110	0.499	1.000

Table 3.5B: CCC estimated correlation Matrix (2012-2019)

	QQQ	FXO	EEM	XLU	DBA	VNQ	XLV
QQQ	1.000	0.733	0.691	0.254	0.141	0.456	0.736
FXO	0.733	1.000	0.644	0.277	0.141	0.527	0.670
EEM	0.691	0.644	1.000	0.272	0.248	0.444	0.560
XLU	0.254	0.277	0.272	1.000	0.047	0.590	0.332
DBA	0.141	0.141	0.248	0.047	1.000	0.105	0.111
VNQ	0.456	0.527	0.444	0.590	0.105	1.000	0.467
XLV	0.736	0.670	0.560	0.332	0.111	0.467	1.000

We can see that the two matrices show a similar pattern of correlations between the assets and that all ETFs are positively correlated in both of them. However CCC estimates correlations are adjusted for the volatility of the assets, whereas the sample are all lower than their sample counterpart. For example, the correlation between QQQ and XLU is 0.25 in the GARCH CCC-model, while it is 0.27 in the sample correlation matrix, similarly for the others. Another point to make is that the GARCH CCC-model tends to estimate lower correlations between assets that have higher volatility. This is because CCC adjusts is correlation component to keep the overall covariance matrix in check, in fact we can see that the biggest differences in the two matrices are in QQQ, XLV and VQN, which are the three assets with highest variance within the estimation sample. The highest divergence corresponds to the pair VNQ-FXO which amounts at 0.046 points, while the lowest is surprisingly XLV-DBA for 0.01 difference, however this is caused by the fact that DBA is

the ETF with lowest volatility in our set, thus filtering the CCC tendency to contain correlation magnitude for the pair.

When selecting the best GARCH model, there are several criteria that can be used to evaluate the goodness of fit and the complexity of the model. The four commonly used criteria are the Bayesian Information Criterion (BIC), the Akaike Information Criterion (AIC), the Hannan-Quinn Information Criterion (HQC), and the negative log-likelihood(NLL). The BIC, AIC, and HQC are all information criteria that balance the fit of the model with its complexity. They provide a trade-off between a good fit to the data and a model that is not too complex. The BIC penalizes more heavily model complexity compared to AIC, which tends to select larger models. The HQC is a modification of AIC that gives more weight to the number of observations in the sample. The negative log-likelihood, on the other hand, is a measure of how well the model fits the data, with lower values indicating better fit.

In **Table 3.6** below, we compare the values of these four criteria MGARCH model to evaluate their relative performance, with the best model being chosen based on the lowest value of the criteria.

TABLE 3.6: Model Selection Criteria

	NLL	BIC	AIC	HQC
SBEKK	-14386	-28544	-28713	-28712
DBEKK	-14592	-28865	-29100	-29099
DVECH	-14388	-27817	-28524	-28520
DCC	-14344	-28353	-28599	-28598
CCC	-14499	-28679	-28914	-28913

As we can see the DBEKK presents the lowest Negative Loglikelihood while the DCC's values is the highest, thus under this evaluation they should be the model that best fits the data and the worst one respectively, moreover due to the extremely low value reached by the loglikelihoods evaluated in their optimal estimated points, the evaluations hardly change along the different metrics. The only information criteria that gives significant weight to the number of observation is the BIC, accordingly it changes the evaluation of the DCC respect to the Loglikelihood, placing it above t the DVECH due to the high number of parameters present in the DVECH model which amounts to 126 scaled by the log of the number of observations. In the BIC's evaluation the DVECH is considered the least optimal models

amongst the ones proposed, while the SBEKK is still inferior to the CCC. The AIC model still penalizes the DVECH enough for the DCC to surpass it, although by a lower margin; keeping the same conclusions of the BIC and the same applies to the HQC.
The information obtained through these criterions are however far from being conclusive, especially taking in consideration the scaling issue brought forward by the significant magnitude of the loglikelihoods which leads to weakened effect of the parameters and weight penalties applied by the different criterions.

Below will be reported a series of graphs with estimated and forecasted conditional volatility and correlations for three of the seven ETFs, due to the high number of correlations combination amongst seven assets three out of the seven were picked to display a graphical overview. The ETFs presented are QQQ, FXO and EEM.

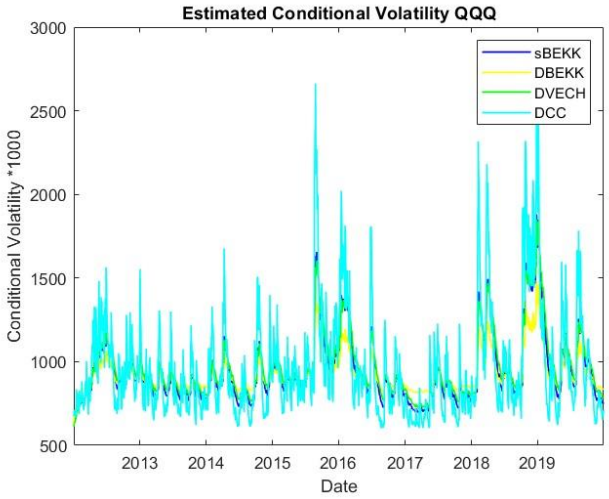


Figure 4.a

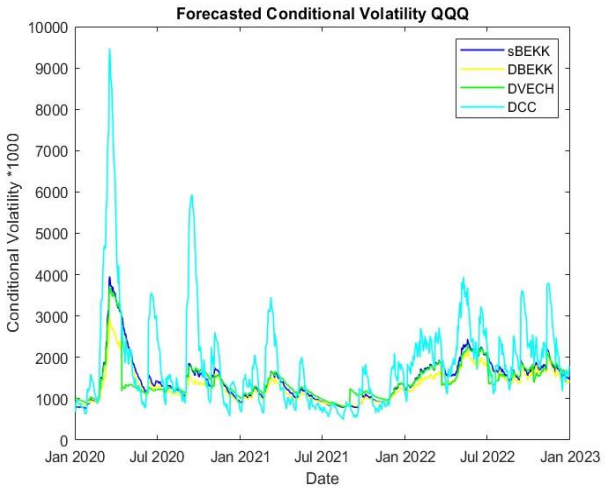


Figure 4.d

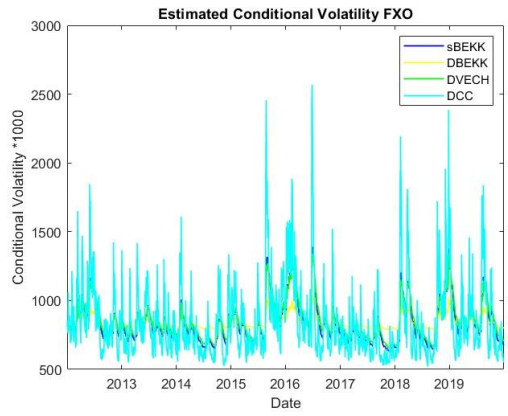


Figure 4.b

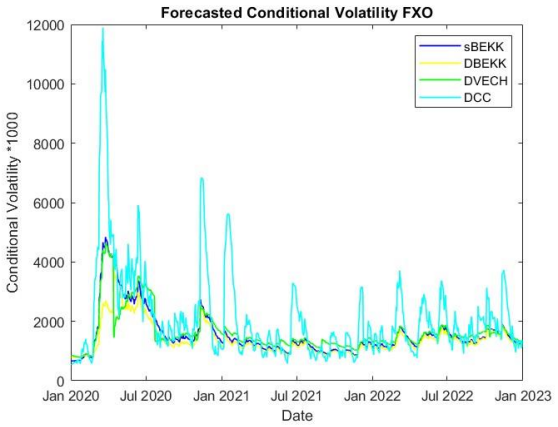


Figure 4.e

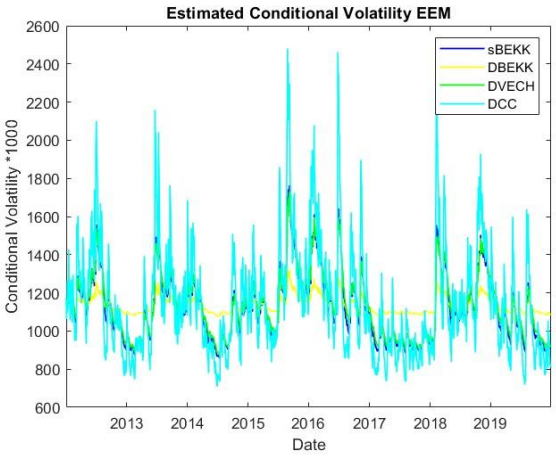


Figure 4.c

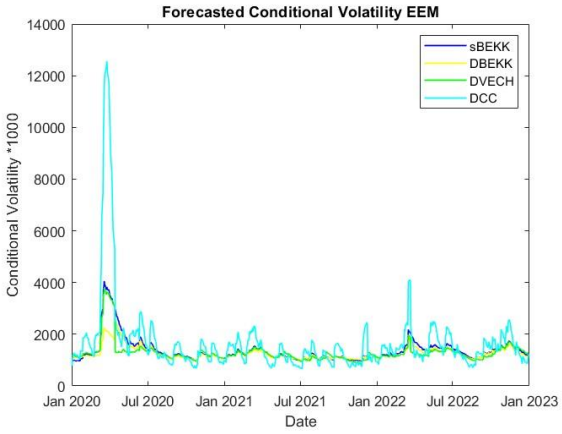


Figure 4.f

In **Figures 4.a, 4.b** and **4.c** the in-sample estimates of the conditional volatilities scaled by 1000 are shown for the three ETFs. CCC was not included as its Variance estimates coincide with DCC by structure. As we can see DCC consistently estimated higher magnitudes of dispersion across all three assets with higher dispersions o the results as well as we can notice for the larger movements it displayed across all assets, it appeared to be the most unstable estimator with relative high peaks and low lows.. DVECH and SBEKK have yielded similar results, with their graphs almost overlapping in all three examples; while the DBEKK gave lower estimates than other, in particular in EEM its projections are much "flatter than the others". **Figures 4.d, 4.e** and **4.f** represent the corresponding forecasts for the three assets along the forecasting sample (2020-2022), we can clearly se that forecasts present a similar pattern to the estimation samples; with DCC application leading to higher and more volatile

projections presenting significant peaks in correspondence to the subsamples in which residuals displayed more extreme behaviours as it can be noticed by its highest peak in the first half of 2020. The other three models displayed similar forecasts, with SBEKK predicting higher variations on average; DBEKK was much closer in predictions to the other two models respect to the estimation sample, this pattern can be clearly seen comparing the EEM forecast with its sample respective, in-sample estimations were much "flatter". Lastly DVECH tracked SBEKK and DBEKK for the whole horizon.

In the same fashion as for the conditional volatilities, below are plotted the conditional correlations of the three assets, the line parallel to the x-axis represents the correlation estimated by the CCC model which is constant; besides our five model it was added the conditional correlation proposed by RiskMetrics and introduced in chapter, the smoothing parameter used here is 0.94. On the LHS we will present the in-sample conditional corelations and on the RHS their forecasted estimates.

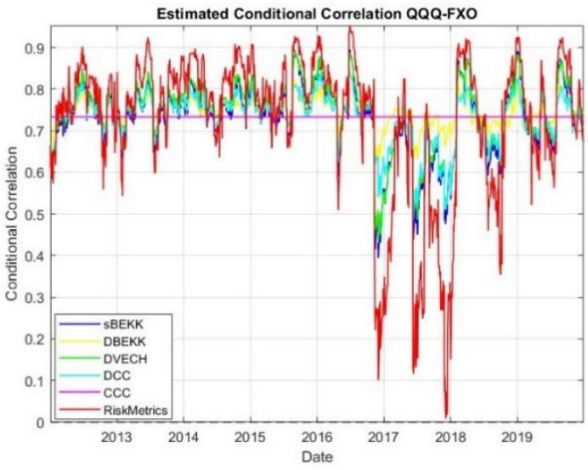


Figure 5.a

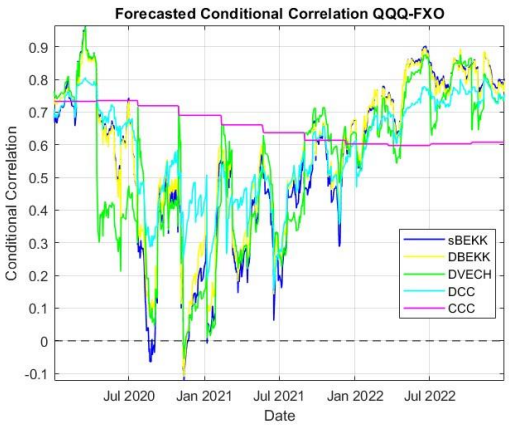


Figure 5.d

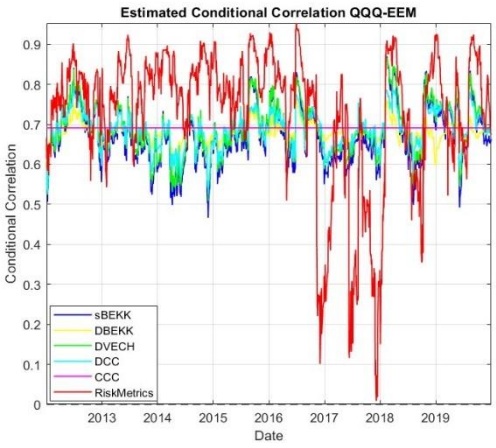


Figure 5.b

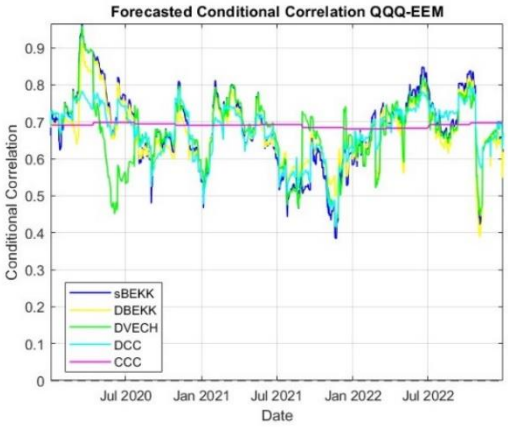


Figure 5.e

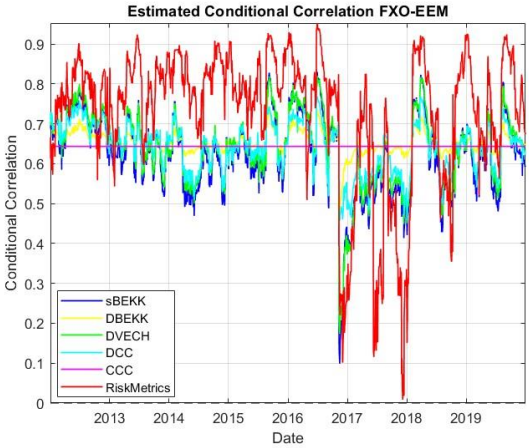


Figure5.c

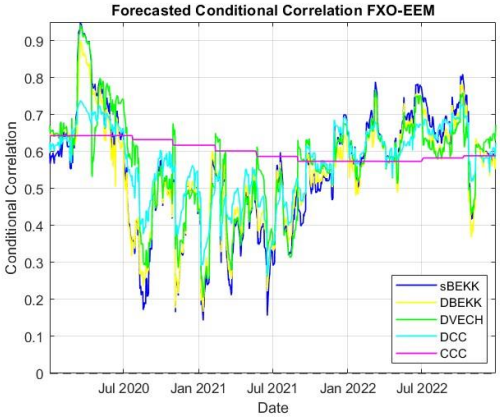


Figure 5.f

In the **Figures 5.a**, **5.b** and **5.c** are presented the different correlation estimations for the the five models, plus the ones obtained through the RiskMetrics estimator. As we can see RiskMetrics' is much noisier than the other models estimates with significant swing up and down along the estimation horizon, it mostly placed significantly above or below the other models. CCC is apereas to be a mean value around which the other oscillate, which makes sense given its time independent nature. SBEKK, DBEKK and DVECH yielded quite similar results, with DVECH and SBEKK almost overlapping each other DBEKK oscillating above and below. All estimators yielded strictly positive correlation, with only RiskMetrics almost hitting the zero mark in 2018. In **Figures 5.d, 5.e** and **5.f** are displayed the respective

forecasts, the noisiest estimator appears to be the DVECH in all three pairs, it still tracks the SBEKK while the DBEKK increases the distance from the other two, keeping an higher expectation on future correlation on average. Negative correlation where reached only on the pair QQQ-FXO in 2020 and 2021, by DVECH, DBEKK and SBEKK with the latter pushing furthest downwards respect to the other two in both occasions. CCC's presents a "stairs" outlook as it was estimated 10 times in the rolling window, leading to ten different constant correlations forecasts with descending trajectory, without any significant jump. The similar behaviour displayed by DVECH, SBEKK and DBEKK in both estimating and forecasting variances and correlations does not come as a surprise; in fact, as shown by Engle and Kroner(1995) in their paper "Multivariate Simultaneous Generalized Arch", every single DVECH model can be expressed as a specific case within the DBEKK models class, These models belong essentially to the same family, with the DBEKK framework being a generalization. We can clearly see this close relationship even graphically, as in all the examples above they're the ones that performed the most similarly even in the forecasting environment.

3.3.1 PORTFOLIO OPTIMIZATION: MINIMUM VARIANCE

In this section, we will use the Markowitz Portfolio Optimization approach to construct two optimal portfolios in a dynamic setting where we rebalance the portfolio at each iterations, for a total of 756 iterations spanning across three years. For both portfolios, we will estimate the covariance matrix of the assets using the multivariate GARCH forecasts from the different models obtained previously. The first portfolio will be constructed using a simple mean-variance optimization procedure, which assumes that the investor is risk-neutral and interested only in maximizing returns. The second portfolio will be constructed with a Value at Risk (VaR) constraint to ensure that the portfolio is constructed with a specified level of risk that is acceptable to the investor, and we will further add two parameters that shape investors risk aversion to dispersion of returns and extreme returns respectively, providing insights into the sensitivity of the portfolios to changes in investor risk preferences.

As discussed in chapter 2.2 Markowitz minimum variance optimal portfolio is the particular portfolio that minimizes risk for a given level of expected return, in this application we will compute the optimal mean-variance portfolio for each period of the forecasting sample through a rolling window. As time goes on, new information's will enter our datasets that will lead to changes in the optimal portfolio structure, in such a sense the portfolio is

dynamic, time conditional and with variables weights. However, as will be discussed in the second application with VaR constraint, this is not an intertemporal application. As our optimization takes int account only one period at a time and doesn't aim to maximize the objective function considering the whole stochastic (as it still unknown) future dataset. The resulting weights are the decision outcome of T optimizations, with T being the number of one-step ahead forecasts in our investment, it can be considered a rolling forecast as new observations will gradually enter the estimation dataset to infer the changing relationship between the variables and estimate new parameters values. In this particular case the size of the forecasting sample is of three years, composed more specifically by 756 daily forecasts. At the end of each day we will run an optimization on the new dataset which includes one new point and loses another at the beginning of the data string, we will use this information to carry out forecasts to use as input in our optimization algorithm in order to choose the optimal weights for the next period. In particular we will run Covariance forecasts through the Multivariate GARCH models encountered in the previous section, which will be the minimization inputs of our objective function. This particular optimization belongs to the first class of portfolio optimization techniques discussed in chapter 2.2, i.e variance minimization under an Expected return constraint.

The model specification is the following:

$$\text{Min}(\mathbf{w_t'H_tw_t}) \qquad (1)$$

Subject to:

$$\mathbf{W_t}'E_{t-1}[\mathbf{r_t}] >= E_{t-1}[r_{mkt,t}] \qquad (2)$$

$$\mathbf{z_t}'\boldsymbol{i} <= 1.66 \qquad (3)$$

$$\mathbf{w_t'w_t}=1 \qquad (4)$$

to iterate for every t=1,..,T

Where $\mathbf{z_t}$ is an auxiliary variable defined as $\mathbf{z_t}=[|w_{1,t}|,..,|w_{N,t}|]'$, and $\boldsymbol{i}$ an Nx1 vector of ones.

(1) Is our minimization objective function, where $\mathbf{w_t'H_tw_t}$ expresses the conditional variance estimated through MGARCH forecasting, scaled by the risk aversion parameter "tau" which we introduced previously and represents investors aversion to dispersion in returns.
(2) Is our constraint on portfolio Expected returns, coinciding with the expected return of the market portfolio which was computed as a sample moving average of the market return proxied by the S&P500 index. (3) Allows for the option of short selling up to a maximum of 33% per unit, this value was chosen as neighbourhood of regulatory constraints on short sales, which are quite heterogeneous, but often lay within the range of the 30%. Lastly (4) imposes the constraint that the capital has to be fully invested in the portfolio.

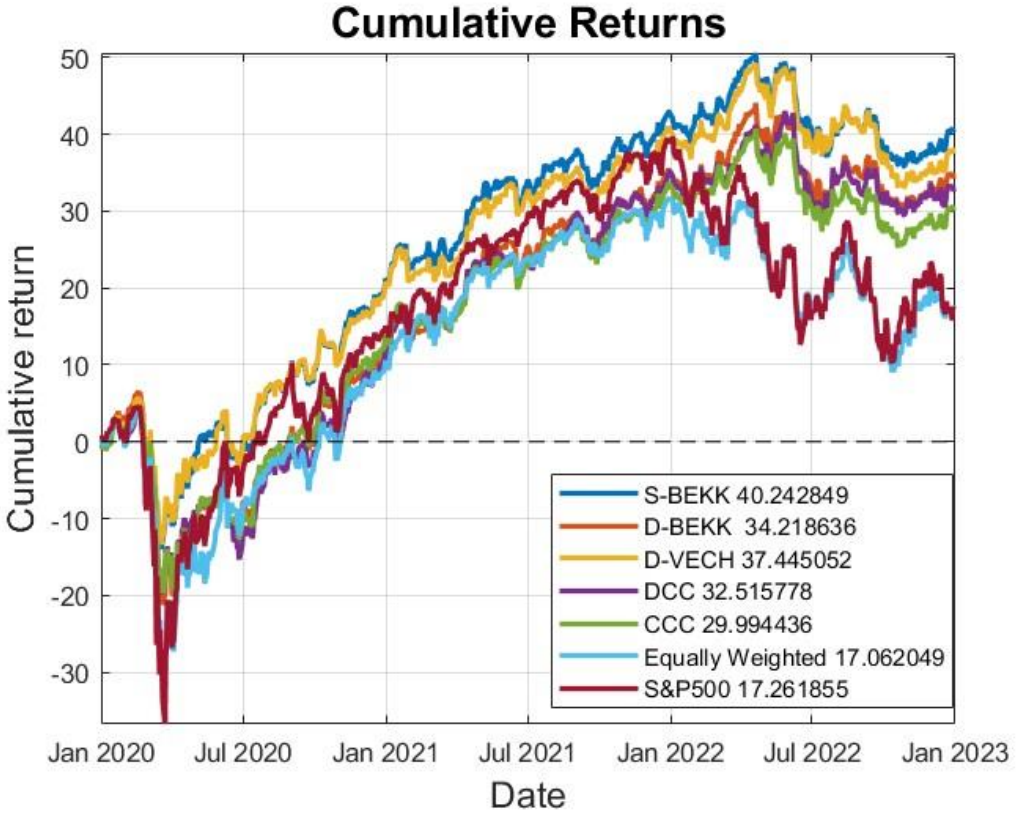


Figure 6

In Figure 6 are presented the cumulative returns of the different GARCH models driven portfolio allocations, SBEKK and DBEKK are the ones who achieve consistently a better performance, with SBEKK coming on top at 40.24%. The worst performer was CCC at 29.94%. Cumulative returns have been negative for a fraction of time spanning the second and third quarters of 2020, accordingly to the market trend. From the first half of 2020 to Jan 2022 returns seem to have kept increasing across all portfolios bouncing back from the lows of the crisis, but seem to have stopped their ascension during the last year, with S&P500 and the

equally weighted portfolio tracking back towards the zero mark. Even the worst portfolio under CCC specification scored 12% more than the market, but it laid below it on the graph for most of the time horizon of reference, in particular during the whole recovery phase.

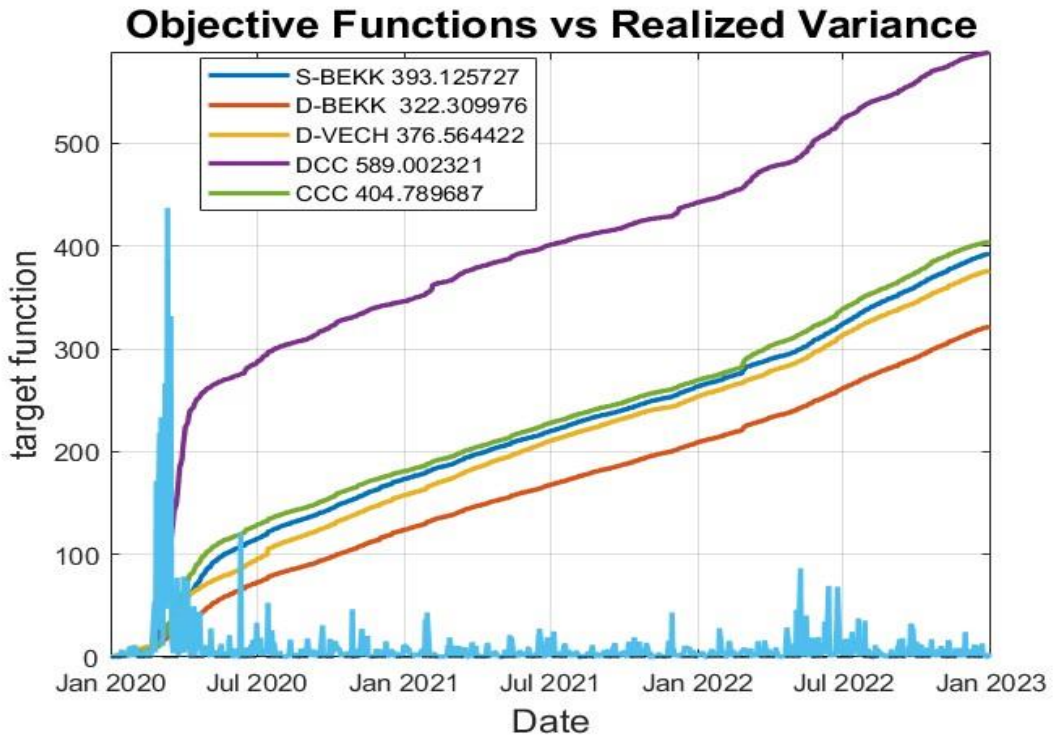


Figure 7

Figure 7 displays the dynamics of the minimization objective function vs the realized variance. Being the objective function only driven by returns dispersion in this particular case, the graph displays a steep increase in the objective function as realized variance shot up in early 2020. It should be remembered that unlike the Model discussed further below, in which we want to maximize a function, here our aim is minimization. Suc an increase in the function is obviously sign of loss of "efficacy" of the models to reach optimum point at the given market conditions. Realized Variance in the graph was scaled by 10. DCC was the worst performer, maintaining significance distance from the others for all three years, SBEKK,DVECH and CCC yielded similar results showcasing small differences on their curves while DBEKK clearly overperformed. This graph would suggest that the model performed best under DBEKK conditional covariance specification for the ETFs.

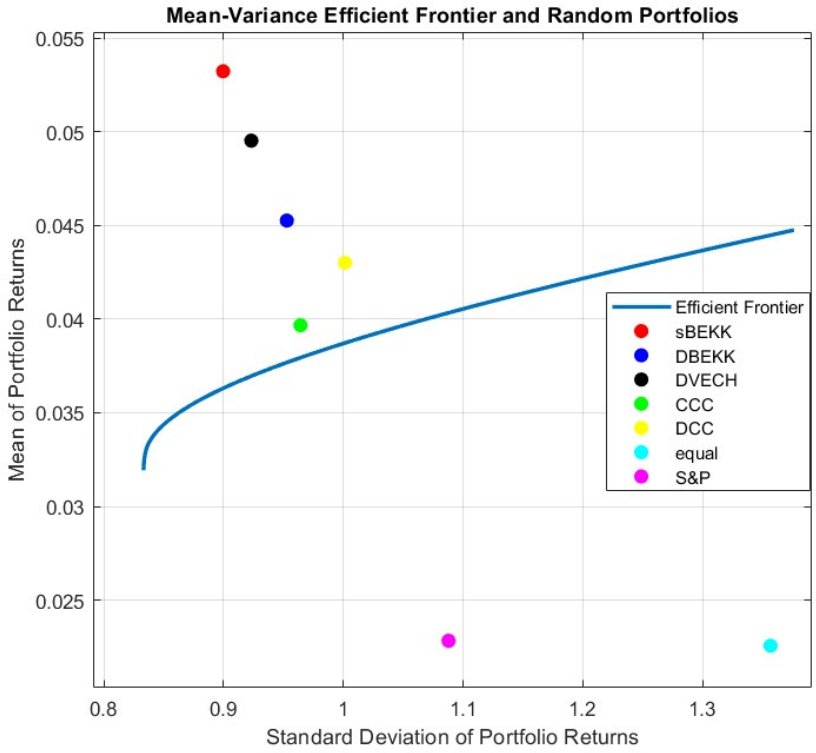


Figure 8

In **Figure 8** the efficient frontier is plotted for the forecasting sample timeline. It's an ex-post efficient frontier as in the standard framework, where the line represents the optimal allocation of portfolios weights in order to maximize mean portfolio return for a given level of risk. As we can see our portfolios lie above the frontier, this is because they're obtained through a dynamic allocation with changing weights at each itineration, while the classical efficient frontier only offers a static snapshot evaluated at a given time (in this case at the end of our forecasting horizon). As we can see, all portfolios performed better than the one suggested by the frontier, hinting that even the most misspecified of our models, if actively managed, can outperform the "optimal" static allocation. S&P500 and the equally weighted portfolio lie within the frontier as their weight are constant and thus are accounted for in its optimization process. As we notice, they're far from being optimal, even in a static environment.

In **Tables 3.7** and **3.8** below are reported the weights of the different asset allocations according to the various portfolio and the average value of the maximization objective function in our minimum variance framework for the various models. A number of

descriptive statistics and performance metrics are presented, amongst which: Realized volatility computed as the absolute value of daily returns, RMSE adjusted for volatility computed as the square root of the average square difference between realized volatility and GARCH forecast estimates normalized by realized volatility in order to present the measure of fit as a percentage of the realized volatility itself, Information ratio using the S&P500 index as benchmark portfolio, the VaR at 95% significance level and the Sharpe Ratio over the Fed 3-months Treasury bill return rate.

The Sharpe ratio was computed with the closing ofthe time varying rates posted on a daily basis.

TABLE 3.7: Portfolios Weights

	S-BEKK	D-BEKK	D-VECH	DCC	CCC
QQQ	0.006	0.020	-0.005	0.024	0.025
FXO	0.032	0.056	-0.012	0.028	0.022
EEM	0.053	0.013	0.072	0.072	0.021
XLU	0.203	0.199	0.172	0.189	0.188
DBA	0.459	0.478	0.453	0.387	0.430
VNQ	0.010	0.016	0.043	0.051	0.051
XLV	0.237	0.218	0.278	0.248	0.262
Obj. Fun.	0.520	0.426	0.498	0.779	0.535

From **Table 3.7** we can observe that for some of the assets the weight allocations s vary significantly across the different models. For instance, the weight allocation for XLV ranges from 0.218 to 0.278 in DBEKK and DVECH respectively, similarly, the weight allocation for DBA ranges from 0.3847 in DCC to 0.478 in DBEKK.

An even bigger divergence was encountered in the average value of the objective function The D-BEKK model has the lowest average value at 0.426, while the DCC has the highest scoring 0.779. This suggests that the DCC model may be the least effective under variance minimizations constraints in the analysed sample, as we're in a minimization framework our objective is to minimize the objective function. We can also notice that the weight allocations for the different models relatively consistent across the assets, they're all centred around DBA with DCC, as the most conservative in this regard, allocating 0.387 of its portfolio to it. In particular, all the models seem to distribute over 80% of the weights to three single assets namely DBA, XLU and XLV; with the second best falling significantly short at 7% (DCC-EEM). This shows that pure variance optimization through MGARCH leads to low

level of diversification in this study, however the assets are ETFs themselves so they're diversified financial instruments even as stand alone. DBEKK scored the lowest in the objective function average suggesting it may be the model best fit to minimize the variance in this framework, while CCC surpassed both S-BEKK and DVECH at 0.5354. Moreover DCC which scored the highest in the objective function evaluation, is the one with the highest level of diversification giving the least weight to DBA compared to all other models and distributing the extra space to other less popular assets as VNQ and EEM. DVECH is the only model which has average negative weights for an asset, implying higher short sales over that asset along the horizon, in QQQ and FXO with weights -0.005 and -0.012 respectively.

TABLE 3.8: Descriptive Statistics and Risk adjusted Measures of portfolios log-returns

	Mean	Max	Min	Total Return	RV	VaR	Sharpe Ratio	Info Ratio	RMSE_Adj
S-BEKK	**0.053**	**5.060**	-5.857	**40.243**	0.901	**-1.426**	**0.057**	**0.025**	0.327
D-BEKK	0.045	4.633	-7.218	34.219	0.953	-1.364	0.046	0.020	0.385
D-VECH	0.050	5.048	-5.954	37.445	0.924	-1.352	0.052	0.023	0.422
DCC	0.043	4.808	-7.360	32.516	**1.002**	-1.467	0.041	0.018	**0.502**
CCC	0.040	4.948	**-7.482**	29.994	0.964	-1.434	0.039	0.015	0.406
S&P	0.023	8.968	-12.765	17.262	1.610	-2.476	0.013		
Equal Weight	0.023	7.893	-12.258	17.062	1.356	-1.985	0.015	-0.001	

Referring to **Table 3.8,** we observe that the mean return and total return vary significantly across the different models. S-BEKK has the highest mean return of 0.053 and the highest total return of 40.24. On the other hand, CCC has the lowest mean return of 0.0397 and the lowest total return of 29.99. This suggests that S-BEKK may be the best model for generating returns and maximizing the performance of the portfolio. Additionally, the maximum and minimum returns for each model also vary significantly. S-BEKK still has the highest maximum return at 5.05, but it loses to both S&P500 and the equally weighted portfolio as they display much higher levels of volatility, while DCC has the lowest maximum return of 4.8077. Similarly, CCC has the lowest minimum return of -7.48, while S-BEKK has the highest minimum return of -5.85. We can also see that the range of returns for each model is different. S-BEKK has the lowest range of returns, with a difference of 10.8 between the maximum and minimum, DCC, on the other hand, is the highest in this regard, with a difference of 12.5. By these measures alone it would appear that S-BEKK is the most stable

and appetible model amongst the ones analysed, as the data suggests that S-BEKK may be the best model for generating returns and maximizing the performance of the portfolio. In terms o realized volatility used as proxy of uncertainty in the marker DCC showcase the highest level at 1.002, while S-BEKK the lowest at 0.901. Moving on to the VaR DCC still brings along the highest loss -1.426 with 5% probability, while the SBEKK's is also quite high following close behind with a VaR of -1.43 and ranking sthirs amongst the models, right behind the CCC at -1.43. These values are in line with the average 5% quantiles of the single assets showcased in the in-sample analysis which is a bit counterintuitive since the forecasting sample was characterized by a quite higher level of volatility. The relatively low VaR of the portfolios is good evidence of the efficacy of the minimum variance framework, at least in balancing risk. The lowest VaR was displayed by the DVECH model, close enough to the others, there doesn't appear to be any significative difference amongst the VaRs of different portfolios. In terms of Sharpe Ratio s-BEKK has to be preferred as well, outclassing the second-best portfolio by 0.06 at 0.0572. CCC scores quite purely compared to the others, with a Sharpe Ratio of 0.039. Information Ratio displays then same dynamics, with SBEKK coming on top 0.0251 and CCC placing last with 0.015, other models are centred around 0.02. Since we computed the Information ratio using the S&P500 Index as benchmark this indicates that all portfolio convincedly beat the market in risk-adjusted terms as well and not just in gains or stability. Finally, we take a look at the Adjusted RMSE, being normalized with respect to the realized volatility it represents some sort of unfitness of the models to the noise of the market in percentage terms. It should be highlighted that using realized volatility computed as absolute value of the returns, although simple in application, can be quite misleading and we shouldn't be too confident in this "fitness results". Regardless, SBEKK had the best performance under this metrics as well with 0.327, which is still quite an high value as it expresses a percentage. The worst performer was DCC at 0.502 which was surprisingly overtaken by the CCC at 0.405. CCC relative performance in this metrics was quite unexpected as it lagged behind in most of the analysis so far, her however he performed better than DVECH as well, placing extremely close to the DBEKK at 0.385.

Overall, by comparing the various metrics we can conclude that SBEKK model is the superior model within a simple variance minimization framework. It dominated all of the performance risk indices by a good margin on average, however it lost to the DCC in terms of average Objective function evaluation which may indicate that this model was more performance oriented rather than purely focused on risk minimization. As it displayed the

lowest Realized volatility this may sound counterintuitive, however its was amongst the highest, suggesting that it was more skewed towards extreme returns than DBEKK and DVECH. CCC was by far the worst performer, but in the RMSE Adjusted, dominated one another in different metrics making it harder and more subjective to judge which one performed better. DCC had the highest Objective function evaluation and highest VaR in absolute terms indicating a relatively risky allocation , it was the model which most differentiated amongst the different assets in terms of weight allocations strangely enough may be one of the main drivers behind its poor performance, proof of is its RMSE which was the highest. Under the assumption that RMSE gives solid evidence of unfitness in this sample the choice to hold a portfolio with higher level of diversification respect to its competitors ,may not be the cause of DDC performance but rather proof and consequence of its poor predictive ability in the sample . All models displayed a really high Adjusted RMSE, which would indicate significant model misspecification. Main drivers o this outcome are probably the gaussianity assumption, which facilitates estimations at the cost of misspecification on the tails f the distribution. This was further accentuated by the global crisis which it the whole economy during our forecasting timeframe, causing a structural change in the data dynamics at least in the short term, far higher volatility was pervaded the market respect to our sample observations with an increase of extreme values; our models failed to capture this distorted price structure, however it should be noted that every single portfolio significantly outperformed the S&P500 and the equally weighted under all metrics suggesting that our volatility estimates, and optimization choice still yielded superior performance to passive investment, even under adverse conditions which weakened further the already weak Normality assumptions.

3.3.2 PORTFOLIO OPTIMIZATION: VaR CONSTRAINT

We now move on to a more complex model, introducing VaR constraints and investors risk aversion. We introduce two different kind of parameters, one parameterizes aversion to dispersion in returns (tau), while the other is linked to the constrained we impose to the VaR(gamma). In particular we constraint our portfolio such that the VaR tolerated by the investor can’t be higher than gamma. In this case our objective function is to maximize a function of the portfolio expected return and a penalty loss for variance exposure, under a Var constraint and a constraint on the expected value of our portfolio such that we want it to be higher than the expected value of the market, indexed by theS&P500.

In “Portfolio Optimization with Conditional Value-at-Risk Objective and Constraints" by R. Kouwenberg and R. Vorst, published in The Journal of Banking and Finance in 2004, used a range of risk aversion parameter values between 0 and 5 in their analysis of portfolio optimization with a conditional value-at-risk (CVaR) objective function. The authors note that the optimal value of the risk aversion parameter will depend on the investor's risk preferences. In this analysis, we chose to use a range of risk aversion parameter values “tau” based on recommendations in the literature. Specifically, we used a value of 0.5 as a benchmark for the limiting case of an highly risk tolerant investor which mostly aims to maximize expected return with much less regard to the risk it carries, a value of 5 as a benchmark for a moderately risk-averse agent based on the recommendation of Markowitz (1952) and other studies (e.g., Jagannathan and Ma, 2003; Kouwenberg and Vorst, 2004), and a value of 10 as a benchmark for a highly risk-averse investor. These values are consistent with the findings of previous studies, which have suggested a range of values for the risk aversion parameter depending on the level of risk aversion and the specific investment goals and constraints, such as “Robust Portfolio Optimization" by S. Boyd et al which, introduces a robust optimization framework for portfolio optimization that takes into account the uncertainty in the inputs. The paper suggests a range of risk aversion parameter values between 0.1 and 10. By using a range of values for the risk aversion parameter, we are able to explore the impact of different levels of risk aversion on the optimized portfolio, and to provide a more robust analysis that takes into account a variety of risk preferences. As for tau, the appropriate value of gamma in portfolio optimization with VaR constraints depends on the investor's risk preferences and investment objectives. In the literature, different authors have used different values of gamma based on their assumptions about the investor's risk tolerance. For example, Meucci (2009) in “Risk and asset allocation (Vol. 1). Springer Science & Business Media” suggests using gamma values ranging from 1% to 10% of the portfolio value, depending on the investor's risk aversion. He notes that a gamma value of 1% would correspond to a 99% confidence level for VaR, while a gamma value of 10% would correspond to a 90% confidence level. Similarly, DeMiguel et al. (2009) “Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy?” consider gamma values ranging from 1% to 5% for different risk tolerances. Pelizzari in “A stochastic dominance approach. European Journal of Operational Research (2015)” considers gamma values ranging from 0.5% to 5. Here due to values of 3% and 7.5% were chosen benchmarks. In our optimization problem, the gamma value is a parameter that we use to set the minimum value we want our VaR to have. This parameter is inversely related to the quantile of the

theoretical distribution of our portfolio. We specify the gamma value based on the desired confidence level for the VaR constraint. Since we want our VaR at a 95% confidence level to be higher than -gamma, this means that the 0.05 quantile of the portfolio return distribution has to be associated with a higher quantile of an unknown distribution. We considered two different scenarios: a risk-averse investor, for which we wanted to associate a low gamma value, and a risk-tolerant investor. To set the gamma value for the risk-averse investor, we took the average of the 7.5% quantiles of the sample return series as the value of gamma. For the risk-tolerant investor, we took the mean of the 3% quantiles of the sample return series as the value of gamma. This choice of gamma coupled with the three choice of tau allow us to study the performance of the portfolio under six different risk scenarios for each model. The maximization problem falls in the third of the three main categories discussed in chapter 2.2 and can be expressed as follows:

$$\text{Max}(\mathbf{w_t}'E_{t-1}[\mathbf{r_t}] - \tau \mathbf{w_t}'\mathbf{H_t}\mathbf{w_t}) \qquad (5)$$

Subject to:

$$\mathbf{w_t}'E_{t-1}[\mathbf{r_t}] + Z_\alpha(\mathbf{w_t}'\mathbf{H_t}\mathbf{w_t})^{1/2} >= -\Upsilon \qquad (6)$$

$$\mathbf{W_t}'E_{t-1}[\mathbf{r_t}] >= E_{t-1}[r_{mkt,t}] \qquad (7)$$

$$\mathbf{z_t}'\boldsymbol{i} <= 1.66 \qquad (8)$$

$$\mathbf{w_t}'\mathbf{w_t} = 1 \qquad (9)$$

to iterate for every t=1,..,T

Where $\mathbf{z_t}$ is an auxiliary variable defined as $\mathbf{z_t}=[|w_{1,t}|,..,|w_{N,t}|]'$, and $\boldsymbol{i}$ an Nx1 vector of ones. (5) Is our maximization objective function, with $\mathbf{w_t}'E[\mathbf{r_t}]$ expressing the expected value of the portfolio return in (t-1), since the optimization is carried on in a dynamical setting in which in each period we want the optimal portfolio weights to hold until the next iteration. $\mathbf{r_t}$ **is** our Nx1 vectors of asset returns**,** in this application N=7, the expected value is computed as the sample moving average for the first 50 observations and through an AR(1) specification from the 51^{th} onwards. $\tau\mathbf{w_t}'\mathbf{H_t}\mathbf{w_t}$ expresses the penalty term of our objective

function, with the conditional variance estimated through MGARCH forecasting , scaled by the risk aversion parameter "tau" which we introduced previously and represents investors aversion to dispersion in returns. (6) Is the analytical formulation of our VaR constraint, where Z_a is the quantile of the portfolio returns distribution corresponding with the (1-α) quantile of the loss distribution function. Since we assumed that our assets are distributed as Multivariate Normal and that the sum of the weights is equal to 1, $\mathbf{w_t}$'E[$\mathbf{r_t}$] is a convex linear combination of the returns which is distributed as an univariate Normal. This implies that Z_a is the z-score value of $F^{-1}(1-\alpha)$, since we're considering $\alpha=0.95$ in our analysis, we obtain that $F^{-1}(0.05)=-1.645$. We can reformulate the VaR constraint as:

$\mathbf{w_t}$'E[$\mathbf{r_t}$] $-1.645(\mathbf{w_t'H_tw_t})^{1/2} >= -\Upsilon$. Where Υ is the risk aversion factor, indexing aversion to extreme realization, we introduced above. In other words this constraint set a lower boundary for the accepted values of our portfolio at 95% confidence, i.e it sets a minimum on the VaR. (7) Introduces a lower boundary to our portfolio Expected returns, coinciding with the expected return of the market portfolio which was computed as a sample moving average of the market return proxied by the S&P500 index. (8) Allows for the option of short selling up to a maximum of 33% per unit, this value was chosen as neighbourhood of regulatory constraints on short sales, which are quite heterogeneous, but often lay within the range of the 30%. Lastly (9) imposes the constraint that the capital has to be fully invested in the portfolio and I can't invest more capital than what I have (unless I resort to short selling, but the overall invested capital still stay the same as the extra capital gets balanced by the counterpart carrying negative weight). It's important to remember that the algorithm gets itinerated and find an optimum set of points for each period of the horizon T (here T=756), it's a conditional one step ahead optimization framework, in which optimum points don't get selected by the algorithm evaluating possible scenarios and interdependencies further than ahead than one period in time. In other words we're running T optimizations, independent from one another, the only layer od intertemporal dependency is driven by our AR(1,1) and MGARCH(1,1) components which are both presents in our objective functions and constraints; however being a one step iteration algorithm, each objective function gets maximized in each period of time without accounting for intertemporal optimization in other words the objective functions and constraint are not bounded by predictions on future events and the maximization doesn't aim to find an optimum for the whole timeframe at each evaluation, but just looks for a "static" optimum point.

We will now present a series of graphs regarding some of the results of the analysis, the graphs will be pertaining the most and least risk averse scenario, to draw graphical remarks in comparison f the two limiting cases.

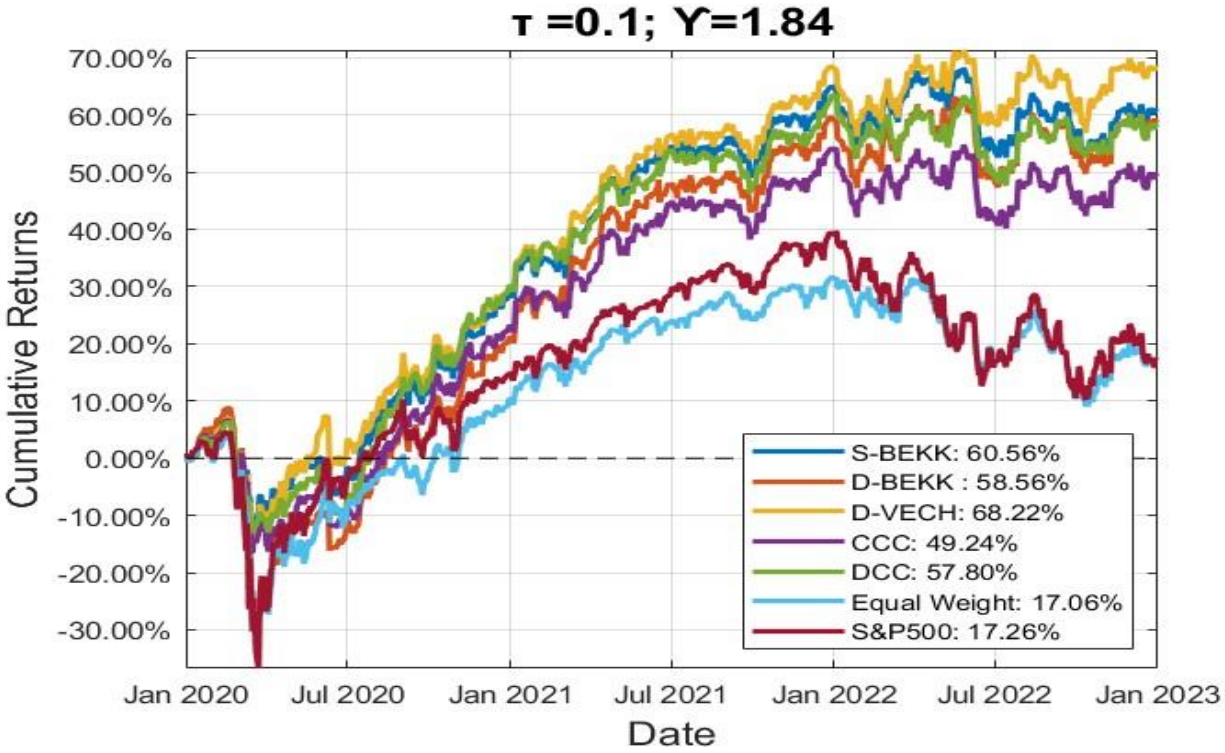


Figure 9.a

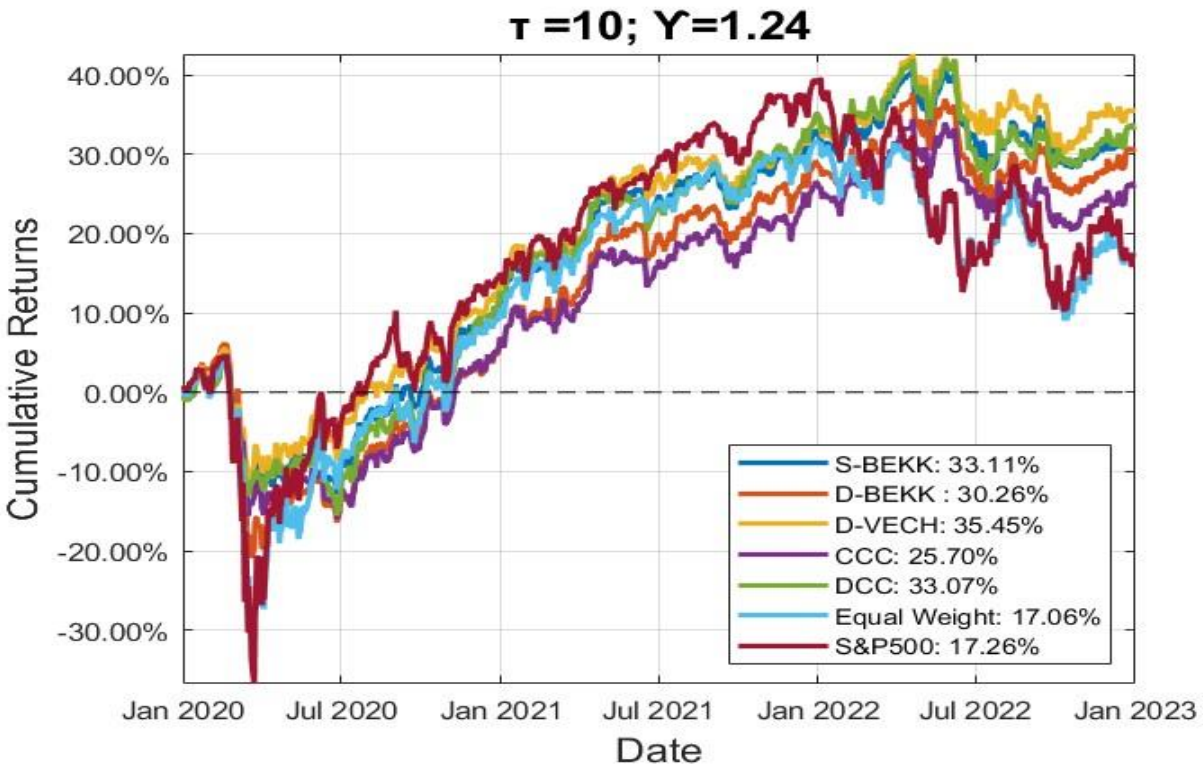


Figure 9.b

In **Figure 9.a**, representing the least risk averse scenario, DVECH and S-BEKK outperformed the rest of the models consistently, netting 68.22% and 60.56% return respectively over the investment horizon. DBEKK and DCC are close behind, at 58.56% and 57.80% respectively, while CCC was the worst performer with a total return of 49.24%. The equally weighted portfolio scored worse than the S&P500 at 17.06%. There is a steep drop in all portfolio's returns in the middle o 2020 in coinciding with the peak of the Covid-19 outbreak which pushed downwards all returns in negative zone, following the decline a bull rend dominated the market until January 2022 when returns started settling and no further significant increase was registered. It was during this window of time that the five portfolios significantly diverged from the market index. **Figure 9.b** instead displays the most risk averse investors, their overall return is largely lower, amounting to roughly 50% of scenario of the other scenario, however DVECH still placed first 35.45% followed by SBEKK at 33.11% and DCC at 33.07%. CCC was still the worst performer. The order of performance of the various strategy did not change however their distance from the market

index was significantly reduced, with CCC scoring merely 7 percentage points above the S&P500.

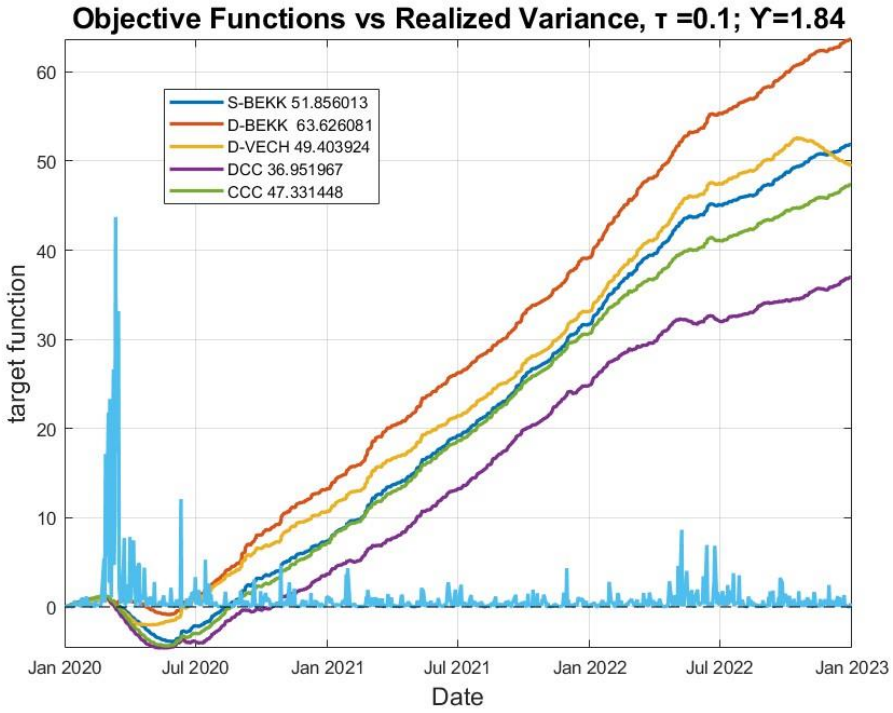


Figure 10.a

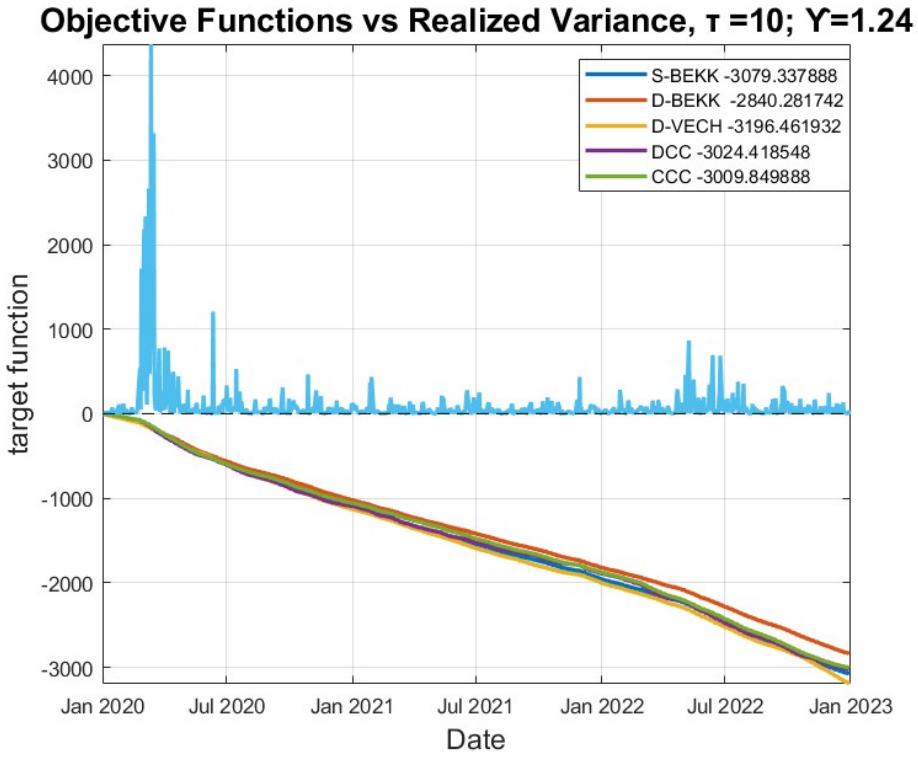


Figure 10.b

Figures 10.a and **10.b**, display the dynamics od the cumulative of the objective in the optimization process versus the Average realized variance of the portfolios, although the maximization algorithm doesn't aim to maximize the function in an intertemporal framework, but looks for the optimum in each interval of time, the cumulative value of the function was shown to embody more meaningful graphical remarks. The Realized variance was multiplied by 100 in figure **10.b** due to the low value of the target function, in order to give a better visual representation of its dynamics. Realized variances in the frameworks follow similar patterns, although in the more risk averse case we see lower relative peaks, as for the values of the objective function we have an explosive decrease in the risk averse framework, where it averages around -3000 with its lowest value displayed by the DVECH at -3196.46. In the most risk neutral scenario instead, the objective function had negative values only for the first quarter of 2020. After the collapse in prices and peak in realized volatility of the second quarter of 2020 due to Covid-19, it had a sharp increase, reaching positive values, and it no longer receded for the whole window, with maximum value at 51.85 for DCC. This does not come as a surprise as more risk averse investors suffer significantly more from the penalty term in the maximization algorithm, while risk neutral ones, giving less importance to dispersion of returns and extreme values, don't get punished as much by the penalty terms as in the objective function that parameter tau is 100 times smaller in **Figure 10.a** respect to **Figure 10.b.** It should be pointed out that DBEKK performed best in both scenarios, with values of 63,62 and -2840.28 respectively.

In the six tables below, ranging from **Table 3.9A** to **3.9F**, the weights associated to the seven ETF under different risk aversion scenarios will be presented. Split for the decision making process in the portfolio optimization algorithm driven by the choice of the MGARCH model used to forecast the variance. The value of the maximization objective function is also displayed; the objective function values represent the level of "success" obtained in the optimization of the portfolio; therefore, higher objective function values indicate that the model was more successful in producing a minimum variance portfolio.

It should be noted that being a multiple period optimization, the weights displayed are average values, computed as the average of the single value assumed in each iteration, counting 756) iterations. It's important to note that these weights reflect also the presence of short selling, which we accounted for in our optimization problem up to 33%, and this weights are thus the resulting compensation along all the timeframe of reference (i.e an high

percentage of short sale on a certain asset will significantly lower the average net position it has in the portfolio)

TABLE 3.9A: Average Weights and Obj. Function for **τ=0.1; ϒ=1.84**

τ=0.1; ϒ=1.84	S-BEKK	D-BEKK	D-VECH	DCC	CCC
QQQ	0.198	0.231	0.189	0.165	0.1876
FXO	0.049	0.059	0.026	0.037	0.0335
EEM	0.006	-0.005	0.024	0.039	0.0046
XLU	0.204	0.200	0.195	0.208	0.2247
DBA	0.267	0.251	0.236	0.226	0.2359
VNQ	0.022	0.038	0.057	0.044	0.0503
XLV	0.254	0.227	0.273	0.266	0.2586
Obj. Fun.	0.069	**0.084**	0.065	**0.049**	0.0626

TABLE 3.9B: Average Weights and Obj. Function for **τ=5; ϒ=1.84**

τ=5; ϒ=1.84	S-BEKK	D-BEKK	D-VECH	DCC	CCC
QQQ	-0.002	0.015	-0.008	0.016	0.018
FXO	0.029	0.055	-0.011	0.025	0.017
EEM	0.054	0.010	0.076	0.074	0.020
XLU	0.205	0.201	0.170	0.190	0.187
DBA	0.468	0.486	0.451	0.391	0.442
VNQ	0.005	0.012	0.047	0.048	0.049
XLV	0.240	0.221	0.273	0.240	0.263
Obj. Fun.	-2.323	**-1.987**	-2.452	-2.420	**-2.265**

TABLE 3.9C: Average Weights and Obj. Function for **τ=10; ϒ=1.84**

τ=10; ϒ=1.84	S-BEKK	D-BEKK	D-VECH	DCC	CCC
QQQ	-0.004	0.013	-0.009	0.014	0.018
FXO	0.029	0.055	-0.010	0.026	0.018
EEM	0.055	0.010	0.077	0.075	0.021
XLU	0.205	0.201	0.170	0.189	0.186
DBA	0.468	0.488	0.451	0.391	0.443
VNQ	0.005	0.013	0.047	0.048	0.049
XLV	0.240	0.221	0.272	0.240	0.262
Obj. Fun.	-4.675	**-4.015**	**-4.942**	-4.892	-4.576

TABLE 3.9D: Average Weights and Obj. Function for **τ=0.1; ϒ=1.24**

τ=0.1; ϒ=1.24	S-BEKK	D-BEKK	D-VECH	DCC	CCC
QQQ	0.113	0.152	0.109	0.113	0.1233
FXO	0.036	0.048	0.002	0.029	0.0202
EEM	0.014	-0.016	0.036	0.042	0.0036
XLU	0.216	0.211	0.192	0.204	0.2179
DBA	0.348	0.343	0.317	0.265	0.2983
VNQ	0.002	0.025	0.042	0.038	0.0441
XLV	0.259	0.234	0.287	0.272	0.275
Obj. Fun.	0.056	**0.073**	0.055	**0.043**	0.0546

TABLE 3.9E: Average Weights and Obj. Function for **τ=5; ϒ=1.24**

τ=5; ϒ=1.24	S-BEKK	D-BEKK	D-VECH	DCC	CCC
QQQ	-0.014	0.011	-0.015	0.013	0.014
FXO	0.025	0.052	-0.016	0.020	0.015
EEM	0.057	0.008	0.080	0.075	0.021
XLU	*0.208*	0.203	0.172	0.189	0.186
DBA	0.472	0.494	0.444	0.382	0.437
VNQ	0.002	0.012	0.045	0.046	0.047
XLV	0.239	0.219	0.273	0.239	0.262
Obj. Fun.	-2.016	**-1.857**	**-2.092**	-1.980	-1.969

TABLE 3.9F: Average Weights and Obj. Function for **τ=10; ϒ=1.24**

τ=10; ϒ=1.24	S-BEKK	D-BEKK	D-VECH	DCC	CCC
QQQ	-0.015	0.008	-0.016	0.012	0.013
FXO	0.024	0.052	-0.016	0.020	0.015
EEM	0.058	0.009	0.081	0.076	0.022
XLU	0.208	0.202	0.172	0.189	0.1859
DBA	0.472	0.495	0.446	0.383	0.4385
VNQ	0.002	0.013	0.046	0.047	0.047
XLV	0.239	0.219	0.273	0.239	0.261
Obj. Fun.	-4.073	**-3.757**	**-4.228**	-4.001	-3.9813

From the tables above we can see that the different GARCH models produced similar results, with only minor variations in the weights assigned to each asset in the portfolios while moving from tau =5 to tau=10, this pattern is present for both gammas and indicates a decreasing marginal impact of risk aversion to dispersion of returns, such patterns will obviously be mirrored in the Descriptive statistics displayed in the next set of tables. E can observe that higher the level of risk aversion (higher tau values), the lower the weights assigned to riskier assets, such as EEM and VNQ. On the other hand, the lower the level of risk aversion (lower tau values), the higher the weights assigned to these assets. Similarly, we can observe that the higher the value of gamma, the higher the weights assigned to riskier assets, and vice versa. This relationship between gamma and the level of risk aversion is expected, as a higher value of gamma means that the investor is less risk-averse, and therefore assigns lower weights to riskier assets. Analysing the objective function values, we can observe that the S-BEKK and D-VECH models produced lower values than the other models beside in the tables with value of tau equal to one, indicating that they performed worse in terms of minimum variance portfolio optimization but still better than other in a more risk neutral environment. This is particularly true for DVECH which scored the worse objective function evaluation in most of the tables. In contrast, DBEKK higher target function levels, evidence that it may be more effective in producing minimum variance portfolios than other models. However, it is worth noting that the differences in objective function values between the models are relatively small, indicating that all models can produce effective minimum variance portfolios. It's also important to highlight that DVECH

and SBEKK scored particularly bad in the most risk averse settings with respect to tau, while in **Table 3.9A** and **3.9A** they placed right after the DBEKK, having higher value in the first table which represents the least risk averse investors space. DCC and CCC have generally been more conservative with weights associated to those models being more equally distributed, CCC diverges from this trend only in the two scenarios associated to a value of 0.1 for tau in which it gives more weight to XLU than any other of the models, but still within a moderate range of 0.22 for gamma=1.84 and 0.21 for gamma=1.24.

In the objective function evaluations we can see a steep decline moving up in the risk aversion scale, such trend seems to be mainly driven by tau rather than gamma. Both **Table 3.9A** and **3.9B** have positive values for all objective functions with DBEKK scoring best in both cases with 0.084 and 0.073 respectively, while DCC hit the lowest at 0.049 and 0.043 respectively. In the second set of tables (**3.9B** and **3.9E**) all values had already turned negative and then doubled in magnitude in the most risk averse frameworks in which the lowest was reached by DVECH in both cases with values of -4.9 and -4.3 for gamma values of 1.84 and 1.24 respectively, while the highest was still reached by DBEKK with values of -4,75 and -4,015.It's also worth noting that in all combinations but the ones with lowest tau all the different models are centred around DBA, with an average weight of around 0.4. This is because it's the ETF with lowest variance and correlation, both in the in-sample analysis than in the forecast estimates as the two are unavoidably linked.

In terms of individual assets, we can observe some interesting trends. We can see that all tables values for the wights of DBA are consistently higher than those of other assets, implying that it may be closer to the optimal ETF within the optimization framework amongst the ones considered in this sample, the model which consistently gave more weight to DBA is the DBEKK which is also the one with highest objective function values, enforcing the idea of the strength of the DBA within this framework regardless of the risk aversion combination. In particular **Tables 3.9A** and **3.9D**, we can see that the weights assigned to DBA are quite lower than the ones assigned to it in the other risk aversion spaces being 0.267 and 0.348 respectively, which would seem counterintuitive as an higher level of risk aversion should lead to a more diversified portfolio. This is however in line with the fact that DBA has lower projected variance and correlation within the GARCH framework, and thus is preferred by the more risk averse investors; in fact, in **Table 3.9A** its value is the lowest amongst all the different weights it was assigned in other risk settings.

Another outlier is VNQ which has been subject to heavy short selling by all models in all settings with weights equal all lower than 0.005, beside in the more risk neutral setting in

which it peaked under DVECH with 0.057. Similarly, FXO reached its maximum of 0.059 in the same environment with DBEKK, however that was also roughly the average it had amongst most models in all the different scenarios, the normalized variance of its allocation was the smallest. These two outliers are assets which have been subject to heavier short selling respect to the others, leading to a convergence of their net position on the portfolio towards zero in this case, it may have been negative. All other ETF have been almost equally split in the different models, with weights centred around 0.2. It's also worth noting that despite VNQ and FXO being close to zero across tables on average, there only negative weights present are QQQ in **Tables 3.9B,3.9C,3.9E,3.9F** in the s-BEKK model and EEM in **Tables 3.9A** and **3,9D** in DBEKK model; which although negative are still close to zero. This implies that our resulting "average portfolio" is indeed a long portfolio under all GARCH models all risk aversion settings considered in this analysis.

In the six tables below (**Table 3.10A** to **3.10F**) a number of descriptive statistics are reported for the daily log-returns of the different portfolios. We have average return, max and min, Total return which is the total return over the three years of the investment horizon, Realized volatility (RV), Value at risk(95%), Sharpe ratio computed as the excessive return of the portfolio over the 3-month Fed Treasury Bill, Information Ratio(IR) computed using the market return (indexed by the performance of the S&P500) as benchmark and Root Mean Squared Error Adjusted(RMSE_Adj) evaluated for the volatility and normalized through realized volatility. RMSE is a measure of the accuracy of a statistical model in predicting values., a low RMSE indicates that the model is better at predicting the future volatility of the asset.Here we use RMSE to deduce the accuracy of the various MGARCH models employed in predicting volatility, as a realized value for daily volatility we take realized volatility, although it's effectiveness is contested in the literature due to its highly noisy nature, it' simple to compute which makes it a handy benchmark; RV was computed as the square root of the square daily realized returns of the various portfolio or as the absolute value of the realized return. RMSE was than normalized by dividing it by the average Realized volatility in order to resolve scale issues and make data easier to compare. In other words the RMSE_Adj is the "unfitness" of our model to volatility in percentage terms.

$$RMSE_{Adj} = \frac{\sqrt{\sum_T \left(\sigma_{t,MGARCH} - |r|_t\right)^2 / T}}{\sqrt{\sum_T r_t^2 / T}}$$

Where the numerator is the standard RMSE for volatility, while the denominator is the Normalization term i.e. the average realized volatility of the forecast sample, while r_t is the return of the generic time-varying portfolio, whose weights are time conditional and depend on the specific GARCH model and risk aversion scenario we're analysing.

TABLE 3.10A: Descriptive statistics of portfolio log-returns for **τ =0.1; ϒ=1.84**.

τ=0.1; ϒ=1.84	Mean	Max	Min	Total Return	RV	VaR	Sharpe Ratio	Info Ratio	RMSE_Adj
S-BEKK	0.080	4.534	-5.772	60.556	1.151	**-1.907**	0.068	0.049	0.344
D-BEKK	0.078	**5.154**	**-14.638**	58.564	**1.384**	-1.900	0.055	0.052	0.420
D-VECH	**0.090**	4.860	-7.111	**68.215**	1.216	-1.838	**0.073**	**0.060**	0.470
DCC	0.077	3.912	-4.837	57.799	1.050	-1.654	0.071	0.04	**0.493**
CCC	0.065	4.948	-6.167	49.243	1.142	-1.702	0.056	0.037	0.466
S&P500	0.023	8.968	-12.765	17.262	1.610	-2.476	0.013		
Equal Weight	0.023	7.893	-12.258	17.062	1.356	**-1.985**	0.015	-0.001	

Table 3.10A represent the scenario with least risk averse investors (low τ and high ϒ), this is the framework in which we expect to see the highest returns overall and highest degree of risk taken. D-VECH model outperformed all other models with a mean return of 0.0902 and a total return of 68.215. The S-BEKK model had the second-highest total return of 60.556, followed by D-BEKK, DCC, and CCC. These results suggest that the D-VECH model is the most profitable, on the other hand, the CCC model had the lowest mean and total returns of 0.065 and 49.231, respectively. With the Sharpe Ratio and D-VECH again performed the best with a Sharpe Ratio of 0.073. The DCC is a close second with a Sharpe Ratio of 0.071. The other models had Sharpe Ratios ranging between 0.0548 to 0.068. These results suggest that D-VECH and DCC models provide better risk-adjusted returns than other models, within the space of risk-taking investors. The D-VECH model has the highest Information Ratio of 0.06, which means it has outperformed the S&P500 in risk-adjusted terms. This

result suggests that the S-BEKK model may be the best choice for investors who want to outperform the market. The other models had Information Ratios ranging from 0.037 to 0.052.The S-BEKK model had the highest VaR(95%) of -1.9074 amongst the actively managed portfolios, which means it is the riskiest portfolio in terms of potential losses. The other models had VaR(95%) ranging from -1.65 to -1.83 with S&P500 hitting-2.476.The S-BEKK model had the lowest RMSE_Adjusted of 0.344, which means that it was the most accurate in forecasting volatility. The other models had RMSE_Adjusted values ranging from 0.42 to 0.493. Comparing the results of the MGARCH models with the S&P500, the S&P500 had a mean return of 0.0228 and a total return of 17.2619, which is lower than all models except the equally weighted portfolio. The S&P500 had the lowest Sharpe Ratio of 0.0131 and the highest VaR(95%) of -2.47, thus despite having low expected return and high variance it's extreme losses were also severe.In conclusion, the D-VECH and S-BEKK models have provided the best risk-adjusted returns and outperformed the S&P500 in terms of risk-adjusted performance. The S -BEKK model was the most accurate in forecasting volatility but had also the highest potential for losses. The CCC model had the lowest performance across all metrics.

TABLE 3.10B: Descriptive statistics for **τ =5; ϒ=1.84**

τ=5; ϒ=1.84	Mean	Max	Min	Total Return	RV	VaR	Sharpe Ratio	Info Ratio	RMSE_Adj
S-BEKK	0.051	4.534	-4.798	38.777	0.854	**-1.387**	**0.058**	0.023	0.345
D-BEKK	0.047	4.622	**-6.929**	35.204	**0.943**	-1.357	0.048	0.021	0.414
D-VECH	**0.053**	**4.861**	-5.124	**39.968**	0.887	-1.288	**0.058**	**0.025**	0.476
DCC	0.043	2.814	-4.195	32.482	0.819	-1.282	0.050	0.016	**0.494**
CCC	0.037	4.948	-6.155	27.907	0.886	-1.340	0.040	0.012	0.475
S&P	0.023	**8.968**	**-12.765**	17.262	1.610	-2.476	0.013		
Equal Weight	0.023	7.893	-12.258	17.062	1.356	-1.985	0.015	-0.001	

Table 3.10B shows the results of portfolio optimization for investors a higher level of aversion to dispersion of returns (tau=5) compared to the models in Table 3.10A. The results demonstrate that the performance of the models varied significantly across different metrics, emphasizing the importance of selecting an appropriate model based on an investor's goals and risk tolerance. The D-VECH model had the highest mean return of 0.0529, indicating its potential to generate higher returns compared to the other models. The S-BEKK model

had the second-highest mean return of 0.0513, while the D-BEKK model had the highest realized volatility of 0.9431, which may result in a higher potential for losses. In terms of volatility parsimony, the DCC model once again demonstrated the highest accuracy, with a realized volatility of 0.82, as expected the realized volatility for all models was lower compared to Table 3.10A caused by the higher variance penalty term in the objective function of our optimization algorithm. Regarding risk-adjusted returns, D-VECH had the highest Sharpe ratio of 0.058 followed by S_BEKK at 0.0575 and by the DCC at 0.050, the CCC model once again had the lowest Sharpe ratio of 0.04, although in table 310A it had won over the DBEKK by 0.001.D-VECH had the highest information ratio (IR) of 0.0251, indicating that it generated the highest excess return over the S&P 500 benchmark compared to the other models. The S-BEKK model had the second-highest IR of 0.0227, followed by the D-BEKK model with an IR of 0.0207. The CCC still stands at the bottom with an IR of 0.0117, while the equally weighted portfolio lost to the market with an IR of -0.001.In terms of total return, the D-VECH model generated the highest return of 39.97 followed closely by the S-BEKK model with a return of 38.78. The D-BEKK model generated a total return of 35.2, while The CCC once again had the lowest total return of 27.90. The maximum return in Table 3.10B was generated by the S&P 500 with a value of 8.96, while the highest maximum return for the models was generated by the equally weighted portfolio with a value of 7.9. DCC had the highest minimum return of -4.19, while S&P 500 hit a value of -12.7652. CCC had the highest maximum return of 4.9, followed by DVECH. As for the VaR SBEKK has the highest one at -1.387, while the DCC displays the lowest by a significant margin with 1.282, closely followed by the DVECH at 1.288. Finally, in terms of accuracy in volatility forecasting, the RMSE adjusted by realized volatility for the S-BEKK model was 0.345, which is the lowest by a significant margin as all the others are centred around 0.45. D-BEKK also fared relatively well with 0.414 in second place.

TABLE 3.10C: Descriptive statistics for **τ =10; ϒ=1.84**

τ=10; ϒ=1.84	Mean	Max	Min	Total Return	RV	VaR	Sharpe Ratio	Info Ratio	RMSE_Adj
S-BEKK	0.051	4.534	-4.858	38.315	0.851	**-1.389**	**0.058**	0.022	**0.344**
D-BEKK	0.047	4.622	**-6.911**	35.223	**0.939**	-1.347	0.048	0.021	0.414
D-VECH	**0.052**	4.868	-5.247	**38.951**	0.886	-1.284	0.056	**0.024**	0.477
DCC	0.042	2.899	-4.195	31.738	0.818	-1.275	0.049	0.015	0.495
CCC	0.035	**4.948**	-6.162	26.522	0.885	-1.327	0.038	0.010	0.474
S&P	0.023	8.968	-12.765	17.262	1.610	-2.476	0.013		
Equal Weight	0.023	7.893	-12.258	17.062	1.356	-1.985	0.015	-0.001	

In **Table 3.10C** we increase investor risk aversion further up to 10 which leads to a decrease of the investor utility obtained from increase in wealth at the cost of risk. In this framework the DVECH still has the highest mean at 0.052, followed by S-BEKK at 0.051, and CCC has the lowest mean at 0.035. The maximum return is once again 4.948 for CCC, while the minimum return is -6.911 for D-BEKK.In terms of total return, D-VECH has the highest total return at 38.951, followed closely by S-BEKK at 38.315. CCC has the lowest total return at 26.522. This indicates that both S-BEKK and D-VECH are able to provide higher returns to investors despite their risk preferences. The realized volatility is lowest for DCC at 0.818 and highest for S&P at 1.610.VaR(95%) is lowest for DCC at -1.275 and highest for SBEKK at -1.389 but even this ceiling amounts roughly to half respect to the one showcased by S&P500 at-2.476.In terms of Sharpe Ratio, S-BEKK has the highest value at 0.058, surpassing fore the first time the D-VECH at 0.056. CCC has the lowest Sharpe Ratio at 0.038. The Information Ratio is highest for D-VECH at 0.024, followed by S-BEKK at 0.022. CCC has the lowest Information Ratio at 0.010 and the equally weighted still loses to the market at -0.001. Finally, the RMSE_Adj is the highest for DCC at 0.495, and the lowest for S-BEKK still at 0.344. This results are quite consistent with the one obtained in **Table 3.10B** which would seem to implicate that beyond a given a threshold an marginal impact of increase of investor aversion to dispersion of the return is negligible.

TABLE 3.10D: Descriptive statistics for **τ =0.1; ϒ=1.24**

τ=0.1; ϒ=1.24	Mean	Max	Min	Total Return	RV	VaR	Sharpe Ratio	Info Ratio	RMSE_Adj
S-BEKK	0.066	3.160	-4.601	50.142	0.922	-1.479	0.070	0.034	**0.340**
D-BEKK	0.064	**4.722**	**-11.435**	48.242	**1.135**	**-1.633**	0.055	0.038	0.418
D-VECH	**0.076**	3.328	-6.342	**57.762**	0.978	-1.494	**0.077**	**0.044**	0.465
DCC	0.066	3.322	-4.097	49.843	0.900	-1.423	0.071	0.033	0.486
CCC	0.052	3.640	-4.527	39.548	0.947	-1.463	0.053	0.024	0.467
S&P	0.023	8.968	-12.765	17.262	1.610	-2.476	0.013		
Equal Weight	0.023	7.893	-12.258	17.062	1.356	-1.985	0.015	-0.001	

In T**able 3.10D** we shifted ϒ to 1.24, indexing higher risk averse investor towards extreme realizations**.** Here the DVECH model had the highest mean return at 0.076. The maximum return was achieved by the S&P 500 at 8.9683, while the D-BEKK model had the highest maximum return among the MGARCH models at 4.722. The minimum return was lowest for the DCC model at -4.09 and highest for the D-BEKK model at -11.435.

In terms of total return, the D-VECH model had the highest among the MGARCH models at 57.7623while the CCC had the lowest total return among the MGARCH models at 39.56.When considering realized volatility, the DBEKK had the highest value at 1.11, while the DCC model had the lowest at 0.9. Looking at VaR, the D-BEKK model had the highest value among the MGARCH models at -1.633, while the DCC model had the lowest at -1.42. When considering Sharpe ratio, the D-VECH model had the highest value among the MGARCH models at 0.0766, while the S-BEKK model was a close second at 0.0702.

The information ratio was highest for the D-VECH as well with a 0.0438 ratio, while the CCC model had the lowest at 0.024. The equally weighted portfolio still displayed negative information ratio at -0.001. Finally, the RMSE_Adj was highest for the DCC model at 0.4860, while the S-BEKK model had the lowest at 0.3404., maintaining its static trend in the sample. The performance of the models varied across the different statistics. While the S-BEKK model had the lowers RMSE_Adj proving its good fit relative to the others, it still had a lower total and mean return compared to the D-VECH, although it displayed lower realized volatility and VaR. The D-BEKK model had the highest maximum return among the MGARCH models, but also had the lowest minimum return and highest VaR and realized

volatility. The D-VECH mean and total return, Sharpe ratio, and information ratio among the MGARCH models. The DCC model had the highest RMSE_Adj value.

TABLE 3.10E: Descriptive statistics for **τ =5; ϒ=1.24**

τ=5; ϒ=1.24	Mean	Max	Min	Total Return	RV.	VaR	Sharpe Ratio	Info Ratio	RMSE_Adj
S-BEKK	0.044	3.077	-4.177	33.131	0.763	-1.263	0.055	0.016	**0.343**
D-BEKK	0.041	**4.622**	**-6.853**	30.908	**0.899**	**-1.349**	0.044	0.015	0.416
D-VECH	**0.047**	3.319	-4.280	**35.870**	0.799	-1.209	**0.057**	**0.019**	0.468
DCC	0.045	2.446	-3.500	33.835	0.762	-1.193	**0.057**	0.016	0.490
CCC	0.034	3.611	-4.421	25.918	0.790	-1.286	0.041	0.009	0.472
S&P	0.023	8.968	-12.765	17.262	1.610	-2.476	0.013		
Equal Weight	0.023	7.893	-12.258	17.062	1.356	-1.985	0.015	-0.001	

In **Table 3.10E** The S-BEKK model has a relatively low maximum return of 3.07, a median low at -4.17, with total return of 33.13, making it a relatively risky investment. The D-BEKK model has the lowest minimum return of -6.8 and VaR of -1.3491, making it a suboptimal portfolio allocation model for extreme value averse investors driven by a lower parameter **ϒ**. However, it displays a relatively low adjusted RMSE value of 0.41, placing right after the CCC as in the other scenarios. The D-VECH model has the highest mean return of 0.0474, highest Sharpe ratio of 0.0572, and highest total return of 35.87 among the MGARCH models, indicating that it has performed well across several metrics. Is VaR of -1.2094 is relatively low as well, and as in all other frameworks it possesses the highest information ratio of 0.0194. The DCC model has the lowest volatility of 0.76 and VaR of -1.2862, his average returns also stays right behind the DVECH in second place while their Sharpe ratio are equal at 0.057, however its adjusted RMSE is quite high at 0.49. The CCC instead fares quite badly under all metrics, with the lowest mean and total return, high volatility and low performance ratio of 0.041 and 0.009 which lace him last in both of them. Overall, the D-VECH model has outperformed the others in this scenario as well in several metrics, including the highest mean return, highest Sharpe ratio, and highest total return among the MGARCH models. However, it is relatively risky with a high VaR and realized volatility

The D-BEKK model is quite unnoticeable in this scenario, its predictive ability according to RMSE stick out respect to the other but it shows high VaR and volatility and no particular performance in other metrics to make up for it.

TABLE 3.10F: Descriptive statistics for **τ =10; ϒ=1.24**

τ=10; ϒ=1.24	Mean	Max	Min	Total Return	RV	VaR	Sharpe Ratio	Info Ratio	RMSE_Adj
S-BEKK	0.044	3.035	-4.177	33.110	0.762	-1.263	0.055	0.016	**0.343**
D-BEKK	0.040	4.622	**-6.567**	30.258	**0.895**	**-1.338**	0.043	0.015	0.415
D-VECH	**0.047**	3.387	-4.280	**35.446**	0.798	-1.204	**0.057**	**0.019**	0.467
DCC	0.044	2.446	-3.498	33.071	0.762	-1.200	0.055	0.016	0.489
CCC	0.034	3.614	-4.520	25.704	0.790	-1.266	0.041	0.009	0.472
S&P	0.023	8.968	-12.765	17.262	1.610	-2.476	0.013	0.000	0.000
Equal Weight	0.023	7.893	-12.258	17.062	1.356	-1.985	0.015	-0.001	0.000

In **Table 3.10F** we are in the scenario with highest risk aversion combination take in consideration in our analysis, but the results are quite similar to the ones obtained under the former combination of gamma and tau. In terms of mean return, the D-VECH model performed the best with a mean return of 0.047, while the CCC model had the worst mean return of 0.034. Looking at the maximum return, the D-BEKK model had the best performance with a maximum return of 4.62. On the other hand, the DCC model had the worst performance in terms of maximum return with a value of 2.44. In terms of the minimum return, the D-BEKK model had the best performance with a minimum return of -6.57. While the DCC performed best, high maximum and low minimum has been a persistent characteristic of DBEKK along all frameworks, which is also explained by the relative high realized volatility and VaR it carried. When looking at the total return, the D-VECH model had the best performance with a value of 35.4464, hardly different for the case with tau=5, implying once again significant decreasing marginal impact of investor aversion to risk on portfolios performance within the models of reference. The CCC had the worst performance with a value of 25.7.Moving on to the risk metrics, the D-BEKK model had the worst performance in terms of risk with a RV. of 0.895 and a VaR of -1.3382, while the DCC

performed on top eith an RV 0.762 and a VaR of -1.2.For the Sharpe ratio and IR, the D-VECH model had the best performance with a ratio of 0.057and 0.019 respectively, while the CCC model had the worst performance with a ratio of 0.041 and 0.009.RMSE_Adj remained pretty much the same also here, with SBEKK on top and DBEKK right after.

Drawing conclusive remarks about the data presented in the tables, the first important thing to notice is the significant decline of the marginal impact of risk aversion on the various portfolio outcomes. While an increase from tau equal 0.1 to tau equal 5 yields significant differences, the same doesn't apply in the next step from tau equal 5 to tau equal 10. Symmetrically, an increase in gamma increases portfolio performance and risk significantly even for the same level of tau.It makes sense that in our framework, with a constraint on the 95%VaR, decreased aversion to extreme outcomes leads to consistently higher returns. Moreover, in the two risk parameters in this sample show oppositive behaviour with IR being the highest for gamma equal 1.84 and tau equal 0.1, while SR being the highest in the least risk aversion setting for gamma in which tau equals 0.1 and gamma equals 1.24. In particular, we observe that the D-VECH model consistently has the highest mean and total returns among the MGARCH models in all six tables. This indicates that the D-VECH model is the best model in terms of returns generation. On the other hand, the CCC model consistently displays the lowest mean and total returns, making it an inferior model for portfolio optimization. The S-BEKK model follows closely behind the D-VECH model. It is worth noting that the models' performance is affected by varying levels of risk aversion. We can see that increasing risk aversion tends to decrease mean and total returns across all models, which is expected since higher risk aversion implies a lower tolerance for investment risk. Considering the models' performance in terms of volatility. We observe that the DCC model consistently displays the lowest realized volatility across all tables, indicating its superiority in forecasting optimization processes with the objective of simply reduce the variance. On the other hand, the D-BEKK model tends to have the highest realized volatility across all tables, making it a relatively riskier investment, while S-BEKK model is in the middle. We observe that the DCC model consistently displays the lowest VaR across all tables, indicating that it is better able to control downside risk and provide downside protection for investors. This is consistent with its performance in realized volatility, however, it sacrifices potential returns for the sake of downside protection The S-BEKK and D-BEKK models, on the other hand, tend to have the highest VaR across most tables, hinting that they are relatively riskier investments with higher potential for losses. D-BEKK model tends to have

mixed results across different scenarios, displaying high maximum returns in some scenarios but also high minimum returns and VaR in others. This indicates that the model's performance is not entirely predictable and can vary based on different factors such as the level of risk aversion and investor preferences. Another important consideration when selecting a model for portfolio optimization is the trade-off between risk and return. We observe that the D-VECH model consistently generates higher returns but is also relatively riskier, with high realized volatility and VaR. On the other hand, the DCC model provides downside protection but also generates lower returns. Therefore, it is important to consider an investor's risk tolerance and preference for downside protection when selecting a model for portfolio optimization. Regarding the Sharpe ratio, we observe that the D-VECH model consistently displays the highest Sharpe ratio across all tables. In particular, the D-VECH model has the highest Sharpe ratio in **Tables 3.10A, 3.10B, 3.10D, 3.10E,** and **3.10F**, indicating that it is the most efficient model in generating risk-adjusted returns for investors. The S-BEKK model also displays a high Sharpe ratio in most tables, following closely behind the DVECH. On the other hand, the CCC model consistently displays the lowest Sharpe ratio across all tables, indicating that it is the least efficient model in generating risk-adjusted returns. As for the Information ratio across all tables, we observe that the D-VECH consistently displays the highest information ratio, indicating that it generates the most excess returns per unit of active risk. The S-BEKK also displays a relatively high information ratio in most while, the CCC consistently displays the lowest information ratio across all tables, indicating that it generates the least amount of excess returns relative to a benchmark index. Finally, we move to the RMSE_Adj, which measures the difference between the predicted and actual volatility of a portfolio, normalized to the realized volatility of the portfolio model. Across all tables, we observe that the S-BEKK model consistently displays the lowest RMSE_Adj among the MGARCH models, indicating that it is the most accurate predictor of volatility. The D-BEKK model also displays a relatively low RMSE_Adj in most tables,. On the other hand, the CCC model consistently displays the highest RMSE_Adj among the MGARCH models, being the least accurate predictor of volatility under this metrics. It’s also worth noting that RMSE values are quite static, with minimal variation for different combination of risk aversion-VECH model consistently displays the highest Sharpe and information ratios, the DCC model consistently displays the lowest VaR, and the S-BEKK model consistently displays the lowest RMSE_Adj. These findings further reinforce the superiority of the D-VECH model in generating risk-adjusted returns, the DCC model in controlling downside risk, and the S-BEKK model in accurately predicting volatility. In

conclusion, the six tables provide a comprehensive comparison of the performance of different MGARCH models under varying levels of risk aversion. We can see that the D-VECH model consistently outperforms the other models in terms of generating returns and providing higher risk-adjusted returns, while the S-BEKK model is more accurate in predicting volatility. The CCC model consistently displays inferior performance across different metrics, making it a less attractive investment. The SBEKK performance is inferior to the DVECH in risk-adjusted metrics as well, but as a standalone it displays significantly lower VAR, RV and RMSE.

4. CONCLUSIONS

In this study we presented and in-depth review of some of the most popular GARVH models in multivariate time series, and an introduction to portfolio optimization in the Markowitz framework. We then proceeded to apply these models to a real dataset. In particular we estimated and forecasted the variance covariance matrices and annexed parameters for seven ETFs through five different MGARCH models specifications under the assumption of gaussianity: SBEKK, DBEKK, DVECH CCC and DCC. We compared and discussed the results drawing conclusions,, and applied the results in to forecast the covariance matrices later used in two different application of portfolio optimization within the Markowitz framework. We introduced and analysed limitations ensuing in estimation of the parameter matrices for large datasets in the DVECH and the biasness of the DCC forecasts. In both applications to portfolio optimization DCC scored the highest Adjsusted RMSE and lowest absolute value for the optimizations objective functions, this is partially caused by the mispefication of the model within the dataset in which the innovations' conditional correlations matrices Qt don't converge to the correlation matrices of the residuals R_t asymptotically, despite being their true population values. In portfolio optimization, the application of more complex models like Black-Litterman under a more fitting distribution such as the Generalized Hyperbolic Distribution (GHD) will be addressed in further research.

In this study, two optimization models were applied to the same dataset to assess their performance, and the results obtained were found to be quite heterogeneous.

The first model aimed to minimize risk, which was proxied by portfolio variance, while the second model aimed to maximize return penalized by variance and under a Value at Risk (VaR) restriction. The investors for both models were different, with the first model designed

to appeal to risk-averse investors, and the second model catering to those who were willing to take on more risk to achieve higher returns, under different degrees of risk aversion scenarios. Despite being linked to a minimum return constrained against the S&P500, the two models had different objectives. The first model aimed to minimize risk by minimizing the portfolio's variance, while the second model aimed to maximize returns while keeping risk under control. Consequently, the models used different approaches to allocate assets and achieve their objectives. The results obtained from the two optimization models were quite different. The cumulative return of the second model was higher in the most risk-neutral scenario and lower in the most risk-averse one. In the second model, DVECH was the outperformer in both scenarios, while SBWKK was the outperformer in the first model. However, in the VaR portfolio DVECH wasn't consistently the over-performer in all metrics, and its objective function evaluation was consistently lower than the SBEKK's in all risk scenarios.The standard variance approach favoured the SBEKK, which had the highest Sharpe ratio and information ratio and the lowest realized volatility and adjusted RMSE. However, it also displayed the highest VaR, which was not fully captured in the evaluation of its objective function and choice of weights. This means that the model with the highest Sharpe ratio and information ratio maynot necessarily be the best-performing model, and investors need to consider other evaluation metrics to make an informed decision.The second methodology emphasized extreme realizations, penalizing the SBEKK, which displayed high extreme values under all risk-aversion scenarios. The second model's results were more heterogeneous and open to interpretation, with DVECH dominating in all risk-adjusted returns. However, it was not consistently the over-performer in all metrics. Its objective function evaluation was consistently lower than the SBEKK's and DBEKK’s in all risk scenarios, but it outperformed the SBEKK in other metrics, such as realized volatility and Sharpe ratio. The optimal choice in the second model was the DBEKK, which displayed the highest average objective function in all six scenarios. The weight portfolio allocation was different for both models, both centred around DBA, followed by XLU and XLV. The VaR models gave significant importance to QQQ as well in the two scenarios with tau=0.1, while the others coincided with a much closer representation to the minimum variance.In risk-adjusted performance metrics, the second model consistently scored above the first, which netted results inferior even to the most risk-averse scenario in Sharpe Ratio. Realized volatilities level of scenarios with tau=0.1 were worse than in the first model, while all the others showcased superior performance. The only metrics in which minimum variance portfolios overcame the competing ones was adjusted RMSE. However, having computed it

with respect to realized volatility, it was an extremely noisy measure. It is important to consider this while evaluating the results of these models. Overall, the study results indicate that the second methodology was significantly superior in performance, even considering different scenarios and spanning a wide range of evaluation criteria. The results also highlight the importance of understanding the objectives and constraints of an optimization model, as well as the underlying assumptions, which may influence the performance of the model. It is, therefore, critical for investors to be aware of these factors and use them to inform their decision-making process. However, regardless of their relative performance both models consistently overperformed the market indexed by the S&P500, under all market conditions proving the profitability of active management. The study also highlights that different investors have different risk-reward trade-offs and hence require different optimization models. Therefore, investors need to identify their investment goals, risk tolerance, and other investment constraints before selecting an appropriate optimization model.

In conclusion, the study shows that the selection of an appropriate optimization model is crucial for investors, and the models need to be evaluated based on multiple metrics, including risk-adjusted returns, realized volatility, Sharpe ratio, and information ratio. The evaluation of these metrics allows investors to compare the performance of different models and make informed investment decisions. Moreover, the study reveals that the weight portfolio allocation was different for both models. thus, the allocation of portfolio weights depended on the investors' risk aversion levels and investment objectives.The study results also indicate that the choice of optimization model can significantly impact the portfolio's performance. Investors need to carefully evaluate the available models and select the one that best aligns with their investment goals and risk tolerance levels. Finally, the results show that there is no one-size-fits-all solution when it comes to portfolio optimization. The choice of optimization model needs to be tailored to the investors' risk-reward trade-off and investment objectives. This highlights the importance of understanding the investors' needs and designing appropriate investment strategies that align with their objectives. The study's results indicate that the second methodology was significantly superior in performance, but the choice of optimization model will depend on the investors' risk aversion levels and investment objectives.

REFERENCES

1. 1.Anderson, T. W. and Olkin, I. (1985). “Maximum-Likelihood Estimation of the Parameters of a Multivariate Normal Distribution”
2. Bauwens, L., Laurent, S., and Rombouts, J. V. K. (2006). “Multivariate GARCH Models: A Survey”
3. Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model. The review of economics and statistics, 498-505
4. Bollerslev, T. (1988). “A Capital Asset Pricing Model with Time Varying Covariances”
5. Boyd, S. et al. (2005). “Robust Portfolio Optimization”
6. Bunnang, T. (2015). “Hedging Petroleum Futures with Multivariate GARCH Models”
7. Cavanaugh, J. E. and Neath, A. A. (2000). “The Akaike Information Criterion: Background, Derivation, Properties, Application, Interpretation, and Refinements”
8. DeMiguel, V., Garlappi, L., and Uppal, R. (2009). “Optimal versus Naive Diversification: How Inefficient is the 1/n Portfolio Strategy?”
9. De Santis, G. and Gerard, B. (1997). “International Asset Pricing and Portfolio Diversification with Time-Varying Risk”
10. Ding, Z. and Engle, R. F. (1996). “Large Scale Conditional Covariance Matrix Modeling, Estimation and Testing”
11. Ding, Z., Granger, C. W. J., & Engle, R. F. (1993). A long memory property of stock market returns and a new model. Journal of empirical finance, 1(1), 83-106.
12. Engle, R. F. and Kroner, K. F. (1995). “Multivariate Simultaneous Generalized Arch”
13. Engle, R. F. (1999). “Dynamic Conditional Correlation – A Simple Class of Multivariate GARCH Model”
14. Fan, J. et al. (2008). “A Large-Dimensional Factor Model for High-Dimensional Time Series”
15. Francq, C. and Zakoian, J.-M. (2004). “Maximum Likelihood Estimation of Pure GARCH and ARMA-GARCH Processes”
16. Hafner, K. M. (2009). “A Generalized Dynamic Conditional Correlation Model: Simulation and Application to Many Assets”
17. Hafsa, H. (2015). “CVaR in Portfolio Optimization: An Essay on the French Market”
18. Jagannathan, R. and Ma, T. (2003). “Estimating the Global Minimum Variance Portfolio”
19. Krokhmal, P., Palmquist, J., and Uryasev, S. (2001). “Portfolio Optimization with Conditional Value-at-Risk Objective and Constraints”
20. Ling, S., & McAleer, M. (2003). Asymptotic theory for a vector ARMA-GARCH model. Econometric theory, 19(2), 278-308.
21. Lam, C. (2010). “Estimation for Latent Factor Models for High-Dimensional Time Series”
22. Markowitz, H. (1952). “Portfolio Selection”
23. Mattheakis, M. and Protopapas, P. (2004). “Model Selection & Information Criteria: Akaike Information Criterion”
24. Michaud, R. O. and Michaud, R. O. (2008). “Efficient Asset Management: A Practical Guide to Stock Portfolio Optimization and Asset Allocation”

25. Ofosu-Hene, E. D. and Siaw, R. O. (2017). “Investment Portfolio Optimization with GARCH Models. Elk Asia Pacific Journal of Finance and Risk Management”
26. Pelizzari, C. (2015). “Portfolio Optimization with Value-at-Risk Constraints: A Stochastic Dominance Approach”
27. Silvennoinen, A., & Terasvirta, T. (2009). Multivariate GARCH models. In Handbook of financial time series (pp. 201-229). Springer, Berlin, Heidelberg.
28. Todor Stoilov1, Krasimira Stoilova1, Miroslav Vladimirov(2021).” The Probabilistic Risk Me asure VaR as Constraint I n Portfolio Optimization Problem.”
29. Uppal, R. (2009). “Optimal versus Naive Diversification: How Inefficient is the 1/n Portfolio Strategy?”
30. Wenqing Hu, W. (2016). “Value at Risk (VaR) and its Calculations: An Overview”
31. Zempleni, A. (2012). “Weighted Bootstrap in GARCH Models”
32. Kouwenberg, R. and Vorst, A. C. F. (2004). “Portfolio Optimization with Conditional Value-at-Risk Objective and Constraints"
33. Hu, H. (2016). “Value-at-Risk Analysis in Risk Measurement and Formation of Optimal Portfolio in Banking Share”
34. Hu, W. (2016). “Value at Risk (VaR) and its Calculations: An Overview”