METODI STATISTICI PER L' ECONOMIA E L'IMPRESA

SCUOLA DOTTORALE / DOTTORATO DI RICERCA IN

XXVI

CICLO DEL CORSO DI DOTTORATO

MULTIVARIATE STATISTICAL ANALYSIS FOR PORTFOLIO SELECTION

Titolo della tesi

ANDREA PIERINI

Nome e Cognome del dottorando

firma

ALESSIA NACCARATO

Docente Guida/Tutor: Prof.

firma

MARCO REALE

CO-TUTOR: PROF.

FIRMA

JULIA MORTERA

Coordinatore: Prof.

firma

Contents

List of Tables

List of Figures

1. CVAR - BEKK - CC

1.1 Introduction

The portfolio selection problem is discussed through the introduction of tree new types of models, called respectively CVAR-BEKK-CC, Factor-BEKK, multiple bi-dimensional CVAR-BEKK, that gradually increase the complexity of the solution while remaining computationally feasible.

The use of bivariate cointegrated vector autoregressive models and Baba-Engle-Kraft-Kroner models (Engle et al. 1995), is proposed for the selection of a stock portfolio (Markowitz type portfolio) based on estimates of average returns on shares and the volatility of share prices. The model put forward envisages the use of explicative variables. This article employs the intrinsic value of shares as a variable, which will make it possible to take the theory of value into account. The model put forward is applied to a series of data regarding the prices of 150 shares traded on the Italian stock market.

The selection of a stock portfolio is broadly discussed in the literature, generally with reference to heteroskedastic regression models (Bollerslev *et al.*, 1994). The model used in the case of multiple time series is of the vector autoregressive (VAR) type and rests on the predictability of the average return on shares (Brown and Reily, 2008, Hamilton, 1994).

In particular this paper suggests the use of cointegrated vector autoregressive models (CVAR) and Baba-Engle-Kraft-Kroner models (BEKK) for the selection of a stock portfolio. In other words, it addresses the problem of estimating average returns and the associated risk on the basis of the prices of a certain number of shares over time. This estimate is then used to identify the assets offering the best performance and hence constituting the best investments. While Campbell *et al.* (2003) proposes the use of a VAR(1) model, it is suggested here that use should be made of VEC models, which make it possible to take into account any cointegration between the series employed and the market trend as measured by means of the Thomson Reuters *Datastream Global Equity Italy Index* (Datastream 2008).

Moreover, while Bollerslev, Engle and Wooldridge (1988) employ diagonal vectorization (DVEC) models to estimate share volatility, the use of a BEKK model, as proposed here, makes it possible to extend the estimation procedure based on DVEC models so as to take into account also the correlation between the volatility of the series and the volatility of the market trend.

The series considered regard the Italian stock market (BIT), and specifically the monthly figures for the top 150 shares in terms of capitalization, from 1 January 1975 to 31 August 2011.The estimation procedure proposed for portfolio selection involves two steps.

In the first step, a two-dimensional CVAR model is developed for all of the 150 shares considered in order to obtain an estimate of the average stock market return. A BEKK model is then applied to the series of residuals thus obtained in order to estimate the volatility of the series. The BEKK model appears particularly suitable because it does not entail the condition of normality for the accidental componentof the model (Hamilton, 1994).

The second step regards the selection of shares for inclusion in the portfolio. Only those identified as presenting positive average returns during the first phase are considered eligible. For the purpose of selecting the most suitable of these, a new endogenous variable is constructed as the product of two further elements, namely the price-to-earnings ratio (P/E) and earnings per share (EPS). This variable, which indicates the "intrinsic value" of the share in question, is not constructed for the entire set of 150 shares but only for those presenting positive average returns in the first phase, as it would be pointless in the case of negative returns.

The CVAR-BEKK model is applied once again to this new series in order to estimate the intrinsic value of the shares, and the top 10, (Evans and Archer, 1968), are selected for inclusion in the portfolio on the basis of the difference between this intrinsic value and the price estimated in the first phase (Brown and Reily, 2008).

A quadratic programming model is then employed to determine the quantities to be bought of each of the 10 shares selected.

It should be noted that the variable P/E · EPS is estimated for each industrial sector (Nicholson, 1960).

2. The Cointegrated Vector Autoregressive Models

2.1 Models description

A concise outline is now given of the phases involved in the selection of shares for inclusion in the portfolio as well as the quantity of shares to be bought for each type selected. The starting point is the $K = 150$ series, regarding the average returns $R_{k,t}$ on the shares, and the average return of the market$R_{M,t}$, $t = t_k, \dots, T$, $k = 1, \dots, K$. It should be noted in this connection that the length of the series considered is not homogeneous because not all of the joint-stock companies are quoted as from the same point in time. This aspect involves further complications in the estimation procedure.

Step one : For each series, the model *CVAR(p)* is defined for the random vector $y_t = [y_{1,t}, y_{2,t}]' = [R_{k,t}, R_{M,t}]'$ as:

$$y_t = \mu_t + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + u_t \tag{1}$$

with $\mu_t = \mu_0 + \mu_1 \cdot t$, A_i $(i = 1, \dots p)$ is a (2×2) of the unknown coefficients, and $u_t = [u_{1,t}, u_{2,t}]'$ is the vector of errors such that $u_t \sim N(0, \Sigma_u)$.

Model (1) can be rewritten as follows to take into account a possible cointegration of the variables considered:

$$\Delta y_t = \mu_t + \Pi y_t + \Gamma_1 \Delta y_{t-1} + \Gamma_2 \Delta y_{t-2} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t \tag{2}$$

where

$$\Gamma_i = -(A_{i+1} + \cdots + A_p) \tag{3}$$

and

$$\Pi = -(I - A_1 - \cdots - A_p) \tag{4}$$

2.2 Specification , test and forecast

It has to be noticed that if Π is singular, $y_1 = [y_{1,1}, \dots , y_{1,T}]'$ and $y_2 = [y_{2,1}, \dots, y_{2,T}]'$ are cointegrated (Johansen, 1995, Lutkepohl, 2007).

In specifying the CVAR model the lag order and the cointegration rank have to be determined. We start by determining a suitable lag length because, in choosing the lag order, the cointegration rank does not have to be known.

The AIC criterion is used to estimate the *lag* $\hat{p}$, with reference to model (1):

$$\hat{p} = \arg\min_m C(m) = \min_m \left\{ \ln\left(\det\left(\tilde{\Sigma}_u(m)\right) + \frac{mc_T}{T} : m = 0,1,\dots, p_{\max} \right) \right\} \qquad (5)$$

where $\tilde{\Sigma}_u(m)$ is the maximum likelihood estimate (MLE) of $\Sigma_u(m)$ for a *VAR(m)* of type (1) with a sample of breadth $T - t_k + m$ and m values of initialization,with $c_T = 8,\ p_{\max} = 10$.

Notice that, while we have considered specifying the VAR order *p*, the criterion is also applicable for choosing the number of lagged differences in a VEC model (2) because *p-1* lagged differences in a VEC correspond to a VAR order *p* (Lutkepohl, 2007).

In practice it is common to use statistical tests in specifying the cointegration rank. In this framework to ascertain the presence of cointegration in the model (2), the likelihood ratio test (LR) is used:

$$LR(r_0) = -T \sum_{j=r_0+1} \log\left(1 - \lambda_j\right) \qquad (6)$$

where: $r_0 = rk(\Pi) = 0, 1$ and λ_j are the eigenvalues of the matrix $S_{11}^{-1/2} S_{10} S_{00}^{-1} S_{01} S_{11}^{-1/2}$ with $S_{00} = \frac{1}{T-p} \sum_{t=p+1}^{T} \hat{u}_t \hat{u}_t'$, $S_{01} = \frac{1}{T-p} \sum_{t=p+1}^{T} \hat{u}_t \hat{v}_t'$, $S_{11} = \frac{1}{T-p} \sum_{t=p+1}^{T} \hat{v}_t \hat{v}_t'$.

The quantities $\hat{u}_t$, $\hat{v}_t$ are the residuals of the regressions of Δy_t and y_{t-1} estimated by maximum likelihood (Johansen, 1995).

However, the asymptotic distribution of the LR statistic is nonstandard, in particular it is not a chi-squared distribution. In other words, the limiting distribution is a functional of a standard Wiener process. Percentage points of the asymptotic distribution, thus, critical values for the LR test can be generated considering multivariate random walks (Johansen *et al.*, 1990).

Assessment of the presence of cointegration between the series by means of the LR test is followed by estimation of the parameters of the model.The result of the LR test is considered in deciding whether to adopt the model in form (1) or (2). In particular, if the test shows that the rank of matrix Π is equal to 0 (hypothesis of stationarity of Δy_t), then $\Pi = 0$ and the method of maximum likelihood is applied directly to (2) in order to estimate the parameters μ_0, μ_1 and $\Gamma_1, \ldots, \Gamma_{p-1}$.

If it instead transpires that the rank of Π is equal to 1 (hypothesis of cointegration of y_1 and y_2), then $\Pi = \alpha\beta$. In this case, it is necessary to estimate model (2) in two stages. First, an MLE of β is obtained by concentrating the log-likelihood with respect to β. Second, this estimate is inserted into (2) in order to obtain the MLE of the other parameters (Johansen, 1995).

If $rk(\Pi) = 2$, the method of maximum likelihoodis applied directly to (1) in order to obtain estimates of the parameters μ_0 , μ_1 and $A_1 \ldots, A_p$.

The portmanteau test is used to ascertain the presence of correlation of residuals, the generalized Lomnicki-Jarque-Bera test for the normality of residuals, and the ARCH test to determine heteroskedasticity

In order to forecast it is convenient in this framework to use the levels VAR representation. Therefore we consider the model type (1) with integrated and possibly cointegrated variables replacing the coefficients with their estimates as calculated before:

$$\hat{y}_{T+1|T} = \hat{\mu}_{T+1} + \hat{A}_1 \hat{y}_{T|T} + \cdots + \hat{A}_p \hat{y}_{T+1-p|T} \quad (7)$$

3. The BEKK model for heteroskedasticity

3.1 BEKK model description

In the event of the latter test revealing the presence of heteroskedasticity, the *BEKK(1,1)* model is used to estimate the conditional variance-covariance matrix

$$\Sigma_{t|t-1} = \mathrm{cov}(u_t | past) = (\sigma_{i,j}(t)) \qquad i,\ j = 1,\ \ldots,\ n$$

which has the following structure:

$$\begin{bmatrix} \sigma_{11,t} & \sigma_{12,t} \\ \sigma_{21,t} & \sigma_{22,t} \end{bmatrix} = \begin{bmatrix} c_{0,11} & 0 \\ c_{0,21} & c_{0,22} \end{bmatrix} \begin{bmatrix} c_{0,11} & c_{0,21} \\ 0 & c_{0,22} \end{bmatrix} +$$

$$+ \begin{bmatrix} c_{1,11} & c_{1,21} \\ c_{1,21} & c_{1,22} \end{bmatrix} \begin{bmatrix} u_{1,t-1}^2 & u_{1,t-1}u_{2,t-1} \\ u_{2,t-1}u_{1,t-1} & u_{2,t-1}^2 \end{bmatrix} \begin{bmatrix} c_{1,11} & c_{1,21} \\ c_{1,21} & c_{1,22} \end{bmatrix} + \qquad (8)$$

$$+ \begin{bmatrix} c_{2,11} & c_{2,21} \\ c_{2,21} & c_{2,22} \end{bmatrix} \begin{bmatrix} \sigma_{11,t-1}^2 & \sigma_{12,t-1} \\ \sigma_{21,t-1} & \sigma_{22,t-1}^2 \end{bmatrix} \begin{bmatrix} c_{2,11} & c_{2,21} \\ c_{2,21} & c_{2,22} \end{bmatrix}$$

The MLE of parameters $c_{k,i,j}$ at time $t = T$ is obtained by maximizing the log-likelihood function:

$$-\ln(2\pi) - \frac{\ln|\Sigma_{t|t-1}|}{2} - \frac{1}{2} u'_t \Sigma_{t|t-1}^{-1} u_t \qquad (9)$$

3.2 ARCH-Test and Markowitz problem estimation

Once the parameters have been estimated, the Generalized Portmanteau test (Hosking, 1980) is applied to ascertain that the model BEKK(1,1) has effectively eliminated the ARCH effects in the residuals of the CVAR model:

$$\widetilde{Q}_m = n^2 \sum_{k=1}^{m} (n-k)^{-1} \hat{g}_k^{'} \left(\hat{G}_0^{-1} \otimes \hat{G}_0^{-1} \right) \hat{g}_k \cong \chi^2_{m-p} \tag{10}$$

where p is the CVAR order, $n = T - t_{\max}$, $t_{\max} = \max\{t_i, t_j\}$, $\hat{g}_k = vec\left(\hat{G}_k\right)$, $\hat{G}_k = L^T C_k L$, $\hat{C}_k = n^{-1} \sum_{t=k+1}^{n} \hat{u}_t \hat{u}_{t-k}^{'}$, $LL^{'} = \hat{C}_0^{-1}$, $m = 1, \ldots, 10$

Step two: The estimates obtained in phase one are used to select the shares for which positive average returns are predicted. For the shares thus selected and for each industrial sector (IS), the model CVAR*(p)*-BEKK*(1,1)* is estimated for the random vector

$y_t = \left[y_{1,t}, y_{2,t}\right]^{'} = \left[(P/E)(EPS)_{h,t}, (P/E)(EPS)_{IS(h),t}\right]^{'}$, where $h = 1, \ldots, H$ is the index that identifies only the series with positive returns selected out of the initial 150.

On the basis of the $(P/E)(EPS)_{h,T+1}$ and $R_{h,T+1}$ forecasts obtained in phase two, the shares are listed for each industrial sector in decreasing order with respect to the values of the difference between *intrinsic value* and *expected price*. The first $n = 10$ shares are thus selected to make up the portfolio.

It should be stressed that the choice of $n = 10$ is made on the basis of the assertion of Evans and Archer (1968) that this quantity is sufficient for diversification of portfolio choices. In actual fact, the number of shares selected will vary in further developments of this work.

In order to determine the quantities to be bought of each of the 10 shares selected, it is necessary to solve the Markowitz problem (Markowitz, 1952) by estimating the matrix of share volatility. To this end, let $\hat{V}_t$ be the estimator of the matrix $n \times n$ of volatility V_t for $t = T + 1$, the elements of which are $v_{i,j}(t) = \text{cov}(R_t | past)$, $i, j = 1, \ldots, n$

The elements of $\hat{V}_t$ are given by:

$$\hat{v}_{i,j}(T+1) = \begin{cases} \hat{\sigma}^{bekk}_{ii,T+1|T} & se\ i = j \\ \hat{c}_{ij} & se\ i \neq j \end{cases} \tag{11}$$

with $\hat{C} = \left(\hat{c}_{i,j}\right)_{i,\, j=1,\ldots,n} = \sum_{t=t_{\max}}^{T} \frac{\left(R_{i,t} - \tilde{R}_i\right)\left(R_{j,t} - \tilde{R}_j\right)}{T - t_{\max}}$, $t_{\max} = \max\left\{t_i,\, t_j\right\}$,
$\hat{\sigma}^{bekk}_{ii,T+1|T}$ and $\hat{\sigma}^{(i)}_{11,T+1}$ in (8), $i,\ j = 1, \ldots, n,\ n = 10$.

On the basis of (11), the solution of the following quadratic problem of Markovitz type, called *global minimum variance portfolio*, for the future time $T+1$ can be obtained with the approximation given by the dual method (Goldfarb, 1983, Higham, 2002):

$$\min_{\varpi \in \mathbf{R}} \left\{ \varpi' \hat{\mathrm{V}}_{\mathrm{T+1}} \varpi : \varpi' \underline{1} = 1,\ \varpi \geq 0 \right\} \tag{12}$$

To obtain a better diversification we have also to find the solution of the quadratic problem of Markovitz type (12) without the constraint $\varpi \geq 0$, for the future time $T+1$ using the explicit solution

$$\hat{\varpi}_{opt,T+1} = \hat{V}^{-1}_{T+1} \mathbf{1} \big/ \left(\mathbf{1}' \hat{V}^{-1}_{T+1} \mathbf{1} \right) \tag{13}$$

Then we put to zero the $\varpi_i < 0$ and reprportionate the remaining ϖ_i, $i = 1, \ldots, n$.

We omit the constraint of a fixed value for the expected return to eliminate the sensitiveness of allocation optimization to errors in predicted returns (Hlouskova *et al*., 2002).

In cases where the matrix $\hat{V}_{T+1}$ proves positive neither in (12) nor in (13), we approximate it with the closest matrixin the sense of Frobenius possessing the same diagonal given by the elements estimated with the BEKK model.

4. CVAR - BEKK - CC Financial application

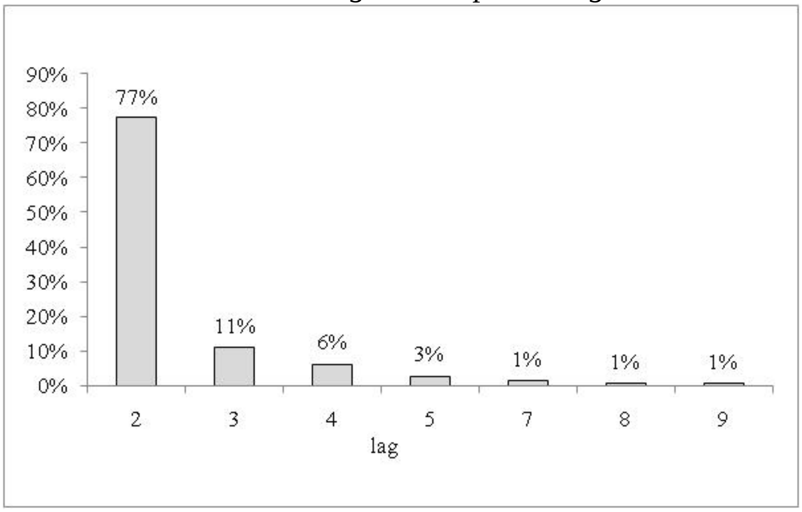
4.1 Results

Application of the model proposed in this work to the monthly figures for the 150 BIT shares with the highest level of capitalization indicates the following results.

On the basis of the minimum AIC, the optimal lag *p* is 2–9 months. Figure 1 shows the empirical distribution of the lag. In particular, while the optimal lag is 2 months for 76% of the entire set of 150 shares it is 2 months for 90% of the shares in the portfolio and 1 month for the remaining 10%. This means that at least 2 months of observations are sufficient to predict the average returns on the vast majority of the shares considered, instead of a random walk model.

It was therefore decided to regard lag 9 as the maximum.

Figure 1: Optimal lag



The results of the LR test for all of the shares considered, yields a degree of cointegration that proves equal to 2 for 91% of the 150 shares, 1 for 7% and 0 for the remaining 2%.

This means that in 7% of cases,when the rank of matrix Π in (2) is equal to 1,use was made of the VEC model, which proves to be estimable (Johansen, 1995) and stationary, unlike the VAR model, which instead proves neither directly estimable nor stationary.

The coefficients of the model estimated in both steps of the procedure prove significant for almost all of the series considered. Table 1 shows, for example, the values of the coefficients estimated for models (1) and (8) for time series number 159.

Table 1: Parameters for models (1) and (8) – stock n°159 selected for the Portfolio.

stock 159					
	VAR		VAR		BEKK
$\mu_{0,1}$	0,014	$a_{5,1,1}$	-1,35	$c_{0,1,1}$	0,037
$\mu_{0,2}$	0,005	$a_{5,2,1}$	-0,604	$c_{0,2,1}$	0
$\mu_{1,1}$	0	$a_{5,1,2}$	0,286	$c_{0,1,2}$	0,023
$\mu_{1,2}$	0	$a_{5,2,2}$	-0,026	$c_{0,2,2}$	0,029
$a_{1,1,1}$	-1,353	$a_{6,1,1}$	-0,935	$c_{1,1,1}$	0,497
$a_{1,2,1}$	-0,472	$a_{6,2,1}$	0,138	$c_{1,2,1}$	-0,107
$a_{1,1,2}$	0,376	$a_{6,1,2}$	0,226	$c_{1,1,2}$	0,122
$a_{1,2,2}$	-0,213	$a_{6,2,2}$	-0,143	$c_{1,2,2}$	0,481
$a_{2,1,1}$	-1,368	$a_{7,1,1}$	-1,108	$c_{2,1,1}$	-0,015
$a_{2,2,1}$	0,023	$a_{7,2,1}$	-0,564	$c_{2,2,1}$	0,001
$a_{2,1,2}$	0,283	$a_{7,1,2}$	0,171	$c_{2,1,2}$	0,033
$a_{2,2,2}$	-0,517	$a_{7,2,2}$	-0,287	$c_{2,2,2}$	-0,001
$a_{3,1,1}$	-1,59	$a_{8,1,1}$	-0,902		
$a_{3,2,1}$	-0,836	$a_{8,2,1}$	-0,692		
$a_{3,1,2}$	0,549	$a_{8,1,2}$	0,391		
$a_{3,2,2}$	0,374	$a_{8,2,2}$	0,265		
$a_{4,1,1}$	-1,522	$a_{9,1,1}$	-0,37		
$a_{4,2,1}$	-0,516	$a_{9,2,1}$	-0,458		
$a_{4,1,2}$	0,203	$a_{9,1,2}$	0,062		
$a_{4,2,2}$	0,162	$a_{9,2,2}$	0,075		

The p value of F statistic is below 0.10 in 83% of cases and below 0.2 in 86%. The model thus proves 90% significant for nearly all of the series. Table 2 shows diagnostic tests for models (1) and (8). The results of the portmanteau Test on the residuals of the CVAR and BEKK models are given in Table 2 (here too, for brevity, they regard only the ten series selected for the portfolio). As regards the presence of ARCH effects in the residuals of the CVAR model (column 5 of Table 2: pAR), the test leads to the conclusion that the hypothesis of the presence of a heteroskedastic component in the model estimated at a confidence level of 99% cannot be ruled out for 60% of the shares chosen. The results of the Hosking test carried out in order to ascertain the presence of ARCH effects in the residuals of the BEKK model, which are shown in column 10 of Table 2 (pHq), lead to the conclusion that the hypothesis of the absence of any heteroskedastic component in the model is to be accepted in 90% of cases. In other words, it can be concluded that the combination of the CVAR and BEKK models picks up the heteroskedastic component and includesthe informationderived from the same in the procedure.

Column 9 (pH) of Table 2 shows the results of the portmanteau test carried out in order to ascertain the presence of autocorrelation of residuals in the BEKK model. They suggest that the hypothesis of autocorrelation can be ruled out and that the maximum lag considered is sufficient.

Table 2: Test statistics - stocks selected for the portfolio

Series	R2.1	pFvalue	pSC	pAR	pJB
4	0,114	0	0,187	0	0
34	0,085	0,006	0,717	0,475	0
63	0,078	0,08	0,785	0	0
130	0,014	0,894	0,347	0	0
49	0,052	0,007	0,664	0,002	0
109	0,041	0,736	0,502	0	0
159	0,725	0,02	0,049	0,458	0,847
85	0,179	0	0,2	0,056	0,001
54	0,12	0,01	0,838	0,131	0,025
60	0,113	0	0,423	0	0
Series	pSK	pKU	pH	pHq	intervalYN
4	0	0	0,142	0,556	true
34	0,127	0	0,233	0,606	true
63	0,002	0	0,822	0,004	true
130	0,054	0	0,303	0,566	true
49	0,015	0	0,752	0,344	false
109	0	0	0,887	0,266	true
159	0,905	0,553	0,176	0,864	true
85	0,161	0	0,459	0,103	true
54	0,415	0,009	0,924	0,793	true
60	0,108	0	0,877	0,042	true

R2.1= goodness of fit stock equation1;
pFvalue = P(F>Foss|H0: A_i=0 ,$i=1,\ldots,p$) zero coefficients of CVAR;
pSC =P $(\chi^2>\chi^2_{oss}$ | H_0 : $E(u_t\, u_{t-i}') = 0,\ i=1,..,h>p)$ autocorrelations in VAR residuals test;
pAR =P $(\chi^2>\chi^2_{oss}$ | H_0 : no *ARCH*) ARCH in CVAR residuals test;
pJB =P $(\chi^2>\chi^2_{oss}$ | H_0 : normality) CVAR normality test ;
pSK = $P(\chi^2>\chi^2_{oss}$ | H_0 : $E(u_t^3)=0$) CVAR skewness test;
pKU=$P(\chi^2>\chi^2_{oss}$ |H_0 : $E(u_t^4)=3$) CVAR kurtosis test;
pH =P $(\chi^2>\chi^2_{oss}$ | H_0 : *no autocorrelations*) autocorrelations in BEKK residuals test;
pHq =P $(\chi^2>\chi^2_{oss}$ | H_0 : *no ARCH*) ARCH in BEKK residuals test;

A OLS-based CUSUM test for stability of the market index was also carried out, and the results suggest that the hypothesis of stability of the series over the period considered is acceptable.

The BEKK estimate of *volatility* for each share is between 0.001 and 0.01 for 93% of the series and never above 0.031.

It can be concluded in the light of this result that for most of the series considered, the estimated value of the share does not differ from its real value at a confidence level of 95%by more than ±0.2.

In actual fact, the value at risk calculation put forward by J.P.Morgan (Longerstaey *et al*., 1995, Duffie *et al*., 1997). could be used in order to include the information deriving from the presence of correlation between the series considered and hence to assess the overall risk rather than the risk of the individual share.

Further confirmation of the adequacy of the CVAR-BEKK model with respect to the series observed was sought before selecting the shares to be included in the portfolio. Specifically, the *confidence interval* at the level of significance of 95% contains the *actual* value $_{T+1}$ in 94 % of the series. The CVAR-BEKK model can therefore be considered reliable for most of the series for the purposes of prediction.

The next step after verification of the suitability of the model was prediction of the prices of the shares as well as their intrinsic values. Table 3 shows the values predicted on the basis of model (7), once again restricted to the ten shares selected for the sake of brevity.

Table 3: Forecast return, volatility, intrinsic value, potential value – stocks included in portfolio

Series	pf	varf	vinf	indm	vin_p
4	0,008	0,082	3,622	10	3,614
63	0,001	0,044	0,732	8	0,731
130	0,011	0,063	0,611	8	0,6
49	0,009	0,047	0,515	4	0,506
109	0,012	0,044	0,229	3	0,217
159	0,019	0,039	0,193	6	0,174
85	0,031	0,09	0,204	3	0,173
54	0,007	0,046	0,109	6	0,102
60	0,006	0,079	0,05	10	0,044
34	0,009	0,081	0,037	2	0,028

pf= forecasted return ;
varf= forecasted volatility
vinf= forecasted intrinsic value;
indm= sector index;
vin_p= forecasted intrinsic value – price.

On the basis of potentials, understood as the difference between share price and intrinsic value, the ten shares with the highest potential returns were then selected.

Two criteria of ranking were used, namely the partial criterion and the total criterion.

The former involves arranging the values of potential of all the shares considered in decreasing order for every industrial sector (Goodman and Peavy III, 1983) and selecting the first share in each.

In the latter, the values of potential of all the shares considered are arranged in decreasing order regardless of industrial sector and the first ten are chosen.

The use of the partial criterion is connected with the relationship between P/E and share performance manifested most strongly in each industrial sector. As Goodman and Peavy III write,“*firms in the same industry tend to cluster in the same relative P/E ranking, detected return differences between P/E groups may be attributable to industry performances rather than P/E level. This bias is eliminated by using P/E relative to its industry.*”

The total criterion has the advantage that the selection is unconnected with the industrial sector of the share in question and therefore not necessarily influenced by the possibly negative trend of individual sectors. Its application thus means selection of the ten best shares in absolute terms (Nicholson**,** 1960).

The choice between these two criteria of ranking is obviously subjective in that it depends on the opinions of investors. This paper considers both, which evidently lead to the selection of different portfolios.

Table 4 lists the optimal allocations, i.e. the solutions of the problems of optimization (12) and (13); column 4 shows the results for the best portfolio. It has a *monthly average return* of 1.9%, a *monthly standard deviation* of 0.655, and a *Sharpe index* of 0.029.

Table 4: portfolios weights, volatility, return, Sharpe

Total Ranking			Partial Ranking		
Series	w(i) prop.	w(i) opt	w(i) opt	w(i) prop.	Series
4	0,462	0	0	0,299	142
63	0	0,581	0	0	34
130	0	0	0	0,225	109
49	0,003	0	0	0	49
109	0,216	0	1	0,344	159
159	0,204	0	0	0,009	65
85	0	0	0	0	63
54	0	0	0	0,123	147
60	0,004	0,419	0	0	4
34	0,112	0	0	0	28
Return	0,011	0,003	0,019	0,011	Return
St. Dev.	1,185	0,77	0,655	0,913	St. Dev.
Sharpe	0,009	0,004	0,029	0,012	Sharpe

4.2 Comments

The selection of a share portfolio has historically constituted a complex problem that has no single solution but depends both on market conditions and on the information available to investors. In other words, the choice of shares to invest in must be based on objective criteria making it possible to assess risk and return without ignoring investors' opinions. To this end, the paper suggests the use of a model for the analysis of multiple historical series with a view to the prediction of share return and associated risk but also taking the indications of the market into account at the same time in the specification of the model itself. Variables obtained as functions of P/E and EPS have thus been used together with the market index as regressors of the combined model (1) and (8).

The innovative choices in the construction of a portfolio selection model put forward here regard two distinct aspects. The first concerns transition from a model of the VAR (1) type for the prediction of a multiple historical series (Campbell, 2003) to one of the CVAR (p) type, which makes it possible to take into consideration any cointegration of the series considered and therefore constitutes an improvement of the information available for estimation purposes. The subsequent use of a combinationof CVAR and BEKK models, which extends the results of Bollerslev, Engle and Wooldrige (2), makes it possible to consider also the temporally variable correlation between the volatility of the series and the volatility of the market index within the estimation procedure.

The second concerns the choice of criterion for the selection of shares,which is addressed here by seeking to insert a typical concept of finance such as intrinsic value into the primarily statistical context of the prediction of a multiple historical series. An intrinsic value estimated by means of the combined CVAR-BEKK model is used to obtain a “potential value” serving as a basis to rankthe different shares and then select the top ten.

The method put forward was applied to the seriesof 150 shares with highest capitalization quoted on the Italian stock exchange and led to the selection of 10 shares constituting a portfolio with an average monthly return of 1.9% and a risk of 0.655.

Comparison of the results of the CVAR-BEKK model put forward here and those obtained by means of models ofthe VAR (1) and DVEC type found in the literature was carried out on the basis of the values of log-likelihood of the models themselves. In other words, since one model could produce a higher value of return than another but nevertheless prove less reliable, it was decided to assess the models’ performance in terms of correspondence to the series observed. The log-likelihood of the CVAR-BEKK model always proves greater than that of the other models, thus indicating more accurate representation of the series observed and hence better predictions.

Further developments of the work will regard the study of “value at risk”, understood as assessment of the greatest loss possible, as well as identification of possible structural breaks of the individual series of share returns with a view to making the model more adaptable.

5. A Factor - BEKK model

5.1 Introduction

As the *CVAR-BEKK-CC* model cannot explain the covariance between pairs of variables, the following *Factor - BEKK* model is used to explain the covariance through the help of two manifest factors, while retaining the computational feasibility of constant covariance.

The use of Factor Model combined with Baba-Engle-Kraft-Kroner (BEKK) model is proposed for the estimation of the volatility for stock portfolio (Markowitz type portfolio) The combined model uses esplicative variables as the instrisic value, which regards the value theory (Brown and Reily, 2008) and the market index (Sharpe 1970). The model put forward is applied to a subset of promising universe among the series of data regarding the prices of the best capitalized 150 shares traded on the Italian stock market (BIT) between 1 January 1975 and 31 August 2011.

The problem that arises in the stock portfolio selection framework is the estimation of the stock volatility, that is to say the estimation of the variance - covariance matrix of the stocks, as through the volatility it is possible to evaluate the risk of investment in the stocks (Markowitz 1952).

The diagonal elements and off - diagonal elements of the variance - covariance matrix are separately estimated: the former through BEKK models while the latter through a type 1 Factor model (Connor 1995).

In particular this paper suggests the use of BEKK model (Engle Kroner 1995) applied to the residuals of the bivariate cointegrated vector autoregressive models (CVAR) (Johansen 1995) for the compound stock values and the market index value (Pierini, Naccarato, 2012) to estimate the diagonal elements of the volatility matrix.

Per la stima degli elementi extra diagonali della matrice di volatilità si propone invece the use of a Factor model of type I (Connor G. 1995) with macroeconomic factor given by the market trend as measured by means of the *Global Equity Italy Index* (Datastream Global Equity Indices, 2008) and fundamental factor given by the mean “intrinsic value”.

The Factor model utilization with respect to the use of sample covariance as an estimation of the off-diagonal elements of the volatility matrix has the following pros: it gives the possibility to estimate with fewer parameters, that's to say using $2n+n+4$ parameters instead of $n(n+1)/2$ required otherwise. In doing so we obtain a more precise result because each parameter is estimted with error and with fewer parameters we accumulate less errors.

Moreover having fewer parameters to estimate is easier to update or add other stocks to the portfolio.

E' comunque il caso di osservare che la stima della matrice di volatilità potrebbe risultare distorta poichè, se le stime degli elementi extradiagonali, derivano da un misspecified factor model, allora esse sono distorte (Ruppert, 2011).

La scelta di utilizzare due differenti procedure di stima per gli elementi della matrice di volatilità è dovuta a problemi di complessità computazionale; sarebbe in effetti auspicabile poter utilizzare il modello BEKK per la stima di tutti gli elementi della matrice di volatilità – poiché esso presenta caratteristiche migliori in termini di stime (Lutkpol new) di altri modelli già proposti in letteratura quali ad esempio modelli DVEC, ARCH, GARCH (Bollerslev, Engle et al. 1994). Tuttavia, al crescere del numero *n* delle azioni incluse nel portafoglio la stima del modello BEKK diviene computazionalmente non risolvibile (lutkepol new); da qui la nostra proposta di utilizzare la combinazione di modelli diversi per la stima della matrice di volatilità.

The series considered regard the Italian stock market (BIT), and specifically the monthly figures for the top 150 shares in terms of capitalization, from 1 January 1975 to 31 August 2011.

Before solving the volatility estimation problem, it is necessary to select the stocks to include in the portfolio and estimate the volatility of each stock during the time span considered. Starting with 150 stocks, in order to select the stocks to include in the portfolio, the procedure in two steps in Pierini, Naccarato, (2012) is uesd. It starts with the estimation of the market value of the stock through a CVAR*(p)* model (Johansen, 1995) and then the estimation – based itself on a *CVAR(p)* model – of its "intrisic value".

The intrinsic value is a new endogenous variable constructed as the product of two elements, namely the price-to-earnings ratio (P/E) and earnings per share (EPS). Si osservi che the "intrinsic value" of the share in question, is not constructed for the entire set of 150 shares but only for those presenting positive average returns in the first phase, as it would be pointless in the case of negative returns.

As far as the single volatilities are concerned – that are the diagonal elements of the volatility matrix – BEKK models are applied.
The n stocks included in the portfolio are then selected using a match between two estimated values, the intrinsic value and the market value.
Only those identified as presenting positive average returns are considered eligible for the portfolio selection.

It is then possible to employ a quadratic programming model to determine the optimal n and the quantities to be bought of each of the n shares selected after estimating the volatility matrix.

5.2 BEKK model Estimates of the volatility matrix diagonal elements

The estimation of the diagonal elements of the volatility matrix is divided in three steps. It is estimated by applying the BEKK model to the *CVAR(p)* residuals of the stocks. However to estimate this model for the stocks it is necessary to have selected the n stocks to include in the portfolio. So in order to obtain the estimation of the diagonal elements of the volatility matrix – in line with this methodology – it is necessary to implement a two step procedure: firstly two *CVAR(p)* models are estimated, the first one applied on the returns and the second one applied on the intrinsic values. Then through a match between the obtained values, the n stocks to include in the portfolio are selected (Pierini, Naccarato, 2012). Lastly the BEKK model is applied to the *CVAR(p)* residuals to estimate the off-diagonal elements of the volatility matrix related to the n selected stocks on the basis of the match between the intrinsic values and returns.

To be noticed that – as the selection of the stocks to include in the portfolio is done on the basis of the match between return and intrinsic value – it is necessary that these two values are comparable.
So the transformation of the intrinsic value in a return intrinsic value is needed. This transformation is defined as the difference between the logarithm of the intrinsic value in the two successive times:

$$R_{k,t} = \ln\left(p_{k,t}\right) - \ln\left(p_{k,t-1}\right) \quad (14)$$

$$(P/E) \cdot (EPS)_{k,t} = \ln\left((p/e) \cdot (eps)_{k,t}\right) - \ln\left((p/e) \cdot (eps)_{k,t-1}\right) \quad (15)$$

In altri termini, il modello CVAR è applicato al rendimento del valore intrinseco e non già al valore intrinseco.

The *CVAR(p)* models are briefly described hereafter as a preliminary step for the estimation of the volatility the diagonal elements.

To estimate the time series of the return , the starting point is the K=150 series, regarding the returns $R_{k,t}$ on the shares, and the average return of the market $R_{M,t}$, $t=t_k,\dots,T,\ k=1,\dots,K$.
For each series the model *CVAR(p)* is considered for the random vector $y_t = [y_{1,t}\ , y_{2,t}]' = [R_{k,t}\ , R_{M,t}]'$ as the equation (1).

To be noticed that the CVAR type model for the estimation of the unknown coefficients in the equation (1) is selected because it can detect the presence of cointegration or integration between the two components of the random vector y_t , for the returns and for the instrisic values too, by considering its alternative reparametrization, called VEC form as the equation (2).

On the same 150 stocks, the bivariate time series of the intrinsic values and the intrinsic sector values is considered.
The sector is intended as the industrial sector to which each stock belogns.

For each series the model *CVAR(p)* like the equation (14) is considered for the random vector $y_t = [y_{1,t}\ , y_{2,t}]' = [\ (P/E)\cdot(EPS)_{h,t}\ ,\ (P/E)\cdot(EPS)_{IS(h),t}]'$, where $h=1,\dots H$ is the index that identifies only the series with positive returns selected out of the initial 150.

On the basis of the $(P/E)\cdot(EPS)_{h,T+1}$ and $R_{h,T+1}$ forecasts obtained in phase 2, the shares are listed in decreasing order with respect to the values of the difference between *intrinsic value* and *expected price*. The first $n \in \{10,11,\dots,H\}$ shares are thus selected to make up the portfolio.

To take into consideration the presence of heteroskedasticity (Pierini Naccarato 2012), the *BEKK(1,1)* model is used to estimate the conditional variance-covariance matrix of u_t given in phase 1 of the equation (1), $\Sigma_{t|t-1} = cov(u_t|past) = ((\sigma_{i,j}(t)))_{i,j=1,\dots,n}$, which has the following structure in the equation (8).

At the time $t=T$, the MLE (maximum likelihood estimations) of the parameters $c_{k,i,j}$, $k=0,1,2,\ i,j=1,2$ are obtained maximixing the function in (9).

5.3 Factor model Estimates of the volatility matrix off-diagonal elements

Moreover the following two Factor model is applied

$$R_{i,t} = \beta_{0,i} + \beta_{1,i}F_{1,t} + \beta_{2,i}F_{2,t} + \varepsilon_{i,t} \text{ , } i=1,\dots,n, \ t=1,\dots,T \qquad (16)$$

where F_1 is the market trend as measured by means of the Thomson Reuters Datastream Global Equity Italy Index and F_2 is the fundamental factor given by the mean "intrinsic value",$\beta_{k,i}$ are parameters and ε_i are uncorrelated mean-zero random variables.

If we define , $\boldsymbol{\beta}_0 = (\beta_{0,1}, \dots, \beta_{0,n})^{\mathrm{T}}$, $\boldsymbol{R}_t = (R_{1,t}, \dots, R_{n,t})^{\mathrm{T}}$, $\boldsymbol{F}_t = (F_{1,t}, F_{2,t})^{\mathrm{T}}$, $\boldsymbol{\varepsilon} = (\varepsilon_{1,t}, \dots, \varepsilon_{n,t})^{\mathrm{T}}$

$$\beta = \begin{bmatrix} \beta_{1,1} & \cdots & \beta_{1,i} & \cdots & \beta_{1,n} \\ & \ddots & & \ddots & \\ \beta_{2,1} & \cdots & \beta_{2,i} & \cdots & \beta_{2,n} \end{bmatrix} \qquad (17)$$

The model (4) can be reexpressed as

$$\boldsymbol{R}_t = \boldsymbol{\beta}_0 + \beta^{\mathrm{T}} \boldsymbol{F}_t + \boldsymbol{\varepsilon}_t \text{ , } t=1,\dots,T \qquad (18)$$

Then, with $\boldsymbol{\beta}_j = (\beta_{1,j}, \beta_{2,j})^{\mathrm{T}}$ and Σ_F the 2×2 covariance matrix of $\boldsymbol{F}_t$

$$d_{i,j} = cov(R_i, R_j) = \boldsymbol{\beta}_i^T \Sigma_F \boldsymbol{\beta}_j \qquad (19)$$

We estimate the $\boldsymbol{\beta}$ with the regression coefficients using (6) and Σ_F with the sample covariance of the factors. By doing so we obtain an estimation $\hat{d}_{i,j}$ of $d_{i,j}$.

In order to determine the quantities to be bought of each of the n shares selected, it to is necessary solve the Markowitz type problem (Markowitz 1952) by estimating the matrix of share volatility. To this end, let $\hat{V}_t$ be the estimator of the matrix $n\times n$ of volatility V_t for $t=T+1$, the elements of which are $v_{i,j}(t)=cov(R_t|past), \ i,j=1,\dots,n$.
We define the elements of $\hat{V}_t$ by:

$$v_{i,j}(T+1) = \begin{cases} \hat{\sigma}^{bekk}_{ii,T+1|T} & if\ i = j \\ \hat{d}_{i,j} & if\ i \neq j \end{cases} \tag{20}$$

with $\hat{\sigma}^{bekk}_{ii,T+1!T}$ $is\ \sigma^{(i)}_{11,T+1}$ in the equation (2) *i,j=1,…,n.*

5.4 Results

Application of the model proposed in this work to the monthly figures for the 150 BIT shares with the highest level of capitalization indicates the following results.

In the following **Table 5** we see the selected stocks with positive forecasted return: the number of this subset is *H= 25*

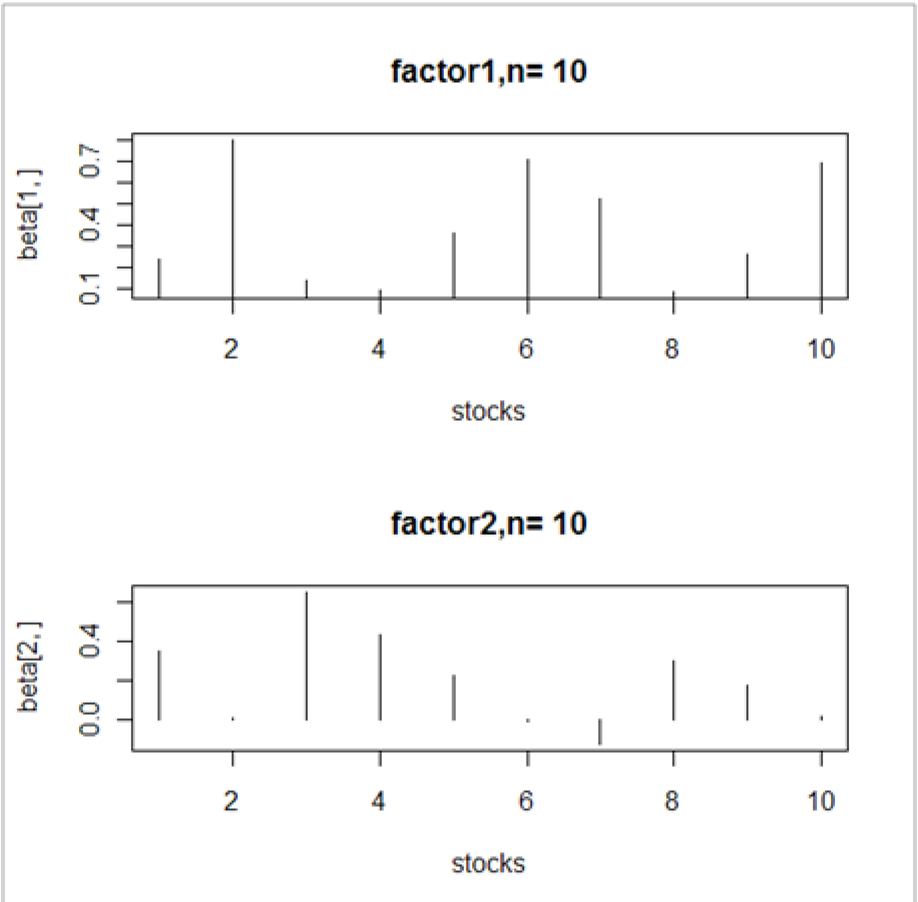
Table 5: Stocks selected for the Portfolio.

id	pf	varpf	vinf	indm	vin_p	var_vinf	var_uf
28	0,04	0,117	0,052	2	0,012	0,236	0,353
113	0,034	0,096	-0,027	4	-0,061	0,162	0,258
85	0,031	0,09	0,204	3	0,173	0,336	0,426
44	0,029	0,085	-0,018	6	-0,047	0,82	0,905
117	0,028	0,063	-0,089	4	-0,117	0,182	0,245
38	0,022	0,071	-0,063	3	-0,085	0,326	0,397
159	0,019	0,039	0,193	6	0,174	0,221	0,26
109	0,012	0,044	0,229	3	0,217	0,237	0,281
130	0,011	0,063	0,611	8	0,6	0,876	0,939
45	0,01	0,103	-0,139	6	-0,149	0,305	0,408
147	0,01	0,047	-0,03	9	-0,04	0,087	0,134
34	0,009	0,081	0,037	2	0,028	3,682	3,763
49	0,009	0,047	0,515	4	0,506	0,865	0,912
4	0,008	0,082	0,943	10	0,935	1,385	1,467
32	0,007	0,081	-0,02	2	-0,027	1,979	2,06
54	0,007	0,046	0,109	6	0,102	0,32	0,366
14	0,006	0,062	-0,027	4	-0,033	0,134	0,196
60	0,006	0,079	0,05	10	0,044	0,114	0,193
57	0,005	0,089	-6,769	4	-6,774	0,103	0,192
156	0,004	0,069	-0,04	6	-0,044	0,341	0,41
7	0,002	0,038	0,019	10	0,017	0,055	0,093
104	0,002	0,095	-0,039	2	-0,041	0,309	0,404
63	0,001	0,044	0,732	8	0,731	1,9	1,944
65	0,001	0,063	-1,176	7	-1,177	1,617	1,68
142	0,001	0,049	0,027	1	0,026	0,212	0,261

column 1, called id, th ere is the identification number of the each stock;
column 2 , called pf, there is the forecasted return of each stock (> 0);
column 3 , called varpf, there is the forecasted return variance of each stock ;
column 4 , called vinf, there is the forecasted intrinsic value of each stock ;
column 5 , called indm, there is the industrial sector of each stock ;
column 6 , called vin_p, there is the difference between the intrinsic value and the return for each stock ;
column 7 , called var_vinf, there is the forecasted intrinsic value variance of each stock ;
column 8 , called var_uf, there is the forecasted potential variance of each stock ;

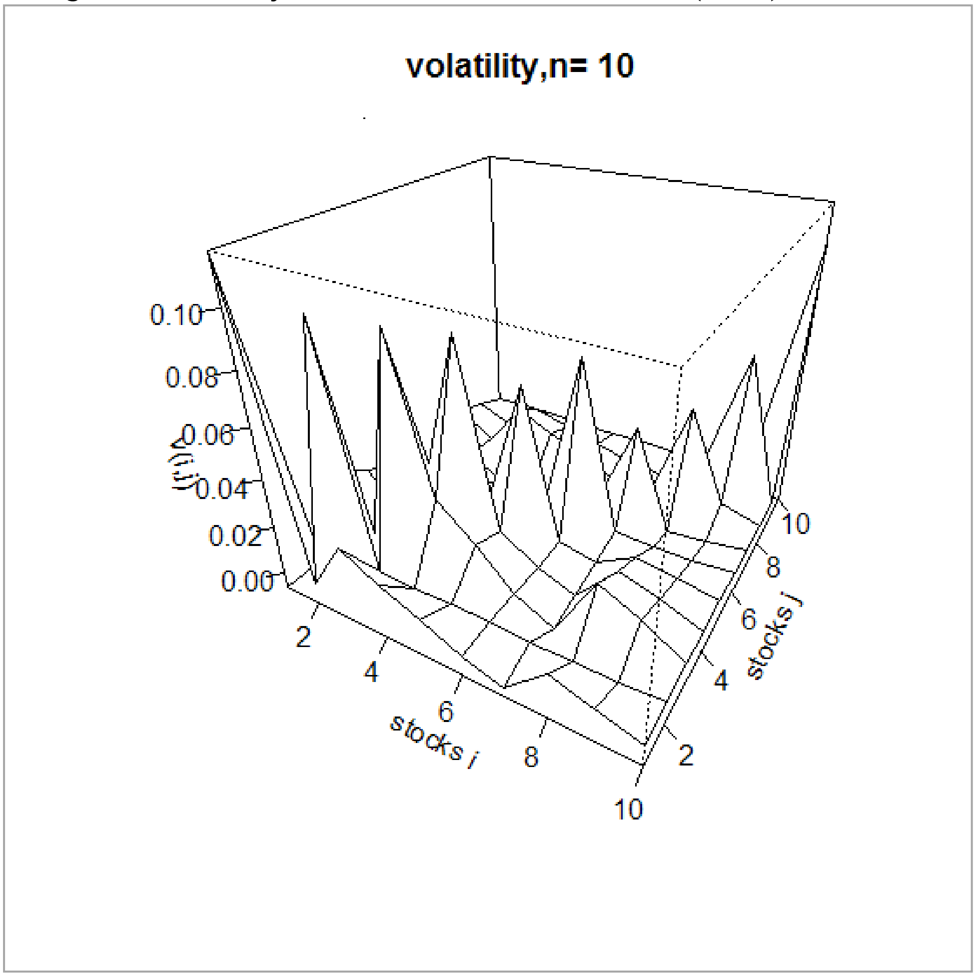
In the following **Figure 2** we see the betas for the two factors in a portfolio of the first *n=10* stocks among the subset selected before .

Figure 2: β_1 , β_2 for factor 1,2 (n=10)



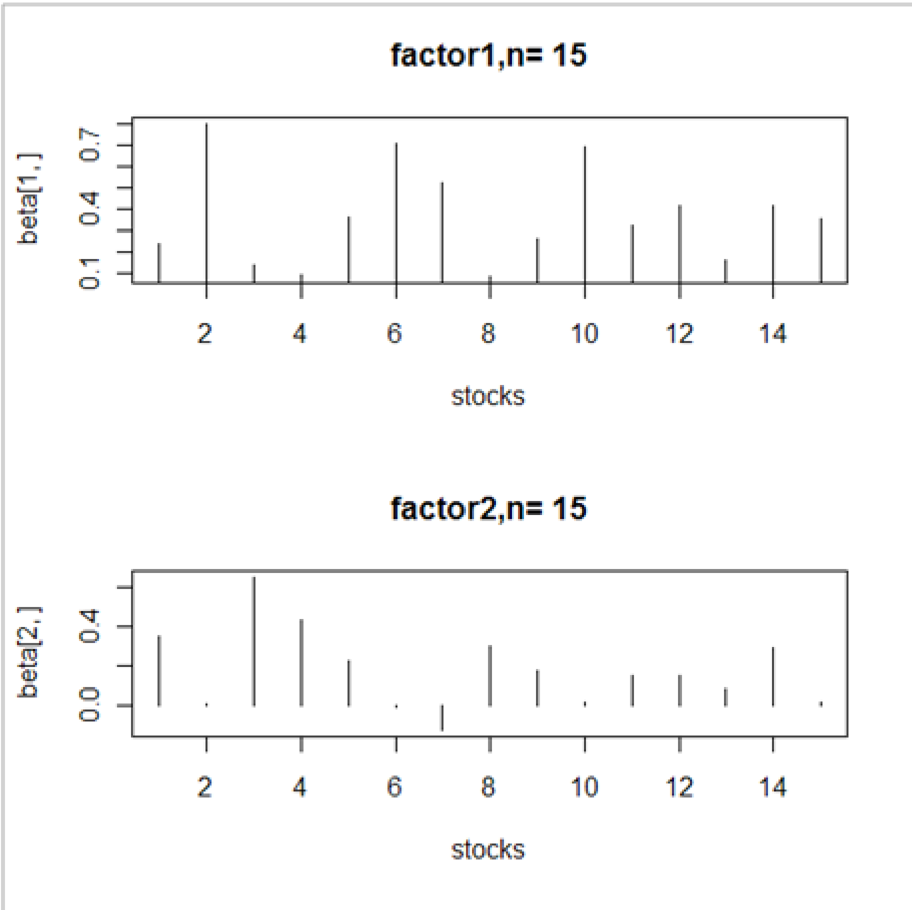
In the following **Figure 3** we see the volatility matrix estimated with the Factor-BEKK model for a portfolio of the first *n=10* stocks among the subset selected before :

Figure 3: Volatility Estimation with Factor-BEKK (n=10)



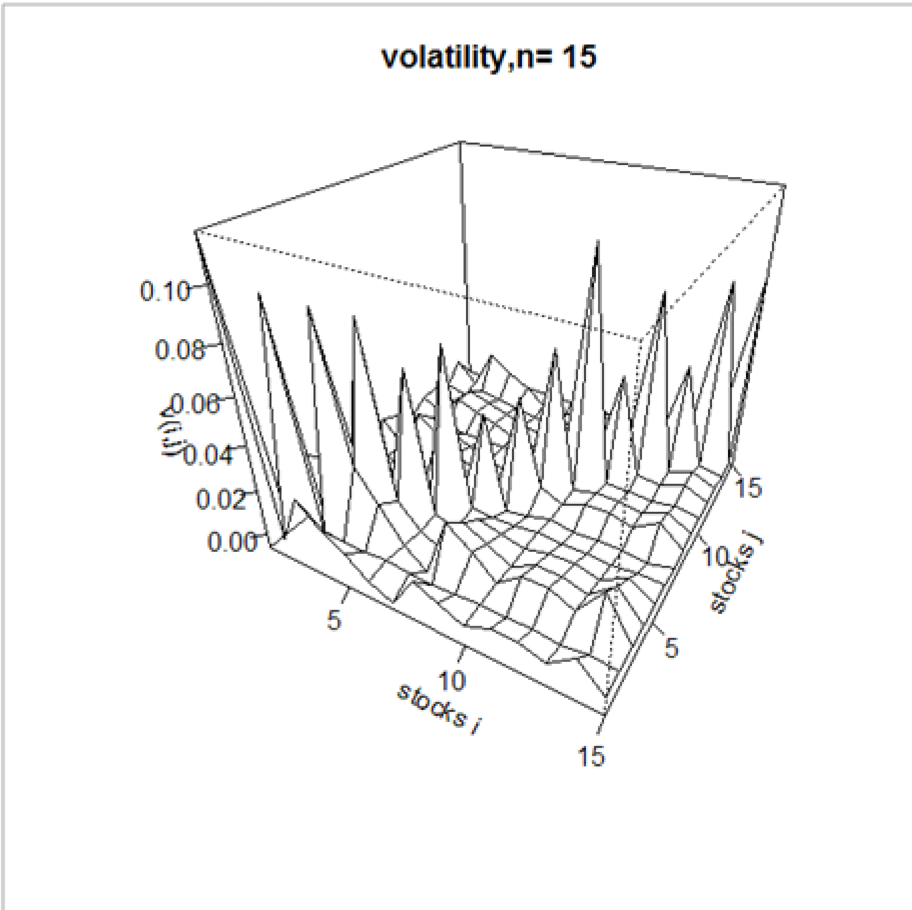
In the following **Figure 4** we see the betas for the two factors in a portfolio of the first *n=15* stocks among the subset selected before:

Figure 4: β_1 , β_2 for factor 1,2 (n=15)



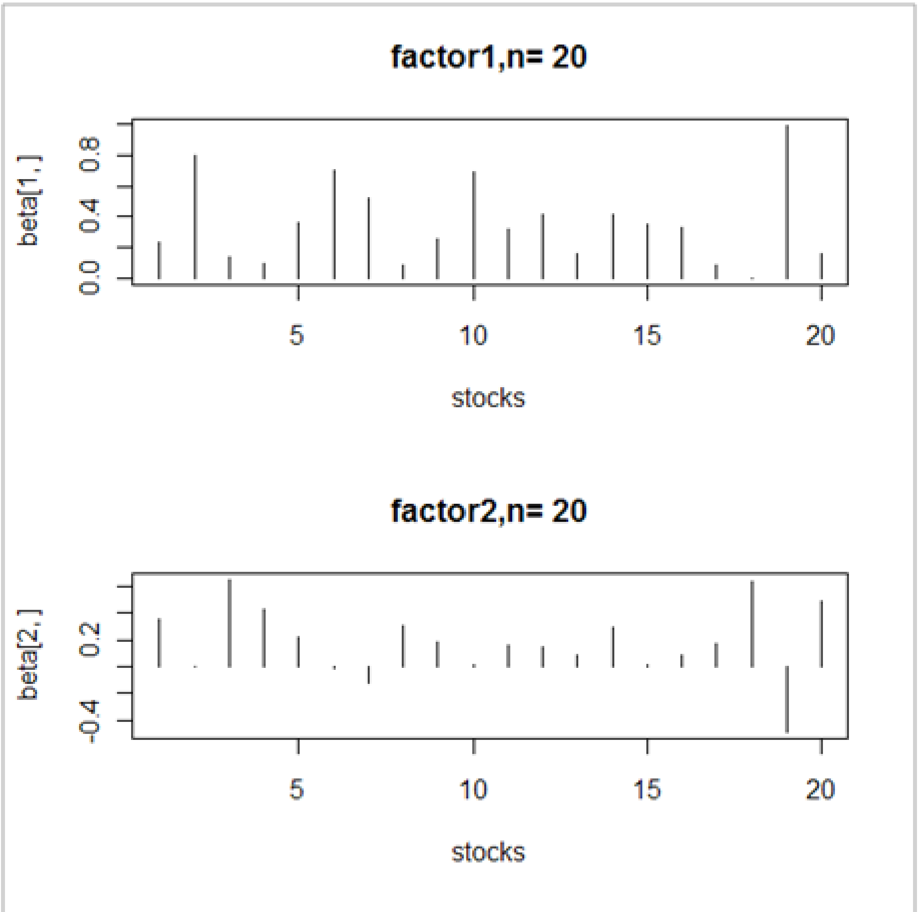
In the following **Figure 5** we see the volatility matrix estimated with the Factor-BEKK model for a portfolio of the first *n=15* stocks among the subset selected before :

Figure 5: Volatility Estimation with Factor-BEKK (n=15)



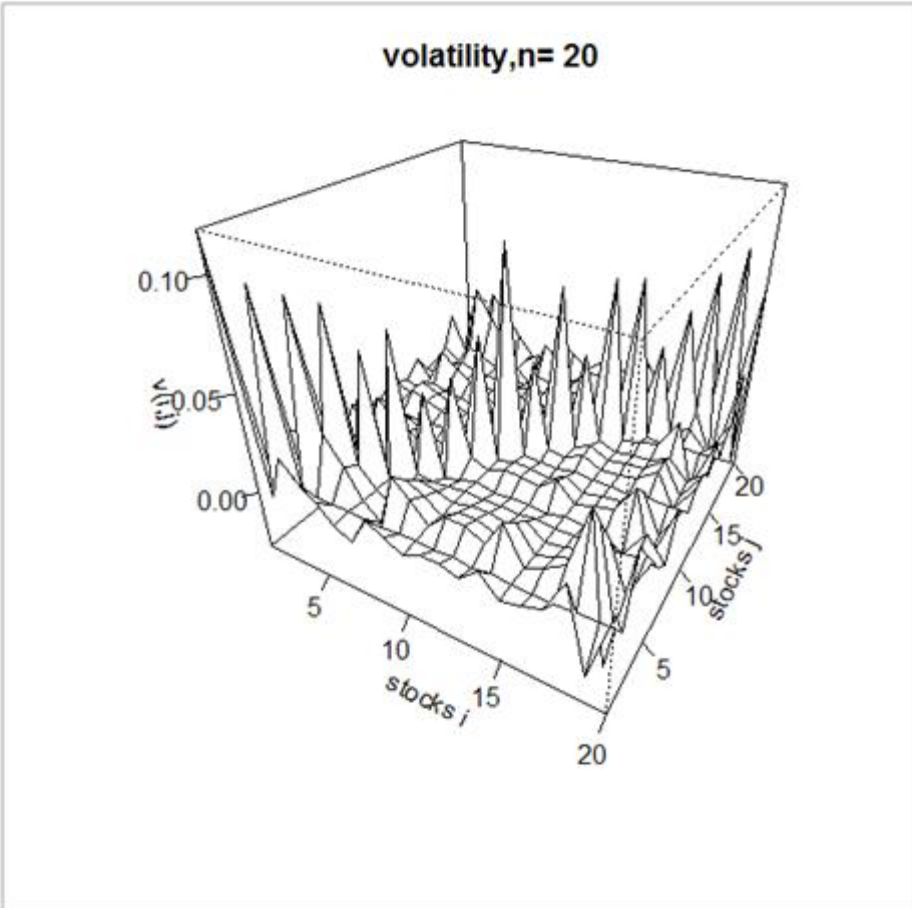
In the following **Figure 6** we see the betas for the two factors in a portfolio of the first *n=20* stocks among the subset selected before:

Figure 6: β_1 , β_2 for factor 1,2 (n=20)



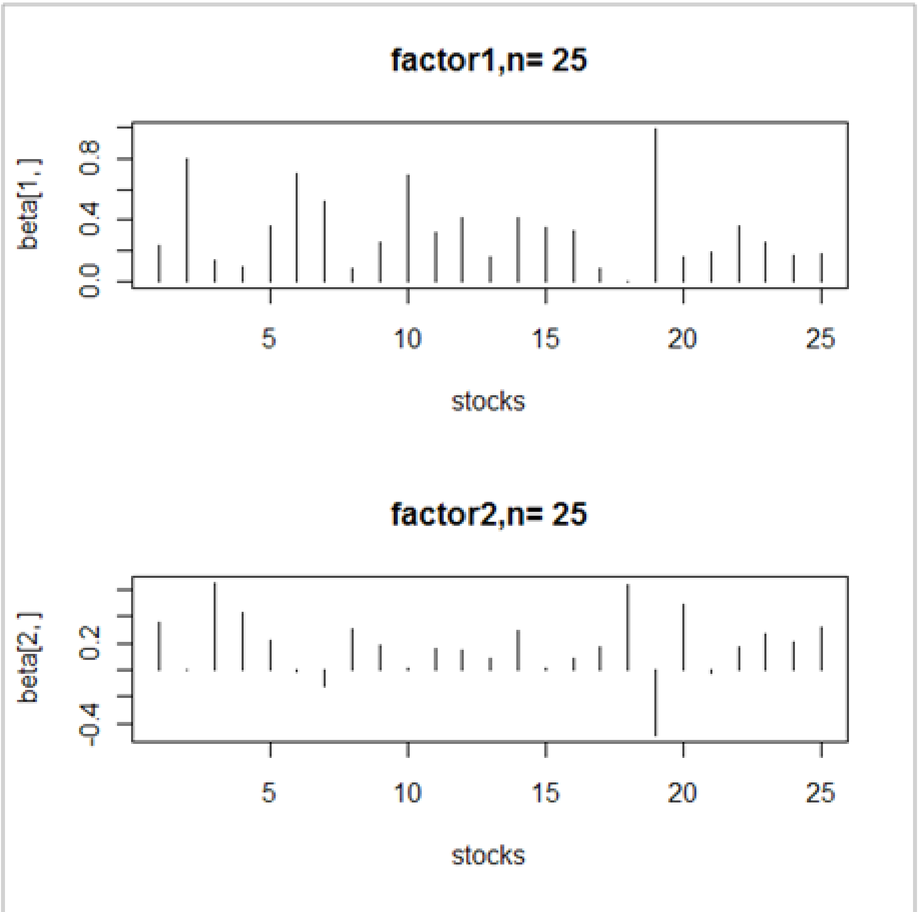
In the following **Figure 7** we see the volatility matrix estimated with the Factor-BEKK model for a portfolio of the first *n=20* stocks among the subset selected before :

Figure 7: Volatility Estimation with Factor-BEKK (n=20)



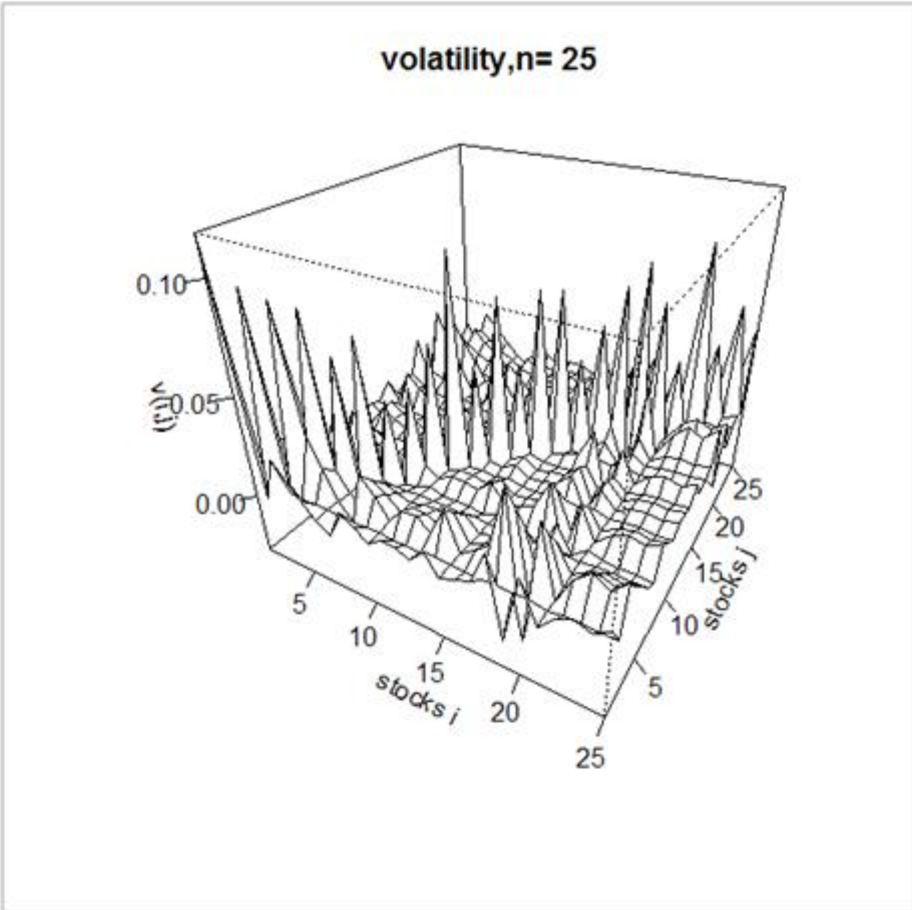
In the following **Figure 8** we see the betas for the two factors in a portfolio of the first *n=25* stocks among the subset selected before:

Figure 8: β_1 , β_2 for factor 1,2 (n=25)



In the following **Figure 9** we see the volatility matrix estimated with the Factor-BEKK model for a portfolio of the first *n=25* stocks among the subset selected before :

Figure 9: Volatility Estimation with Factor-BEKK (n=25)



The Figure 2,4,6,8 show that the sensitivity of the stocks to the two factors changes with different number of stocks *n* in the portfolio. Sometimes the $\beta_{i,j}$ may be near to zero showing no sensitivity of their relative stocks.

The Figure 3,5,7,9 show that increasing the number of stocks *n* in the portfolio has the effect of increasing also the covariance among the stocks (smaller or higher picks). Therefore it can give a considerable amount of decreasing in the overall portfolio variance.

To be notice that we start from *n=10* as indicated in Evans Archer 1968.

5.5 Comments

The application of the Factor-BEKK model is an improvement of the multivariate estimation of the volatility matrix in (Pierini Naccarato 2012) because it gives an insight of the variability between the stocks through the macroeconomic market movements by using the market index factor. Moreover by using the intrinsic value factor it gives an insight also through movements in the fundamentals P/E (Nicholson, S. F., 1960) and EPS which make the intrinsic value.
Nevertheless this model maintains the dimension tractability of its predecessor avoiding the strong numerical intractability of a crude BEKK application to the overall *n* stocks (Lutkephol 2007) or the unreal semplicity of a sample covariance application to the overall *n* stocks.

6. Multiple bi-dimensional BEKK model

6.1 Model description

The previous two models have the advantage of being computationally feasible while retaining a certain amount of power in the description of the real data involved. If the variances are regarded as time-dependent, however, then it is more plausible to regard the covariances as time-dependent too. This is supported by empirical results as well as the literature (Tsay 2010). Everything has its cost, however, and consideration also of time-varying covariances entails a considerable computational burden making BEKK infeasible for a dimension superior to 5 (Ding, Engle, 1994). A new method is put forward below to tackle this problem.

A combination of bi-dimensional BEKK models is proposed for estimation of the volatility matrix of the Markowitz stock portfolio. Each diagonal element of volatility is estimated by taking the variance given by a bi-dimensional BEKK model with stock i and the market index as variables. Each off-diagonal element of volatility is estimated by taking the covariance given by a bi-dimensional BEKK model with stocks i and j as variables.

The model is applied to a subset of promising universes among the series of data regarding the prices of the 150 shares with the highest market capitalization traded on the Italian stock exchange between 1 January 1975 and 31 August 2011. The volatility matrix of returns is required in order to select an optimal stock portfolio.

The diagonal and off-diagonal elements are estimated separately, as stated above, and the efficient frontier given by the solution of the estimated Markowitz problem is then simulated, thus providing the optimal number of stocks, fractions and returns in order to obtain the minimum-risk portfolio.

This approach gives a time-dependent overall estimation of the stock variances-covariances while resolving the computational burden through decomposition of the original problem without losing the strength of BEKK.

The application of a multiple bi-dimensional BEKK model is an improvement on the multivariate estimation of previous volatility-matrix models. Its purpose is to solve the infeasible problem of estimating the entire volatility matrix by breaking it down into multiple feasible bi-dimensional estimations, as in the other two previous models. The full strength of the BEKK model is used here in such a way as to make every estimation time-dependent. The estimation of the variances is not different from the previous time-dependent estimations, as each variance of stock i at time t depends on the variance of the market index at time $t – 1$, its own variance at time t - 1 and the covariance between the stock and the market index at time t - 1 (Ruey Tsay, 2010).
The market-index time series thus helps the forecast for each stock variance of interest.

The *new idea* is that the estimation of the covariance elements is now different from the previous time-independent ones in that each covariance between stock i and stock j at time t depends on the variance of stock i at time t - 1, the variance of stock j at time t - 1 and their covariance at time t - 1.
In this way the covariance forecast for each pair of stock is guided by past correlation information.

The multiple bi-dimensional BEKK model is applied to the residuals of a bi-dimensional CVAR*(p)* model for each item in a set of n stocks. The CVAR*(p)* model also gives the predicted return for each stock.
In order to select the n most promising stocks, they are sorted in decreasing order with respect to their potential values. The potential value of a stock is the difference between its return and its compound intrinsic value, where the intrinsic value is the P/E times EPS.
The Markowitz portfolio is then solved and the dimension of the portfolio n is increased by 1 each time to find the minimum-risk choice among the best Markowitz portfolios.

The starting point taken in order to estimate the time series of market returns is the K=150 series with respect to the returns on shares $R_{k,t}$ and the average return of the market $R_{M,t}$, $t=t_k,\dots,T$, $k=1,\dots,K$.
For each series, the model *CVAR(p)* is considered for the random vector $y_t = [y_{1,t}, y_{2,t}]' = [R_{k,t}, R_{M,t}]'$ as in equation (1).

The CVAR model makes it possible to deal with cases of the integration and cointegration of y_t components whenever present.

For each series, as in equation (1), the model *CVAR(p)* is considered for the random vector $y_t = [y_{1,t}, y_{2,t}]' = [(P/E)\cdot(EPS)_{h,t}, (P/E)\cdot(EPS)_{IS(h),t}]'$, where $h=1,\ldots H$ is the index that identifies only the series with positive returns selected out of the initial 150.

On the basis of the $(P/E)\cdot(EPS)_{h,T+1}$ and $R_{h,T+1}$ forecasts obtained in phase 2, the shares are listed in decreasing order with respect to the values of the difference between *intrinsic value* and *expected price*. The first $n\in\{10,11,\ldots,H\}$ shares are thus selected to make up the portfolio.

BEKK1

In order to take into consideration the presence of *variance heteroskedasticity* (Pierini Naccarato 2012), the *BEKK(1,1)* model is used to estimate the conditional variance matrix of $u_t = [u_{1,t}\, u_{2,t}]' = =[u_{k,t}\, u_{M,t}]'$ given in phase 1 of equation (1) for stock k and market index M, $\Sigma_{t|t-1}=cov(u_t|past)=((\sigma_{i,j}(t))_{i,j=1,\ldots,n}$, which has the same structure as in equation (8).

In order to take into consideration the presence of *time-varying autocorrelated disturbances* (Pierini, Naccarato CFE 2013, Naccarato, Pierini 2014) for each series previously selected, the model *CVAR(p)* is considered for the random vector $y_t = [y_{1,t}, y_{2,t}]' = [R_{k,t}, R_{s,t}]'$, as in equation (1).

BEKK2

The *BEKK(1,1)* model is then used to estimate the conditional variance matrix of $u_t = [u_{1,t}\, u_{2,t}]' = [u_{k,t}\, u_{s,t}]'$ given by the residuals of equation (1) for stock k and s, $\Sigma_{t|t-1}=cov(u_t|past)=((\sigma_{i,j}(t))_{i,j=1,\ldots,n}$, which has the following structure:

$$\begin{bmatrix}\sigma_{11,t} & \sigma_{12,t}\\ \sigma_{21,t} & \sigma_{22,t}\end{bmatrix} = \begin{bmatrix}d_{0,11} & 0\\ d_{0,21} & d_{0,22}\end{bmatrix}\begin{bmatrix}d_{0,11} & d_{0,21}\\ 0 & d_{0,22}\end{bmatrix} + \\ + \begin{bmatrix}d_{1,11} & d_{1,21}\\ d_{1,21} & d_{1,22}\end{bmatrix}\begin{bmatrix}u_{1,t-1}^2 & u_{1,t-1}u_{2,t-1}\\ u_{2,t-1}u_{1,t-1} & u_{2,t-1}^2\end{bmatrix}\begin{bmatrix}d_{1,11} & d_{1,21}\\ d_{1,21} & d_{1,22}\end{bmatrix} + \\ + \begin{bmatrix}d_{2,11} & d_{2,21}\\ d_{2,21} & d_{2,22}\end{bmatrix}\begin{bmatrix}\sigma_{11,t-1} & \sigma_{12,t-1}\\ \sigma_{21,t-1} & \sigma_{22,t-1}\end{bmatrix}\begin{bmatrix}d_{2,11} & d_{2,21}\\ d_{2,21} & d_{2,22}\end{bmatrix} \qquad (21)$$

At time $t=T$, the maximum likelihood estimations (MLE) of the parameters $d_{k,i,j}$, $k=0,1,2$, $i,j=1,2$ are by obtained maximizing the following function:

$$\sum_{t=1}^{T} l_t(\theta), l_t(\theta) = -\ln(2\pi) - \frac{\ln|\Sigma_{t|t-1}|}{2} - \frac{1}{2}u'_t\Sigma_{t|t-1}^{-1}u_t \qquad (22)$$

It should be noted that the (Q)ML estimator of the parameter vector

$$\theta = vec(D_0, D_1, D_2) \qquad (23)$$

where the vec operator transforms a matrix into a vector by stacking its columns, $(D_m)_{i,j=1,2} = d_{m,ij}$, $m = 0,1,2$, cannot be obtained analytically. An iterative optimization algorithm is therefore required.

Use can be made of an algorithm of the Newton-Raphson type, which requires the first and second derivatives of (22), (Comte, Lieberman, 2003):

$$\frac{\partial l_t}{\partial \theta_i} = -\frac{1}{2} tr(\dot{\Sigma}_{t|t-1,i}\, \Sigma_{t|t-1}^{-1} - u_t u_t'\, \Sigma_{t|t-1}^{-1} \dot{\Sigma}_{t|t-1,i}\, \Sigma_{t|t-1}^{-1}) \qquad (24)$$

and

$$\begin{aligned}\frac{\partial^2 l_t}{\partial \theta_i \partial \theta_j} = -\frac{1}{2} tr(&\ddot{\Sigma}_{t|t-1,j,i}\, \Sigma_{t|t-1}^{-1} - \dot{\Sigma}_{t|t-1,j}\, \Sigma_{t|t-1}^{-1} \dot{\Sigma}_{t|t-1,i}\, \Sigma_{t|t-1}^{-1} + \\ &+ u_t u_t'\, \Sigma_{t|t-1}^{-1} \dot{\Sigma}_{t|t-1,i}\, \Sigma_{t|t-1}^{-1} \dot{\Sigma}_{t|t-1,j}\, \Sigma_{t|t-1}^{-1} + \\ &+ u_t u_t'\, \Sigma_{t|t-1}^{-1} \dot{\Sigma}_{t|t-1,j}\, \dot{\Sigma}_{t|t-1,j}\, \Sigma_{t|t-1}^{-1}\, \dot{\Sigma}_{t|t-1,i}\, \Sigma_{t|t-1}^{-1} + \\ &- u_t u_t'\, \Sigma_{t|t-1}^{-1}\, \ddot{\Sigma}_{t|t-1,i,j}\, \Sigma_{t|t-1}^{-1})\end{aligned} \qquad (25)$$

where $\forall$ $A=(a_{ik})_{i,k=1,\dots,n}$, $tr(A) = a_{11}+\dots+ a_{nn}$, $\dot{\Sigma}_{t|t-1,j}$ is the matrix containing the first order derivatives of each element of $\Sigma_{t|t-1}$ with respect to θ_j , $\ddot{\Sigma}_{t|t-1,i,j}$ is the matrix containing the second order derivatives of each element of $\Sigma_{t|t-1}$ with respect to θ_i , θ_j .

A BHHH algorithm type (Brandt, Hall, Hall, Hausman, 1974) can be implemeted for the *s*-step as :

$$\hat{\theta}^{(s)} = \hat{\theta}^{(s-1)} + \delta \left(\sum_{t=1}^{T} \left. \frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta'} \right|_{\theta=\hat{\theta}^{(s-1)}} \right)^{-1} \sum_{t=1}^{T} \left. \frac{\partial l_t(\theta)}{\partial\theta} \right|_{\theta=\hat{\theta}^{(s-1)}} \tag{26}$$

where δ is the step length.

Moreover the (Q)ML estimator $\hat{\theta}$, as obtained above, is almost certainly consistent and

$$\sqrt{T}(\hat{\theta} - \theta) \longrightarrow N(0, I^{-1} J\, I^{-1}) \tag{27}$$

where $I = -E\left[\frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta'}\right]$, $J = E\left[\frac{\partial l_t(\theta)}{\partial\theta}\frac{\partial l_t(\theta)}{\partial\theta'}\right]$, even if the distribution of u_t is not normal as long as some regularity conditions are met (Comte, Lieberman, 2003).

Finally, the procedure of estimation in two steps – where the parameters of the CVAR are estimated in the first and the residuals of this model are then used to estimate the BEKK parameters in the second – is justified because the estimators of CVAR end BEKK are asymptotically independent (Lutkepohl 2007).

In order to determine the quantities to be bought of each of the *n* shares selected, it is necessary to solve the Markowitz problem iteratively with no shorting (Markowitz 1952) by estimating the matrix of share volatility:

$$min_{\omega\in\Re}\left\{\omega'\,\hat{V}\omega : \omega'\cdot\bar{1} = 1, \omega'\cdot\hat{r} = R_p, \omega \geq 0\right\} \tag{28}$$

where $\omega = (\omega_1, \dots, \omega_n)$, $R_p \in \left[min_{\{i=1,\dots,n\}}\, \hat{r_i}\ , max_{\{i=1,\dots,n\}}\, \hat{r_i}\right]$.

To this end, let $\hat{V}_t$ be the *estimator* of the matrix $n{\times}n$ of volatility V_t for $t=T+1$, the elements of which are $v_{i,j}(t)=cov(R_t|past),\ i,j=1,\dots,n$.
The elements of $\hat{V}_t$ are defined as follows:

$$v_{i,j}(T+1) = \begin{cases} \hat{\sigma}_{ii,T+1|T}^{bekk1} & if\ i = j \\ \hat{\sigma}_{ij,T+1|T}^{bekk2} & if\ i \neq j \end{cases} \tag{29}$$

where $\hat{\sigma}^{bekk1}_{ii,T+1|T}$ and $\hat{\sigma}^{bekk2}_{ij,T+1|T}$ are the estimates obtained respectively from equations (1) and (2) with *i,j=1,...,n*:

$$\begin{pmatrix} v_{1,1} & \cdots & v_{1,n} \\ \vdots & \ddots & \vdots \\ v_{n,1} & \cdots & v_{n,n} \end{pmatrix}_{T+1} \approx \begin{pmatrix} \hat{\sigma}^{bekk1}_{11,T+1|T} & \cdots & \hat{\sigma}^{bekk2}_{1n,T+1|T} \\ \vdots & \ddots & \vdots \\ \hat{\sigma}^{bekk2}_{n1,T+1|T} & \cdots & \hat{\sigma}^{bekk1}_{nn,T+1|T} \end{pmatrix} \tag{30}$$

Whenever $\hat{V}$ is not a positive definite $n \times n$ matrix, the Goldfarb-Idnani (1983) numerically stable dual method for finding the nearest positive definite $n \times n$ matrix $\hat{V}_{def}$ to $\hat{V}$ is used to ensure positive portfolio risk $\omega' \hat{V}_{def} \omega$.

6.2. Multiple 2D- BEKK Results

Application of the model put forward here to the monthly figures for the 150 BIT shares with the highest level of capitalization gives the following results. The selected stocks with positive predicted returns are a subset of maximal dimension $n_{max} = 25$.

Figure 10 shows the volatility and the return obtained by solving the Markowitz optimization problem for variation in the expected return $R_{p;T+1}$ and the dimension n of the portfolio (n = 10, 11,…, n_{max}), where n_{max} = 25 is the maximum number of shares with positive predicted returns. The portfolio risk tends to decrease as n increases.
A dimension *n=12* would give the maximum return 0.01423 but with a high risk of 0.0859.

Figure 10: Markowitz portfolio as n increases:
sdP=portfolio volatility, muP portfolio return
dimension n

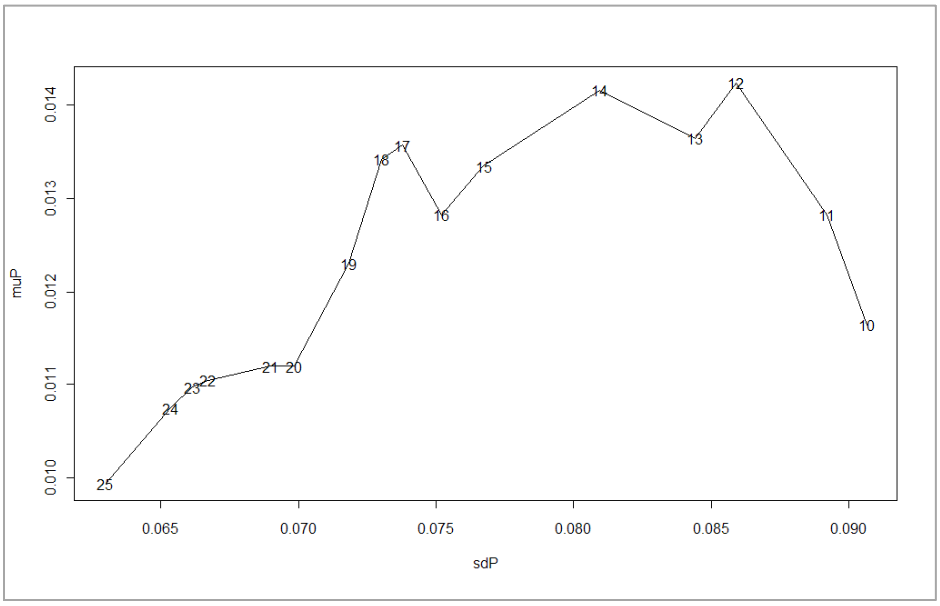
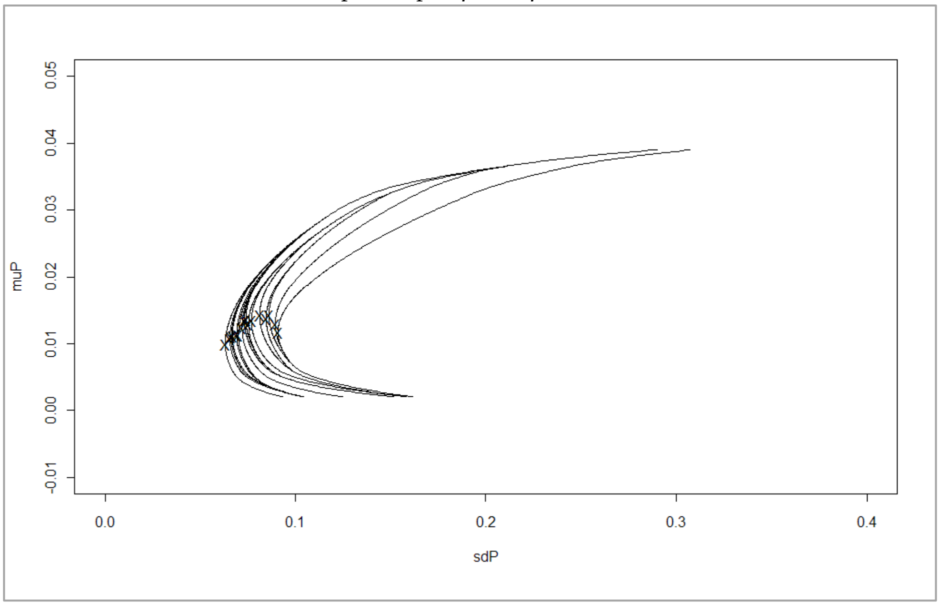


Figure 11 shows the efficient frontiers obtained by solving the Markowitz optimization problem for variation in the expected return $R_{p;T+1}$ and the dimension n of the portfolio (n = 10, 11,…, n_{max}), where n_{max} = 25 is the maximum number of shares with positive predicted returns.

Figure 11: Frontiers of Markowitz portfolio as n increases:
sdP=portfolio volatility, muP portfolio return
X=optimal portfolio of dimension n



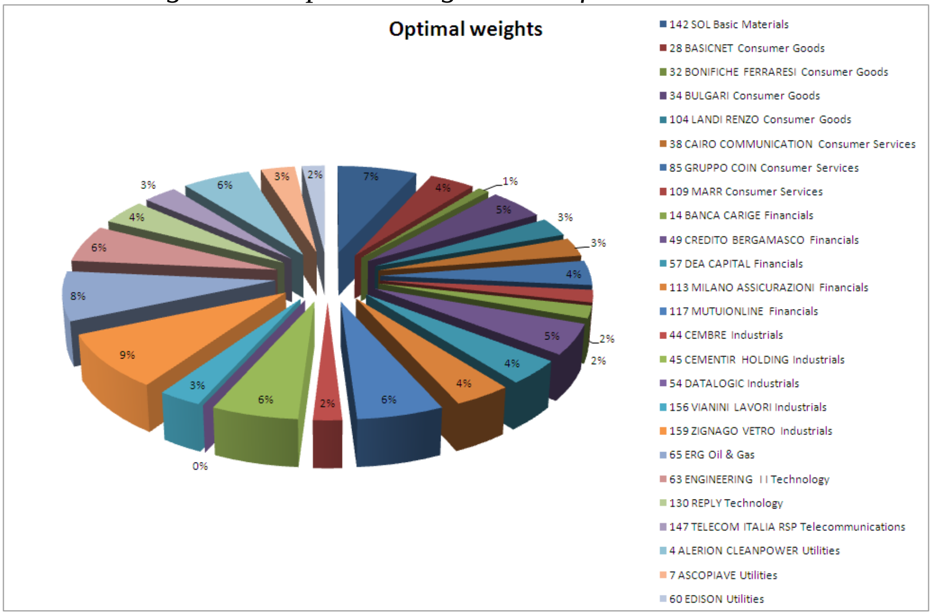
In overall terms, the portfolio risk tends to decrease as n increases. The optimal risk from a risk-averse standpoint (i.e. the least of all those calculated) corresponds to n = 25. This point is located closer to the origin of the axes and marked in figure 2 as X. This portfolio has a monthly average return of 0.00993, a monthly standard deviation of 0.0630 and a Sharpe index 0.15771. A portfolio is therefore selected of n = 25 shares with the optimal allocations shown in **Table 6**.

Table 6: Optimal weights n=25

id	op. w
4	0,06
7	0,03
14	0,02
28	0,04
32	0,01
34	0,05
38	0,03
44	0,02
45	0,06
49	0,05
54	0
57	0,04
60	0,02
63	0,06
65	0,08
85	0,04
104	0,03
109	0,02
113	0,04
117	0,06
130	0,04
142	0,07
147	0,03
156	0,03
159	0,09

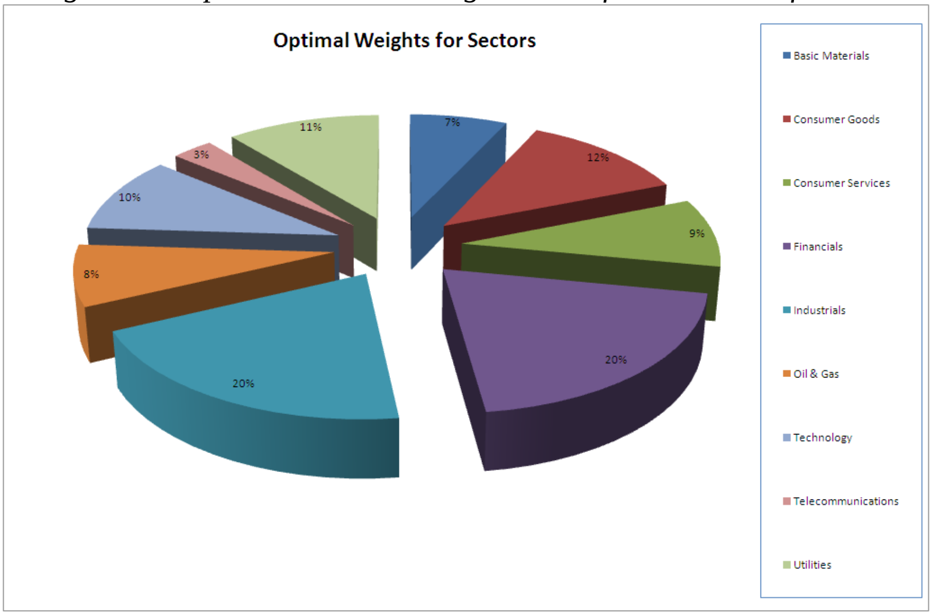
The selection for the portfolio of n = 25 shares with the optimal allocations is also shown in **Figure 12**. It can be seen that the allocation is quite diversified over the stocks.

Figure 12: Optimal weights n=25 for each stock



The optimal selection for the portfolio made up of n = 25 shares for each industrial sector is shown in **Figure 13**. It can be seen that the allocation is quite diversified over the stocks with a prevalence for the industrial (20%) and financial (20%) sectors.

Figure 13: Optimal Selection weights n=25 for each stock for Sectors



The empirical evidence of time-varying covariance can be seen, for example, by considering two stocks as in figures 14 and 15. **Figure 14** shows the sample covariance between UCG and PMI using a moving window of 100 observations. This smoothing suggests that covariance changes over time. **Figure 15** shows the same covariance estimated with a BEKK model so that it varies each time.

Figure 14: Sample covariance UCG , PMI (moving window of 100)

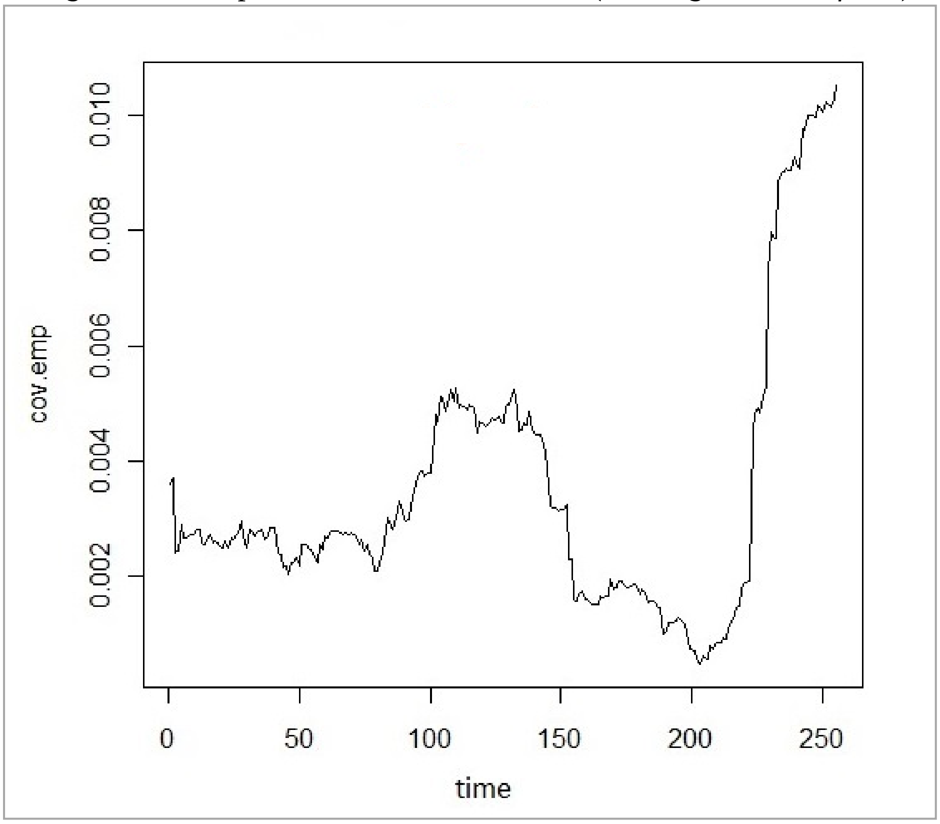
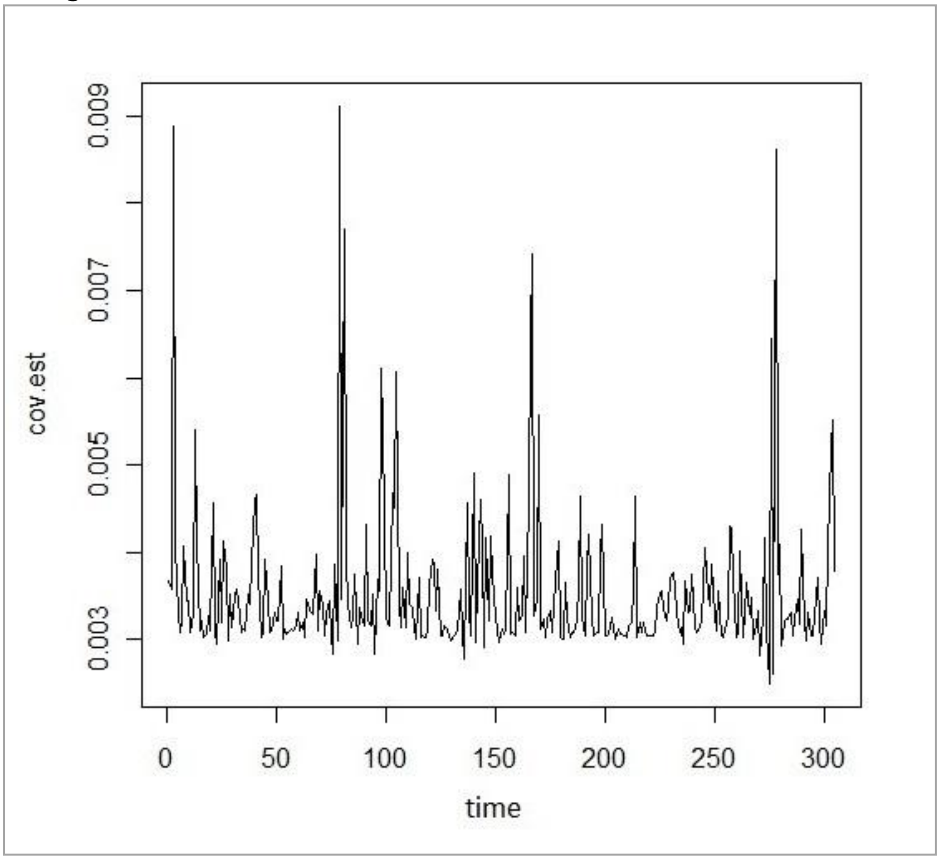
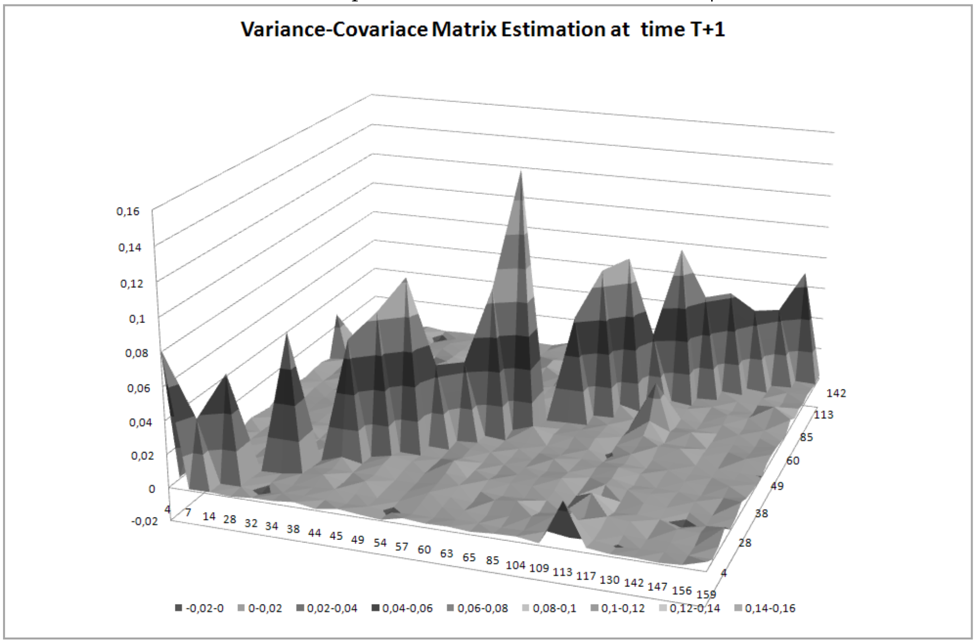


Figure 15:Estimated covariance UCG , PMI with BEKK



The estimation with equation (30), which is used to calculate the previous portfolio, is shown in **Figure 16**.

Figure 16: Variance-Covariance estimation
with multiple bi-dimensional BEKK at T+1|T

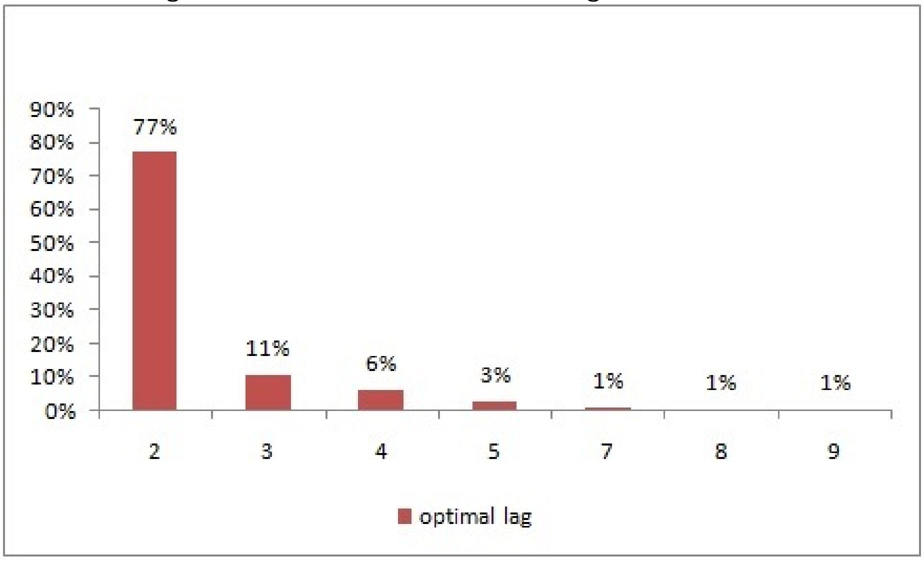


Some elements of the estimated covariance are larger than others and comparable with the variance. Let us call this subset A_s. Most of the covariances are instead lower than the ones in A_s and negligible with respect to the variance.

This suggests that it would be possible to separate the statistical analysis of the shares belonging to set A_s from the analysis of those belonging to the complementary set CA_s.

On the basis of minimum AIC, the optimal CVAR lag is 2–9 months. There is thus empirical evidence of market inefficiency, as the past has information to explain the future. **Figure 17** shows the empirical distribution of the optimal lags. As prediction is the primary objective, AIC is chosen as the lag criterion because it is asymptotically equivalent to the FPE (Final Prediction Error).

Figure 17: CVAR Minimum AIC lag distribution



The results of the LR test for all of the shares considered yield a degree of cointegration between 0 and 2. The 9% of instable time series (integrated and cointegrated) are taken into due consideration. **Figure 18** shows the empirical distribution of the Π matrix, defined in (4), in ranks.

Figure 18: CVAR rank distribution

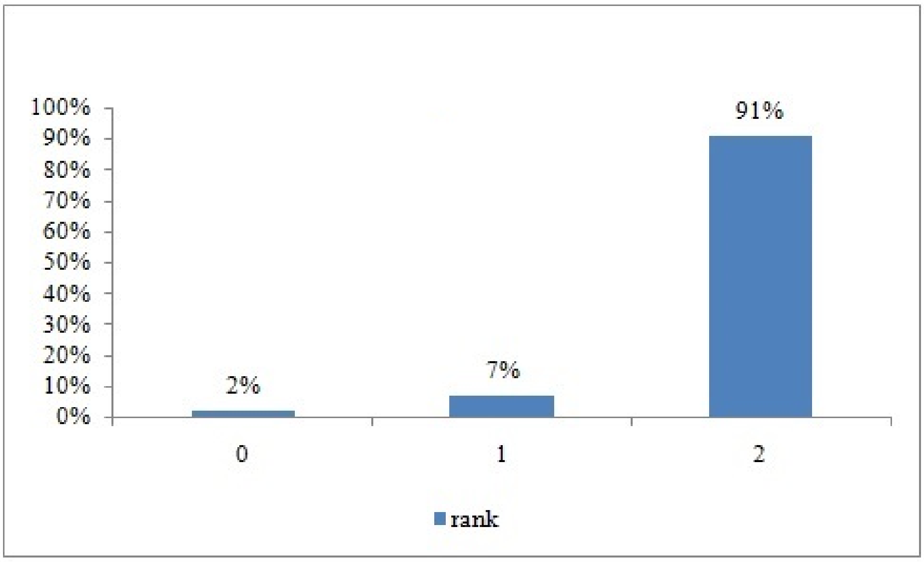


Figure 19 shows the empirical distribution of the *p* values of *F* statistics for testing $H_0 : a_{h,ij} = 0$ in equation (1). This is equivalent to verifying the hypothesis that the model estimated does not exist.

It can be seen that H_0 is rejected at a confidence level of 90% in 83% of the time series. The CVAR model is therefore adequate for most of the cases.

Figure 19: CVAR p values F distribution

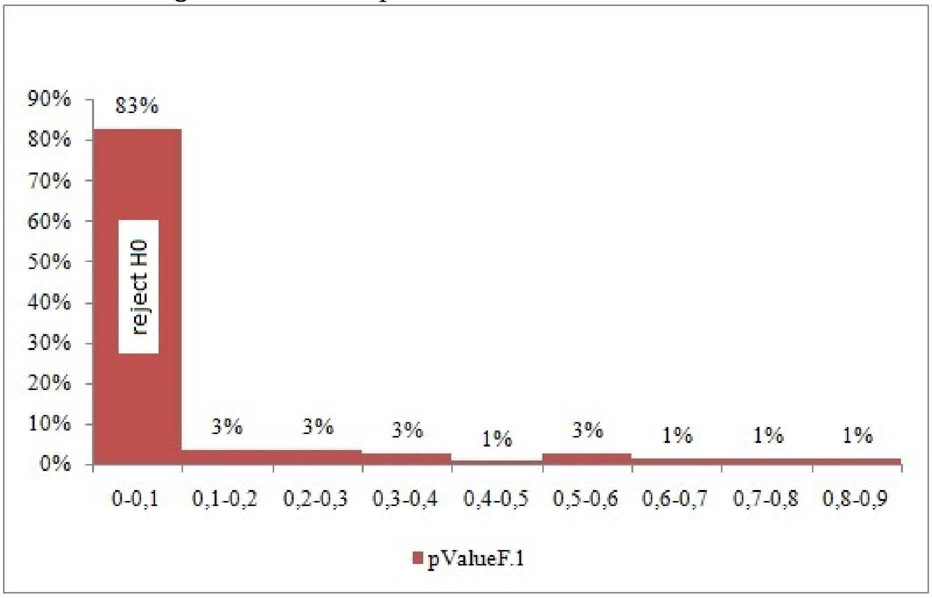


Figure 20 shows the empirical distribution of the *p* value of the portmanteau statistics for testing H_0: no correlation in the CVAR residuals. It can be seen that H_0 is rejected at a confidence level of 90% in 6% of the time series. The CVAR model is therefore adequate for most of the cases even though some lags could be added. This is not done here because the prevailing criterion for lag selection is AIC.

Figure 20: p values portmanteau distribution
for no correlation in CVAR res

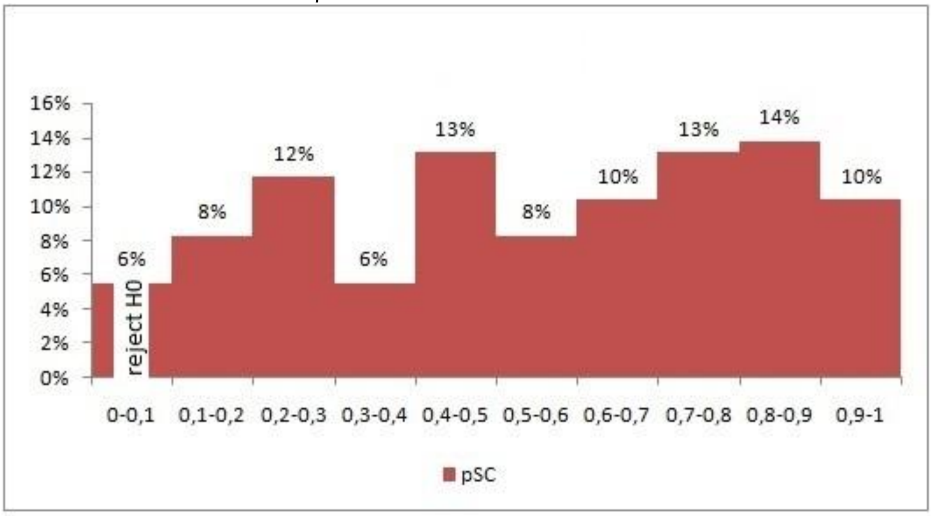


Figure 21 shows the empirical distribution of the *p* values of the portmanteau statistics for testing H_0: no ARCH in the CVAR residuals.
It can be seen that H_0 is rejected at a confidence level of 90% for 79% of the time series. The CVAR residuals are therefore affected with ARCH effects in the majority of cases and GARCH modelling is required because it can represent this effects.

Figure 21: p values portmanteau distribution
for no ARCH in CVAR res

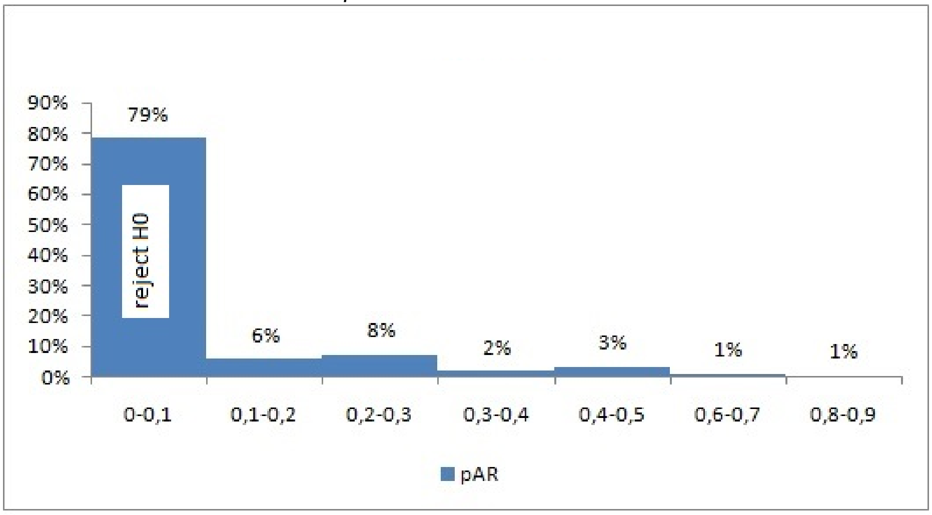
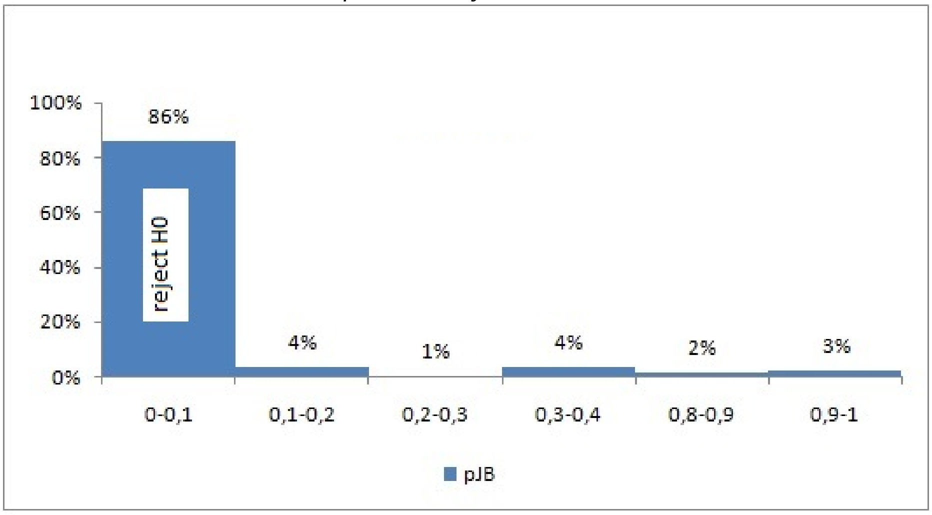


Figure 22 shows the empirical distribution of the *p* values of the Jarque-Bera statistics for testing H_0: normality of the CVAR residuals.
It can be seen that H_0 is rejected at a confidence level of 90% for 86% of the time series considered. The CVAR residuals therefore have a non-Gaussian distribution (mostly due to the leptokurtic effect) for the majority of cases and GARCH modelling is required because it can represent this.

Figure 22: p values distribution
for normality in CVAR res



Let $\hat{\varepsilon}_t = \hat{\Sigma}_{t|t-1}^{-1/2}\hat{u}_t$ where $\hat{u}_t\,|\Omega_{t-1} \sim N\left(0, \hat{\Sigma}_{t|t-1}\right)$ and $\hat{\Sigma}_{t|t-1}$ as in equation (8), which are called BEKK residuals here.
If A is a positive definite $n \times n$ and symmetric matrix then its spectral decomposition $P'AP = diag(\lambda_1,..., \lambda_n) = \Lambda$ exists, where are the eigenvalues of A (with $\lambda_i \geq 0$) and the columns of P are the corresponding eigenvectors.
Therefore $\lambda_i^{1/2}$ exists and the matrix $Q= P\Lambda^{1/2} P'$ is the square root of A and it is denoted as $A^{1/2}$. As the BEKK builds a positive 2×2 and symmetric matrix $\hat{\Sigma}_{t|t-1}$, the matrix $\hat{\Sigma}_{t|t-1}^{1/2}$ can be defined as above and $\hat{\Sigma}_{t|t-1}^{-1/2}$ as its inverse matrix.

Figure 23 shows the empirical distribution of the portmanteau statistics of the *p* values for testing H_0: no correlation in the $\hat{\varepsilon}_t$. It can be seen that H_0 is rejected at a confidence level of 90% for 10% of the time series considered. While there is thus no correlation in the BEKK residuals in most of the cases, correlation does persist in some cases. More lags can be added but this means increasing the computational burden. A tentative solution to this problem will be put forward in the section on further development.

Figure 23: p values portmanteau distribution
for no correlation in BEKK res

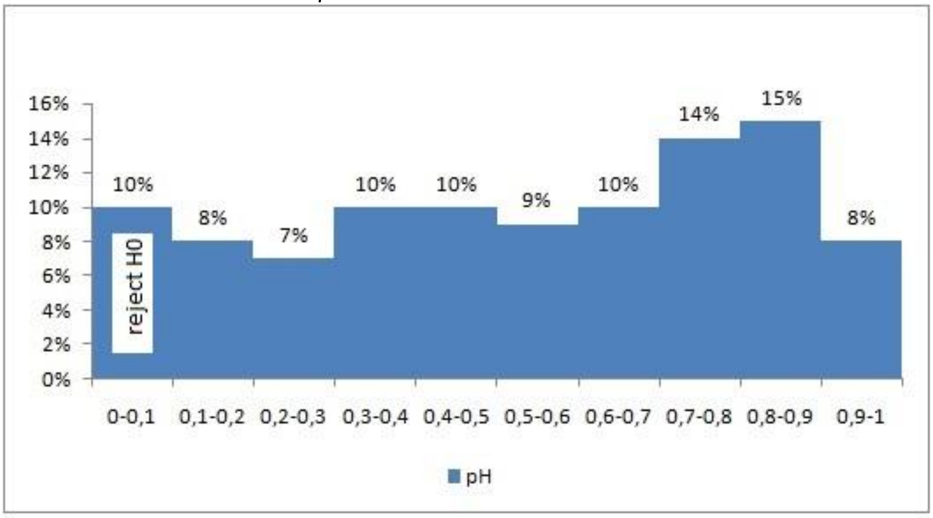


Figure 24 shows the empirical distribution of the portmanteau statistics of the p values for testing H_0: no ARCH in $\hat{\varepsilon}_t$.
It can be seen that H_0 is rejected at a confidence level of 90% for 24% of the time series considered. While there is thus no ARCH effect in the BEKK residuals in most majority of the cases, ARCH effects do persist in some cases. More lags can be added but these involve worsening the computational aspect. A multivariate Student-t distribution can also be tentatively used as the conditional distribution of u_t (Fiorentini et al 2003). This approach has the disadvantage that if the assumption is wrong, the ML estimates are generally not even consistent. On the other hand, using a Gaussian likelihood, also known as quasi-maximum likelihood (QML), retains consistency under misspecification (Hafner, Herwartz 2013). A

tentative solution to this problem will be put forward in the section on further development.

All in all, the models obtained prove adequate for most of the time series considered.

Figure 24: p values portmanteau distribution
for no ARCH in BEKK res

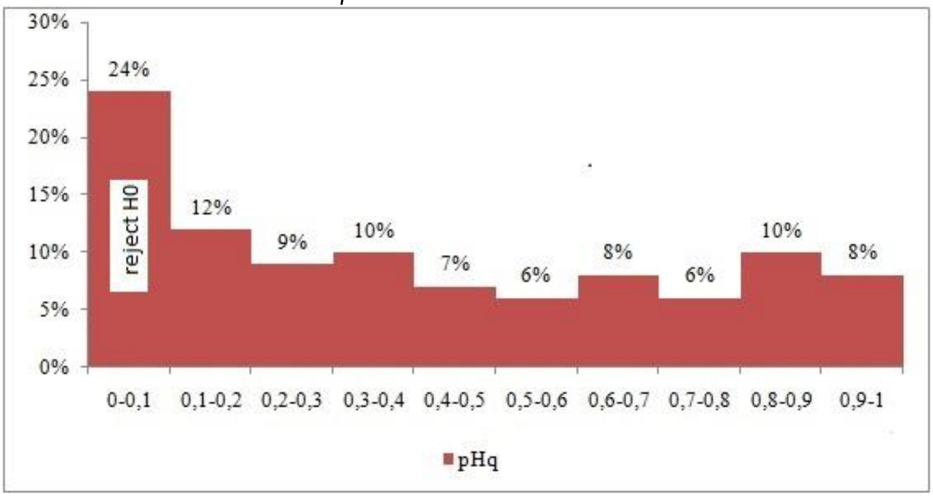
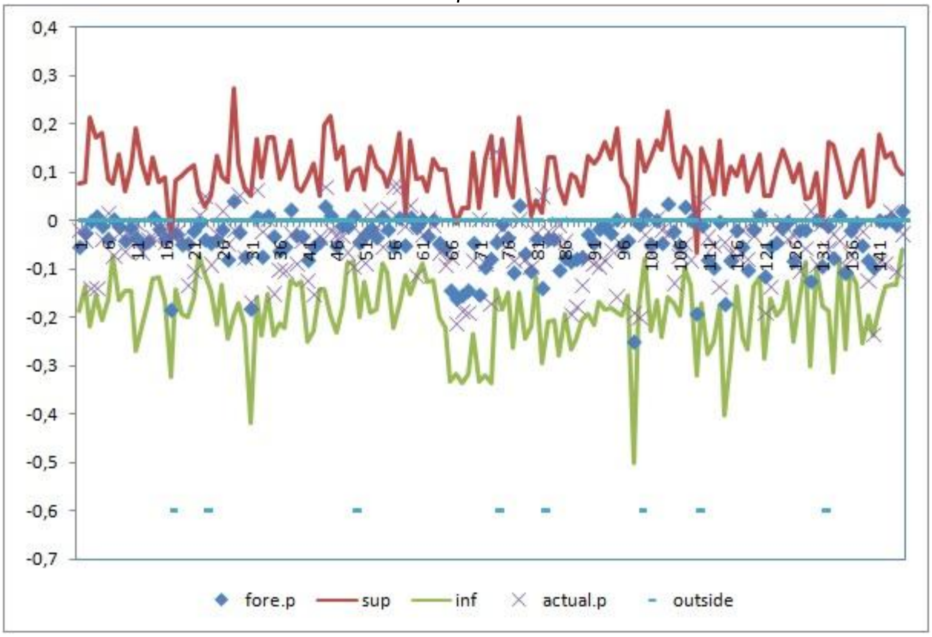


Figure 25 shows the 1 – α approximate interval forecast
$\hat{r}_{k,t} \pm z_{\alpha/2}\hat{\sigma}_{k,t}$ together with the forecast and the actual values $r_{k,t}$ with α = 0,05. The model's adequacy is confirmed by the fact that only 6% of the actual values are outside the forecast intervals. It should be noted that the conditional normal quantile hypothesis could be relaxed by means of a bootstrap for the interval.

Figure 25: 95% approximate interval forecast
vs actual values for all the time series



7. Further Development

7.1 Introduction

As noticed before, the computational burden is one of the main problem one must deal with stock portfolio selection.
To reduce the *VAR* parameter space it is possible to apply the following undirected gaussian graphical models.
However, this new approach is tentatively applied to 31 EU index values and in a successive effort will be applied to the hole set of 150 stocks. This is due to the fact that indexes are less in number and smoother, being mean of stocks, than stocks themselves.

7.2 Undirected Gaussian Graphical models for VAR parameters reduction

Undirected Gaussian Graphical models are applied to non-parametric resampling of a multivariate time series of EU index returns, in order to construct graphs such that the edges included in the graphs are definitely present indicating a significant relationship between the variables.
Moreover a convex optimization technique and a forward minimum AIC search is used to construct alternative *UG* graphs. *VAR(1)* models are constructed based only on the previous specified relationships
and *GARCH(1,1)* models are applied to their residuals to estimate the immediate future return and variances for each index.
To estimate the covariances between each pair of return a *Latent Factor* model is applied to the set of EU index returns. Finally the simulated portfolio returns and a probability of gain is calculated by
solving the corresponding Markowitz problem.

Stock Index portfolio selection is wildly discussed in the literature with reference to heteroskedastic models , Bollerslev et al. (1994). The model used in the case of multiple time series is of the vector autoregressive (*VAR*) type and rests on the predictability of the average
return on stocks, Tiao G., Box G. (1981).
The *VAR* type models suffer of computational burden as the number of parameters to be estimated are large with respect to univariate models.

This paper suggests the use of Gaussian Graphical Models (*GM*) to identify only the significant parameters to be estimated for the *VAR* models whenever a subset of the variables are considered connected.

1. the simultaneous *p*–value (*SIN*), which is a thresholding approach ;
2. the glasso, which is a penalized log-likelihood maximization approach ;
3. the decomposable stepwise search, which is a forward search through decomposable graphs minimizing *AIC* .

A stepwise backward search from saturated graphs was tentatively used but it did not give results due to the NP–hard complexity so that it was discarded.

If case of connected variables, *Restricted VAR(1)* models are estimated taking into account only the connections identified before.

As the residuals exhibit heteroskedasticity, *tGarch(1,1)* models are then applied to them in order to take this particular aspect into due consideration.

For the subset of variables which are not connected at all, univariate *AR(p) –tGarch(p,0)* are estimated.

In this way a great reduction of calculation is obtained and a more efficient estimation is reached with respect to multidimensional full models.

The orders *p,P,Q* of *VAR(p)* , *tGarch(P,Q)* can be augmented leaving the approach unchanged.

Moreover the application of multidimensional models for the residuals, as BEKK or DVEC models, Engle (1982), in the case of detected connection, can be prohibitive due to the dimension of the problem.

Lastly a *Latent Factor Model* (LMF), Connor (1995), is applied to estimate the covariances between each pair of index return. LMF has the advantage of being capable to deal with large dimension by considering only few latent factors to represent the whole dimension.

However it may happen that the some *(V)AR–tGarch* models result unstable.

In this cases the LMF variance estimation is considered and the unstable models are discarded.

Thus the problem of estimating future returns of indexes and the associated risk on the basis of past returns of the European Stock Indexes over time is tackled via the combination of *GM*
and *(V)AR–tGarch,* LFM models.

Through this estimation the Markowitz problem, Sharpe (1970), is solved with quadratic optimization technique.

A resampling is then applied for g = 500 times to obtain an empirical distribution of the optimal portfolio returns so that an estimation of the probability of gaining is calculated.

The time series considered regard the European Stock Indexes monthly figures from 1 January 1995 to 31 December 2012 taken by the Eurostat database.

Let's consider a multivariate time series where the log return for the stock index i at time t is $r_{i,t}$ $i = 1,...,k$ and $t = 1,...,T$, that's to say $r_{i,t} = log(p_{i,t} / p_{i,t-1})$, $p_{i,t}$ is the index value at time t. Then a back shift in time of length 1 is done, obtaining $r_{i,t}^{lag1} = r_{i,t-1}$, $i = 1,...,k$ and $t = 1,...,T$ where $r_{i,0} = NA$. A graph is defined as a couple $G = (V,E)$ where V is a set of vertices or nodes and E is a set of edges, Lauritzen (1996). Undirected edges only are used and each edge is associated with a pair of nodes.
A graph $G_0 = (V_0,E_0)$ is called subgraph if $V_0 \subset V$, $E_0 \subset E$.
Let A,B,C be subset of V and f a joint density function of $(X_v)v \in V$ then $A \perp B \mid C$ if $f(x_A,x_B|x_C)= f(x_A|x_C) f(x_B|x_C)$ for each possible value x_A of X_A x_B of X_B , x_C of X_C.
In the hypothesis of multivariate gaussian variables , $f = N(\mu,\Sigma)$, $K = \Sigma^{-1}$, if V represents the set of this variables, it is possible to define a dependence graph $G_K =(V,E_K)$ with $k_{u,v}$ =0 whenever there is no edge between a pair of vertices $u,v \in V$ .So if two vertices $u,v \in V$ are not adjacent, that's to say there's no edge, it holds that $u \perp v | V \setminus \{u,v\}$.
The choice here is that V represents the set of variables Y given by r_i , $i = 1,...,k$ followed by r_i^{lag1} , $i = 1,...,k$ with $r_i = (r_{i,1},..., r_{i,T})$, $r_i^{lag1} = (r_{i,1}^{lag1}, \dots, r_{i,T}^{lag1})$, $f = N(\mu,\Sigma)$.

The first method to select an undirected gaussian graphical model is to specify a threshold a for the partial correlations, defined as $\rho_{uv|V\setminus\{u,v\}} = -k_{uv} / (k_{uu} k_{vv})^{1/2}$, so edges are not present for all u,v : $|\rho_{uv|V\setminus\{u,v\}}| < \alpha$.
The special form of threshold here is the following: consider the set of hypotheses $H = \{ H_{u,v} : Y_u \perp Y_v | Y_{V\setminus\{u,v\}} \}$ and let $P = \{p_{u,v}\}$ be the p–values. Using Fisher's z transformation, P is transformed in simultaneous p–values $\underline{P} = \{\underline{p}_{u,v}\}$. So if we reject $H_{u,v}$ whenever $\underline{p}_{u,v} < \alpha$, the probability of rejecting one or more true $H_{u,v}$ is less or equal to α.
It follows that the graph G_α which include those edges with $\underline{p}_{u,v} < \alpha$ has a probability of not being the true one less or equal to α.
As this approach attempts to include only the edges that are definitely present, it empirically proved to give a lot of unconnected vertices. For this reason other approaches are applied too.

The second method to select an undirected gaussian graphical model is to maximize a log-likelihood for K which penalized by $|K|$:

$$L_{pen}(K) = log(det(K)) - tr(KS) - \rho|K| \tag{31}$$

where $det(K)$ is the determinant of K,
$S_{u,v} = \sum_{t=1,..n} (y_{u,t} - y_v)(y_{v,t} - y_v)' / T$, $tr(KS)$ is the trace of KS, $|K| = \sum_{u,v \in V} |k_{uv}|$, $\rho \geq 0$.

The the number of non-zero elements of K is penalized by the last term in L_{pen}. So the smaller the ρ value the denser the graph that results. In this way we can have more connected vertices than the first method with a relatively fast calculation.

The third method to select an undirected gaussian graphical model, that can deal with high dimension in a better way than the approaches before, is to restrict the search among the decomposable graph.
Decomposable models are graphs that are triangulated. Triangulated graphs have no chordless cycles of length ≥ 4. A cycle is a sequence of nodes $v_1, v_2, \ldots, v_{n-1}, v_1$ of length n. If a cycle has adjacent, elements, that's to say with an edge between them , v_i-v_j and $j \notin \{i-1, i+1\}$ then it is said to have a chord. If it has no cords it is said chordless.
In this case the starting graph is the minimal AIC forest. A forest is an acyclic undirected graph, that is, an undirected graph with no cycle. Then the search the edge giving the minimal AIC reduction is added until further reduction are not possible. Here AIC = $-2logL+$
$2dim(G)$, $dim(G)$ is the number of independent parameters of G, L is the likelihood.

Any method gives a final graph $\underline{G}^{(2k)}$ of $2k$ nodes that can be represented by its adjacency matrix. The adjacency matrix is an $2k \times 2k$ symmetric matrix A of 0,1 elements where $a_{i,i}$ =0 (no loop), $a_{i.j}$ = 1 if there's an edge between vertex i and j, $a_{i,j}$ = 0 otherwise.
To get only the connections among the variables r_i, $i = 1,\ldots,k$ and r_i^{lag1}
$i = 1,\ldots,k$, the subgraph $\underline{G}^{(k)} \subset \underline{G}^{(2k)}$ of k nodes identified by the adjacency matrix A_0, where A_0 is the $k \times k$ given by the first k rows of A and the last k columns of A.
Then it is possible to split the set of variables $\Omega_t = \{r_i : i = 1,\ldots,k\}$ in two parts:

1. a subset of $m < k$ variables C = $\{r_i : i = i_1,\ldots,i_m\}$ which are connected to some element of $\Omega_{t-1} = \{ r_i^{lag1} : i = 1,\ldots,k \}$ with the exception of their own past;
2. a subset of $k-m$ variables $\underline{C} = \{r_j : j = j_1,\ldots,j_{k-m}\}$ which are not connected to any element of $\Omega_{t-1} = \{ r_i^{lag1} : i = 1,\ldots,k \}$, with the exception of their own past.

Let's start with the time series modeling for the subset C. In this case it is possible to apply a restricted vector autoregressive model to the variables in C, as they are connected to their past.
The following model is estimated:

$$r_t = \phi_0 + \Phi_1 r_{t-1} + u_t$$

where $r_t = (r_{1,t},\ldots,r_{m,t})'$, $u_t = (u_{1,t},\ldots,u_{m,t})'$ is a conditional gaussian error, ϕ_0 is a m-dimensional vector, Φ_1 is a $m \times m$ matrix.
The elements of Φ_1 are restricted to 0 whenever no connection was founded between pair of variables and their past with the graph approach, Eichler (2012) .
To *restrict* the *VAR* model above, it is necessary to rewrite and re parametrize it. It can be rewritten as follow:

$$\underline{Y} = BZ + U \qquad (32)$$

where $\underline{Y_t} = (r_1,\ldots,r_T)$, $r_t = (r_{1,t},\ldots,r_{m,t})'$, $Z = (Z_0,\ldots,Z_{T-1})$, $Z_t = (1, y_t)'$, $B = (\phi_0, \Phi_1)$, $U = (u_1,\ldots,u_T)$, $u_t = (u,\ldots,u_{m,t})'$.
The linear constraint for f1 can be given in the form

$$\beta = vec(B) = R\gamma + \underline{r} \qquad (33)$$

for an opportune R fixed matrix of dimension $m(m+1) \times \underline{M}$, with $\underline{M}$ the number of restrictions, $\underline{r} = 0$ a vector of dimension $m(m+1)$, γ a $\underline{M}$ vector of unknown unrestricted parameters.
The *vec* operator transforms a matrix B into a vector β by stacking its columns.
So it is possible to impose the constraints for β, by re parametrizing as follow

$$\underline{y} = vec(\underline{Y}) = (Z' \otimes I)vec(B) + vec(U) =$$

$$= (Z' \otimes I)(R\gamma + \underline{r}) + u =$$

$$= (Z' \otimes I)\, R\gamma \; + u \tag{34}$$

Thus it possible to minimize the generalized sum of squared errors $S(\gamma) = u(I \otimes \Sigma)u'$, obtaining a generalized estimator $\hat{\gamma}_g$ and so $\hat{\beta}_g$ which has smaller asymptotic variance than the unrestricted estimator $\hat{\beta}$.

As the financial data exhibit volatility cluster, that is, variance may be high for certain time periods and low for other periods, and variance jump are rare, for each $\hat{u}_i$, *i* = *1,...,m* the *t – Garch(P,0)* model is applied . To simplify the notation the index *i* of the time series is
suppressed. For each time series let $u_t = \sigma_t \varepsilon_t$ then

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \cdots + \alpha_P u_{t-P}^2 \tag{35}$$

where a standardized Student-t distribution with *n* degree of freedom is assumed.
So the parameters α_i can be estimated by maximizing recursively the conditional likelihood function $f(u_{s+1}, \dots, u_T | \alpha_0, \dots, \alpha_P, u_1, \dots, u_s)$

$$\prod_{t=s+1}^{T} \frac{\Gamma[(\nu+1)/2]}{\Gamma(\frac{\nu}{2})\sqrt{(\nu-2)\pi}} \; \frac{1}{\sigma_t} \left[1 + \frac{u_t^2}{(\nu-2)\sigma_t^2}\right]^{-(\nu+1)/2} \tag{36}$$

For the time series belonging to <u>C</u> the variables are connected at most to their own past.
So *m – k* unidimensional $AR(p_i)$ model are separately applied:

$$r_{i,t} = \phi_{i,0} + \phi_{i,1} r_{i,t-1} + \cdots + \phi_{i,p_i} r_{i,t-p_i} + u_{i,t} \tag{37}$$

The parameters $\phi_{i,0}, \phi_{i,1} r_{i,t-1}, \dots, \phi_{i,p_i}$ are estimated by maximizing the conditional log-likelihood or equivalently the following

$$-\sum_{t=p_i+1}^{T} \frac{(r_{i,t} - \phi_{i,0} - \phi_{i,1} r_{i,t-1} - \cdots - \phi_{i,p_i} r_{i,t-p_i})}{2\sigma_i^2} \tag{38}$$

$\hat{u}_i$, *i* = *1,...,m* the *t – Garch(P,0)* model is applied .

At last the estimation of $r_{i,T+1}$ is $E(r_{i,T+1}|\Omega_T)$ accordingly to the preceding modeling

$$\hat{r}_{i,T+1} = \begin{cases} \hat{\phi}_{i,0} + \hat{\phi}_{i,1} r_{i,T} + \cdots + \hat{\phi}_{i,p_i} r_{i,T-p_i} & if \;\; r_i \in \underline{C} \\ \\ \hat{\phi}_{i,0} + \hat{\phi}_{i,11} r_{i,T} + \cdots + \hat{\phi}_{i,k1} r_{k,T} & if \;\; r_i \in C \end{cases} \tag{39}$$

The estimation of $\sigma^2_{i,T+1} = var(r_{i,T+1}|\Omega_T)$ is

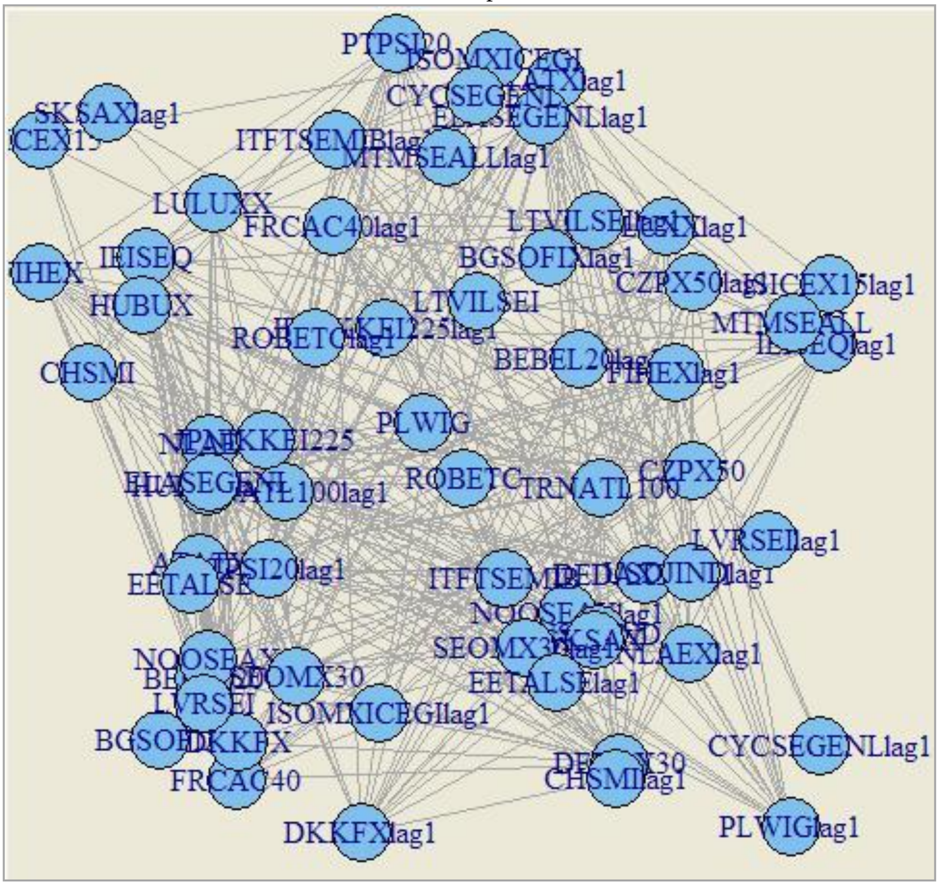
$$\hat{\sigma}^2_{i,T+1} = \hat{\alpha}_{i,0} + \hat{\alpha}_{i,1} u^2_{i,T} + \cdots + \hat{\alpha}_{i,P_i} u^2_{i,T-P_i} \tag{40}$$

Let V be the conditional variance covariance matrix of $r_{i,t}$ with dimension $k \times k$.
The diagonal elements of V are estimated by the *tGarch* models.
For the estimation of the off-diagonal elements of V, it is assumed a time-constant association and a multivariate Factor model is applied as in equation (18), (19) but with 2 latent factors.
Resampling of $r_{i,t}$ is done for $g \in \boldsymbol{N}$ times and the empirical distribution of the g iteration optimal portfolio return $R_p^{(g),opt}$ is built. Thus the probability of $R_p^{(g),op} > 0$ is estimated.

The combination of *GM* and *(V)AR-tGarch*, *LFM* models is applied to the time series of the European Stock Indexes monthly figures from 1 January 1995 to 31 December 2012 taken by the Eurostat database.
As the EU indexes are in number of 31, each graph has 62 nodes (present and past).
The SIN approach (α = 0,3) gives only few (two if g = *1*) edges that connect the past with the present variables, as expected. For the Glasso (ρ = 0,4) and the Decomposable search , **Figure 26**, there are a lot of edges that makes impossible to detect by eye the relationships among the variables. The empirical distributions of optimal portfolio returns given by the procedures are calculated for g = 500 iterations of resampling.
Thus the analysis shows a probability of gain between 56% and 91% depending on the graphical method used. The structure of dependence among the variables and their past seems to play a key role in portfolio making, changing the results of the forecast. However the SIN approach pushes the successive modeling towards a unidimensional prevalence. At the contrary the Glasso and Decomposable Search push the successive modeling towards a multidimensional prevalence.
So it seems more appropriate Glasso and Decomposable Search approaches to get multidimensionality into the analysis.

Figure 26:gaussian graphical model

Glasso and Decomposable Search



7.3 Mixture of Gaussian distribution

As noticed before, the distributional aspect is the other main problem one must deal with stock portfolio selection.
To get the normality we postulate that the whole dataset consists of a number of clusters each having different variance-covariance matrix and mean with multivariate normal density function.
This hypothesis results in a finite mixture of gaussian density for the whole dataset (Everitt, Hand 1981)

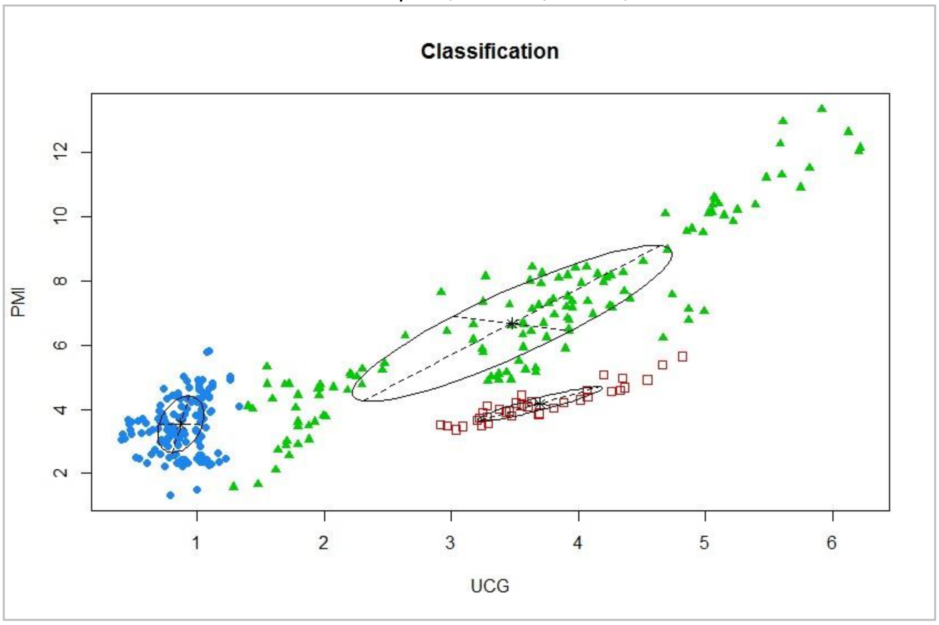
$$f(u_t, p, \theta) = \sum_{j=1}^{c} p_j g_j(u_t, \mu_j, \Sigma_j) \qquad (41)$$

where $\theta = (\mu_1, \Sigma_1, \dots, \mu_c, \Sigma_c)$ and $p = (p_1, \dots, p_{c-1})$ with $\sum_{j=1}^{c} p_j = 1$.
The parameters (θ, p) can be estimated with log-likelihood maximization through EM or MCMC (Marin et al 2005). Then each observation can be associated to a cluster on the basis of the maximum value of the following estimated probability , *j=1,...,c*

$$\hat{P}(j|u_i) = \frac{\hat{p}_j g_j(u_i, \hat{\mu}_j, \hat{\Sigma}_j)}{f(u_j, \hat{p}, \hat{\theta})} \qquad (42)$$

However, this new approach is tentatively applied to bidimensional stock vector r_t = [PMI_t , UCG_t] (*PMI = Popolare di Milano, UCG = Unicredit*) and in a successive effort will be applied to the hole set of 150 stocks. There is the possibility to lose the time dependency, even if the experiments show that the clusters are almost formed by consecutive observations so representing the states of the system. This could be no longer true for bigger dimensions.
However the successive application of the CVAR modeling in each cluster separately has shown a smaller AIC than the whole CVAR modeling. So the clustering seems a promising approach.
In the **Figure 27** *c = 3* with a bidimensional u_t and r_t = [PMI_t ; UCG_t]:

Figure 27: Gaussian Finite Mixture
for r_t = [PMI_t , UCG_t]



7.5 Multivariate Student-t conditional distribution

A multivariate Student-t distribution is tentatively used as the conditional distribution of u_t (Fiorentini et al 2003) for the bidimensional u_t and r_t = [PMI_t ; UCG_t] (Popolare di Milano, Unicredit) in order to check experimentally the difference with a Gaussian conditional distribution of u_t In this case the conditional distribution of $\varepsilon_t = \Sigma_{t|t-1}^{-1/2} u_t$ is

$$f_\nu(\varepsilon_t) = \frac{\Gamma(\frac{\nu+n}{2})}{\Gamma\left(\frac{\nu}{2}\right)(\pi(\nu-2))^{n/2}}\left(1+\frac{\varepsilon_t{}'\varepsilon_t}{\nu-2}\right)^{-(\nu+n)/2} \quad (43)$$

where $\nu > 2$, Γ is the gamma function and the log likelihood is

$$\sum_{t=1}^{T} l_t(\theta,\nu)\, , l_t(\theta,\nu) = \ln\left|\Sigma_{t|t-1}(\theta)\right|^{-1/2} + \ln\left(f_\nu(\Sigma_{t|t-1}^{-1}(\theta)u_t)\right) \quad (44)$$

The estimated variance covariance matrix $\Sigma_{T+1|T}(\theta)$ at time *T+1* , indicated as $(\hat{\sigma}_{i,j})_{i,j=1,2}$, with the maximization of (29), (30) are shown in the following **table 7**

Table 7: estimation of variance covariance matrix $\Sigma_{T+1|T}(\theta)$

		Student-t		Gauss		Relative difference
Log-likelihood		808,2082		802,2919		0,0074
$\hat{\sigma}_{1,1}$	$\hat{\sigma}_{1,2}$	0,0100	0,0042	0,0099	0,0038	
$\hat{\sigma}_{2,1}$	$\hat{\sigma}_{2,2}$	0,0042	0,0072	0,0038	0,0071	

It can be seen from the table above that the relative difference in log-likelihood is of 0,7% . So it is possible to conclude that there is no difference in the use of the two kind of conditional distributions in getting the leptokurtosis of the data. Moreover there is almost no difference in the estimation of the variance covariance matrix too.

Bauwens and Laurent (2005) found that the multivariate Student t density provides better, or at least not worse, out-of-sample VaR forecasts (Value at Risk, that is the maximal forecasted loss of a financial position) than a symmetric density.
So there is no evident improvement in using the Student t instead of Gaussian conditional distribution for u_t .

7.6 Resampling and asymptotic test statistic distributions

Asymptotic theory is applicable to the properties of the test statistics used here. As the observations range from about 200 to 400, the use of asymptotic theory, as is usual in similar problems (Lutkephol, 2010), is justified.
A rigorous treatment of the properties of the statistics is not, however, available at present. This difficulty can be overcome through simulation of the properties, a method known as *resampling* or *bootstrap*, which can be used to investigate the distribution of functions of the time series involved.
The values $y_1,\ldots, y_T$ of the time series and a pre-sample are available.The model (1) is fitted in the first step and the coefficient estimates $\hat{\mu}, \hat{A}_1, \ldots, \hat{A}_p$ and a series of residuals $\hat{u}_1, \ldots, \hat{u}_T$ are obtained.

The quantity of interest is the function $q = q(\mu, A_1, \ldots, A_p) = \tilde{q}_T(y_1, \ldots, y_T) = \tilde{\tilde{q}}_T(u_1, \ldots, u_T)$ which is estimated with
$\hat{q} = q(\hat{\mu}, \hat{A}_1, \ldots, \hat{A}_p)$.
The *bootstrap* residuals $\hat{u}_1^*, \ldots, \hat{u}_T^*$ are obtained in the second step by means of random extraction with repetition from the set of the residuals obtained in the first step $\hat{u}_1, \ldots, \hat{u}_T$.
The bootstrap time series are then computed recursively in the third step as

$$y_t^* = \hat{\mu} + \hat{A}_1 y_{t-1}^* + \cdots + \hat{A}_p y_{t-p}^* + u_t^* \qquad (45)$$

for *t=1,...,T.*
The parameters $\mu, A_1, \ldots, A_p$ are re-estimated in the forth step using the bootstrap time series of the third step, say $\hat{\mu}^{(1)}, \hat{A}_1^{(1)}, \ldots, \hat{A}_p^{(1)}$.
A bootstrap statistic, say $\hat{q}^{(1)}$, is calculated in the fifth step by using the parameter estimates obtained in the forth : $\hat{q}^{(1)} = q(\hat{\mu}^{(1)}, \hat{A}_1^{(1)}, \ldots, \hat{A}_p^{(1)})$.
The steps 2-5 are repeated *N* times and the distribution *f* of the function $q = q(\mu, A_1, \ldots, A_p)$ is investigated by using the empirical distribution $\hat{f}$ of the bootstrap statistic $(\hat{q}^{(1)}, \ldots, \hat{q}^{(N)})$.

In order to understand the quantile of the test statistics used here, let us consider the bi-dimensional time series $r_t = [PMI_t \; ; \; UCG_t]$, that is the stocks Popolare di Milano and Unicredit.
The idea is to compare the bootstrap quantile with the asymptotic theoretical quantile of the test statistic in order to see if there is difference that would result in a different decision in the test of interest.
In order to lighten the computational complexity, parallel computation is employed with 2 CPUs at 1.8GHz.

The first test statistic of interest is the normality test or *Jarque-Bera test* statistic. In this case, the function q for each time series is as follows:

$$JB = \frac{\hat{S}^2(u)}{6/T} + \frac{\left(\hat{K}(u) - 3\right)^2}{24/T} \qquad (46)$$

where $\hat{S}(u), \hat{K}(u)$ are the sample skewness and kurtosis and $u = (u_1, \ldots, u_T)$ and *T* the time series length.
In the case of H_0: normality of u, the asymptotic distribution of the test statistic *JB* is a chi-square with 2 degrees of freedom.

The comparison for the pair of stocks between the asymptotic and the bootstrap 95% quantile of *JB* is given in the following **Table 8,** where the bootstrap residuals u are obtained with *N=1000* and the VAR model equation (1):

Table 8: Resampling and asymptotic JB 95% quantile

	JB statistic	*JB* sim. q.95%	*JB* asy. q. 95%
PMI	290.0622	5.5090	5.7780
UCG	12.3813	5.9971	5.7780

It can be seen from table 8 that there is no substantial difference between the asymptotic and the simulated *JB* quantile.
The decision as regards the hypothesis H_0 is therefore the same in both cases: the *JB* statistic is greater than the asymptotic and the simulated 95% quantile and the null hypothesis of normality is rejected for both the stocks with a type I error of 0.05. A multivariate Garch model is therefore needed to obtain the unconditional non-normality of *u*.

The second test statistic of interest is the correlation test or *Ljung-Box test* statistic. In this case, the function q for each time series is as follows:

$$Q_m = T(T+2)\sum_{i=1}^{m}\frac{\hat{\rho}_i^2}{T-i} \tag{47}$$

where $\hat{\rho}_i$ is the lag-*i* sample autocorrelation of u and $u = (u_1, \dots, u_T)$, T the time series length, m =20.
In the case of H_0: no autocorrelation of u up to lag m, the asymptotic distribution of the test statistic Q_m is a chi-square with m degrees of freedom.
The comparison for the pair of stocks between the asymptotic and the bootstrap 95% quantile of Q_m is given in the following **Table 9,** where the bootstrap residuals u are obtained with *N=1000* and the VAR model equation (1):

Table 9: Resampling and asymptotic Q_m 95% quantile

	Q_m statistic	Q_m sim. q.95%	Q_m asy. q. 95%
PMI	20.0354	33.5425	31.4104
UCG	15.7584	32.3931	31.4104

It can be seen from the table 9 that there is no substantial difference between the asymptotic and the simulated Q_m quantile.
The decision as regards the hypothesis H_0 is therefore the same in both cases: the Q_m statistic is smaller than the asymptotic and the simulated 95% quantile and the null hypothesis of no autocorrelation cannot be rejected for both the stocks with a type I error of 0.05. This means that the VAR model adequately represents the time dependency of the two stocks.

The third test statistic of interest is the *ARCH effects* test or *Lagrange multiplier* test statistic of *Engle*.
It should be noted that the Ljung–Box statistic Q_m to the u^2 series could be used to detect ARCH effects, say $Q_{m,2}$. Both test statistics are therefore simulated.
In the case of the Lagrange multiplier, the function q for each time series is the F statistic for testing the null hypothesis H_0 : $\alpha_1 = ... = \alpha_m = 0$ (no ARCH effects for u) in the auxiliary linear regression

$$u_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \cdots + \alpha_m u_{t-m}^2 + e_t \qquad (48)$$

where e_t is an error term, $t=m+1,...,T$. The F statistic has a chi-squared asymptotic distribution with m degree of freedom.
The comparison for the pair of stocks between the asymptotic and the bootstrap 95 % quantiles of $Q_{m,2}$ and F are given in the following **Table 10,** where the bootstrap residuals u are obtained with $N=1000$ and the VAR model equation (1):

Table 10: Resampling and asymptotic $Q_{m,2}$, F 95% quantile

	$Q_{m,2}$ statistic	$Q_{m,2}$ sim. q.95%	F statistic	F sim. q. 95%	$Q_{m,2}$,F asy. q. 95%
PMI	67.2705	31.3891	66.4221	29.6457	31.4104
UCG	44.6883	30.8424	35.3372	29.4828	31.4104

It can be seen from table 10 that there is no substantial difference between the asymptotic and the simulated $Q_{m,2}$, F quantiles.
The decision as regards the hypothesis H_0 is therefore the same in both cases: the $Q_{m,2}$, F statistics are larger than the asymptotic and the simulated 95% quantiles and the null hypothesis of no ARCH effects for u is rejected for both the stocks with a type I error of 0.05. This means that a multivariate Garch model is needed.

The fourth test statistic of interest is the *cointegration* test of *Johansen*. It should be noted that the *Engle-Granger* statistic could also be used to detect cointegration. Both test statistics are therefore simulated.
In the Johansen case the function q for a cointegration rank $\tilde{m}$ is as follows:

$$LK_{\tilde{m}} = (p - T) \sum_{i=\tilde{m}+1}^{k} \ln(1 - \hat{\lambda}_i) \quad (49)$$

where $\hat{\lambda}_i$ are the solution of a particular eigenvalue problem. Its distribution is non-standard but it is tabulated in Johansen, Juselius (1990).
The comparison for the pair of stocks between the asymptotic and the bootstrap 95 % quantile of $LK_{\tilde{m}}$ is given in the following **Table 11,** where the bootstrap residuals u are obtained with *N=1000* and the VAR model equation (1):

Table 11: Resampling and asymptotic $LK_{\tilde{m}}$ 95% quantile

$\tilde{m}$	$LK_{\tilde{m}}$ statistic	$LK_{\tilde{m}}$ asy. q.95%	$LK_{\tilde{m}}$sim. q. 95%
0	123.1729	15.4948	147.4535
1	40.2240	3.8415	61.35

It can be seen from the table 11 that there is substantial difference between the asymptotic and the simulated $LK_{\tilde{m}}$ quantile.
If the asymptotic theory is applied, the $LK_{\tilde{m}}$ statistic for $\tilde{m} = 0$ is larger than the $LK_{\tilde{m}}$ critical value so that the null hypothesis of rank $\tilde{m} = 0$, i.e. no cointegration, is rejected. $\tilde{m} = 1$ is also rejected. Therefore $\tilde{m} = 2$.
If empirical simulation is applied, however, the $LK_{\tilde{m}}$ statistic for $\tilde{m} = 0$ is smaller than the $LK_{\tilde{m}}$ simulated critical value so that the null hypothesis of rank $\tilde{m} = 0$, i.e. no cointegration but integration, cannot be rejected. In the latter $\tilde{m} = 0$ but in the former $\tilde{m} = 2$.
This different in decision as regards cointegration could be due to the Gaussian assumption of the $LK_{\tilde{m}}$ statistic test.

The problem is, however, better solved by means of the *Engle-Granger* statistic.
In the Engle-Granger case, the following regression is estimated:

$$y_{1t} = \beta_1 y_{2t} + z_t \quad (50)$$

where $\beta = (1,-\beta_1)$ is the cointegrating vector, z_t is the error term.
If z_t is stationary whenever y_t is integrated, then y_{1t} and y_{2t} are cointegrated.
The time series z_t is stationary if it has constant mean and time-independent lag-covariances. If z_t is assumed to follow an AR(p) process, Dickey Fuller (1981) suggests testing the null hypothesis $H_0 : \pi = 0$ (unit root) against the alternative $H_1 : \pi < 0$ (stationary) by estimating the following linear regression by means of OLS (ordinary least-squares) and a Student *t* ratio, say t_{cadf}

$$\Delta z_t = \pi z_{t-1} + \sum_{j=1}^{p-1} \Delta z_{t-j} + e_t \tag{51}$$

In this case, the function q is the usual Student t ratio even though its distribution is not Student t but non-standard and should also take into consideration the fact that z_t are estimated values. It is, however, tabulated in Dickey Fuller (1981).
The comparison for the pair of stocks between the asymptotic and the bootstrap 95 % quantile of t_{cadf} is given in the following **Table 12,** where the bootstrap residuals u are obtained with $N=1000$ and the VAR model equation (1):

Table 12: Resampling and asymptotic t_{cadf} 95% quantile

$\widetilde{m}$	t_{cadf} statistic	t_{cadf} asy. q.95%	t_{cadf} sim. q. 95%
1	-11.9254	-3.3563	-10.8860

It can be seen from table 12 that there is substantial difference between the asymptotic and the simulated t_{cadf} quantile.
In this case, however, the decision for the null hypothesis $H_0 : \pi = 0$ (no cointegration) as against the alternative $H_1 : \pi < 0$ (cointegration) is the same in both the asymptotic and the simulated cases. The statistic value is in fact smaller than the critical value in both cases and the null hypothesis is therefore rejected.
The decisions of the Engle-Granger test and the above Johansen test are respectively cointegration and no cointegration. The Johansen test thus appears to decide for cointegration less frequently than empirically required.
The simulated and asymptotic decisions are the same when the Engle-Granger test is used.

The fifth and sixth test statistics of interest are the correlation test for the BEKK residuals and BEKK squared-residuals or *Ljung-Box test* statistic for BEKK correlation and ARCH effects, say $\tilde{Q}_m$ and $\tilde{Q}_{m,2}$. In this case the functions q for each time series are the same as before ($Q_m$and $Q_{m,2}$) but the residuals u_t must be replaced with $\varepsilon_t = \Sigma_{t|t-1}^{-1/2} u_t$ in the formula.

The comparison for the pair of stocks between the asymptotic and the bootstrap 95 % quantiles of $\tilde{Q}_m$ and $\tilde{Q}_{m,2}$ are given in the following **Table 13,** where the bootstrap residuals ε are obtained with $\tilde{N}$=*500* and the BEKK model equation (8):

Table 13: Resampling and asymptotic $\tilde{Q}_m$ and $\tilde{Q}_{m,2}$ *95% quantile*

	$\tilde{Q}_m$ Statistic	$\tilde{Q}_m$ sim. q.95%	$\tilde{Q}_m$, $\tilde{Q}_{m,2}$ asy. q.95%	$\tilde{Q}_{m,2}$ statistic	$\tilde{Q}_{m,2}$ sim. q.95%
PMI	13.3723	33.2883	31.4104	25.2894	32.2578
UCG	13.9801	31.2223	31.4104	25.3153	30.3538

It can be seen from the table 13 that there is no substantial difference between the asymptotic and the simulated for both $\tilde{Q}_m$ and $\tilde{Q}_{m,2}$ quantiles.

The decision as regards the hypothesis H_0 is therefore the same in both cases: the $\tilde{Q}_m$ and $\tilde{Q}_{m,2}$ statistics are smaller than the asymptotic and the simulated 95% quantiles and the null hypothesis of $H_0^{(a)}$: no correlation and $H_0^{(b)}$: no Arch effects for ε cannot be rejected for both the stocks with a type I error of 0.05 . This means that the used multivariate Garch model employed adequately represents the dynamic of the dataset.

All in all, it appears that with regard to the type of test used, the simulated and asymptotic critical values are the same, or at least that the decisions based on them are the same.

7.7 Full and proposed model comparison

The last aspect of interest as regards the model put forward is matching with the full VAR–BEKK model (Lutkepohl, 2007) applied to the multivariate time series of interest as a whole.
The latter entails a considerable computational burden. As a result, BEKK is unfeasible for a dimension superior to 5 (Ding, Engle, 1994) and cannot be applied to the universe of 150 time series involved here.

Let us, however, now consider a comparison for the tri-dimensional time series r_t = [PMI_t ; UCG_t, $Market_t$], that is the Popolare di Milano and Unicredit stocks and the Bank Market Index.
The Bank Market Index is obtained as the mean of five blue chip bank stocks, namely Unicredit, Popolare di Milano, Credito Emiliano, Intesa San Paolo and Mediobanca.

The full model VAR–BEKK for $\tilde{r}_t$= [PMI_t ; UCG_t] consists of the following equations (52)-(53)-(54) :

$$r_t = \begin{bmatrix} PMI_t \\ UCG_t \\ M_t \end{bmatrix} = \begin{bmatrix} a_{11,1} & a_{12,1} & a_{13,1} \\ a_{21,1} & a_{22,1} & a_{23,1} \\ a_{31,1} & a_{32,1} & a_{33,1} \end{bmatrix} \begin{bmatrix} PMI_{t-1} \\ UCG_{t-1} \\ M_{t-1} \end{bmatrix} + \cdots \tag{52}$$

$$\dots + \begin{bmatrix} a_{11,p} & a_{12,p} & a_{13,p} \\ a_{21,p} & a_{22,p} & a_{23,p} \\ a_{31,p} & a_{32,p} & a_{33,p} \end{bmatrix} \begin{bmatrix} PMI_{t-p} \\ UCG_{t-p} \\ M_{t-p} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{bmatrix}$$

$$u_t = \Sigma_{t|t-1}^{1/2} \varepsilon_t \tag{53}$$

$$\Sigma_{t|t-1} = \begin{bmatrix} \sigma_{11,t} & \sigma_{12,t} & \sigma_{13,t} \\ \sigma_{21,t} & \sigma_{22,t} & \sigma_{23,t} \\ \sigma_{31,t} & \sigma_{32,t} & \sigma_{33,t} \end{bmatrix} =$$

$$\begin{bmatrix} d_{0,11} & 0 & 0 \\ d_{0,21} & d_{0,22} & 0 \\ d_{0,31} & d_{0,32} & d_{0,33} \end{bmatrix} \begin{bmatrix} d_{0,11} & d_{0,21} & d_{0,31} \\ 0 & d_{0,22} & d_{0,32} \\ 0 & 0 & d_{0,33} \end{bmatrix} + \tag{54}$$

$$+\begin{bmatrix} d_{1,11} & d_{1,12} & d_{1,13} \\ d_{1,21} & d_{1,22} & d_{1,23} \\ d_{1,31} & d_{1,32} & d_{1,33} \end{bmatrix} \begin{bmatrix} u_{1,t-1}^2 & u_{1,t-1}u_{2,t-1} & u_{1,t-1}u_{3,t-1} \\ u_{2,t-1}u_{1,t-1} & u_{2,t-1}^2 & u_{2,t-1}u_{3,t-1} \\ u_{3,t-1}u_{1,t-1} & d_{1,32} & u_{3,t-1}^2 \end{bmatrix} \cdot$$

$$\cdot \begin{bmatrix} d_{1,11} & d_{1,12} & d_{1,13} \\ d_{1,21} & d_{1,22} & d_{1,23} \\ d_{1,31} & d_{1,32} & d_{1,33} \end{bmatrix} +$$

$$+\begin{bmatrix} d_{2,11} & d_{2,12} & d_{2,13} \\ d_{2,21} & d_{2,22} & d_{2,23} \\ d_{2,31} & d_{2,32} & d_{2,33} \end{bmatrix} \begin{bmatrix} \sigma_{11,t-1} & \sigma_{12,t-1} & \sigma_{13,t-1} \\ \sigma_{21,t-1} & \sigma_{22,t-1} & \sigma_{23,t-1} \\ \sigma_{31,t-1} & \sigma_{32,t-1} & \sigma_{33,t-1} \end{bmatrix} \cdot$$

$$\cdot \begin{bmatrix} d_{2,11} & d_{2,12} & d_{2,13} \\ d_{2,21} & d_{2,22} & d_{2,23} \\ d_{2,31} & d_{2,32} & d_{2,33} \end{bmatrix}$$

where $a_{ij,k}$ and $d_{ij,k}$ are parameters, $\varepsilon_t \sim N(0, I_3)$

The proposed VAR–BEKK model for $\tilde{r}_t$= [PMI_t ; UCG_t] consists instead of the following (55)-(60):

$$\begin{bmatrix} PMI_t \\ M_t \end{bmatrix} = \begin{bmatrix} \tilde{a}_{11,1} & \tilde{a}_{12,1} \\ \tilde{a}_{21,1} & \tilde{a}_{22,1} \end{bmatrix} \begin{bmatrix} PMI_{t-1} \\ M_{t-1} \end{bmatrix} + \cdots$$

(55)

$$\dots + \begin{bmatrix} \tilde{a}_{11,p_1} & \tilde{a}_{12,p_1} \\ \tilde{a}_{21,p_1} & \tilde{a}_{22,p_1} \end{bmatrix} \begin{bmatrix} PMI_{t-p_1} \\ M_{t-p_1} \end{bmatrix} + \begin{bmatrix} v_{1t} \\ v_{2t} \end{bmatrix}$$

$$\begin{bmatrix} UCG_t \\ M_t \end{bmatrix} = \begin{bmatrix} b_{11,1} & b_{12,1} \\ b_{21,1} & b_{22,1} \end{bmatrix} \begin{bmatrix} UCG_{t-1} \\ M_{t-1} \end{bmatrix} + \cdots$$

(56)

$$\dots + \begin{bmatrix} b_{11,p_2} & b_{12,p_2} \\ b_{21,p_2} & b_{22,p_2} \end{bmatrix} \begin{bmatrix} UCG_{t-p_2} \\ M_{t-p_2} \end{bmatrix} + \begin{bmatrix} w_{1t} \\ w_{2t} \end{bmatrix}$$

$$v_t = \tilde{\Sigma}_{t|t-1}^{1/2} \varepsilon_t \, , w_t = \tilde{\tilde{\Sigma}}_{t|t-1}^{1/2} \varepsilon_t \, , \bar{\Sigma}_{t|t-1} = cov(\tilde{r}_{1t}, \tilde{r}_{2t} | \Omega_{t-1}) \qquad (57)$$

$$\tilde{\Sigma}_{t|t-1} = \begin{bmatrix} \tilde{\sigma}_{11,t} & \tilde{\sigma}_{12,t} \\ \tilde{\sigma}_{21,t} & \tilde{\sigma}_{22,t} \end{bmatrix} = \begin{bmatrix} \tilde{d}_{0,11} & 0 \\ \tilde{d}_{0,21} & \tilde{d}_{0,22} \end{bmatrix} \begin{bmatrix} \tilde{d}_{0,11} & \tilde{d}_{0,21} \\ 0 & \tilde{d}_{0,22} \end{bmatrix} +$$

$$+ \begin{bmatrix} \tilde{d}_{1,11} & \tilde{d}_{1,21} \\ \tilde{d}_{1,21} & \tilde{d}_{1,22} \end{bmatrix} \begin{bmatrix} v_{1,t-1}^2 & v_{1,t-1}v_{2,t-1} \\ v_{2,t-1}v_{1,t-1} & v_{2,t-1}^2 \end{bmatrix} \begin{bmatrix} \tilde{d}_{1,11} & \tilde{d}_{1,21} \\ \tilde{d}_{1,21} & \tilde{d}_{1,22} \end{bmatrix} + \quad (58)$$

$$+ \begin{bmatrix} \tilde{d}_{2,11} & \tilde{d}_{2,21} \\ \tilde{d}_{2,21} & \tilde{d}_{2,22} \end{bmatrix} \begin{bmatrix} \tilde{\tilde{\sigma}}_{1,t-1}^2 & \tilde{\tilde{\sigma}}_{12,t-1} \\ \tilde{\tilde{\sigma}}_{21,t-1} & \tilde{\tilde{\sigma}}_{2,t-1}^2 \end{bmatrix} \begin{bmatrix} \tilde{d}_{2,11} & \tilde{d}_{2,21} \\ \tilde{d}_{2,21} & \tilde{d}_{2,22} \end{bmatrix}$$

$$\tilde{\tilde{\Sigma}}_{t|t-1} = \begin{bmatrix} \tilde{\tilde{\sigma}}_{11,t} & \tilde{\tilde{\sigma}}_{12,t} \\ \tilde{\tilde{\sigma}}_{21,t} & \tilde{\tilde{\sigma}}_{22,t} \end{bmatrix} = \begin{bmatrix} q_{0,11} & 0 \\ q_{0,21} & q_{0,22} \end{bmatrix} \begin{bmatrix} q_{0,11} & q_{0,21} \\ 0 & q_{0,22} \end{bmatrix} +$$

$$+ \begin{bmatrix} q_{1,11} & q_{1,21} \\ q_{1,21} & q_{1,22} \end{bmatrix} \begin{bmatrix} w_{1,t-1}^2 & w_{1,t-1}w_{2,t-1} \\ w_{2,t-1}w_{1,t-1} & w_{2,t-1}^2 \end{bmatrix} \begin{bmatrix} q_{1,11} & q_{1,21} \\ q_{1,21} & q_{1,22} \end{bmatrix} + \quad (59)$$

$$+ \begin{bmatrix} q_{2,11} & q_{2,21} \\ q_{2,21} & q_{2,22} \end{bmatrix} \begin{bmatrix} \tilde{\tilde{\tilde{\sigma}}}_{1,t-1}^2 & \tilde{\tilde{\tilde{\sigma}}}_{12,t-1} \\ \tilde{\tilde{\tilde{\sigma}}}_{21,t-1} & \tilde{\tilde{\tilde{\sigma}}}_{2,t-1}^2 \end{bmatrix} \begin{bmatrix} q_{2,11} & q_{2,21} \\ q_{2,21} & q_{2,22} \end{bmatrix}$$

$$\bar{\Sigma}_{t|t-1} = \begin{bmatrix} \bar{\sigma}_{11,t} & \bar{\sigma}_{12,t} \\ \bar{\sigma}_{21,t} & \bar{\sigma}_{22,t} \end{bmatrix} = \begin{bmatrix} \bar{d}_{0,11} & 0 \\ \bar{d}_{0,21} & \bar{d}_{0,22} \end{bmatrix} \begin{bmatrix} \bar{d}_{0,11} & \bar{d}_{0,21} \\ 0 & \bar{d}_{0,22} \end{bmatrix} +$$

$$+ \begin{bmatrix} \bar{d}_{1,11} & \bar{d}_{1,21} \\ \bar{d}_{1,21} & \bar{d}_{1,22} \end{bmatrix} \begin{bmatrix} u_{1,t-1}^2 & u_{1,t-1}u_{2,t-1} \\ u_{1,t-1} & u_{2,t-1}^2 \end{bmatrix} \begin{bmatrix} \bar{d}_{1,11} & \bar{d}_{1,21} \\ \bar{d}_{1,21} & \bar{d}_{1,22} \end{bmatrix} + \quad (60)$$

$$+ \begin{bmatrix} \bar{d}_{2,11} & \bar{d}_{2,21} \\ \bar{d}_{2,21} & \bar{d}_{2,22} \end{bmatrix} \begin{bmatrix} \bar{\sigma}_{1,t-1}^2 & \bar{\sigma}_{12,t-1} \\ \bar{\sigma}_{21,t-1} & \bar{\sigma}_{2,t-1}^2 \end{bmatrix} \begin{bmatrix} \bar{d}_{2,11} & \bar{d}_{2,21} \\ \bar{d}_{2,21} & \bar{d}_{2,22} \end{bmatrix}$$

where $\tilde{a}_{ij,k}$, $b_{ij,k}$, $q_{ij,k}$, $\tilde{d}_{ij,k}$, $\bar{d}_{ij,k}$ are parameters, $\varepsilon_t \sim N(0, I_2)$, Ω_{t-1} is the information until the time $t-1$.

The mean of the squared differences between the T back forecasts and actual values for the two stocks in question is shown in the following **Table 14** for the VAR−BEKK 3D and the multiple VAR−BEKK 2D. In other words, the following quantities are calculated for both methods:

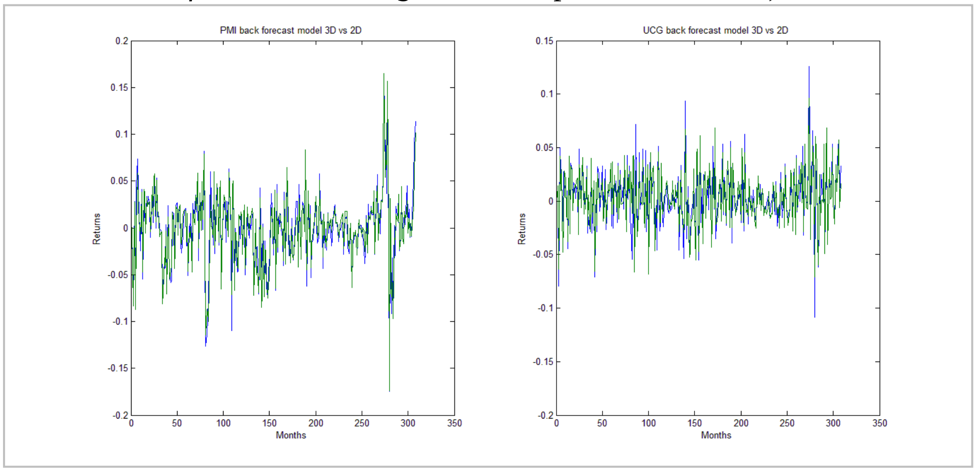
$$e_{1,mean} = \frac{1}{T}\sum_{t=1}^{T}(\hat{r}_{1t} - r_{1t})^2 \quad , e_{2,mean} = \frac{1}{T}\sum_{t=1}^{T}(\hat{r}_{2t} - r_{2t})^2 \quad (61)$$

Table 14: VAR–BEKK multi. 2D vs 3D: $e_{j,mean}$ values

	3D $e_{j,mean}$	*Multiple 2D* $e_{j,mean}$
PMI	0.0015	0.0015
UCG	0.0007695	0.0007260

It can be seen from table 14 that the mean squared errors for both cases are very similar. The return forecasts are therefore as reliable in the multiple 2D model proposed as in the 3D model. This can be also seen in the following figure:

Figure 28: VAR–BEKK multi. 2D vs 3D: PMI (left), UCG (right) forecast values green: multiple 2D, blue: 3D)



As regards variance-covariances, the following **Table 15** shows that there is a slight difference in the estimated variance of *PMI* and the covariance and a perfect match for the variance of *UCG*.

Table 15: VAR–BEKK 2D vs 3D: forecasted variance-covariance

	3D	*Multiple 2D*
$Var(PMI_t \mid \Omega_{t-1})$	0.0077	0.0058
$Var(UCG_t \mid \Omega_{t-1})$	0.0060	0.0060
$Cov(PMI_t , UCG_t \mid \Omega_{t-1})$	0.0045	0.0061

To sum up, the two types of model give very similar results for the return forecasts and similar results for the volatility forecasts.
The usefulness of the proposed model in terms of computation feasibility therefore seems to overweigh the possible slight difference in terms of forecasts.

8. Conclusions

Suitable application of the BEKK model is capable of generating time series with higher unconditional kurtosis than the normal density. In other words, the conditional distribution of errors u_t has more mass around 0 and on the tails than normal. Outliers are therefore more frequent than implied by normal random variables following the data-generating process.
The phenomenon of volatility clusters, whereby variance may be high for certain periods and low for others, is embedded in the model specification.
This approach gives a time-dependent overall estimation of the stock variances-covariances $\sigma_{k,h}(t)$ as well as of the returns $r_{k,t}$.
The computational burden of direct estimation of the original problem is solved by decomposition into feasible bi-dimensional problems without losing the strength of the *CVAR – BEKK* model.

While the BEKK model estimator does not have normal density, it is almost certainly consistent and possesses asymptotic normal density.
The asymptotic information matrix of VAR parameters and BEKK parameters is a block diagonal. The estimators of VAR and BEKK are therefore asymptotically independent. The two-step procedure adopted is therefore equivalent to the overall procedure.
If an alternative density could be reasonably assumed, ML estimators outperform the QML used here in terms of efficiency.
As the assumption of normality is not satisfied, it is not possible to provide exact forecast intervals by means of normal quantiles.
The R^2 goodness-of-fit index proved to be low for almost every CVAR model, as is customary for financial data. Explanatory variables should be found and included. The $P/E \cdot EPS$ time series is used to rank the stocks.

Employment of the bi-dimensional multiple procedure using CVAR makes the problem of estimating 150 × 150 matrices of a *VAR(p)* manageable.
It gives a description of the phenomenon by using the past stock return *i* and market index return to represent the time-dependent dynamics of the return *i*.
The 7% of cases of cointegration are also taken into due consideration.
Employment of the bi-dimensional multiple procedure using BEKK makes the resulting 25 × 25 volatility matrix estimation problem (70 × 70 volatility matrix estimation in the case of shorting) manageable.

Multiple 2D-BEKK gives a complete description of the phenomenon by using the past stock return *i* and market index return to represent the time-dependent dynamics of its diagonal element *i* and the past stock *h* and stock *k* to represent the time-dependent dynamics of its off-diagonal element *h*, *k*.

Large datasets can therefore be handled with this method.

A ranking of stocks to be included in the portfolio is obtained by using the *P/E · EPS* time series in conjunction with the multiple *CVAR - BEKK* approach.

The optimal dimension of the portfolio is found by subsequently increasing the size and simulating the efficient frontier.

The solution of the best (minimum-risk) Markowitz portfolio is finally obtained.

References

Bauwens, L. and Laurent, S. (2005). A new class of multivariate skew densities, with application to GARCH models. *Journal of Business and Economic Statistics*, 23(3), 346–354.

Bollerslev T., Engle R.F., Nelson D.B. (1994), ARCH Model, *Handbook of Econometrics*, Elsevier Science, Amsterdam, vol. IV, pp. 2959–3038.

Bollerslev T., Engle R.F., Wooldridge J.M. (1988), A Capital Asset Pricing Model with Time-Varying Covariance, *Journal of Political Economy*, vol. 96, n. 1, pp. 116–131.

Brandt M.W., Goyal A., Santa-Clara P., Stroud H. R., (2005), A simulation approach to dynamic portfolio choice with an application to learing aout return predictability, *The Review of Financial Studies*, vol.18 , n. 3, pp. 831–873 .

Berndt, E. K., Hall, B. H., Hall, R. E. & Hausman, J. A. (1974), Estimation and inference in nonlinear structural models, *Annals of Economic and Social Measurement* 3/4: 653–665.

Brown K., Reily F. (2008), *Investment Analysis and Portfolio Management*, South-Western College Publishing.

Campbell, J.Y., Chan Y.L., Viceira L.M. (2003), A Multivariate Model of Strategic Asset Allocation, *Journal of Financial Economics*, vol. 67, n.1, pp. 41–80.

Comte, F., Lieberman, O. (2003). Asymptotic theory for multivariate GARCH processes, *Journal of Multivariate Analysis* 84: 61–84.

Connor G. (1995), The three types of factor models: a comparison of their explanatory power. *Financial Analysis Journal*, 42-46.

Datastream Global Equity Indices (2008), user guide, issue 5, Thomson Reuters Ltd.

Dickey D., Fuller W. (1981), Likelihood ratio statistics for autoregressive time series with unit root, *Econometrica*, n. 49, pp. 1057-1072.

Ding, Z. , Engle, R. F. (1994), Large scale conditional covariance matrix modeling, estimation and testing, mimeo, UCSD.

Duffie D., Pan J. (1997), An overview of value at risk, *Journal of Derivatives*, pp. 7–48.

Eichler M. (2012): Graphical modelling of multivariate time series, *Probability Theory Related Fields*, Vol. 1, N. 153, pp. 233-268.

Engle R. (1982): Autoregressive conditional heteroscedasticity with estimates of the variance of the United Kingdom inflations, *Econometrica* , Vol. 1, N. 50, pp. 987-1007.

Engle R.F., Kroner, K.F. (1995), Multivariate Simultaneous Generalized ARCH, *Econometric Theory*, vol. 11, n.1, pp. 122–50.

Evans J. L., Archer S. H., (1968), Diversification and the Reduction of Dispersion: an Empirical Analysis, *The Journal of Finance*, vol. 23, pp. 761–767.

Everitt B., Hand D. (1981), Finite Mixture Distributions, Chapman & Hall, p.186.

Fiorentini G., Sentana E., Calzolari G. (2003), Maximum Likelihood Estimation and Inference in Multivariate Conditionally Heteroscedastic Dynamic Regression Models With Student t Innovations, Journal of Business & Economic Statistics

Goldfarb D., Idnani A. (1983), A numerically stable dual method for solving strictly convex quadratic programs, *Mathematical Programming*, vol. 27, pp. 1–33.

Goodman D.A., Peavy III J.W. (1983), Industry Relative Price-Earnings Ratios as Indicators of Investment Returns, *Financial Analysts Journal*, vol. 39, n. 4, 60–66 .

Hafner C., Herwartz H. (2013), Analytical quasi maximum likelihood inference in multivariate volatility models, Econometric Institute Report EI 2003–21, Econometric Institute, Erasmus University Rotterdam, Netherlands.

Hamilton J.D., (1994), *Time Series Analysis*, Princeton University Press.

Higham N., (2002), Computing the nearest correlation matrix - a problem from finance, *IMA Journal of Numerical Analysis*, vol. 22, pp. 329–343.

Hlouskova J., Schmidheiny K., Wagner M., (2002), Multistep Predictions from Multivariate ARMA-GARCH Models and their Value for Portfolio Management, *Diskussionsschriften*, Universität Bern.

Hojsgaard S., Edwards D., Lauritzen S.L. (2012): Graphical models with R, Springer.

Hosking J. R. M., (1980), The multivarate portmanteau statistic, *Journal of American Statitical Association*, vol. 75, pp. 602–608.

Johansen S., Juselius, K. (1990), Maximum likelihood estimation and inference on cointegration with applications to the demand for money, *Oxford Bulletin of Economics and Statistics*, vol. 52, pp. 169–210.

Johansen S. (1995), *Likelihood-Based Inference in Cointegrated Vector Autoregressive models*, Oxford University Press, Oxford.

Lauritzen SL. (1996): Graphical models, Calderon Press, Oxford.

Longerstaey J., More L. (1995), *Introduction to RiskMetrics 4th ed*., Morgan Guaranty Trust Comapany, New York.

Lutkepohl H., (2007), *New Introduction to Multiple Time Series Analysis*, Springer.

Marin J., Mengersen K., Roberts C. (2005), Bayesian modelling and inferences on mixtures of distributions, *Bayesian Thinking, Modeling and Computation*, Elsevier, pp. 15840-15845.

Markowitz H. (1952), Portfolio Selection, *Journal of Finance*, n.7, pp. 77–91.

Nicholson S. F., (1960), Price-earnings ratios, *Financial Analysts Journal*, vol. 16, n. 4, pp. 43-50.

Pierini Naccarato, (2012), A multivariate VEC-BEKK model for portfolio selection, *Working Paper* del Dipartimento di Economia, Roma Tre.

References

Ruppert D. (2011), Statistics and Data Analysis for Financial Engineering, Springer;

Sharpe (1970), Portfolio Theory and Capital Markets. McGraw-Hill, New York.

Tiao G., Box G. (1981): Modelling multiple time series with applications, *Journal of the American Statistical Association*, Vol. 1, N. 76, pp. 802-816.