Models for risk and forecasting

Project Work

Lorenzo Battista [1], Gianlorenzo Gattinara [2], Manfredi Tirri [3]

Student ID number: 1809959 [1], 1979008 [2], 1957973 [3]

Abstract

The aim of this study is to investigate a pool of assets: FTSE, IBEX 35, CAC 40; evaluate the set of models which best estimates the covariance matrix, build the global minimum variance portfolio (by using the best model) and forecast the Value at Risk and Expected Shortfall of the portfolio. The models have been compared using the MCS procedure which make use of a loss function. For our purposes we used two robust loss function, one symmetric and the other one asymmetric. In order to forecast the *VaR* and *ES* of the portfolio both the parametric, non-parametric and semi-parametric approaches have been employed. All the models have been evaluated in order to check their prediction ability, by using 2 methods: the backtest procedure and furthermore the MCS procedure again, but this time by employing a quantile loss function and the Fissler and Ziegel loss function, which allow to jointly evaluate VaR and ES forecast.

1. Introduction and Literature review

Everything in the financial markets is interrelated, and it is now widely recognized that assets and market volatilities move together over time. Using a multivariate modeling framework to recognize this trait leads to more relevant empirical models than using distinct univariate models. These type of models were devoloped in the late 80's, beginning of the 90's, but suffered a period of tranquillity due to the curse of dimensionality and computational requirements.
In this paper, we will review different class of models: non parametric multivariate models, multivariate generalization of univariate GARCH models of Bollerslev, linear combination of GARCH models and dynamic conditional correlation models. The most important aspect of MGARCH modeling is to ensure the positivity of the covariance matrix by providing a realistic but parsimonious specification. Flexibility and parsimony are at odds. Engle and Kroner (1995) proposed BEKK models which are flexible, but they have too many parameters for multiple time series. However, there are two variants: the Diagonal BEKK and the Scalar BEKK, which are significantly more sparse, but quite restrictive for cross-dynamics. Factor models allow to overcome the computational difficulty encountered by BEKK models, while ensuring the positive definiteness of the conditional covariance matrix. More specifically, Carol (2011) proposed an Orthogonal specification of the factor models that well performs in a higly correlated system as the financial market. DCC models allow for different persistence between variances and correlations, but impose common persistence in the latter (although this may be relaxed). However, they make it possible to handle a larger number of series than before. Considering the DCC model to forecast the volatility and correlations among the assets of Engle (2002), in this paper we employ it's corrected version, the cDCC model by Aielli (2013). Furthermore we employed the DCC-MIDAS model, adding a long-run correlation component by using a mixed data sampling approach Ghysels et al. (2007), Colacito et al. (2011). In the literature, several models have been proposed to include macroeconomic variables as additional volatility determinants because of their influence, through expectations and announcements, on volatility, among others: Conrad et al. (2014), Schwert (1989) and Officer (1973).
All the volatility and co-volatility estimates have been compared using the MCS procedure by Hansen et. al (2011), following we decided to pick one model in order to build the "global minimum variance portfolio" by calculating the optimal weights according to the modern portfolio theory of Markowitz. From this portfolio, Value at Risk and Expected Shortfall estimates were calculated by going out of sample (20%) and subsequently backtested and compared with the MCS procedure.
The rest of this paper is organized as follows: the methodology used in further details (section 2), the empirical application (section 3) and following the conclusion (section 4).

2. Methodology

2.1 Stylized facts

Much of the literature on volatility modelling relies on stationary time series such as returns, rather than on asset price, which empirical economics has shown are non-stationary. Moreover, returns give us information on price changes and, thus, on profits and losses from a given financial instrument. Studying returns, therefore, also has a greater informational advantage than studying asset price, which may give us an idea about the price level, but tell us nothing about the profitability performances of financial products. From a mathematical point of view, daily returns are nothing more than the logarithmic variation of closing prices:

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

Statistically, returns correspond to the first-order difference that eliminates non-stationarity in the mean, and later we will see this characteristic of the mean of returns equal to zero and constant over time. In recent decades, numerous researches conducted on large samples of financial time series have led to the discovery of certain statistical properties of stock and commodity prices, and, more specifically, of the returns of these financial instruments, regardless of the reference market or the nature of the financial product. Cont (2001) refers to *stylized facts* as a set of "qualitative statistical properties of asset returns". The first stylized fact is precisely the stationarity of returns. As mentioned earlier, unlike asset price times series, returns are stationary. Stationarity implies that the statistical properties of time series, or rather the process that generates them, are stable over time. Thus, the mean of returns will be constant over time and assumed to be zero. In order to test stationarity, the *Augmented-Dickey-Fuller test* is one of the most widely used in empirical works and it is based on the following regression:

$$\Delta y_t = \eta + \rho y_{t-1} + \sum_{j=1}^{p} \delta_j \Delta y_{t-j} + \varepsilon_t \tag{2}$$

where the dependent variable Δy_t represents the first difference of the series at time t, η is simply the intercept, ρ is the coefficient of the past observation y_{t-1}, ε_t is the error term, and finally the regression contains p first difference of lagged observations of y and δ_j represent their coefficients. The null hypothesis of this test assumes that $\rho = 1$ and, thus, the presence of unit roots, a characteristic of time series that makes them non-stationary, while under the alternative hypothesis the time series has no unit roots, and therefore it is stationary.

Another characteristic of the returns is that they are homoscedastic, i.e., the fact that the conditional variance is constant over time. The *ARCH-LM test* can be used to check whether a series is homoschedastic or not. Also this test is based on a regression:

$$r_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + \cdots + \alpha_p r_{t-q}^2 \tag{3}$$

The null hypothesis establishes that $\alpha_1 = \cdots = \alpha_q = 0$, that is, there is no ARCH-effect and therefore homoscedasticity. The test statistic is TR^2, where T represents the length of the time series and R^2 is obtained from the regression. Under the null hypothesis this test statistic is asymptotically χ_q^2 distributed.

Moreover, the unconditional distribution of returns is far from Normality, usually it is leptokurtic and asymmetric. A distribution is leptokurtic (kurtosis is greater than 3) when its tails are heavier than the Gaussian distribution, which implies that extreme returns may occur with a higher probability than in the Normal case. With regard to asymmetry, typically the skewness of the returns is $SK \neq 0$, which indicate that the two tails are not equal, but one is heavier than the other. The *Jarque-Bera test* make use of both the kurtosis and skewness to verify whether the distribution is Normally distributed. Under the null hypothesis the distribution is Normal, and its test statistic can be expressed as follows:

$$JB = \frac{T}{6}\left(S^2 + \frac{(K-3)^2}{4}\right) \overset{H_0}{\sim} \chi_2^2 \tag{4}$$

where T is the length of the series, and K and SK are respectively the sample kurtosis and sample skewness.

The fourth stylized fact characterising returns is the uncorrelation. It is well known that returns do not exhibit any significant autocorrelation, implying that returns at time t do not linearly depend on past returns, so they follow a *White Noise* process. There are numerous correlation tests that assess whether a series is *WN*. The *Box-Pierce test* belongs to this class of tests and under the null hypothesis the first h autocorrelations are jointly equal to zero and therefore the returns are uncorrelated.

Formally, the test statistic is formulated in the following way:

$$BP_h = T \sum_{k=1}^{h} \hat{\rho}^2(k) \overset{H_0}{\sim} \chi^2_h \tag{5}$$

where T is always the sample size and $\hat{\rho}$ is the autocorrelation function for the first h returns. The *Ljung-Box test* is a better correlation test than the *Box-Pierce test* because it has a faster convergence to the asymptotic distribution. The null hypothesis is the same of the previous test, but *LB test* differs in terms of the test statistic:

$$LB_h = T(T+2) \sum_{k=1}^{h} \left(\frac{\hat{\rho}(k)^2}{T-k} \right) \overset{H_0}{\sim} \chi^2_h \tag{6}$$

As empirical works have shown, for financial returns, this tests should not reject the null hypothesis. The absence of autocorrelations in returns does not imply independence. The latter indicates that any non-linear function of returns will also have no autocorrelation. In fact, simple non-linear functions of returns, such as squared returns, exhibit significant positive autocorrelation or persistence. Returns, therefore, are not *Random Walk*, and this persistence is a quantitative signature of the *volatility clustering* phenomenon: large price variations are more likely to be followed by large price variations. This persistence implies some degree of predictability for the conditional variance of returns.

2.2 Models Specification

Models for volatility try to replicate the stylized facts embedded in returns, therefore, a good model should be able to capture the features of financial data. In the multivariate framework, models take into account both the volatility and the *co-volatility*, so we consider the variances of each financial position and its relationships with the other assets. The object of interest, therefore, will no longer be the conditional variance, but the *Conditional Covariance Matrix*, which, from now on, we will denote by H_t. The latter is a square matrix presenting the conditional variances of each asset on the main diagonal and the conditional covariances off of it. By construction H_t is symmetric, and must respect certain characteristics, mainly, being positive definite, covariance stationary and having all elements of the main diagonal, as variances, positive.
Next, we propose some multivariate specifications, widely known in the literature, which ensure these characteristics of H_t.

2.2.1 Moving Covariance Model

The *Moving Covariance Model* belongs to the family of non-parametric multivariate models. They are so called because no assumption about the distribution of returns is necessary. The *MC* model only uses what has already been observed, as can be seen from the formula:

$$H_t = \frac{1}{V} \sum_{j=1}^{V} \mathbf{r}_{t-j}\mathbf{r}'_{t-j}, \quad \text{with} \quad V > k \tag{7}$$

The *MC* model performs a simple average of the cross-products of past returns to obtain the conditional covariance matrix. The parameter V represents the length (kept constant) of the moving window in which the average cross-product of the returns is calculated, and it is constrained to be greater than the number of assets k. The first V matrices of H_t will be replaced by the sample covariance. The main drawbacks of this model is that it assigns equal weight to past observations and H_t is strongly influenced by extreme returns.

2.2.2 Exponentially Weighting Moving Average Model

The *Exponentially Weighted Moving Average model*, unlike *MC* model, assigns smaller weights as observations move away in time.

$$H_t = \lambda H_{t-1} + (1-\lambda)\mathbf{r}_{t-1}\mathbf{r}'_{t-1} \tag{8}$$

H_t depends on λ, the conditional covariance matrix and the cross-product of the returns of the day before. The λ parameter usually takes value between 0.9 and 0.99. The decreasing weights are the result of the regressive nature of the formula. Hence, regressing the initial formula leads to this expression:

$$(1-\lambda) \sum_{i=1}^{V} \lambda^V \mathbf{r}_{t-i}\mathbf{r}'_{t-i} + \lambda^V H_{t-V} \tag{9}$$

Since λ takes value between 0 than 1, λ^V will decrease as one moves away in time, giving less and less importance to past observations.

2.2.3 BEKK models

This family of models was proposed by Engle and Kroner (1995), and it belong to the "Generalization of univariate GARCH models" class. The formulation for the first order model is the following:

$$H_t = CC' + A(r_{t-1}r'_{t-1})A' + GH_{t-1}G' \tag{10}$$

where: A, G are k x k matrices and C is lower triangular, so its transpose would be upper triangular. The fact that each matrix is multiplied by its transposed counterpart ensures that positive definiteness of H_t, which is one of the main advantages of this model. The parameters are estimated by solving (numerically) the following optimum problem: $\text{argmax}_{\Theta} L_T(\Theta)$, where $L_T(\Theta)$ under the Multivariate Normal assumption is :

$$L_T(\Theta) = -\frac{1}{2}\sum_{t=1}^{T} log|H_t| - \frac{1}{2}\sum_{t=1}^{T} \mathbf{r_t'} H_t \mathbf{r_t} \tag{11}$$

If this assumption of normality is too restrictive, it is possible to use also other multivariate distribution, however we will not cover them in this paper. It should be noted that the estimation procedure is quite complicated, since in the previous equation, H_t is inverted for every time period.
It is possibile to reduce the number of parameters by introducing some restrictions in the model, which lead to 2 variants: DBEKK and SBEKK. In the diagonal BEKK (DBEKK), the matrices A and G are constrained to be diagonal; while in the scalar BEKK, A and G are replaced by 2 scalars: α and β, which are then multiplied for an identity matrix.
The difficulty when estimating a BEKK model is the high number of unknown parameters, even after imposing several restrictions.

2.2.4 Orthogonal GARCH

The "*curse of dimensionality*" encountered with some multivariate GARCH models is overcome by *factor models*, a class of multivariate models that exploits the linear combination of several univariate models. In this section, we show how an orthogonal factor model can be used to simplify the estimation process of multivariate models. The model proposed here is the *Orthogonal GARCH* (*OGARCH*), developed by Carol (2011), which is based on orthogonal factorisation by means of *Principal Component Analysis* (*PCA*) of assets. The basic idea of OGARCH is that, in highly correlated systems, such as financial markets, it is possible to represent a large part of the variability of the system through a few factors that influence, to a greater or lesser extent, all the variables within it, leaving out a small part of the system variability that is referred to as "*noise*", and models that do not know how to filter out this noise may lack robustness. The advantages of OGARCH over multivariate GARCH models are, therefore, manifold. The computational difficulty of the model estimation process is reduced, since only a few univariate GARCH models are needed to construct the conditional covariance matrix; no restriction on the dimensionality of the original data is necessary, and finally OGARCH, unlike MGARCH, provides the possibility of omitting this noise from the data.
As mentioned earlier, OGARCH relies on PCA to extract the factors driving all the variables in the system. From a set of k initial variables, PCA will give up to k orthogonal variables, our factors, which are called the *principal components*. PCA is sensitive to re-scaling of the data, and so it is common practice to standardise the data before PCA. Let denote X the standardised matrix of original data. The input of PCA is the correlation matrix $X'X$ between the variables of X. The PCA decomposes, by means of the *eigendecomposition*, $X'X$ into three matrices:

$$X'X = WLW' \tag{12}$$

where W is the orthogonal matrix of eigenvectors of $X'X$ and L is the diagonal matrix of eigenvalues of $X'X$. The eigenvalues in L, are arranged in a decreasing order and the columns of W, the eigenvectors, are arranged according to the order of the eigenvalues. The m^{th} principal component of the system will be given by:

$$p_m = w_{1m}x_1 + w_{2m}x_2 + \dots + w_{km}x_k \tag{13}$$

where x_i represents the columns of X and w_{im} the columns of W. Therefore, each principal component is a linear combination of the columns of X, where the weights came from the eigenvectors of $X'X$.
Principal components are uncorrelated with each other and are arranged according to the share of total variation they explain. Thus, the first principal component will explain most of the total variation of the system, the second principal component most of the remaining variability, and so on. If each principal component is placed as the columns of a Txk matrix P we have:

$$P = XW \tag{14}$$

Since W is an orthogonal matrix, $W' = W^{-1}$, $P'P = L$. Moreover, given that L is a diagonal matrix, the columns of P are uncorrelated. Since $W' = W^{-1}$, 14 is equivalent to $X = PW'$, that is:

$$X_i = w_{i1}P_1 + w_{i2}P_2 + \ldots\ldots + w_{ik}P_k \tag{15}$$

where X_i and P_i refer to the $i-th$ columns of X and P. Hence, each variable in X is an orthogonal linear combination of the principal components. In order to construct the conditional covariance matrix, one must, first of all, re-scale the principal components, again performing standardization. Therefore, we re-write the principal components as follows:

$$y_i = \mu_i + \omega^*_{i1}p_1 + \omega^*_{i2}p_2 + \ldots\ldots + \omega^*_{im}P_m + \varepsilon_i \tag{16}$$

where y_i is the original non-standardized variable, μ_i is the mean of y_i, $\omega^*_{ij} = w_{ij}\sigma_i$, σ_i is the standard deviation of y_i and, finally, ε_i is the error term in case we consider $m < k$ principal components. Once the principal components have been standardised, the conditional covariance matrix is obtained through this multiplication:

$$H_t = AD_tA' + H_\varepsilon \tag{17}$$

where A is the matrix of standardized eigenvectors of the initial W matrix, D_t is the diagonal matrix of variances of principal components and H_ε is the covariance matrix of the error terms. However, the only drawback of OGARCH is that the positive definiteness of H_t is not ensured when considering $m < k$ principal components, because H_ε could only be postive semi-deifinite. The only way to ensure the positive definiteness of H_t, therefore, is to use k principal components and, thus, $H_\varepsilon = 0$. The estimation of univariate models takes place on each standardised principal component and serves to estimate the conditional variances of each of them, that will enter the D_t matrix. Estimating k univariate GARCH models is certainly easier and faster than directly estimating complex multivariate GARCH models.

2.2.5 DCC

As the last class of models that we are going to consider we focus on the family of MGARCH based on correlation models.
These models are based on the decomposition of the covariance matrix, where, in order to obtain the conditional covariance matrix, we specify separately the conditional variances and the conditional correlations. In fact, we split the estimation into 2 different steps (in order to reduce the time of estimation): we first specify a model for the univariate part, meaning for the individual conditional variances, and then a model for the correlation part.
We decompose the covariance matrix as:

$$H_t = D_tR_tD_t \tag{18}$$

Where H_t ($H_t = E_{t-1}[r_tr_t^{'}]$) is the conditional covariance matrix, $D_t = diag(\sqrt{h_{1,t}}, \sqrt{h_{2,t}}, ..., \sqrt{h_{k,t}})$ (with k number of assets) is the diagonal matrix of the conditional standard deviations and R_t is the conditional correlation matrix. By construction, R_t is the conditional covariance matrix of the vector of the standardized returns, that is $E_{t-1}[\eta_t\eta_t^{'}] = R_t$, where $\eta_t = r_t/\widehat{D}_t$.
As said, we first specify D_t, where the diagonal elements are modeled as univariate GARCH models, and then we model R_t through the standardized returns obtained from the previous step, $\eta_t = \widehat{D}_t^{-1}r_t$.
The Dynamic Conditional Correlation (*DCC*) model presented by Engle (2002), with respect to the first Constant Conditional Correlation (*CCC*) model of Bollerslev (1990), assumes that the conditional correlation matrix (R_t) is time-varying, capturing the dynamics in the correlation.
For H_t to be positive definite, all the conditional variances $h_{i,t}$ with $i = 1,...,k$ are well-defined (meaning that the univariate *GARCH* models are correctly specified) and the conditional correlation matrix R_t is positive definite at each point in time. The specification for the matrix R_t proposed by Engle (2002) is constructed as follows:

$$\begin{gathered} R_t = (Q_t^*)^{-1}Q_t(Q_t^*)^{-1} \\ Q_t = (1-\alpha-\beta)R + \alpha\eta_{t-1}\eta_{t-1}^{'} + \beta Q_{t-1} \\ Q_t^* = diag(\{q_{ii,t}^{1/2}\}_{i=1}^k) \end{gathered} \tag{19}$$

In this one we have a sort of "sandwich" representation of the correlation matrix R_t, composed of Q_t, representing a sort of *MGARCH* type dynamics, and by the diagonal matrix Q_t^* given by the square root of the *ii* entries of Q_t. This latter diagonal matrix is introduced to endure that R_t is a correlation matrix, Billio (2006). For R_t to be positive definite it suffices that Q_t is positive definite, which is the case if $\alpha \geq 0$, $\beta \geq 0$, and $\alpha + \beta < 1$, and R is positive definite, Aielli (2013). Having only to estimate α and β for the correlation part, the *DCC* model solves the curse of dimensionality.
Furthermore, Engle proposed a generalized version of the DCC, where Q_t is formulated as:

$$Q_t = (I_k - A - B) \odot R + A \odot \eta_{t-1}\eta_{t-1}^{'} + B \odot Q_{t-1} \tag{20}$$

Where $\odot$ is the Hadamard product, I_k is an Identity matrix of dimension k and A and B are (n×n) matrices of coefficients. For Q_t to be positive definite, it suffices that one of matrices between A, B or $(I_k - A - B)$ is positive definite, while the others are positive semi-definite, Engle (2002). However, in this case, due to the number of estimated parameters, the obstacle of the curse of dimensionality comes back, Otranto (2010).
The *DCC* correlation driving process, Q_t, is often treated as a linear *MGARCH* process. As a result, standard interpretation of α, β and R might lead to potentially incorrect findings. As a matter of fact, applying a standard result on linear *MGARCH* processes, R is replaced by the estimated standardized returns. Aielli (2013) pointed out that the *DCC* model was less tractable than expected, concluding that the Engle's model was a biased estimator. After recognizing that Q_t is not a linear *MGARCH*, finding a tractable representation of Q_t proves to be difficult. R turned out to be neither straightforward to interpret nor to estimate by means of a moment estimator. Aielli, afterwards, presented himself with a corrected version of the model, the *cDCC* model, being more tractable that its predecessors. The reformulation of the correlation driving mechanism as follows lead to the improvement in the tractability of this model:

$$Q_t = (1 - \alpha - \beta)R + \alpha[Q_{t-1}^{*1/2}\eta_{t-1}\eta_{t-1}^{'}Q_{t-1}^{*1/2}] + \beta Q_{t-1} \tag{21}$$

Where $R = [r_{ij}]$ is commonly assumed to be a unit-diagonal matrix ($r_{ii} = 1$, $i = 1, ..., k$). Necessary and sufficient conditions for Q_t to be positive definite are $\alpha > 0$, $\beta > 0$, $\alpha + \beta < 1$ and the positive definiteness of the R matrix.
A more explicit representation of the *cDCC* correlation process $\rho_{ij,t}$ can be considered but we will not discuss it in this paper; see Aielli (2013) for further information.

2.2.6 DCC-MIDAS

The *DCC-MIDAS* was proposed by Colacito et al. (2011). The model's estimation follows the same two-step procedure as the *DCC* model, with the only difference that in the correlation equation we have a further subdivision of two different dynamics: a short-run dynamics and a long-run dynamics. So, we consider the double time index i,t where t refers to the period and i refers to a particular day in that period. The model is defined as:

$$\begin{aligned} H_{i,t} &= D_{i,t}R_{i,t}D_{i,t} \\ D_{i,t} &= \text{diag}\left(\sqrt{h_{1,i,t}}, \cdots, \sqrt{h_{j,i,t}}, \cdots, \sqrt{h_{k,i,t}}\right) \\ R_{i,t} &= (\text{diag}(Q_{i,t}))^{-1/2} Q_{i,t} (\text{diag}(Q_{i,t}))^{-1/2} \\ Q_{i,t} &= (1 - \alpha - \beta)\bar{R}_{i,t}(\omega) + \alpha\left(\eta_{i-1,t}\eta_{i-1,t}'\right) + \beta Q_{i-1,t} \end{aligned} \tag{22}$$

As before, $H_{i,t}$ is the conditional covariance matrix of day i of period t, the matrix $D_{i,t}$ includes in the main diagonal the conditional standard deviations, obtained using a univariate *GARCH* model. The real difference from the previous *DCC* model is $\overline{R}_{i,t}(\omega)$, which represents the long-run correlation. The main assumption is that the long run correlation $\overline{R}_{i,t}(\omega)$ can be filtered from empirical correlations, in fact it is formulated as:

$$\bar{R}_{i,t}(\omega) = \sum_{l=1}^{K_c} \Phi_l(\omega) \odot C_{t-l} \tag{23}$$

This long-run correlation (which, reminding, is a matrix) is given by a summation weighted according to a *Beta function* $\Phi_l(\omega)$ of the Hadamard product of K_c lagged realizations of C_t. This latter is given by:

$$C_t = \begin{bmatrix} v_{1,i,t} & 0 & 0 \\ \vdots & \ddots & 0 \\ 0 & 0 & v_{k,i,t} \end{bmatrix}^{-1/2} \left(\sum_{j=t-N_c}^{t} \eta_j \eta_j' \right) \begin{bmatrix} v_{1,i,t} & 0 & 0 \\ \vdots & \ddots & 0 \\ 0 & 0 & v_{k,i,t} \end{bmatrix}^{-1/2} \tag{24}$$

Where C_t is a correlation matrix and, for simplicity, lets consider $C_t = G^{-1/2}FG^{-1/2}$. G is a diagonal matrix where, in the main diagonal, we have $\upsilon_{k,i,t}$ which are the scalar of the last N_c lagged squared standard returns $\upsilon_{k,i,t} = \sum_{j=t-N_c}^{t} \eta_j^2$. F is also a matrix and considers the summation of the cross-product of the standardized returns coming from the first step. This one considers N_c standardized returns. So, in this correlation model, the key parameters are K_c and N_c.
So, to sum up, the long-run dynamics filters the information coming from the K_c lagged realizations of a particular matrix C_t, this latter consisting of N_c cross-products of the standardized returns coming from the first step.

2.3 MCS Procedure

The Model Confidence Set (MCS) procedure introduced by Hansen et al. (2011) allows to determine the set of models, $\mathscr{M}^*$, that consists of the best model(s) according to a user-specified criterion, starting from a collection of models, $\mathscr{M}^0$. The procedure yields a "superior set of model", $\widehat{\mathscr{M}^*}$, which contain the best model(s) at a given significance level α, and it is based on an equivalence test, $\delta\mathscr{M}$, and an elimination rule, $e\mathscr{M}$. Since the MCS procedure make use of a loss function, which play a key role in the equivalence tests, before moving further with the explanation of the procedure, it's necessary to shift our focus first on loss functions.

2.3.1 Loss functions: QLIKE and RMSE

In this study we are going to use four different loss function, two for evaluating volatility and co-volatility forecast and the other two for the evaluation of the *VaR* forecast of a portfolio of assets. The latter will be approached in the next paragraph, while now we will focus on the volatility estimates. However, it should be noted that the general framework is the same. Since in the first part of the paper we are going to evaluate different multivariate models, our object of interest is the true but unobservable covariance matrix $\Sigma_{t,\delta}$. In order to evaluate a forecast, we make use of a loss function $\mathscr{L}$, which maps the distance between the actual and forecasted value H_t. The main issue here is that volatility is a latent variable, which means that we need to use a proxy instead and this will bring in some "noise". In general, the selection of models performed by the MCS depends on the orderings implied by a loss function and an incorrect choice of the loss function, although not invalidating per se the test, may result in an incorrect identification of the set of superior models. As matter of fact, the ranking based on an inconsistent loss function, together with an uninformative proxy, would be severely biased. For this reason, in order to avoid a distorted outcome, it's necessary to use "robust" or "consistent" loss functions, whose rankings are invariant to the noise in the proxy even when using a less accurate proxy, although the expected losses using the proxy will be larger than those obtained using the "true" but unobserved value. Our proxy for the true but unobservable covariance matrix $\Sigma_{t,\delta}$ will be the realized covariance estimator $\widehat{\Sigma}_{t,\delta}$ with δ = 1 day, which is calculated as the sum of the outer product of intra-daily returns. Our choice for the loss functions fell on the followings:

$$\mathscr{L}(\widehat{\Sigma}_t, H_t) = log|H_t| + vec(H_t^{-1}\widehat{\Sigma}_t) \qquad \text{(QLIKE)} \tag{25}$$

$$\mathscr{L}(\widehat{\Sigma}_t, H_t) = [norm(H_t - \widehat{\Sigma}_t)^{1/2}] \qquad \text{(RMSE)} \tag{26}$$

Vec is an operator that transform a matrix k x k into a vector K^2 x 1, while norm assigns a positive length. Both QLIKE and RMSE are consistent, however the first one is asymmetric which means that it penalises more under-predictions with respect to over-predictions. On the other hand, RMSE is symmetric, which means that it weights the same under and over-predictions.

2.3.2 Quantile loss and Fissler & Ziegel Loss

In this part, by employing a loss function which maps the distance between the observed returns, r_t, and the estimated *VaR* from model l, we will use again the MCS procedure. The loss functions are the Quantile Loss (QL) and the Fissler and Ziegel (FZ Loss), the first one is formally defined as follows:

$$l(r_t, \widehat{VaR}_{l,t,\tau}) = (\tau - I(r_t < VaR_{l,t,\tau}))(r_t - VaR_{l,t,\tau}) \tag{27}$$

Where $\widehat{VaR}_{l,t,\tau}$ is the the VaR estimate of the day t obtained from model l at level τ and it penalize more heavily negative returns which overcome V*aR*.
With the so-called revelation principle, Osband (1985) was one of the first to demonstrate that a functional can be a component of an elicitable vector-valued functional even if it is not itself elicitable. For example, the pair (mean, variance) is elicitable,

despite the fact that variance is not. However, the presence of a bijection between this pair and the first two moments is critical for the validity of the revelation principle. Starting from this result, Fissler and Ziegel (2016) proved that VaR and ES are jointly elicitable, even though ES is not elicitable individually.

2.3.3 The Superior Set of Models

Let's now consider $\mathscr{M}^0$, which contains a finite number of elements indexed by $i = 1, \ldots, m_0$ that represent the loss that is associated with object i in period t as $L_{i,t}$ (with $t = 1, ..., n$). Then the loss differential is defined as follow:

$$d_{ij,t} = L_{i,t} - L_{j,t} \quad \text{for all} \quad i, j \in \mathscr{M} \tag{28}$$

Following the paper by Hansen, we can rank alternatives in terms of the expected loss $E(d_{ij,t})$; which is finite and doesn't depend on the time t. Whenever $\mu_{i,j} < 0$ it means that the alternative i is preferred to the alternative j. Now, we are able to define $\mathscr{M}^*$, the set of superior objectives as:

$$\mathscr{M}^* \equiv i \in \mathscr{M}^0 : \mu_{ij} \leq 0 \quad \text{for all} \quad j \in \mathscr{M}^0 \tag{29}$$

The superior set of models can be determined, as noted before, using a sequence of equal predicting ability (EPA) test, where objects that are significantly inferior or "not equally good" compared to the other elements contained in $\mathscr{M}^0$ are discarded. The hypothesis of the test are:

$$H_{0,\mathscr{M}} : \mu_{i,j} = 0$$

$$H_{1,\mathscr{M}} : \mu_{i,j} \neq 0$$

Where $\mathscr{M} \subset \mathscr{M}^0$. A peculiarity of the MCS procedure is that it acknowledges the limitations of the data: with more informative data, the MCS will return only the best model, however with less informative data, the MCS will return multiple models. Once got the superior set of models, that best estimate the covariance matrix, H_t, we will then use it to calculate the global minimum variance portfolio.

2.4 Portfolio Optimization

A key component of Modern Portfolio theory is diversification (Markowitz, 1959), to yield an investment opportunity with a set of different portfolio compositions, we must vary the allocation of the underlying assets. A blend of various combinations of risky assets manifest specific portfolio risk-return characteristics. The global minimum variance portfolio (GMVP) allocates a given budget among k financial assets such that the risk for the rate of expected portfolio return is minimized, which means that an investor cannot hold a portfolio of risky (note: risk-free assets are excluded at this point) assets with a lower risk than the global minimum-variance portfolio. Once calculated the weights ($\mathbf{w}_{i,t}$) for each asset of the portfolio, we can calculate both the portfolio returns ($\mathbf{r}_{p,t}$) and variance ($\sigma_{p,t}$) as follows:

$$\mathbf{r}_{p,t} = \sum_{i=1}^{k} \mathbf{w}_{i,t} \mathbf{r}_{i,t} \tag{30}$$

$$\sigma_{p,t} = \mathbf{w}_{i,t} H_t \mathbf{w}'_{i,t} \tag{31}$$

2.5 Risk Measures

Let M be the set of all risks and $L \in M$ a financial loss, a risk measure is a map $\rho : M \to \mathbb{R}$ such that $\rho(L)$ identifies the amount of capital necessary to back a financial position with loss L. Hence, it summarizes the amount of capital necessary to compensate loss occurrences for as risky position, so that the risk is measured in monetary terms.

2.5.1 Value at Risk (VaR) and Expected Shortfall (ES)

Value at Risk is a widely used risk measure in finance and it tells us the maximum future loss that can occur over a specific period with a given confidence level τ. Be aware that there is a probability equal to (1-τ) of losing more than $VaR_\tau(L)$.

Formally, the Value at Risk at a confidence interval τ is given by the smallest number $l \in L$ such that the probability that the loss L exceeds l is no larger than $(1-\tau)$.

$$VaR_\tau = inf\{l \in \mathbb{R} : P(L > l) \leq 1-\tau\} = inf\{l \in \mathbb{R} : F_L(l) > \tau\} \tag{32}$$

In probabilistic terms, *VaR* is thus simply a quantile of the loss distribution, so we get:

$$VaR_\tau(L) = q_\tau(L) \tag{33}$$

The main drawback about *VaR* is that it is not a coherent risk measure, because the sub-addictivity property does not hold. So the *VaR* of a portfolio is greater than the sum of the *VaR* of the single financial positions. However, it is elicitable, which means that it is the unique minimizer of the expected loss given by a scoring function.
However, even though 2 positions may have the same *VaR*, they can present different level of tail risk, or the risk related to the 1-τ% worst cases. Due to the crucial importance for the Risk Management function to understand what happens in these extreme cases that could lead to tremendous losses, another risk measure has to be consider. The risk measure that help us answering this question is the Expected Shortfall, *ES*. The *ES* is the conditional expectation of loss given that the loss is beyond the *VaR* level, Yamai (2005); that is:

$$ES_\tau(L) = \mathbb{E}[L|L \geq VaR_\tau(L)] \tag{34}$$

So, if we compute the *ES* at a given level τ, this is nothing more than the expected value of the losses that lie above the *VaR* at τ level. With respect to the *VaR*, the *ES* is a coherent risk measure, since it satisfies all the 4 axioms of coherent risk measure. However, the *ES* is not elicitable, meaning that it cannot be used as a source of comparison between models, since no loss function is computed.

2.5.2 Parametric approach

The parametric approach assumes a given distribution of the returns, $r_t|\mathscr{F}_t \sim N(\mu_t, \sigma_t^2)$, and it calculates *VaR* and *ES* in an indirect way: the first step consist in the estimation of a volatility model, then the one-step-ahead *VaR* forecast can be calculated using the conditional volatility σ_t obtained from step one as follows:

$$VaR_{\tau,t} = \mu_t + \sigma_t \Phi^{-1}(\tau) \tag{35}$$

Where Φ^{-1} is the inverse of the CDF of the Standard Gaussian. If we assume a Student's-t distribution, then:

$$VaR_{\tau,t} = \mu_t + \sigma_t G^{-1}(\tau)\sqrt{\frac{v-2}{v}} \tag{36}$$

Where G^{-1} is the inverse of the CDF of Student's-t with v degrees of freedom and $\sqrt{\frac{v-2}{v}}$ is a correction factor. Keep in mind that usually μ_t is zero, or very close to zero, so we can assume $\mu_t = 0$. It is also possible to account for a low-frequency variable, by employing a GARCH-MIDAS or DAGM model and, in this case, the conditional volatility can be expressed as follow: $\sigma_{i,t} = \sqrt{\tau_t g_{i,t}}$; where τ_t is the long run component, or the low-frequency variable, and $g_{i,t}$ is the short run component. Moving on to the Expected Shortfall, for the Gaussian distribution it can be expressed in the following way:

$$ES_{\tau,t} = \mu_t - \sigma_t \frac{\phi(\Phi^{-1}(\tau))}{\tau} \tag{37}$$

Where ϕ is the probability density function of a standard Gaussian distribution. The ES for the Student's-t distribution can be formulated as follows:

$$ES_{\tau,t} = \mu_t - \sigma_t \left(\frac{g_v(G_v^{-1}(\tau))}{\tau}\right)\left(\frac{v+(G_v^{-1}(\tau))^2}{v-1}\right)\sqrt{\frac{v-2}{v}} \tag{38}$$

Where g_v is the probability density function of a standard Student's-t with v degrees of freedom.

2.5.3 Non-parametric approach

In this case, no parametric assumption concerning the distribution of returns should be made and we can "let the data speak for themselves". The approach used is the Historical Simulation method (Hendricks, 1996), which calculates the $VaR_(\tau,t)$ as the sample quantile of past returns computed over a moving-window of length ω. The *VaR* at level τ is the τ-th sample quantile of the returns vector:

$$\widehat{VaR}^{\omega}_{\tau,t} = \widehat{Q}_{r_t,\omega}(\tau) \tag{39}$$

The Expected Shortfall is given by the summation of the series of returns times the series of violations, divided by the latter:

$$\widehat{ES}^{\omega}_{\tau,t} = \frac{\sum_{i=1}^{\omega} r_{t-\omega-2+i} I(r_{t-\omega-2+i} \leq \widehat{Q}_{r_t,\omega})}{\sum_{i=1}^{\omega} I(r_{t-\omega-2+i} \leq \widehat{Q}_{r_t,\omega})} \tag{40}$$

The main advantage of the non-parametric approach is that it is easy to implement, however there are a few drawbacks: it cannot extrapolate beyond the maximum observed data point; difficulty in the estimation of extreme quantiles (underestimate the risk); the i.i.d. assumption of the returns within the moving window doesn't hold; and the choice of the length of the moving window ω.

2.5.4 Semi-parametric approach

The last estimation method we propose is the *semi-parametric approach*. It is a compromise between parametric and non-parametric approaches because, although we do not make a distributional assumption for the returns, there is still a finite-dimensional unknown parameter vector to be estimated. Since our aim is to estimate *VaR* and *ES* jointly, we employ here a joint model of conditional *VaR* and *ES*, proposed by Taylor (2019), to be estimated by maximizing an *Asymmetric Laplace* (*AL*) likelihood:

$$f(r_t \mid VaR_{\tau,t}, \sigma, \tau) = \frac{\tau(1-\tau)}{\sigma} \exp\left(-\frac{(r_t - VaR_{\tau,t})(\tau - I(r_t \leq VaR_{\tau,t}))}{\sigma}\right) \tag{41}$$

where $VaR_{\tau,t}$ is the locator parameter, σ is the scale parameter and τ is the skew parameter, but also the significance level of *VaR*, and it is usually set at 1% or 5%. Re-writing *ES* as:

$$ES_{\tau,t} = E(r_t) - \frac{1}{\tau} E((r_t - VaR_{,t})(\tau - I(r_t \leq VaR_{\tau,t}))) \tag{42}$$

we define *ES* estimates as a byproduct of quantile regression. We assume $E(r_t) = 0$ and consider the conditional maximum likelihood estimator for the scale σ of the AL density of expression 41 as follows:

$$\hat{\sigma}_t = E_t((r_t - VaR_{\tau,t})(\tau - I(r_t \leq VaR_{\tau,t}))) \tag{43}$$

Exploiting equations 42 and 43, we can express *ES* in terms of the scale estimator:

$$ES_{\tau,t} = -\frac{\hat{\sigma}_t}{\tau} \tag{44}$$

and finally use expression 44 into the *AL* distribution:

$$f(r_t \mid VaR_{\tau,t}, \sigma_t, \tau) = \frac{(\tau-1)}{ES_{\tau,t}} \exp\left(\frac{(r_t - VaR_{\tau,t})(\tau - I(r_t \leq VaR_{\tau,t}))}{\tau ES_{\tau,t}}\right) \tag{45}$$

A model for the conditional scale σ_t can be estimated, along with a model for the conditional quantile $VaR_{\tau,t}$, using the maximum likelihood based on the last expression of the AL density function. Several models can be employed for *VaR* estimation in the semi-parametric approach, while for *ES* estimation the model here chosen is the *Multiple of VaR* that expresses *ES* as product of the quantile and a constant multiplicative factor:

$$ES_{\tau,t} = (1 + e^{\gamma_0}) VaR_{\tau,t} \tag{46}$$

this formulation avoids that *ES* estimates cross the corresponding *VaR* estimates by ensuring that the constant multiplicative factor is greater than 1. The exponential function ensures that, and γ_0 has no constraint. The next step is to define the models for estimating *VaR*.

2.5.5 Quantile Linear ARCH and ARCH-MIDAS

It is important to stress that what follows is the first step in a broader process described above and concerns the methods for estimating *VaR*. Here we explain the *Quantile Linear ARCH* model (*Q-L-ARCH*), of Koenker and Zhao (1996), the first contribution where the *VaR* is estimated within the *Quantile Regression* framework. It is called Linear ARCH because it has the same ARCH representation with the exception of the lagged terms that are in absolute value instead of squared:

$$r_t = \sigma_t z_t = \left(\beta_0 + \beta_1 |r_{t-1}| + \cdots + \beta_q \left|r_{t-q}\right|\right) z_t, \quad \text{with } t = 1, \ldots, T \tag{47}$$

where $\beta_0 > 0$, $\beta_1, ..., \beta_q \geq 0$ and $z_t \overset{iid}{\sim} (0,1)$ with an unspecified distribution. $VaR_{\tau,t}$ is defined as the $\tau - th$ conditional quantile of returns r_t, given the information set F_{t-1}. Thanks to the link between *VaR* and the quantile of the conditional distribution we can define the *VaR* as:

$$VaR_{\tau,t} = Q_{r_t}(\tau \mid F_{t-1}) \tag{48}$$

where $Q_{r_t}(\cdot)$ is the *quantile regression function* that directly calculates the quantile at any level (in our case τ) of the dependent variable, which in such respect is daily returns. Expression 48 allows us to use the ARCH representation of 47 within it. Therefore, let $x_t = \left(1, |r_{t-1}|, \ldots, \left|r_{t-q}\right|\right)'$ and $\gamma(\tau) = (\alpha_0 Q_z(\tau), \ldots, \alpha_q Q_z(\tau))'$, then, according to quantile regression, $VaR_{\tau,t} = x_t'\gamma(\tau)$. Where $\gamma(\tau)$ is a vector of unknown parameters estimated, under the *QR* framework, by minimizing the *quantile loss function*:

$$\hat{\gamma}(\tau) = \underset{\gamma \in \mathbb{R}^k}{\arg\min} \left[\sum_{t \in \{r_t \geq x_t'\gamma\}} \tau \left|r_t - x_t'\gamma\right| + \sum_{t \in \{r_t < x_t'\gamma\}} (1-\tau) \left|r_t - x_t'\gamma\right| \right], \tag{49}$$

The quantile loss function gives asymmetric penalties to underpredictions $\tau |r_t - x_t'\gamma|$ and overpredictions $(1-\tau)|r_t - x_t'\gamma|$.
An extension of this model is the *Quantile Linear ARCH-MIDAS* model (*Q-L-ARCH-MIDAS*) that includes to the set of regressors also a MIDAS-term, variable observed at different frequencies with respect to that of the dependent variable. Expression 47 changes to:

$$r_{i,t} = \left(\beta_0 + \beta_1 |r_{i-1,t}| + \cdots + \beta_q \left|r_{i-q,t}\right| + \theta \sum_{j=1}^{K} \delta_k(\omega) \left|X_{t-j}\right| \right) z_{i,t} \tag{50}$$

where i indicates the day within the t period that could be a week, a month and so forth, X_{t-j} could be a macro-economic variable which is weighted according to the *Beta function* $\delta(\omega)$, and θ is the coefficient that indicates the impact of the weighted variables. *VaR* is estimated in the same way as in the *Q-L-ARCH* model, by minimizing the quantile loss function, the only difference being that θ cannot be estimated unless a value of ω_2 is fixed within the beta function. Fixing ω_2 serves to make observable the part of the regressors that are a function of it, i.e. the lagged observations of the macroeconomic variable.

2.5.6 CAViaR

Another type of model which considers the modelization of the quantile of the return distribution through the quantile regression approach, Laporta (2019), is the one established by Engle and Manganelli (2004) who, reminding that return's volatility tends to cluster over time, considered this type of behaviour inside the formulation of the *VaR* with an autoregressive specification. Proposing the so-called *Conditional Autoregressive Value at Risk* (*CAViaR*).
Taking into account the first order of the general specification of the *CAViaR*, we are going to consider three different specifications of *CAViaR* processes; all of which consider as the dependent variable directly the τ-th quantile at time t and as independent variables its lagged value and other relevant exogenous variables; in this paper we are going to use only lagged returns, Laporta (2019). Namely:

- *Symmetric Absolute Value* (*SAV*)
$$VaR_{\tau,t} = \beta_0 + \beta_1 VaR_{\tau,t-1} + \beta_2 |r_{t-1}| \tag{51}$$
having symmetric response for positive and negative returns on the dependent variable.

- *Asymmetric Slope* (*AS*)
$$VaR_{\tau,t} = \beta_0 + \beta_1 VaR_{\tau,t-1} + (\beta_2 I(r_{t-1} > 0) + \beta_3 I(r_{t-1} < 0)) |r_{t-1}| \tag{52}$$
where the model allows positive and negative returns to impact differently on *VaR*, Laporta (2019). This asymmetry enters the model through β_2 if the returns are positive or by means of β_3 if the returns are negative.

- *Indirect* $GARCH(1,1)$ (IG)

$$VaR_{\tau,t} = \sqrt{\beta_0 + \beta_1 VaR^2_{\tau,t-1} + \beta_2 r^2_{t-1}} \tag{53}$$

  which is correctly specified if the underlying data were truly a $GARCH(1,1)$ with an $i.i.d.$ error distribution, Engle and Manganelli (2004). It responds symmetrically to past returns as the *SAV* model.

The autoregressive term $\beta_1 VaR_{\tau,t-1}$ ensures that the *VaR* changes "smoothly" over time, Engle and Manganelli (2004).
For any fixed τ-th quantile, the vector of parameters β is estimated by minimizing the quantile loss function of Koenker and Bassett (1978), Laporta (2019).

2.6 Backtesting

Financial institutions have the ability to specify their own model to compute their Value at Risk under the 1996 Market Risk Amendment to the Basel Accord, which took effect in 1998. As a result, regulators must scrutinize the quality of the models in use by assessing forecast accuracy, a process known as "backtesting", Berkowitz (2011).
Considering *VaR* backtesting, this technique takes into account the comparison of the *VaR*'s forecasts with the actual returns during the selected time horizon. As a result, the assessment is based on a binary variable that represents a series of *VaR* violations:

$$L_{T+j,1-\tau} = I \begin{cases} 1 \; if \; r_{T+j} < \widehat{VaR}_{T+j,\tau} \\ 0 \; if \; r_{T+j} \geq \widehat{VaR}_{T+j,\tau} \end{cases} , \quad \text{with} \quad j = 1, \ldots, H \tag{54}$$

Where I is an indicator function, taking value 1 if there is an exceedance in that particular day and 0 otherwise, H represents the length of the period under consideration. If the model is correctly specified, we expect that the number of violations will coincide with those of the significance level considered (1- τ):

$$\mathbb{E}\left[L_{T+j,1-\tau}\right] = \Pr(I = 1) = 1 - \tau \tag{55}$$

Four main tests are considered in this paper to check the prediction ability:

- *I)* Actual over Expected ratio: it's the ratio between the observed number of violations obtained and the expected *VaR* violations considering the specific 1-τ level. The closer the ratio is to 1, the better the model estimates the *VaR*. If the ratio is greater than one, the model is underestimating the risk; otherwise, it is overestimating the risk.
- *II)* Kupiec's unconditional coverage (*UC* test) (1995): this test aims at verifying if the violations have been in line with the 1-τ level of *VaR*. In particular, it performs a likelihood ratio tets where the null hypothesis states that the probability of a violation is equal to the significance level 1-τ. Under the null, the *UC* statistics converges in distribution to a χ^2 with one degree of freedom.
- *III)* Christoffersen's conditional coverage (*CC* test) (1998): this test considers togheter the UC test and the Independence test. It performs a likelihood ratio test where the null verifies simultaneously if the *VaR* violations are independent and the average number of violations is correctly considered. Under the null, the CC statistics is asymptotically distributed as a χ^2 with two degree of freedom.
- *IV)* Egle and Manganelli's Dynamic Quantile (*DQ* test) (2004): it verifies the same hypothesis of the *CC* test by means of a regression-based method, in which a linear regression model is created with the Hit sequence ($Hit = L_{t+1,1-\tau} - (1-\tau)$) as the dependent variable and past Hits and any other relevant variable as explanatory variables. The test statistic is distributed as χ^2 with q degrees of freedom. In this paper we consider 4 degrees of freedom, which corresponds to the number of lagged Hits and lagged *VaR*s, Laporta (2018).

Moreover, after discovering disparities across banks in October 2013, the Basel Committee proposed to switch from the traditional *VaR* technique of calculating losses to an alternative, known as "Expected Shortfall". This approach was questioned, however, since it's well known lack of elicitability property, leading some to conclude that *ES* could not be backtested as well, Acerbi (2014). However, some methods have been proposed, demonstrating that even if the *ES* is not elicitable, it can be backtested. Here we are going to consider the practice of McNeil & Frey (2000), which proposed a bootstrap hypothesis test based on the so-called violation residuals. These measure the distance between the realized losses and the expected shortfall estimates when *VaR* violations occur, where the null hypothesis considers that the residuals have zero mean, Kratz (2018). The test is a one sided t-test against the alternative that the expected shortfall is systematically underestimated, Chinhamu (2015).

3. Empirical Application

The empirical part consists of the analysis of the stock market indices of 3 of the top 4 European economies by GDP: Italy, France, and Spain. Data are provided by *Oxford Man Institute* and the analysis was conducted in R. The analysis runs from June 1, 2009 to December 30, 2021. In this period we collected 3180 observations. As you can see, the time series of the closing prices of the three financial indicators are not stationary. The *FTSE MIB* has ups and downs over time, indicating the presence of economic cycles. The *CAC 40* shows an increasing trend. And the time series of the *IBEX 35* also shows fluctuations typical of economic cycles. We also conducted tests to check the stationarity of the series. In particular, the Augmented Dickey-Fuller

Figure 1. CAC 40 closing price

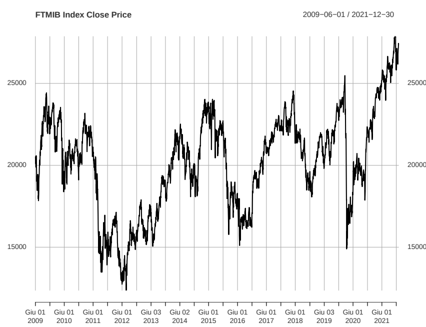

Figure 2. FTSE MIB closing price

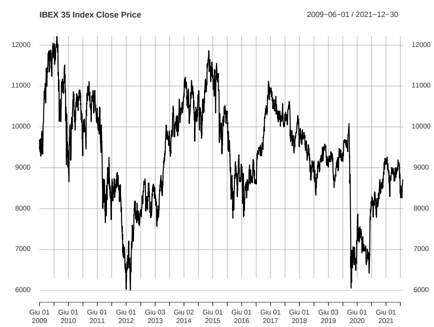

Figure 3. IBEX 35 closing price

test, which returned for all series a p-value greater than 0.05. So we do not reject the null hypothesis and therefore the series are non-stationary.

Table 1. Augmented Dickey-Fuller test

	CAC 40	FTSE MIB	IBEX 35
ADF	0.1504	0.2724	0.162

Unlike closing prices, returns are stationary, the average is constant over time and equal to zero. As mentioned in the methodology part, returns show the so-called stylized facts, one of which is already recognisable graphically, namely volatility clustering. A cluster of high volatility is visible at the end of 2019 for all series. This cluster is most likely due to the onset of the pandemic, which led to strong shocks also in the financial sector. To support graphical representations, again the Augmented

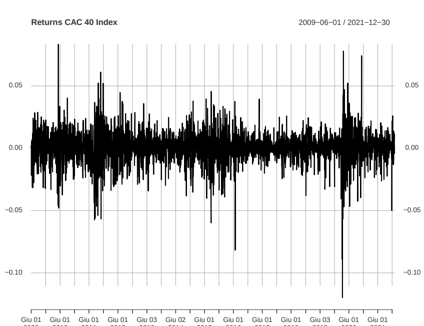

Figure 4. CAC 40 returns

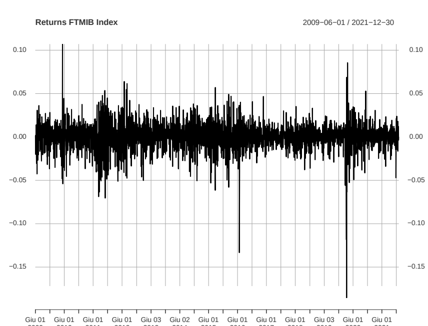

Figure 5. FTSE MIB returns

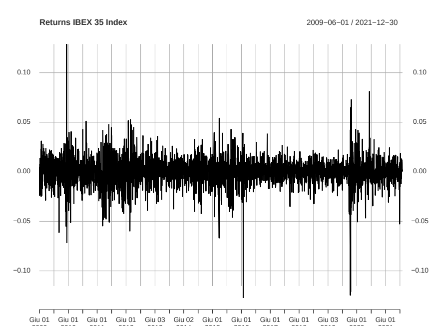

Figure 6. IBEX 35 returns

Dickey-Fuller test was employed. This time the null hypothesis is rejected and the series are therefore stationary. We also tested if the mean of returns was equal to zero and the p-values are greater than 0.05, so we do not reject the null hypothesis that the mean is equal to zero. Finally, the ArchTest checks for the presence of heteroschedasticity in the series. The ArchTest results reject the null hypothesis of heteroschedasticity, so our returns are homoschedastic and the variance is constant over time.

Table 2. tests on returns

	CAC 40	FTSE MIB	IBEX 35
ADF	0.01*	0.01*	0.01*
t.test	0.2937	0.7434	0.9006
ArchTest	$2.2e^{-16}$	$2.2e^{-16}$	$2.2e^{-16}$

*R returned "p-values are smaller than the ones printed"

We calculated descriptive statistics on returns such as mean, maximum and minimum, but kurtosis and skewness are our statistics of interest. As we know, the distribution of returns is not Gaussian, but it is leptokurtic and asymmetric. Therefore, the excess of kurtosis over the Normal distribution is reported in table 3. All three assets are leptokurtic in accordance with theory. With respect to skewness, returns are negatively skewed, so there is a left skewness of the distributions indicating that left tails are heavier than right tails. This means that negative shocks have a greater impact on volatility than positive shocks, a phenomenon called *leverage effect*.

Table 3. Descriptive statistics of returns

	Mean	Minimum	Maximum	Standard deviation	Skewness	Kurtosis
CAC 40	0.0002370371	-0.1199717	0.0834047	0.01272733	-0.520493	6.831773
FTSE MIB	$9.044e^{-05}$	-0.1854114	0.1068403	0.01558057	-0.9279276	10.93105
IBEX 35	$3.160011e^{-05}$	-0.1271772	0.1287164	0.01426623	-0.5141538	9.352784

We tested the non-normality of our returns with the Jarque Bera test, whose null hypothesis is the Normality of the series, but the p-values reject the Gaussian hypothesis.

Table 4. Normality test on returns

	CAC 40	FTSE MIB	IBEX 35
JB test	$2.2e^{-16}$	$2.2e^{-16}$	$2.2e^{-16}$

In addition to the tests, a graphical comparison of our returns against the Gaussian is available below. Kurtosis and skewness are hardly visible from the density function, but a slight deviation from the Normal is detectable. It is more visible with the *quantile-quantile Plot*, which zooms in on the tails of the distribution and emphasises the deviation of these tails from the Normal case. The red line represents the Guassian distribution, while the dots represent the distribution of returns, and the kurtosis and left skewness of the series are evident. Another feature of returns is the uncorrelation. Figures 13, 14, 15 report the correlograms of the three stock market indices. As can be seen, this feature is respected for all three indices. The lines are within the confidence intervals, indicating that the returns on day t do not depend on the returns on day $t-1$ and so on. Therefore, returns seem uncorrelated and follow a *White Noise* process. Also the correlation tests suggest the uncorrelation of returns. Both the Ljung-Box test and the Box-Pierce test do not reject the null hypothesis of uncorrelation. This means that the returns cannot be modelled.

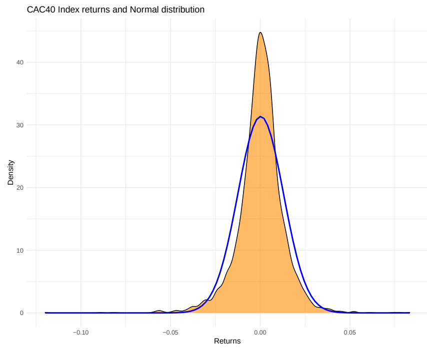

Figure 7. CAC 40 ret. density func.

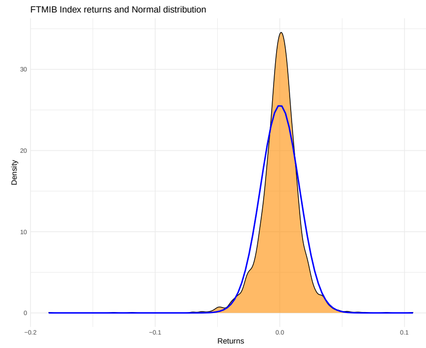

Figure 8. FTSE MIB ret. density func.

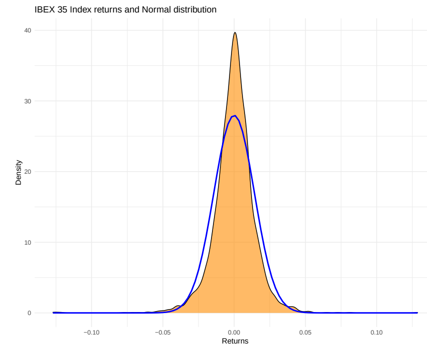

Figure 9. IBEX 35 ret. density func.

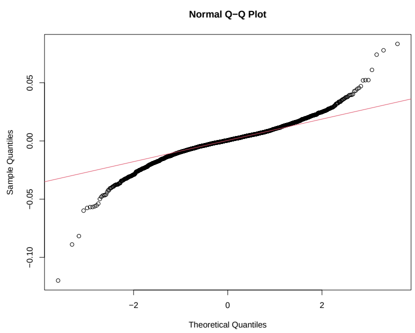

Figure 10. CAC 40 returns QQ-plot

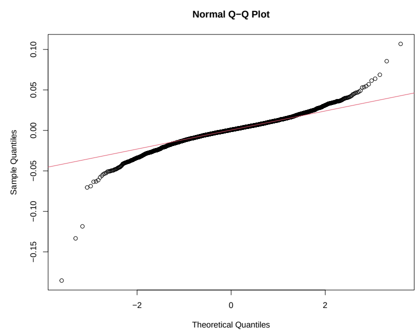

Figure 11. FTSE MIB returns QQ-plot

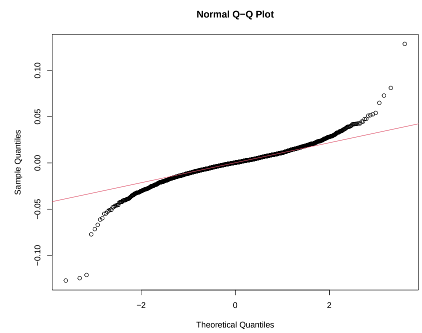

Figure 12. IBEX 35 returns QQ-plot

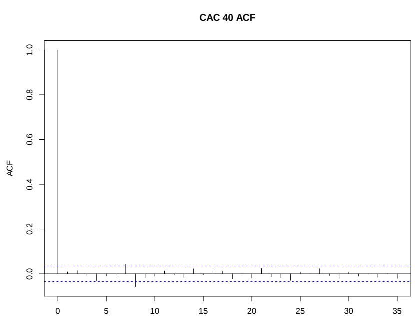

Figure 13. CAC 40 ACF returns

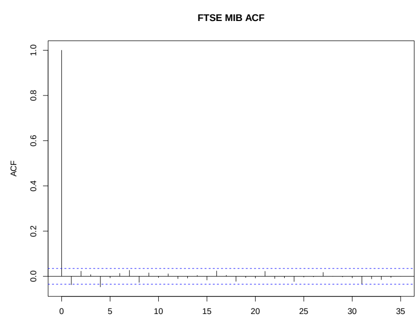

Figure 14. FTSE MIB ACF returns

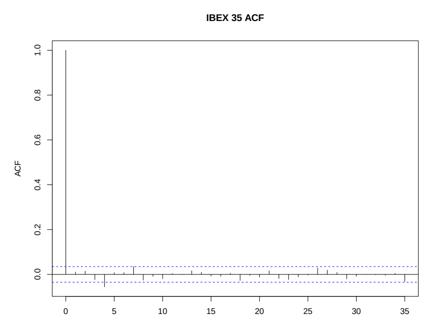

Figure 15. IBEX 35 ACF returns

Table 5. Correlation tests on returns

	CAC 40	FTSE MIB	IBEX 35
LB test	0.08375	0.1974	0.1817
BP test	0.08568	0.2	0.1844

But if we look at the correlograms of squared returns we notice a certain dependence. The lines of the correlograms exceed the confidence interval, which means that the squared returns of day t depend on the squared returns of day $t-1$ and so on. Correlation tests support persistence of squared returns. Both Box-Pierce and Ljung-Box tests return a p-value of $2.2e^{-16}$, for

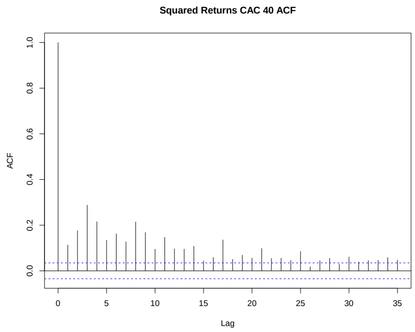

Figure 16. CAC 40 ACF squared ret.

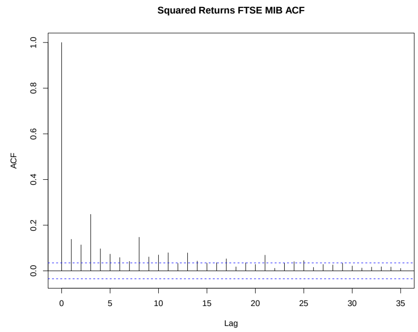

Figure 17. FTSE MIB ACF squared ret.

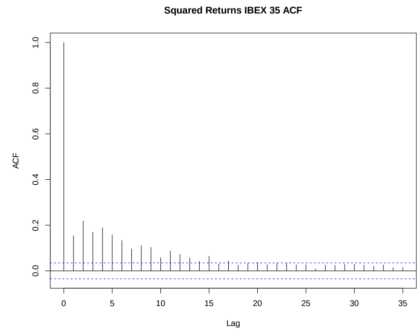

Figure 18. IBEX 35 ACF squared ret.

all assets, implying the rejection of the null hypothesis. This behaviour of the squared returns allows us to study and model the variance and therefore also of the standard deviation, which in finance is called volatility; recall that since we are in a multivariate framework, we will consider also the relationships between the assets.

The first models implemented are the *Moving Covariance* models. Since $V > k$, where k is the number of assets and it is equal to 3 in our case, we have chosen a moving window length of 22 and 252 days, corresponding to the number of trading days, respectively, in one month and in one year.
In EWMA models, instead, the parameter to be set is λ. Usually λ varies between 0.90 and 0.99, so we implemented two EWMA models with different plausible values of λ, 0.94 and 0.97, giving different weights to past observations.
For EWMA models, we observe that all the parameters are statistically significant and the total number of parameters estimated are 8 for the scalar version and 12 for the diagonal variant of the model.
OGARCH model relies on *PCA* to extract the principal components. Carol (2011) has shown that when the system is highly correlated model calibration will need much less care. Therefore, first of all, we checked how much our system is correlated, and the table 6 show a high correlation between returns. This is also an indicator of high noise in the data that could be filtered by PCA procedure.

Table 6. Sample correlation matrix

	CAC 40	FTSE MIB	IBEX 35
CAC 40	1	0.88	0.83
FTSE MIB	0.88	1	0.84
IBEX 35	0.83	0.84	1

Before proceeding with PCA, we standardized returns by column, becasue PCA is sensible to re-scaling. After that, we employed `prcomp()` function in `R` to perform PCA on standardized returns. Figure 19 shows the share of variation of the system explained by each principal component. The first principal component, alone, accounts for 90% of the total variation of the system and it is quite enough to represent the whole system, but the conditional covariance matrix given by a single principal component was not positive definite, and the same using two principal components. Therefore, we employed all three principal components to build the conditional covariance matrix.

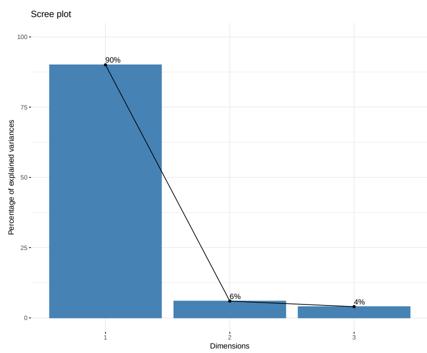

Figure 19. Variability explained by PC

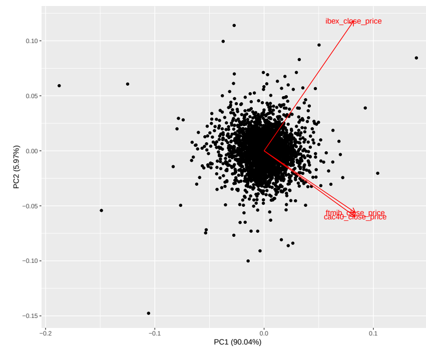

Figure 20. Principal component space

The figure 20 graphically represents what the PCA has done mathematically. The axes are rotated according to the first and second principal components. The black dots represent the original observations while the vectors are referred to each asset and are determined by the weights of eigenvectors. The direction of each vector gives us information on the relationship between assets and principal components and indicates how much an asset contributes to a principal component. For instance, *FTSE MIB* and *CAC 40* point towards the first principal component, meaning that the two assets contribute the most to it. The length of the vector instead indicates how much variability of the assets is explained by the two principal components. For example, the vector of *IBEX 35* is the longest, meaning that the two principal components explain more of its variability than those of the other two assets.
To conclude, another standardization is needed, this time on the principal components, in order to reset the results to the original scale. The univariate specifications employed to estimate the conditional variance of each principal component are the sGARCH and gjrGARCH models with two different distribution, Normal and Student's-t distributions. At this point, to compute the conditional covariance matrix of the returns, we multiplied the matrix of standardized factor weights by the diagonal matrix of principal components variances and, in turn, by the transposed of the former.
Passing now to the cDCC and DCC-MIDAS we considered 4 different GARCH specifications: sGARCH, gjrGARCH, GARCH-MIDAS and DAGM, and 2 different distributions, the Gaussian and Student's-t. Coming to a total of 12 different specifications for each one of the two models. Following the paper of Colatico (2011), we selected as parameters $K_c = 36$ and $N_c = 144$.
For the univariate specifications GARCH-MIDAS and DAGM we selected the *Long-Term Government Bond Yield*: 10-*year* as the macroeconomic variable, collected monthly from the Federal Reserve Economic Data of St. Louis (*FRED*) from the period January 1st 2005 to January 1st 2022 for all the three countries considered (Italy, Spain and France).
We further performed the Augmented Dickey-Fuller test on the macroeconomic variables to check for stationarity. For all three countries the macroeconomic variable didn't reject the null hypothesis of presence of a unit root. For this reason, we applied the first difference to all yields, rejecting this time the null hypothesis of non stationarity. In this case we consider $K = 48$.

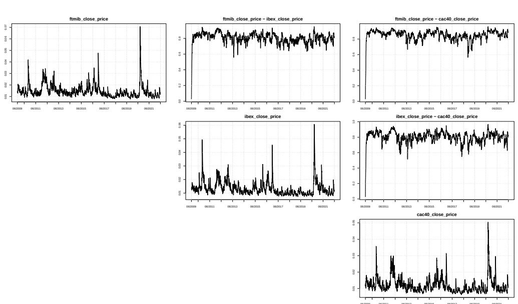

Figure 21. cDCC sGARCH-norm

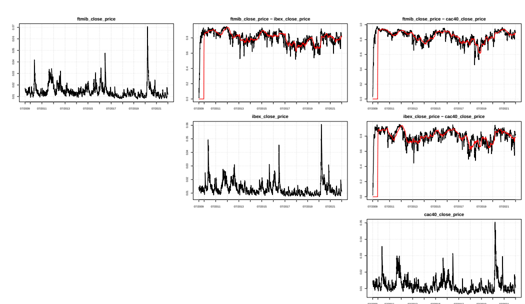

Figure 22. DCC-MIDAS sGARCH-norm

Here we report the plots of the cDCC and DCC-MIDAS model with univariate sGARCH specification with Normal distribution. On the main diagonal we have the conditional variances coming from the univariate specification while on the off diagonal we report the conditional correlations; where, in figure 22, the red line represents the long-run correlation. In addition, Colatico at al. (2011) find that the DCC-MIDAS model fits better than the ordinary DCC model in their article. With this in mind, we check if this is the case with the cDCC.

Table 7. BIC Results

	cDCC	DCC-MIDAS		cDCC	DCC-MIDAS
sGARCH-norm	2149.419	2391.928	GM-skew-norm	2657.221	2914.758
sGARCH-std	2164.015	2334.394	GM-skew-std	3085.657	3270.068
GJR-norm	2660.477	2908.128	DAGM-noskew-norm	2210.559	2449.1
GJR-std	2914.667	3112.488	DAGM-noskew-std	2215.631	2378.892
GM-noskew-norm	2202.859	2443.922	DAGM-skew-norm	4290.643	4185.504
GM-noskew-std	2233.477	2406.492	DAGM-skew-std	2898.863	3117.15

Consider the BIC (Bayesian Information Criterion) as a quick 'approximation' for determining whether the DCC-MIDAS is superior than the cDCC, as shown in the above table, we have that for 11 out of the 12 different specifications the BIC of the cDCC is lower than the BIC of the DCC-MIDAS; with the cDCC-GARCH-norm model having the lowest BIC, the most 'basic' model. Moreover, the models that considered the macroeconomic variable inside their univariate part performed no better than the sGARCH and gjrGARCH, only the models without the skew performed relatively well.

Following the analysis of the BIC information criteria, we employed 2 times the MCS procedure with a level of significance set to $\alpha = 0.25$ using 2 different loss functions (QLIKE and RMSE). The results of both the MCS indicate that no model was discarded and they seem to be in line with the previous considerations using the information criteria. To develop further the analysis with the global minimum variance portfolio and risk measures, we decide to pick the first model from MCS which employ the QLIKE loss function, so Moving Covariance with $V = 22$.

Table 8. MCS (QLIKE loss function)

	Rank_M	v_M	MCS_M	Rank_R	v_R	MCS_R	Loss
MC 22	1	-0.098	1	1	-0.016	1	-29.372
MC 252	19	-0.093	1	11	2.779	0.208	-28.680
EWMA 94	34	0.666	0.551	3	0.554	1	361.455
EWMA 97	33	0.013	0.999	2	0.016	1	39.318
OGARCH-sGARCH-norm	17	-0.093	1	22	3.152	0.082	-28.738
OGARCH-gjrGARCH-norm	15	-0.094	1	19	3.055	0.104	-28.762
OGARCH-sGARCH-std	18	-0.093	1	28	3.321	0.050	-28.714
OGARCH-gjrGARCH-std	16	-0.093	1	23	3.204	0.068	-28.742
SBEKK	2	-0.095	1	4	2.652	0.270	-28.965
DBEKK	3	-0.094	1	21	3.083	0.098	-28.878
GM skew-norm-DCCMIDAS	26	-0.092	1	18	2.995	0.125	-28.527
GM skew-std-DCCMIDAS	22	-0.092	1	6	2.688	0.251	-28.569
GM noskew-norm-DCCMIDAS	30	-0.092	1	30	3.336	0.049	-28.489
GM noskew-std-DCCMIDAS	24	-0.092	1	20	3.061	0.103	-28.560
DAGM skew-norm-DCCMIDAS	32	-0.090	1	33	3.755	0.014	-28.165
DAGM skew-std-DCCMIDAS	25	-0.092	1	10	2.761	0.217	-28.556
DAGM noskew-norm-DCCMIDAS	28	-0.092	1	27	3.285	0.055	-28.508
DAGM noskew-std-DCCMIDAS	21	-0.092	1	15	2.916	0.152	-28.590
sGARCH-norm-DCCMIDAS	29	-0.092	1	25	3.224	0.065	-28.503
sGARCH-std-DCCMIDAS	20	-0.093	1	14	2.846	0.180	-28.605
gjrGARCH-norm-DCCMIDAS	27	-0.092	1	16	2.944	0.142	-28.517
gjrGARCH-std-DCCMIDAS	23	-0.092	1	7	2.732	0.251	-28.562
sGARCH-norm-cDCC	10	-0.094	1	26	3.243	0.061	-28.799
sGARCH-std-cDCC	4	-0.094	1	17	2.977	0.131	-28.873
gjrGARCH-norm-cDCC	9	-0.094	1	8	2.747	0.251	-28.800
gjrGARCH-std-cDCC	7	-0.094	1	5	2.678	0.270	-28.810
GM skew-norm-cDCC	8	-0.094	1	12	2.797	0.199	-28.808
GM skew-std-cDCC	11	-0.094	1	9	2.754	0.251	-28.791
GM noskew-norm-cDCC	14	-0.094	1	31	3.376	0.043	-28.784
GM noskew-std-cDCC	6	-0.094	1	29	3.328	0.050	-28.820
DAGM skew-norm-cDCC	31	-0.090	1	34	4.990	0	-28.173
DAGM skew-std-cDCC	12	-0.094	1	13	2.808	0.195	-28.790
DAGM noskew-norm-cDCC	13	-0.094	1	32	3.436	0.037	-28.786
DAGM noskew-std-cDCC	5	-0.094	1	24	3.206	0.068	-28.837

Table 9. MCS (RMSE loss function)

	Rank_M	v_M	MCS_M	Rank_R	v_R	MCS_R	Loss
MC 22	9	-1.364	1	2	0.589	1	0.022
MC 252	22	-1.055	1	32	3.599	0.012	0.023
EWMA 94	20	-1.159	1	33	5.460	0	0.023
EWMA 97	21	-1.136	1	34	9.502	0	0.023
OGARCH-sGARCH-norm	12	-1.335	1	14	1.747	0.751	0.022
OGARCH-gjrGARCH-norm	1	-1.390	1	1	-0.589	1	0.022
OGARCH-sGARCH-std	13	-1.321	1	26	2.491	0.234	0.022
OGARCH-gjrGARCH-std	7	-1.378	1	15	1.794	0.715	0.022
SBEKK	11	-1.351	1	17	2.179	0.421	0.022
DBEKK	10	-1.354	1	3	0.836	1	0.022
GM skew-norm-DCCMIDAS	23	1.300	0.400	9	1.333	0.968	0.040
GM skew-std-DCCMIDAS	24	1.300	0.400	10	1.333	0.968	0.040
GM noskew-norm-DCCMIDAS	31	1.335	0.390	24	2.435	0.264	0.040
GM noskew-std-DCCMIDAS	33	1.340	0.389	27	2.508	0.226	0.040
DAGM skew-norm-DCCMIDAS	27	1.306	0.398	13	1.465	0.923	0.040
DAGM skew-std-DCCMIDAS	26	1.302	0.400	12	1.334	0.967	0.040
DAGM noskew-norm-DCCMIDAS	29	1.334	0.390	23	2.291	0.345	0.040
DAGM noskew-std-DCCMIDAS	32	1.338	0.389	21	2.230	0.385	0.040
sGARCH-norm-DCCMIDAS	30	1.335	0.390	25	2.474	0.245	0.040
sGARCH-std-DCCMIDAS	34	1.341	0.388	30	2.654	0.226	0.040
gjrGARCH-norm-DCCMIDAS	25	1.301	0.400	11	1.334	0.968	0.040
gjrGARCH-std-DCCMIDAS	28	1.307	0.398	19	2.204	0.421	0.040
sGARCH-norm-cDCC	16	-1.318	1	20	2.218	0.421	0.022
sGARCH-std-cDCC	19	-1.306	1	29	2.652	0.226	0.023
gjrGARCH-norm-cDCC	5	-1.384	1	8	1.201	1	0.022
gjrGARCH-std-cDCC	8	-1.374	1	31	2.884	0.226	0.022
GM skew-norm-cDCC	3	-1.387	1	6	0.943	1	0.022
GM skew-std-cDCC	2	-1.388	1	4	0.926	1	0.022
GM noskew-norm-cDCC	15	-1.319	1	18	2.196	0.421	0.022
GM noskew-std-cDCC	18	-1.310	1	28	2.521	0.226	0.022
DAGM skew-norm-cDCC	4	-1.384	1	5	0.940	1	0.022
DAGM skew-std-cDCC	6	-1.383	1	7	0.967	1	0.022
DAGM noskew-norm-cDCC	14	-1.320	1	16	2.041	0.715	0.022
DAGM noskew-std-cDCC	17	-1.314	1	22	2.257	0.385	0.022

In the period, from 2009 to the end of 2016 there is a predominance in the allocation in favour of *CAC 40* index, with some exceptions; while after 2017 the asset allocation appear to be more diversified and we could perhaps highlight that *FTSE MIB* has a greater influence on the allocation weights in this last period.

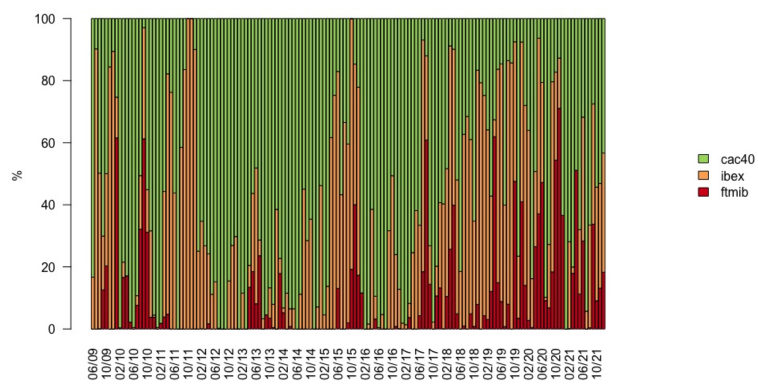

Figure 23. Global Minimum Variance Portfolio

Using the weights estimated within the portfolio optimization approach, we derived the series of the returns for the whole portfolio, as well as the variance, and we employed them for the estimation concerning risk measures. Specifically, we estimated the Value at Risk and the Expected Shortfall considering three different approaches: the parametric approach, the non-parametric approach and the semi-parametric approach, by going *out-of-sample* of 20% of the length of the series.
For the parametric approach, we calculated the *VaR* and *ES* inderectly from the volatility estimates, which were calculated by employing both the GARCH(1,1) with a Normal distribution and the Student's-t distribution.

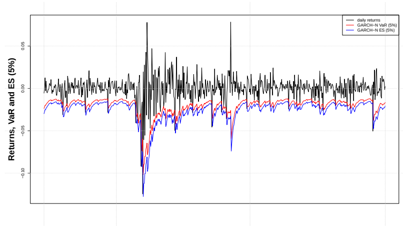

Figure 24. GARCH-N

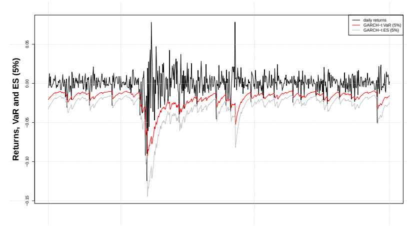

Figure 25. GARCH-T

For the non-parametric approach we have selected three different window lengths: ω = 100, 252 and 500. From the figures, we can observe that if ω is small, then the quantile follows better the data. However, in this case, the sample quantile will not be consistent. If ω is large, then the sample quantile is constant over time and doesn't follow to much the data. The sample quantile will be a consistent and asymptotically normal estimator, but the i.i.d assumption falls (structural breaks).

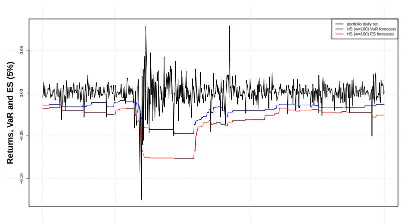

Figure 26. HS (w=100)

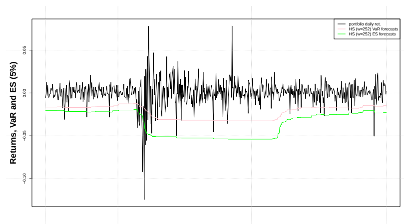

Figure 27. HS (w=252)

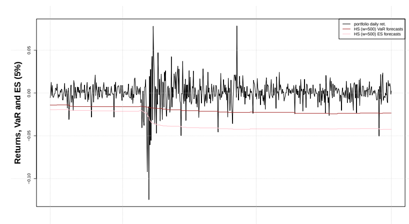

Figure 28. HS (w=500)

For the semi-parametric approach, we estimated *VaR* employing Q-L-ARCH model, Q-L-ARCH-MIDAS model and three CAViaR specifications: SAV, AS and IG. We specified $q = 5$ for both Q-L-ARCH and Q-L-ARCH-MIDAS model. Moreover for the latter we employed $M1$ (money supply in the Eurozone) as macroeconomic variable with $K = 6$; data collected monthly from *FRED* from the period January 1st 2005 to January 1st 2022. First difference has been applied also to *M1* as it wasn't stationary. It is worth to stress that Q-L-ARCH-MIDAS standard errors of the parameters estimated are not reliable because not based on the bootstrap standard errors, but, in this context, we are not interested in the standard errors, we will simply evaluate *VaR* and *ES* estimates by other procedures. While, in estimating the parameters of the the CAViaRs, we have considered a set of 100 initial values (R = 100) and a bootstrap replication equal to 1000 for the AS and equal to 5000 for the SAV and IG in order to obtain the standard errors. Finally *ES* estimates are obtained jointly by employing the *Multiple of VaR* model and maximizing the AL likelihood. Here we report the plots:

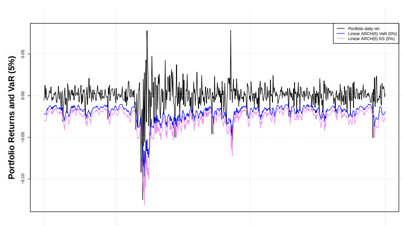

Figure 29. Q-L-ARCH

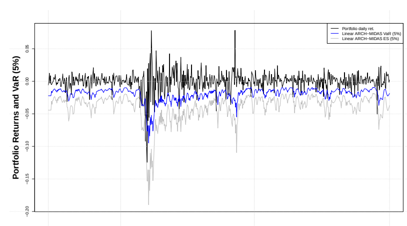

Figure 30. Q-L-ARCH-MIDAS

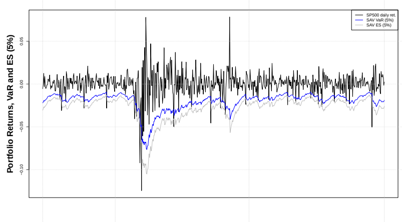

Figure 31. SAV

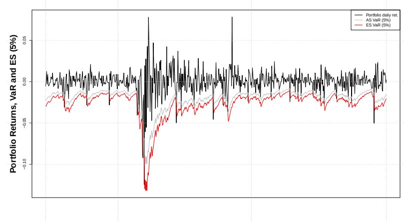

Figure 32. AS

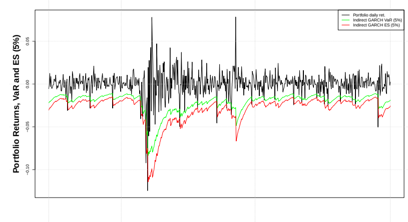

Figure 33. IG

Once the models delivered the estimates we move to evaluate the performance of such models by backtesting and *MCS* procedures.

After having estimated the vectors of *VaR* and *ES* for each of the 10 procedures considered, these are then compared to the actual values through the backtesting procedure. Procedure that applies 4 main tests for the *VaR*: Actual over Expected ratio, Unconditional Coverage test, Conditional Coverage test and the Dynamic Quantile test; and the McNeil & Frey (2000) test for the *ES*, accounting for a number of bootstrap equal to 2000. Considering their p-values, we reject the null at 5% for the *VaR* and *ES* at 95% confidence level. Below we report the results:

Table 10. Backtesting *VaR*

	AE	UC	CC	DQ
GARCH-N	1.036	0.835	0.624	0.616
GARCH-t	1.099	0.573	0.637	0.511
HS (w=100)	1.162	0.361	0.042	0.000
HS (w=250)	1.036	0.835	0.085	0.000
HS (w=500)	1.099	0.573	0.007	0.000
SAV	1.068	0.699	0.300	0.134
AS	1.068	0.699	0.919	0.902
IG	1.005	0.978	0.581	0.540
Lin. ARCH	0.973	0.877	0.888	0.821
Lin. ARCH MIDAS	1.099	0.573	0.852	0.961

Table 11. Backtesting *ES*

	Expected Exc.	Actual Exc.	P-value
GARCH-N	31.000	33.000	0.001
GARCH-t	31.000	35.000	0.083
HS (w=100)	31.000	37.000	0.217
HS (w=250)	31.000	33.000	0.138
HS (w=500)	31.000	35.000	0.099
SAV	31.000	34.000	0.006
AS	31.000	34.000	0.004
IG	31.000	32.000	0.006
Lin. ARCH	31.000	31.000	0.002
Lin. ARCH MIDAS	31.000	35.000	0.999

Examining the results of table 10, for the *AE* we have that almost all models, excluding the HS (w=100), are really close to the value 1, implying that they can be seen as a good estimate of the *VaR*; where the majority among them underestimate the risk, while only the Q-L-ARCH overestimates it. Focusing now on the 3 tests, we can see that only the HS models reject the null hypothesis of correct number of violations and independence among the exceedances. So, in conclusion, the best estimators of the *VaR* are the AS, IG and Q-L-ARCH, while the worst estimators are the HS models.
Taking a glance at table 11, we have quite the opposite situation, where almost all of the models reject the null hypothesis. The models that performed the worst in the *VaR* backtesting here performed relatively well, in particular HS (w=100). With respect to the *VaR* backtesting, we have that the best model is the Q-L-ARCH-MIDAS, while the worst are the GARCH-N, AS and Q-L-ARCH.
Finally, we employed the MCS procedure again to determine the 2 SSM: the first one evaluates the models for the Value at Risk, by employing a quantile loss function (Table 12); and the second one evaluate jointly *VaR* and *ES* by using the Fissler and Ziegel loss function (Table 13). On the right of the tables, in the third column, we indicated which models were included in the superior set of models, while the second column represent the averages of the loss function.

Table 12. MCS VaR

Models	QL	SSM
GARCH-N	0.17	included
GARCH-t	0.171	included
HS (w=100)	0.193	
HS (w=250)	0.204	
HS (w=500)	0.205	
SAV	0.174	included
AS	0.165	included
IG	0.175	
Lin. ARCH	0.169	included
Lin. ARCH MIDAS	0.168	included

Table 13. MCS VaR and ES

Models	FZ	SSM
GARCH-N	-3.421	included
GARCH-t	-3.421	included
HS (w=100)	-3.3	included
HS (w=250)	-3.156	
HS (w=500)	-3.077	
SAV	-3.371	
AS	-3.453	included
IG	-3.383	included
Lin. ARCH	-3.442	included
Lin. ARCH MIDAS	-3.448	included

Results from table 12 and 13 shows that HS models are excluded from the MCS *VaR*, these worst performances are in line with the consideration made in the back-tests. HS (w=250) and HS (w=500) were excluded also in the joint *VAR* and *ES* MCS, for this latter procedure we can observe that also the SAV is excluded from the superior set of models.

4. Conclusion

As in the survey on multivariate models by Bauwens et al. (2006), we reviewed some of the most important models in the multivariate framework, by comparing 34 different multivariate volatility models. The conclusion are that *Moving Covariance* model with $V = 22$, despite its simplistic assumptions, outperformed the DCCs and other more sophisticated models, even if MCS procedure selection did not perform at the most as we expected because no models were rejected from the Superior Set of Models. The main contributions regarding multivariate models are: Engle and Kroner (1995) for BEKK models; Carol, A. (2011) for the OGARCH; Engle (2002), Aielli (2013) and Colacito et al. (2011) for the DCC models.
The evaluation of the models have been made following the MCS procedure by Hansen et. al (2011) and using robust loss functions, which are invariant to the noise in the proxy, as demonstrated by Patton (2011) and by Laurent et al. (2013).
The economic evaluation of the *Moving Covariance* model was based on the estimation of the global minimum variance portfolio (GMVP), and then, on the analysis of the resulting Value-at-Risk (*VaR*) and Expected Shortfall (*ES*). So, in the first step, we calculated the weights of the GMVP in order to derived the portfolio returns and volatility using this weights, and then we estimated the risk measures, such as *VaR* and *ES*, employing 10 different models. In particular, we used both the parametric, non-parametric and semi-parametric approaches to calculate the *VaR* and *ES* of the Global minimum variance portfolio. The most relevant contributions regarding this part were: Taylor, J. W. (2019) for the AL distribution; Corazza et al. (2020) for Q-L-ARCH e Q-L-ARCH-MIDAS; Engle and Manganelli (2004) for CAViaR models.
Finally, for the evaluation of these estimates, we employed backtesting and MCS procedures, following the example of Rocciolett (2015) for backtesting the *VaR* and of McNeil & Frey (2000) for backtesting the *ES*. For the MCS procedure, we employed two different loss functions: the quantile loss function by González-Rivera et al. (2004), which penalize more heavily negative returns which overcome *VaR*, and the Fissler and Ziegel loss function, which allows to jointly evaluate *VaR* and *ES*, as shown in their contributions: Fissler and Ziegel (2016), Fissler et al. (2015).
The results from backtesting and MCS are coherent: overall, the *Quantile-Linear ARCH-MIDAS* model is the one that gave us the best estimates according to both the two validation procedures. On the other hand, as expected, the worst performing model in both the backtesting and MCS procedures is the *Historical Simulation* with any window length, given its simplistic estimation structure relative to other models.
Future studies could include or a broader range of assets or other portfolio strategies, as well as different rolling forecasting techniques. In general, despite the results of this work, multivariate volatility modeling is a field in continuous evolution and new models could be employed as the theory improves.

5. References

[1] Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. Quantitative Finance vol. 1, issue 2, pag. 223-236

[2] Schlitzer, G. (1995). Testing the stationarity of economic time series: further Monte Carlo evidence. Ricerche Economiche vol. 49, issue 2, pag. 125-144.

[3] Carol, A. (2011). A Primer on the Orthogonal GARCH model.

[4] Aielli, G. P. (2013). Dynamic conditional correlation: on properties and estimation. Journal of Business Economic Statistics, 31(3), 282-299.

[5] 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.

[6] Colacito, R., Engle, R. F., Ghysels, E. (2011). A component model for dynamic correlations. Journal of Econometrics, 164(1), 45-59.

[7] Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. Journal of Business Economic Statistics, 20(3), 339-350.

[8] Berkowitz, J., Christoffersen, P., Pelletier, D. (2011). Evaluating value-at-risk models with desk-level data. Management Science, 57(12), 2213-2227.

[9] Laporta, A. G., Merlo, L., Petrella, L. (2018). Selection of value at risk models for energy commodities. Energy Economics, 74, 628-643.

[10] Serrano Bautista, R., Nunez Mora, J. A. (2021). Value-at-risk predictive performance: a comparison between the CaViaR and GARCH models for the MILA and ASEAN-5 stock markets. Journal of Economics, Finance and Administrative Science, 26(52), 197-221.

[10] Engle, R. F., Manganelli, S. (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles. Journal of business economic statistics, 22(4), 367-381.

[11] Sarma, M., Thomas, S., Shah, A. (2003). Selection of Value-at-Risk models. Journal of Forecasting, 22(4), 337-358.

[12] Roccioletti, S. (2015). Backtesting value at risk and expected shortfall. Springer.

[13] Taylor, J. W. (2019). Forecasting Value at Risk and Expected Shortfall Using a Semiparametric Approach Based on the Asymmetric Laplace Distribution. Journal of Business Economic Statistics 37(1), 121-133.

[14] Corazza M., Grilli M, Perna C et al. (2020) Mathematical and Statistical Methods for Actuarial Sciences and Finance - Springer

[15] Kratz, M., Lok, Y. H., McNeil, A. J. (2018). Multinomial VaR backtests: A simple implicit approach to backtesting expected shortfall. Journal of Banking Finance, 88, 393-407.

[16] McNeil, A. J., Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. Journal of empirical finance, 7(3-4), 271-300.

[17] Billio, M., Caporin, M., Gobbo, M. (2006). Flexible dynamic conditional correlation multivariate garch models for asset allocation. Applied Financial Economics Letters, 2(02), 123-130.

[18] Otranto, E. (2010). Identifying financial time series with similar dynamic conditional correlation. Computational Statistics Data Analysis, 54(1), 1-15.

[19] Yamai, Y., Yoshiba, T. (2005). Value-at-risk versus expected shortfall: A practical perspective. Journal of Banking Finance, 29(4), 997-1015.

[20] Chinhamu, K., Huang, C. K., Huang, C. S., Chikobvu, D. (2015). Extreme risk, value-at-risk and expected shortfall in the gold market. International Business Economics Research Journal (IBER), 14(1), 107-122.

[21] Engle, R. F. and F. K. Kroner (1995). Multivariate simultaneous generalized ARCH. Econometric Theory 11, 122–150

[22]Bauwens, Laurent and Rombouts, Journal of Applied Econometrics 2006 - Multivariate Garch Models: A Survey

[23] Hansen et. al. 2011, “Model Confidence Set” p.2

[24] Laurent et al. On loss functions and ranking forecasting performances of multivariate volatility models, 2013

[25] Markowitz, 1959 - Portfolio selection, efficient diversification of investment

[26] Engle, R. F. and F. K. Kroner (1995). Multivariate simultaneous generalized ARCH. Econometric Theory 11, 122–150.

[27] Eric Ghysels , Arthur Sinko Rossen Valkanov, ”MIDAS Regressions: Further Results and New Directions” 2007

[28] Conrad, ”On the macroeconomic determinants of long-term volatilities and correlations in U.S. stock and crude oil markets” 2014

[29] G. William Schwert, ”Why Does Stock Market Volatility Change Over Time?” 1989

[30] R. R. Officer, ”The Variability of the Market Factor of the New York Stock Exchange” 1973

[31] Fissler, T., Ziegel, J.F., (2016). ”Higher order elicitability and Osband’s principle.” The Annals of Statistics 44, 1680-1707.

[32] Fissler, T., Ziegel, J.F., Tilmann, G. (2015). ”Expected Shortfall is jointly elicitable with Value at Risk - Implications for backtesting.”

[33] Patton, A. J., Ziegel, J.F., Chen, R. (2017). ”Dynamic semiparametric models for expected shortfall (and Value-at-Risk).”