UNIVERSITÀ DI SALERNO

DIPARTIMENTO di SCIENZE ECONOMICHE e STATISTICHE

SCUOLA DOTTORALE A.GENOVESI

DOTTORATO di RICERCA

in INGEGNERIA ed ECONOMIA dell'INNOVAZIONE

XIV CICLO

DOCTORAL THESIS

A novel approach to forecasting from non scalar DCC models

Author:

Giancarlo CESALE

Supervisor:

Prof. Giuseppe STORTI

PhD Coordinator:

Prof. Alessandra AMENDOLA

Donnie: Why are you wearing that stupid bunny suit?

Frank: Why are you wearing that stupid man suit?

Donnie Darko, 2001

Abstract

Estimating and predicting joint second-order moments of asset portfolios is of huge importance in many practical applications and, hence, modeling volatility has become a crucial issue in financial econometrics. In this context multivariate generalized autoregressive conditional heteroscedasticity (M-GARCH) models are widely used, especially in their versions for the modeling of conditional correlation matrices (DCC-GARCH). Nevertheless, these models tipically suffer from the so-called curse of dimensionality: the number of needed parameters rapidly increases when the portfolio dimension gets large, so making their use practically infeasible. Due to these reasons, many simplified versions of the original specifications have been developed, often based upon restrictive *a priori* assumptions, in order to achieve the best tradeoff between flexibility and numerical feasibility. However, these strategies may implicate in general a certain loss of information because of the imposed simplifications. After a description of the general framework of M-GARCH models and a discussion on some specific topics relative to second-order multivariate moments of large dimension, the main contribution of this thesis is to propose a new method for forecasting conditional correlation matrices in high-dimensional problems which is able to exploit more information without imposing any *a priori* structure and without incurring overwhelming calculations. Performances of the proposed method are evaluated and compared to alternative predictors through applications to real data.

Acknowledgements

Il mio ringraziamento va a tutto il personale, docente e non docente, del Dipartimento di Scienze Economiche e Statistiche dell'Università degli Studi di Salerno, e in particolare: alla coordinatrice del corso di dottorato, prof.ssa Amendola, per la fiducia concessami e per avermi così dato la possibilità di vivere esperienze di crescita intellettuale che altrimenti mi sarebbero state prescluse; alla dott.ssa Maria Rizzo per la disponibilità e la gentilezza con le quali mi è sempre venuta in aiuto nella risoluzione di problemi di ogni tipo; al dott.Antonio Naimoli per il suo costante e amichevole supporto che mi ha spesso reso meno ostiche le situazioni difficili o spiacevoli; e, sopra tutti, al supervisor, prof.Giuseppe Storti, per aver generosamente condiviso con me alcune delle sue idee, per aver sopportato con pazienza i miei limiti e senza la cui guida questa tesi non avrebbe visto la luce.

Contents

List of Figures

List of Tables

Chapter 1

Multivariate volatility models

1.1 Introduction

Since the seminal work of Engle (1982) modeling volatility has become one of the main issue in financial econometrics. Starting from the mentioned paper a large field of literature has developed that has faced dynamic modeling of univariate second-order moments of financial returns. Almost simultaneously many authors have begun to realize that returns often show comovements, therefore motivating the research on multivariate volatility models. Indeed, understanding and predicting multivariate second-order moments of financial returns is particularly crucial in practical applications, for example in asset pricing, in portfolio allocation or in risk management. In this context the so-called multivariate (M-)GARCH models were born (Bollerslev, Engle and Wooldridge [1988], Ng [1991], Engle and Kroner [1995], Hansson and Hordahl [1998], Kroner and Ng [1998])[1]. A few years later, a paper of Bollerslev (1990) and especially a new article by Engle (2002) have given birth to models to be applied not directly to the covariances but rather based on separate modeling of conditional variances and correlations. Nowadays M-GARCH models are at the same time a cornerstone and an object of further development in this field of econometrics. One of the crucial limitation of

[1]Every contribution in this field focuses on the analysis of the dynamics governing the time evolution of conditional covariance matrices

this kind of models which scholars have had and have yet to tackle is that they suffer from the so-called curse of dimensionality: the number of parameters rapidly increases when the dimension of the problem (that is the number of assets) gets larger; so the main effort of researchers has been particularly aimed to find M-GARCH models that combine flexibility and parsimony. Indeed, a good model should be flexible enough to accurately track the the dynamics of covariances and correlations, but at the same time it is also supposed to be sufficiently parsimonious: a model with too many parameters would turn out to be hard to handle and practically useless.

The chapter is organized as follows: Section 1.2 defines the general framework of multivariate GARCH models; in Section 1.3 some fundamental M-GARCH models (namely: exponentially weighted moving average (EWMA); Vectorized (VEC-) GARCH; Baba-Engle-Kraft-Kroner (BEKK-) GARCH) are discussed which can be derived as generalizations of standard univariate GARCH specifications, along with some of related typical features (curse of dimensionality, positive definiteness, stationarity, covariance targeting); Section 1.4 describes Factor Volatility (FV-) GARCH models; in Section 1.5 models for conditional correlations are presented with a particular detail on Engle's DCC and its noteworthy aspects; Section 1.6 illustrates some of more recent contributions in DCC literature (Diagonal DCC, Generalized DCC, Rotated DCC).

1.2 The M-GARCH framework

Let us consider a stochastic vector $\boldsymbol{r}_t \in \mathbb{R}^N$ representing N asset returns. We can describe the vector process as follows (see for example Tsay [2010], Chapter 10):

$$\boldsymbol{r}_t = \boldsymbol{\mu}_t + \boldsymbol{\epsilon}_t \tag{1.1}$$

where, given the information set I_{t-1}, $\boldsymbol{\mu}_t$ is the conditional expectation $E(\boldsymbol{r}_t|I_{t-1})$[2] and $\boldsymbol{\epsilon}_t$ is a vector of innovations. Denoting by $\boldsymbol{\theta}$ a vector of parameters, a multivariate GARCH model relies on the assumption that:

$$\boldsymbol{\epsilon}_t = H_t^{1/2}(\boldsymbol{\theta})\boldsymbol{z}_t \quad , \quad \boldsymbol{z}_t \sim \text{i.i.d.}(\mathbf{0}, I) \tag{1.2}$$

where $H_t^{1/2}$ is a positive definite $N \times N$ matrix and I is the $N \times N$ identity matrix. Consequently, the matrix $H_t^{1/2}$ must be a convenient decomposition[3] of the conditional variance of $\boldsymbol{r}_t$, so that:

$$\text{var}(\boldsymbol{r}_t|I_{t-1}) = \text{var}(\boldsymbol{\epsilon}_t|I_{t-1}) = H_t^{1/2}(H_t^{1/2})' = H_t. \tag{1.3}$$

More in detail, an M-GARCH model is aimed to describe H_t's dynamics and to estimate its parameters. Typical issues related to M-GARCH models are: the curse of dimensionality (*i.e.* the over-parametrization); the flexibility (*i.e.* the ability to describe the real dynamics of covariances); the positive definiteness of the estimated covariances (*i.e.* a natural property of covariances matrices). These three problems will recur repeatedly in the remainder of this chapter.

1.3 Generalizations of univariate standard GARCH models

This section presents a discussion of exponentially weighted moving average (Subsection 1.3.1), VEC (Subsection 1.3.2) and BEKK (Subsection 1.3.5) models along with a review of a related specification called Cholesky-factor multivariate GARCH (Subsection 1.3.8). These probably represent the most intuitive and straightforward generalizations of univariate GARCH to the multivariate context, even though also the so-called Factor Volatility models (which are examined in Section 1.4) are extensions of univariate conditional volatility models

[2] The conditional expectation $E(\boldsymbol{r}_t|I_{t-1})$ is tipically modeled with VARMA models (see again Tsay [2010], Chapter 8)

[3] For instance, $H_t^{1/2}$ may be an ordinary square root matrix or a Cholesky decomposition (about the Cholesky decomposition see, among others, Lutkepohl [1996], Chapter 6)

(see Bauwens, Laurent and Rombouts [2006]). In the section some of the typical related issues are also discussed: positive definiteness and stationarity conditions of VEC and BEKK (Subsections 1.3.3 and 1.3.6); quasi maximum likelihood estimation procedure for the VEC model (Subsection 1.3.4); covariance targeting (Subsection 1.3.7).

1.3.1 Exponentially Weighted Moving Average

One of the simplest and classical method of estimation of conditional time-varying covariances is the exponentially weighted moving average (EWMA) model (Riskmetrics [1996]):

$$H_t = \frac{1-\lambda}{1-\lambda^{\tau}} \sum_{i=1}^{\tau} \lambda^{i-1} \boldsymbol{\epsilon}_{t-i} \boldsymbol{\epsilon}_{t-i}^{'} \qquad 0 < \lambda < 1. \tag{1.4}$$

Since $\frac{1-\lambda}{1-\lambda^{\tau}} \sum_{i=1}^{\tau} \lambda^{i-1} = 1$, it is easy to realize that the definition (1.4) is a moving average of the previous τ sample covariances $\{\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1}^{'}, ..., \boldsymbol{\epsilon}_{t-\tau}\boldsymbol{\epsilon}_{t-\tau}^{'}\}$, where the covariances are less relevant as i (namely the time distance from the present moment t) increases: indeed, the weights involved by the moving average exponentially decline as the lag from current time t grows. An important shortfall of EWMA is that it assumes the same parameter ruling the dynamics of each entry of the conditional covariances H_t. In addition, if we set the maximum time distance τ large enough ($\lambda^{\tau-1} \approx 0$), we can rewrite (1.4) as an iterative equation which allows us for recursively estimating the sequence of covariances, given a value of λ and a starting point H_0:

$$H_t = (1-\lambda)\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1}^{'} + \lambda H_{t-1}. \tag{1.5}$$

If we define as $\boldsymbol{\vartheta}$ the vector of parameters ruling the mean equation (see Equation (1.1)) and assume that the innovations $\boldsymbol{\epsilon}_t = \boldsymbol{r}_t - \boldsymbol{\mu}_t(\boldsymbol{\vartheta})$ have a multivariate Gaussian distribution, the related log-likelihood function (up to an additive constant) is:

$$\ln L(\boldsymbol{\vartheta}, \lambda) = -\frac{1}{2}\sum_{t=1}^{T}|H_t| - \frac{1}{2}\sum_{t=1}^{T}\boldsymbol{\epsilon}_t' H_t^{-1}\boldsymbol{\epsilon}_t'. \tag{1.6}$$

Hence, once fixed the initial point H_0[4], we can jointly estimate the parameter vector $(\boldsymbol{\vartheta}, \lambda)$ by maximizing the function in (1.6) while using the relationship in (1.5). As we will see, expressions similar to (1.6) are employed to estimate the parameters also when using the more complex models that we are going to discuss hereinafter. Obviously, in case the innovation does not have a Gaussian distribution, the maximization process of (1.6) must be seen in the context of quasi maximum likelihood (see Subsection 1.3.4 about using the quasi maximum likelihood estimator in multivariate volatility models).

1.3.2 VEC model

The EWMA model seen in the previous subsection is a simple estimation method commonly used by practictioners. However it can be derived as a special case from a wider class of models which generalize the univariate GARCH to the multivariate case. The univariate classical formulation is that introduced by Engle (1982) and can be summarized as follows:

$$\begin{aligned} &E(\epsilon_t|I_{t-1}) = 0 \\ &\text{var}(\epsilon_t|I_{t-1}) = h_t = \omega + \sum_{i=1}^{p}\alpha_i\epsilon_{t-i}^2 + \sum_{i=1}^{q}\beta_i h_{t-i}. \end{aligned} \tag{1.7}$$

A natural way to extend the model in (1.7) to multivariate series of dimension N is:

[4] A natural way to set an initial matrix in Equation (1.5) may be to estimate it as the unconditional sample covariance matrix: $H_0 = \text{var}(\boldsymbol{\epsilon}_t)$

$$
\begin{aligned}
&E(\boldsymbol{\epsilon}_t|I_{t-1}) = 0 \quad , \quad \text{var}(\boldsymbol{\epsilon}_t|I_{t-1}) = H_t \\
&h_{klt} = \omega_{kl} + \sum_{i=1}^{p} \sum_{a,b}^{1,\dots N} \alpha_{klab,i}\epsilon_{a,t-i}\epsilon_{b,t-1} + \\
&\qquad\qquad + \sum_{i=1}^{q} \sum_{a,b}^{1,\dots,N} \beta_{klab,i} h_{ab,t-i}.
\end{aligned}
\tag{1.8}
$$

where h_{klt} is the generic entry of H_t. For reasons of symmetry, it must obviously be $\omega_{kl} = \omega_{lk}$, $\alpha_{klab,i} = \alpha_{lkab,i}$ and $\beta_{klab,i} = \beta_{lkab,i}$ while the terms $\alpha_{klab,i}$ and $\alpha_{klba,i}$ can be summed up like also $\beta_{klab,i}$ and $\beta_{klba,i}$. Taking into account these features, the general model in (1.8) can be rearranged in its VEC (vectorized) formulation (Bollerslev *et al.* [1988]) which makes use of the vech operator[5]:

$$
\text{vech}\, H_t = \text{vech}\, \Omega + \sum_{i=1}^{p} A_i \,\text{vech}(\boldsymbol{\epsilon}_{t-i}\boldsymbol{\epsilon}_{t-i}^{'}) + \sum_{i=1}^{q} B_i \,\text{vech}\, H_{t-i} \tag{1.9}
$$

where coherently the A_i's and B_i's are $(N(N+1)/2) \times (N(N+1)/2)$ matrices and Ω is the intercept matrix $[\omega_{kl}]$. The model (1.9) is practically employed only in the bivariate case because of its very large parametrization size, considering that the number of parameters is of order $\text{O}(N^4)$. To overcome this issue, in the same paper the authors propose a simplified version called DVEC (diagonal VEC) where Ω, A_i and B_i have not null entries only on their diagonal:

$$
h_{klt} = \omega_{kl} + \sum_{i=1}^{p} \alpha_{kl,i}\epsilon_{k,t-i}\epsilon_{l,t-i} + \sum_{i=1}^{q} \beta_{kl} h_{kl,t-i} \tag{1.10}
$$

[5] The vech operator is a map $\mathbb{R}^{N\times N} \to \mathbb{R}^{N(N+1)/2}$ that stacks all the unique entries of a symmetric matrix into a vector (see Gentle [2007], Chapter 3)

in which variances depend solely on past own values and squared residuals, while covariances depend solely on past own values and cross products of residuals. The diagonal model can be rearranged as follows[6]:

$$H_t = \Omega + \sum_{i=1}^{p} \tilde{A}_i \odot (\boldsymbol{\epsilon}_{t-i}\boldsymbol{\epsilon}_{t-i}^{'}) + \sum_{i=1}^{q} \tilde{B}_i \odot H_{t-i} \tag{1.11}$$

where $\tilde{A}_i$ and $\tilde{B}_i$ are matrices built with the diagonal elements of A_i and B_i. Nevertheless the DVEC keeps suffering from the curse of dimensionality: indeed it needs $N(1+p+q)(N+1)/2$ parameters which means a number of parameters of order $\mathrm{O}(N^2)$ at least. This is the main reason why this kind of parametrization is often used with only one lag, $p = q = 1$, and for moderate numbers of assets. Finally, note that we will obtain again the EWMA described by (1.5) if the DVEC equation in (1.11) is modified by setting $[\Omega]_{ij} = 0$, $[\tilde{A}]_{ij} = (1 - \lambda)$, $[\tilde{B}]_{ij} = \lambda$ for each pair (i, j) and $p = q = 1$.

1.3.3 Positive definiteness and stationarity conditions in VEC models

From mathematical statistics it is known that the covariance matrix operator must be positive semidefinite. In order to show that, let us consider that we have a random vector $\boldsymbol{x}$ with expected value $\bar{\boldsymbol{x}}$. The covariance of the random vector is:

$$\Sigma = E[(\boldsymbol{x} - \bar{\boldsymbol{x}})(\boldsymbol{x} - \bar{\boldsymbol{x}})^{'}]. \tag{1.12}$$

For Σ to be positive semidefinite, it must be for every vector $\boldsymbol{w}$:

$$\boldsymbol{w}^{'}\Sigma\boldsymbol{w} = E[\boldsymbol{w}^{'}(\boldsymbol{x} - \bar{\boldsymbol{x}})(\boldsymbol{x} - \bar{\boldsymbol{x}})^{'}\boldsymbol{w}] = E[(\boldsymbol{w}^{'}(\boldsymbol{x} - \bar{\boldsymbol{x}}))^2] \geq 0 \tag{1.13}$$

[6]The symbol $\odot$ denotes the element-wise (or Hadamard) matrix product (see, among others, Lutkepohl [1996], Chapter 10)

where $\boldsymbol{w}'(\boldsymbol{x} - \bar{\boldsymbol{x}})$ is a scalar random variable, say X, and the linearity of the expectation operator has been taken into account; the expected value of a squared scalar random variable must not be negative: $E(X^2) \geq 0$; moreover, $E(X^2) = 0$ if and only if $X = \boldsymbol{w}'(\boldsymbol{x} - \bar{\boldsymbol{x}}) = 0$ which means that one of the random variable x_i is a linear combination of the others $x_1, ..., x_{i-1}, x_{i+1}, ..., x_N$[7]. As regards the VEC model (1.9), sufficient conditions for H_t to be positive semidefinite for all t can be imposed on the parameter matrices (see Gourieroux [1997], Chapter 6). If we consider that by definition $h_{klt} = E(\epsilon_{kt}\epsilon_{lt}|I_{t-1})$, the general formulation of multivariate conditional covariance models in (1.8) can be rewritten as follows:

$$h_{klt} = \omega_{kl} + \sum_{i=1}^{p} \boldsymbol{\epsilon}_{t-i}' A_{kli} \boldsymbol{\epsilon}_{t-i} + \sum_{i=1}^{q} E[\boldsymbol{\epsilon}_{t-i}' B_{kli} \boldsymbol{\epsilon}_{t-i} | I_{t-1}] \tag{1.14}$$

where A_{kli} and B_{kli} are the $N \times N$ coefficient matrices relative to h_{klt}; therefore, Equation (1.14) can be expressed in a compact matrix shape[8]:

$$H_t = \Omega + \sum_{i=1}^{p} (I \otimes \boldsymbol{\epsilon}_{t-i}') A_i^* (I \otimes \boldsymbol{\epsilon}_{t-i}) + \sum_{i=1}^{q} E[(I \otimes \boldsymbol{\epsilon}_{t-i}') B_i^* (I \otimes \boldsymbol{\epsilon}_{t-i}) | I_{t-1}] \tag{1.15}$$

where as usual I denotes an $N \times N$ identity matrix and A_i^* and B_i^* are the global parameter $N^2 \times N^2$ matrices built by the blocks A_{kli} and B_{kli}. The recursive definition (1.15) can be used to verify that the H_t's are positive semidefinite if so are Ω and each A_i^* and B_i^* (as long as the starting point of the recursive equation H_0 is positive semidefinite itself); moreover, the H_t's are positive definite if at least one among Ω and each A_i^* and B_i^* is also positive definite (Gourieroux [1997], Section 6.1.2).

On the other hand, it has been demonstrated (Attanasio [1991]) that necessary and sufficient condition to guarantee the positive semidefiniteness of the conditional covariance matrices in the DVEC specification (1.11), with $p = q = 1$, is the positive semidefiniteness of each parameter matrix (Ω, $\tilde{A}$, $\tilde{B}$) and of the starting covariance matrix H_0. Indeed, given that the

[7] In other terms, this would mean that a random variable can be exactly predicted as a linear combination of two or more other random variable

[8] The symbol $\otimes$ denotes the Kronecker product, that is the product in which every entry of the left matrix is multiplied for every entry of the right matrix (see for example Horn and Johnson [1999])

outer product of innovations must be positive semidefinite, the proof descends from Schur's product theorem (namely: the Hadamard product of two positive semidefinite matrices is a positive semidefinite matrix) and from the clear fact that the sum of positive semidefinite matrices is positive semidefinite itself. In addition it is easy to show[9] that the H_t's are positive definite if also H_0 is. In order to ensure positive the semidefiniteness of the parameter matrices in (1.11), Attanasio (1991) proposes to reparameterize Ω, $\tilde{A}$ and $\tilde{B}$ *via* the Cholesky decomposition.

Another crucial issue associated with the conditional covariance modeling is the covariance stationarity, that is the independence on time and the finiteness of the unconditional covariance of the vector stochastic process: this is an important feature for the stability of the models. In Engle and Kroner (1995, Proposition 2.7) a sufficient condition for the stationarity of VEC models is proved: the condition is that all the eigenvalues of $\sum_{i=1}^{p} A_i + \sum_{i=q}^{m} B_i$ must be less than 1 in modulus (see also Gourieroux [1997], Equation 6.5). As a consequence, the unconditional covariance matrix of VEC is stated by the following relationship:

$$\operatorname{vech} H = E[\operatorname{vech} H_t] = \left(I - \sum_{i=1}^{p} A_i - \sum_{i=1}^{q} B_i \right)^{-1} \operatorname{vech} \Omega \tag{1.16}$$

Obviously the same results and relationships also hold for the DVEC specification, bearing in mind that in this case the matrices A_i and B_i are diagonal.

1.3.4 Quasi maximum likelihood estimation of M-GARCH models

The estimation of the parameters $\boldsymbol{\theta}$ ruling an M-GARCH model is usually carried out by the quasi maximum likelihood estimation procedure in which a Gaussian likelihood function is maximized, even though the related random vector of returns' residuals $\boldsymbol{\epsilon_t}$ does not have a Gaussian distribution:

[9] See for example Marcus and Minc (1964)

$$\begin{aligned}
&\boldsymbol{\epsilon}_t = H_t^{1/2}\boldsymbol{z}_t \quad , \quad \boldsymbol{z}_t \sim \text{i. i. d.}(\mathbf{0}, I) \\
&\ln L_T(\boldsymbol{\epsilon}_1, ..., \boldsymbol{\epsilon}_T; \boldsymbol{\theta}) = \sum_{t=1}^{T} l_t = -\frac{1}{2}[TN \ln 2\pi + \sum_{t=1}^{T} \ln |H_t(\boldsymbol{\theta})| + \sum_{t=1}^{T} \boldsymbol{\epsilon}_t' H_t^{-1}(\boldsymbol{\theta})\boldsymbol{\epsilon}_t] \\
&\hat{\boldsymbol{\theta}}_T^{QML} = \underset{\boldsymbol{\theta}\in\Theta}{\text{argmax}} \ \ln L_T(\boldsymbol{\epsilon}_1, ..., \boldsymbol{\epsilon}_T; \boldsymbol{\theta})
\end{aligned} \tag{1.17}$$

where Θ is the parametric space. The underlying assumption that $\boldsymbol{\epsilon}_t | I_t \sim N(\mathbf{0}, H_t)$ is not generally justified by empirical findings (for istance, *fat tails* are usually observed in the empirical distributions). However, there exist some theoretical results that can ensure consistency and asymptotic normality of $\hat{\boldsymbol{\theta}}_T^{QML}$ where the Gaussian assumption is relaxed. In detail, Hafner and Preminger (2009b) prove consistency and asymptotic normality of $\hat{\boldsymbol{\theta}}_T^{QML}$ for the VEC model: this is a very important result as many different M-GARCH specifications (BEKK, O-GARCH, GO-GARCH, FF-GARCH, *etc.*)[10] can be reviewed as special cases of the VEC formulation, which therefore turns out to be the most general framework of multivariate models for conditional covariances. Let us assume a VEC-GARCH(1,1) specification[11]:

$$\text{vech}(H_t) = \text{vech}(\Omega) + A\,\text{vech}(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t') + B\,\text{vech}(H_{t-1}); \tag{1.18}$$

hence it will be: $\boldsymbol{\theta}' = [\text{vech}(\Omega)', \text{vech}(A)', \text{vech}(B)']$. Let us define the true parametric vector as $\boldsymbol{\theta}_0$. The strong consistency[12]:

$$\text{prob}(\lim_{T\to\infty} \hat{\boldsymbol{\theta}}_T^{QML} = \boldsymbol{\theta}_0) = 1 \tag{1.19}$$

[10] Many of these M-GARCH versions are discussed in the next sections

[11] In the cited paper and here a VEC-GARCH(1,1) specification is considered for the sake of simplicity: this does not cause any loss of generality about the results

[12] The strong consistency refers to almost sure convergence: a series of random variable X_n converges almost surely to a random X if, for every $\epsilon > 0$, the probability that $\lim_{n\to\infty} |X_n - X| < \epsilon$ is equal to 1; an estimator $\hat{\phi}_n$ is defined strongly consistent if it converges almost surely to the real parameter, ϕ, as the dimension of the sample, n, goes to infinity (see Casella and Berger [2002])

is proved by Hafner and Preminger (2009b) under the following assumptions:

(A1) the parametric space Θ is compact

(A2) $|\lambda_{B,max}| < 1$ ($\lambda_{B,max}$: maximum eigenvalue of B)

(A3) the process $\boldsymbol{\epsilon}_t$ is strictly stationary in distribution and ergodic[13]

(A4) $E(\|\boldsymbol{\epsilon}_t\|^s) < \infty$[14] for some $s > 0$

(A5) $E(\|\boldsymbol{z}_t\|^2) < \infty$

(A6) $\forall \theta, \theta_0 \in \Theta$, if $H_t(\boldsymbol{\theta}) = H_t(\boldsymbol{\theta}_0)$ then $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ (model identifiability).

In order to prove the asymptotic normality of $\hat{\boldsymbol{\theta}}_T^{QML}$ the authors assume that $\tilde{H}_t$ is the covariance process whose initial values are drawn from its stationary distribution[15]. Defined $\tilde{l}_t$ as the t-th term of the related likelihood function (see Equations (1.17)), the two following matrices are introduced:

$$V = E[\frac{\partial \tilde{l}_t}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}_0)\frac{\partial \tilde{l}_t}{\partial \boldsymbol{\theta}'}(\boldsymbol{\theta}_0)] \quad , \quad J = E[\frac{\partial^2 \tilde{l}_t}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}(\boldsymbol{\theta}_0)] \tag{1.20}$$

That being said, Hafner and Preminger (2009b) prove the following convergence in distribution (asymptotic normality):

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_T^{QML} - \boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, J^{-1}VJ^{-1}) \tag{1.21}$$

under the additional assumptions:

(A7) $\boldsymbol{\theta}_0$ is an interior point of Θ

[13] In the article the authors give some conditions for the process $\boldsymbol{\epsilon}_t$ to be strictly stationary and ergodic (see Hafner and Preminger [2009b], Assumptions 2.1-2.3)

[14] The symbol $\|\ \|$ here denotes the euclidean norm: $\sqrt{\sum_{i=1}^{N} \epsilon_{it}^2}$

[15] The stationary distribution is a feature of Markov chains and, indeed, in Hafner and Preminger (2009b) the vector $[\boldsymbol{\epsilon}_t', \text{vech}(H_t)']$ is considered as a homogeneous Markov chain (see Meyn and Tweedie [2006])

(A8) $E(\|\boldsymbol{\epsilon}_t\|^8) < \infty$.

The proofs given in Hafner and Preminger (2009b) generalize many previous results reached for special cases of the VEC model like, among others, Jeantheau (1998), Ling and McAleer (2003), Comte and Lieberman (2003) and Hafner and Preminger (2009a).

1.3.5 BEKK model

A way to implicitly ensure the positive definiteness of the estimated conditional covariance matrices is to use the BEKK (Baba-Engle-Kraft-Kroner) parametrization (Engle and Kroner [1995]). The BEKK dynamic updating equation for H_t is given by:

$$H_t = C^{'}C + \sum_{k=1}^{K}\sum_{i=1}^{l} A_{ik}^{'}\boldsymbol{\epsilon}_{t-i}\boldsymbol{\epsilon}_{t-i}^{'}A_{ik} + \sum_{k=1}^{K}\sum_{i=1}^{m} B_{ik}^{'}H_{t-i}B_{ik} \tag{1.22}$$

with C upper triangular. The BEKK model involves problems about the parameter identification (uniqueness of representation) which need a bunch of restrictions (*e.g.* positivity of the entries, determination of the *right* K, etc.). In the most simple case, namely for $K = l = m = 1$, the conditions able to eliminate all other equivalent structures are formulated in Engle and Kroner (1995, Proposition 2.1): all diagonal elements of C and the upper-left entries of A and B, say a_{11} and b_{11}, must be positive. Other restrictions ensure the uniqueness of representation in more complex parametric specifications. The BEKK model can be viewed as a restricted version of VEC. In order to show that, let us take the vectorization of Equation (1.22) with $K = l = m = 1$ (for the sake of simplicity); bearing in mind that $\operatorname{vec}(XYZ) = (Z^{'} \otimes X)\operatorname{vec} Y$ (see for example Lutkepohl [1996]), Equation (1.22) becomes[16]:

$$\operatorname{vec} H_t = \operatorname{vec}(C^{'}C) + (A_1^{'} \otimes A_1^{'})\operatorname{vec}(\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1}^{'}) + (B_1^{'} \otimes B_1^{'})\operatorname{vec} H_{t-1}. \tag{1.23}$$

[16]The vec operator stacks all the elements of a $N \times N$ matrix into a vector of dimension N^2

As H_t, H_{t-1}, $CC^{'}$ and $\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1}^{'}$ are symmetric matrix, their vec transformation can be reformulated in terms of the vech operator by means of a convenient $N^2 \times N(N+1)/2$ duplication matrix D:

$$D \operatorname{vech} H_t = D \operatorname{vech}(C^{'}C) + (A_1^{'} \otimes A_1^{'}) D \operatorname{vech}(\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1}^{'}) + (B_1^{'} \otimes B_1^{'}) D \operatorname{vech} H_{t-1}. \quad (1.24)$$

The matrix D is a full rank matrix made up with 1 and 0 and admits a generalized inverse $\tilde{D}$ such that $\tilde{D}D = I$ (see for example Gentle [2007], Section 3.6):

$$\operatorname{vech} H_t = \operatorname{vech}(C^{'}C) + \tilde{D}(A_1^{'} \otimes A_1^{'}) D \operatorname{vech}(\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1}^{'}) + \tilde{D}(B_1^{'} \otimes B_1^{'}) D \operatorname{vech} H_{t-1}. \quad (1.25)$$

Equation (1.25) is clearly a VEC representation as long as $\tilde{D}(A_1^{'} \otimes A_1^{'})D = A$ and $\tilde{D}(B_1^{'} \otimes B_1^{'})D = B$ (*cf.* Equation (1.9)). This result means that, given a BEKK parametrization, there always exists an equivalent VEC representation but this one is not unique and, hence, the opposite proposition is not true; further, a tranformation which shifts a VEC specification into a BEKK one does not exist necessarily: in this sense, VEC is more general than BEKK. It is apparent that in the BEKK model the elements of the matrices A_{ik} and B_{ik} do not have a direct ecomometric interpretation; in addition to this, the model needs $N(N+1)/2 + K(l+m)N^2$ parameters as A_{ik} and B_{ik} are not symmetric. A common method to reduce the number of parameters is to force A_{ik} and B_{ik} to be diagonal.

1.3.6 Positive definiteness and stationarity conditions in BEKK models

As mentioned in the previous subsection, the BEKK representation yields positive definite covariance matrices. This is true if at least one matrix among C and B_{ik} is full rank. Indeed, all the terms $A_{ik}^{'}\boldsymbol{\epsilon}_{t-i}\boldsymbol{\epsilon}_{t-i}^{'}A_{ik}$ are surely positive semidefinite because so are $\boldsymbol{\epsilon}_{t-i}\boldsymbol{\epsilon}_{t-i}^{'}$. Further, as the product between a non singular matrix and its transpose is positive definite (see for example Horn and Johnson [1999]), for every vector $\boldsymbol{w}$ different from the null vector, it must be:

$$\begin{aligned} \boldsymbol{w}^{'}(C^{'}C+\sum_{k=1}^{K}\sum_{i=1}^{m}B_{ik}^{'}H_{t-i}B_{ik})\boldsymbol{w} = \\ \boldsymbol{w}^{'}(C^{'}C+\sum_{k=1}^{K}\sum_{i=1}^{m}B_{ik}^{'}H_{t-i}^{1/2'}H_{t-i}^{1/2}B_{ik})\boldsymbol{w} > 0 \end{aligned} \tag{1.26}$$

provided that C or one of the B_{ik}'s is full rank. That being said, if the starting covariance matrix H_0 is supposed to be positive definite, so will be all the H_t's for recursion.
As regards the covariance stationarity, the BEKK model is stationary if and only if all the eigenvalues of $\sum_{k=1}^{K}\sum_{i=1}^{l}(A_{ik}\otimes A_{ik})+\sum_{k=1}^{K}\sum_{i=1}^{m}(B_{ik}\otimes B_{ik})$ are less than one in modulus (Engle and Kroner [1995], Proposition 2.7). The unconditional covariance of a BEKK model, if it exists, is:

$$\operatorname{vech} H = E[\operatorname{vech} H_t] = \left(I-\sum_{k=1}^{K}\sum_{i=1}^{l}(A_{ik}\otimes A_{ik})-\sum_{k=1}^{K}\sum_{i=1}^{m}(B_{ik}\otimes B_{ik})\right)^{-1}\operatorname{vech}(C^{'}C) \tag{1.27}$$

1.3.7 Covariance targeting

In Pedersen and Rahbek (2013) the so-called variance targeting is discussed. Despite this discussion is carried in the BEKK framework, the variance targeting is both a property and an estimation technique which can apply also to other types of specifications provided that certain assumptions hold. In Subsection 1.3.6[17] it has been seen that the BEKK model is covariance stationary if all the eigenvalues of $\sum_{k=1}^{K}\sum_{i=1}^{l}(A_{ik}\otimes A_{ik})+\sum_{k=1}^{K}\sum_{i=1}^{m}(B_{ik}\otimes B_{ik})$ are less than one in modulus. The covariance stationarity of the demeaned returns $\boldsymbol{\epsilon}_t$ is equivalent to the existence of the unconditional second-order moment, say Γ:

$$\Gamma = \operatorname{var}(\boldsymbol{\epsilon}_t) = E[E(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^{'}|I_{t-1})] = E[H_t]. \tag{1.28}$$

[17] *Cf.* also Subsection 1.3.3

which is a positive definite solution of equation (1.22)[18]:

$$\Gamma = C^{'}C + A^{'}\Gamma A + B^{'}\Gamma B. \tag{1.29}$$

This result is ensured by Proposition 4.3 in Boussama, Fuchs and Stelzer (2011). That being said, the formulation:

$$H_t = \Gamma - A^{'}\Gamma A - B^{'}\Gamma B + A^{'}\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1}^{'}A + B^{'}H_{t-1}B \tag{1.30}$$

is called *variance* (or *covariance*) *targeting BEKK representation*. The variance targeting estimation procedure considered in Pedersen and Rahbek (2013) is a two-step method where Γ is estimated by the sample unconditional covariance matrix of $\boldsymbol{\epsilon}_t$ and, next, the remaining parameters are estimated with the quasi maximum likelihood estimation by optimizing the resulting log-likelihood with respect to A and B. If $\boldsymbol{\epsilon}_t$ is strictly stationary and ergodic, the ergodic theorem guarantees that the moment-based estimator:

$$\hat{\Gamma} = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^{'} \tag{1.31}$$

is consistent for Γ. Then the authors show that under certain assumption (Pedersen and Rahbek (2013), Assumptions 4.1, 4.2 and 4.3) the second-step quasi maximum estimates, $(\hat{A}, \hat{B}|\hat{\Gamma})$, along with the first-step estimate $\hat{\Gamma}$, converge almost surely to the true parameters:

$$(\hat{\Gamma}, \hat{A}, \hat{B}) \xrightarrow{a.s.} (\Gamma, A, B) \text{ as } T \to \infty. \tag{1.32}$$

Furthermore in Pedersen and Rahbek (2013) also the asymptotic normality of $(\hat{\Gamma}, \hat{A}, \hat{B})$ is shown provided that some additional conditions are satisfied (Assumptions 4.4 and 4.5 in the paper).

[18] Equation (1.22) is here considered for $K = l = m = 1$ for simplicity

1.3.8 Kawakatsu parametrization

Another useful solution to the issue of the positivity of the H_t's is the one proposed by Kawakatsu (2003) who suggests to directly reparameterize the conditional covariances in the VEC model by means of lower triangular matrices L_t (Cholesky Factor [FC-] GARCH model). As is known from matrix algebra, a necessary and sufficient condition for a matrix H to be positive definite is that it can be expressed as the product of a lower triangular matrix L and its tranpose: $H = LL^{'}$; so the general econometric model becomes:

$$\text{vech}(L_t) = \boldsymbol{h}_t = \boldsymbol{c}_0 + \sum_{i=1}^{l} A_i \boldsymbol{h}_{t-i} + \sum_{i=1}^{m} B_i \boldsymbol{\epsilon}_{t-i} \tag{1.33}$$

where A_i is of dimension $N(N+1)/2 \times N(N+1)/2$ while B_i is of dimension $N(N+1)/2 \times N$. However, the Cholesky decomposition of a positive definite matrix is not uniquely defined and for this reason the author introduces some identification restrictions. Bearing in mind that the operator vech stacks in a column vector the column elements of L_t, say l_{ijt}, as long as $i \geq j$, then the elements of $\boldsymbol{h}_t$ which correspond to the diagonal entries of L_t are in place $k(k+1)/2$ with $k = 1, ..., N$. A common way to uniquely identify the decomposition factor L_t is to set all of its diagonal elements positive; this is equivalent to requiring that simultaneously: the entries $(k(k+1)/2, k(k+1)/2)$ of A_i are positive and those $((k(k+1)/2, j)$ are zeros (with $\{j = 1, ..., N(N+1)/2\} \cap \{j \neq k(k+1)/2\}$); all the entries $(k(k+1)/2, j)$ (with $j = 1, ...N$) of B_i are zeros. This simple identification rule forces the diagonal elements of the Cholesky factor L_t to depend only on its past diagonal elements and not on the lagged innovations. The general specification in (1.33) can be modified as follows:

$$\text{vech}(L_t) = \boldsymbol{h}_t = \boldsymbol{c}_0 + \sum_{i=1}^{l} A_i \boldsymbol{h}_{t-i} + \sum_{i=1}^{m} B_i |\boldsymbol{\epsilon}_{t-i}| \tag{1.34}$$

where the symbol $|\boldsymbol{\epsilon}_{t-i}|$ denotes the vector built with the absolute values of $\boldsymbol{\epsilon}_{t-i}$'s components. This specification needs no restrictions on matrices B_i, but it clearly cannot take into account the sign effect of innovations. Despite the identification restrictions partially

limits the number of parameters, the model keeps being over-parameterized even in case of small dimension. Therefore, other restrictions are introduced in order to further simplify the specification[19]. The log-likelihood function for Kawakatsu's model is:

$$\ln L = -\frac{TN}{2}\ln 2\pi - \sum_{t=1}^{T}\sum_{i=1}^{N}\ln l_{ii,t} - \frac{1}{2}\sum_{t=1}^{T}(L_t^{-1}\boldsymbol{\epsilon}_t)'(L_t^{-1}\boldsymbol{\epsilon}_t) \tag{1.35}$$

where the starting value L_0 can be assumed such that $L_0 L_0' = 1/T\sum_{t=1}^{T}\hat{\boldsymbol{\epsilon}}_t\hat{\boldsymbol{\epsilon}}_t'$. Since the modeling of the conditional variances (*i.e.* the diagonal entries of $H_t = L_t L_t'$) depends on how the assets are ordered[20], the log-likelihood in (1.35) is used in the empirical part of the paper to find the ordering that best fits the data: practically, the author computes all the $N!$[21] possible orderings and selects the one with the highest log-likelihood value.

1.4 Factor volatility models

Factor volatility models rely on the intuition that comovements of stock returns may be due to a restricted number of common underlying variables, the factors indeed. These can be observable or latent. These models have been proposed mainly to overcome the two typical problems of over-parametrization and positive definiteness, but they may also have an economic motivation. Without pretension of completeness, this section includes a terse review of some of the most important contributions relative to this class of models: the general model introduced in Engle, Ng and Rothschild (1990) that has been the first example of a Factor-ARCH covariance structure (Subsection 1.4.1); a model with uncorrelated factors proposed by Vrontos, Dellaportas and Politis (2003) (Subsection 1.4.2); the Generalized Orthogonal (GO-) GARCH model presented in van der Weide (2002) (Subsection 1.4.3); the Orthogonal (O-) GARCH model by Alexander (2001) (Subsection 1.4.4).

[19] See namely Equations (4), (5) and (6) in Kawakatsu (2003)

[20] It is easy to check that the (i,i)-th entry of H_t is equal to $\sum_{j=1}^{i} l_{ijt}^2$

[21] Consider that in Kawakatsu (2003) the empirical application is carried out on a porfolio of three indices

1.4.1 A model with Factor-GARCH covariance structure

The first example of factor volatility models is the one by Engle, Ng and Rothschild (1990). In this paper the authors specify the following conditional covariance structure:

$$H_t = \sum_{k=1}^{K} \boldsymbol{\beta}_k \boldsymbol{\beta}_k^{'} \lambda_{kt} + \Omega \tag{1.36}$$

where $\boldsymbol{\beta}_k$ are K linearly independent non-stochastic vectors, Ω is a positive semidefinite constant matrix and λ_{kt} are positive random variables. The dimension of H_t, N, is of course supposed to be greater than the number of factors, K. This volatility model can be regarded as the covariance counterpart of a typical factor model for asset excess returns (of the same kind employed by Ross [1976] to derive the arbitrage pricing theory). In detail, a model for returns related to (1.36) is:

$$\boldsymbol{r}_t = \boldsymbol{\mu}_t + \sum_{k=1}^{K} \boldsymbol{\beta}_k f_{kt} + \boldsymbol{\eta}_t \tag{1.37}$$

where:

$$\begin{aligned} &E(\boldsymbol{f}_t | I_{t-1}) = \mathbf{0} \ , \ \ E(\boldsymbol{f}_t \boldsymbol{f}_t^{'} | I_{t-1}) = \Lambda_t \\ &E(\boldsymbol{\eta_t}) = \mathbf{0} \ , \ \ E(\boldsymbol{\eta}_t \boldsymbol{\eta}_t^{'}) = \Omega. \end{aligned} \tag{1.38}$$

In (1.38) the covariance matrix $E(\boldsymbol{f}_t \boldsymbol{f}_t^{'} | I_{t-1})$ is assumed to be diagonal: $\Lambda_t = \text{diag}(\lambda_{1t}, ..., \lambda_{Kt})$, while the vectors $\boldsymbol{f}_t$ and $\boldsymbol{\eta}_t$ are considered uncorrelated: $E(\boldsymbol{f}_t \boldsymbol{\eta}_t^{'}) = 0_{K \times N}$. The authors show how to express the covariance model in (1.36) in terms of (scalar) conditional variances of the K so-called factor-representing portfolios: these are determined by weights $\boldsymbol{\alpha}_i$ such that $\boldsymbol{\alpha}_i^{'} \boldsymbol{\beta}_k = 0$ if $i \neq k$ and $\boldsymbol{\alpha}_i^{'} \boldsymbol{\beta}_k = 1$ if $i = k$. By doing so, they obtain the following expression:

$$
\begin{aligned}
H_t &= \sum_{k=1}^{K} \boldsymbol{\beta}_k \boldsymbol{\beta}_k^{'} \theta_{kt} + \Omega^* \\
\theta_{kt} &= \boldsymbol{\alpha}_k^{'} H_t \boldsymbol{\alpha}_k \\
\Omega^* &= \Omega - \sum_{k=1}^{K} \boldsymbol{\beta}_k \boldsymbol{\alpha}_k^{'} \Omega \boldsymbol{\alpha}_k \boldsymbol{\beta}_k^{'}.
\end{aligned}
\tag{1.39}
$$

In the first line of (1.39) the covariance H_t is formulated in terms of the scalar random factors $\theta_{kt} = \lambda_{kt} - \boldsymbol{\alpha}_k^{'} \Omega \boldsymbol{\alpha}_k$ which can be described by univariate GARCH models. As shown in the second line of (1.39), the θ_{kt}'s are nothing but the conditional variances of the factor-representing portfolios. The Factor-GARCH model is a special case of the BEKK parametrization. In order to check this equivalence, let us assume that θ_{kt} follows a GARCH(1,1) model with a vector of parameters $(\omega_k, \gamma_k, \delta_k)$:

$$
\begin{aligned}
\theta_{kt} =& \omega_k + \gamma_k (\boldsymbol{\alpha}_k^{'} \boldsymbol{\eta}_{t-1})^2 + \delta_k \theta_{k,t-1} = \\
=& \omega_k + \gamma_k (\boldsymbol{\alpha}_k^{'} \boldsymbol{\eta}_{t-1} \boldsymbol{\eta}_{t-1}^{'} \boldsymbol{\alpha}_k) + \delta_k \boldsymbol{\alpha}_k^{'} H_{k,t-1} \boldsymbol{\alpha}_k;
\end{aligned}
\tag{1.40}
$$

then the model in (1.39) can be rearranged as:

$$
H_t = [\Omega^* + \sum_{k=1}^{K} \boldsymbol{\beta}_k \omega_k \boldsymbol{\beta}_k^{'}] + \sum_{k=1}^{K} \gamma_k \boldsymbol{\beta}_k \boldsymbol{\alpha}_k^{'} (\boldsymbol{\eta}_{t-1} \boldsymbol{\eta}_{t-1}^{'}) \boldsymbol{\alpha}_k \boldsymbol{\beta}_k^{'} + \sum_{k=1}^{K} \delta_k \boldsymbol{\beta}_k \boldsymbol{\alpha}_k^{'} H_{k,t-1} \boldsymbol{\alpha}_k \boldsymbol{\beta}_k^{'}. \tag{1.41}
$$

So, if we set $\Gamma_k = \gamma_k^{1/2} \boldsymbol{\beta}_k \boldsymbol{\alpha}_k^{'}$ and $\Delta_k = \delta_k^{1/2} \boldsymbol{\beta}_k \boldsymbol{\alpha}_k^{'}$, the specification in (1.41) becomes:

$$
H_t = \Omega^{**} + \sum_{k=1}^{K} \Gamma_k \boldsymbol{\eta}_{t-1} \boldsymbol{\eta}_{t-1}^{'} \Gamma_k^{'} + \sum_{k=1}^{K} \Delta_k H_{k,t-1} \Delta_k^{'} \tag{1.42}
$$

which is clearly a BEKK-type representation. Finally Engle *et al.* (1990) propose an application of their model in which the number of factors (namely two) are preemptively determined by the application of the classical principal component analysis (PCA)[22] on the unconditional covariance between U.S. T-bills of different maturity and a stock index. Despite it would have been possible to estimate the weight vectors, in the paper two factor-representing portfolios with prespecified weights are constructed: a portfolio with equal weights on each of the bills and a zero weight on the stock index; a portfolio with zero weights on the bills and all weight on the stock index.

1.4.2 A full-factor multivariate GARCH model

The factors chosen in Engle *et al.* (1990) are observed and obviously correlated, while in order to efficiently capture different patterns hidden *under* the data it may be desirable that they have no correlation. Due to this observation many alternative factor models with uncorrelated factors have been proposed in the literature. They generally rely on the assumption that unobserved underlying factors may be identified that drive movements of returns. Let us suppose that (Vrontos, Dellaportas and Politis [2003]):

$$\begin{aligned} \boldsymbol{\epsilon}_t &= W\boldsymbol{x}_t \\ \boldsymbol{x}_t | I_{t-1} &\sim N(\mathbf{0}, \Sigma_t) \end{aligned} \tag{1.43}$$

with W a parameter matrix and $\boldsymbol{x}_t$ a vector of factors. Σ_t is supposed to be diagonal (that is, the factors are orthogonal), so that the conditional volatilities (the terms on the diagonal of Σ_t) may be estimated by means of univariate GARCH models: $\sigma_{it}^2 = \alpha_i + \beta_i x_{i,t-1}^2 + \gamma_i \sigma_{i,t-1}^2$. We can write:

$$\begin{aligned} \operatorname{var}(\boldsymbol{\epsilon}_t | I_{t-1}) = H_t = W\Sigma_t W' = \\ = W\Sigma_t^{1/2}\Sigma_t^{1/2}W' = L_t(\sigma_{it}, w_{ij})L_t'(\sigma_{it}, w_{ij}) \end{aligned} \tag{1.44}$$

[22] For a clear and slightly unconventional discussion about principal component analysis see Shlens (2014) published on arXiv.org

where in parentheses we have stressed the dependence of L_t on the parameters. It is well known that the decomposition of a positive definite matrix into the product of a triangular matrix and its transpose always exists and this decomposition is unique if the diagonal elements are restricted to be positive. So we can take W triangular, with elements $w_{ij} = 0$ when $i > j$ and $w_{ij} > 0$ when $i = j$, and build a conditional covariance matrix $H_t(\sigma_{it}, w_{ij})$ surely positive definite and driven by the volatilities of the orthogonal factors. Note that the $\boldsymbol{x}_t$'s are not parameters since they are equal to $W^{-1}\boldsymbol{\epsilon}_t$. Vrontos *et al.* (2003) show an application of the model to eight stocks from the U.S. market by forcing the number of parameters with the setting $w_{ii} = 1$ and by using a Bayesian approach; however, in theory also a classical QML estimation can be carried out.

1.4.3 Generalized Orthogonal GARCH model

The previous model is a so-called Full-Factor (FF-) multivariate GARCH model because the number of factors is equal to that of returns. Another version of the FF-GARCH model is the one introduced by van der Weide (2002): the Generalized Orthogonal (GO-) GARCH model. The fundamental assumptions are still the same as those summarized in (1.43). Also in this case the factors $\boldsymbol{x}_t$ are assumed to be incorrelated (Σ_t diagonal) with the σ_{it}'s governed by univariate GARCHs; in addition to this, in van der Weide (2002) the unconditional variance of the latent variables is set equal to the identity matrix I so that:

$$\Sigma = E(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^{'}) = WW^{'} . \tag{1.45}$$

By exploiting some algebraic properties, van der Weide (2002) stresses that there exists an orthogonal matrix U such that:

$$P\Lambda^{1/2}U = W \tag{1.46}$$

where P and Λ are respectively the matrix of eigenvectors and the matrix of eigenvalues resulting from the singular value decomposition (see Horn and Johnson [1999]) of the unconditional variance Σ; consequently P and Λ can be estimated from unconditional information. van der Weide (2002) shows a method to parameterize U by means of *Euler angles*, taking advantage of the fact that, from an algebraic point of view, orthogonal matrices can be expressed by the product of *rotational* matrices[23]; the Euler angles can be estimated by ML. Another example of GO-GARCH model can be found in Lanne and Saikkonen (2007).

1.4.4 Orthogonal GARCH model

The GO-GARCH models are by definition a generalization of another class of models, the Orthogonal GARCH models (O-GARCH). The O-GARCH models (Alexander, 2001) allows to generate $N \times N$ conditional covariances from m (with $m \leq N$) univariate GARCH models. Also the O-GARCH models take advantage of the principal component analysis (PCA) to dermine the factors to be employed. The factors identified by PCA based on unconditional information are assumed to be also conditionally uncorrelated and this produces a certain loss of accuracy that needs a careful initial calibration of the model. In more detail the mathematical structure is the following:

$$\begin{aligned} D^{-1/2}\boldsymbol{\epsilon}_t &= \boldsymbol{z}_t = \Lambda \boldsymbol{f}_t \\ \Lambda &= Q_m \operatorname{diag}(l_1^{1/2}, ..., l_m^{1/2}) \end{aligned} \tag{1.47}$$

where: $D = \operatorname{diag}(\sigma_1^2, ..., \sigma_N^2)$ is the diagonal matrix built with the unconditional variances of $\boldsymbol{\epsilon}_t$; $\boldsymbol{f}_t$ is a vector of heteroskedastic, uncorrelated and zero-mean random processes; $l_1, ..., l_m$ are the first larger m eigenvalues of $\operatorname{corr}(\boldsymbol{z}_t)$; Q_m is the $N \times m$ matrix of the related eigenvectors. Once indentified the first m factors able to explain the most part of the variability, m univariate GARCH process can be applied: $\sigma_{f_{it}}^2 = \omega_i + \alpha_i f_{i,t-1}^2 + \beta_i \sigma_{f_{i,t-1}}^2$. Hence, the conditional variance of $\boldsymbol{\epsilon}_t$ will have the following shape:

[23] Rotational matrices govern rotational tranformations, that is tranformations which preserve angles, and have the shape of orthogonal matrices

$$\text{var}(\boldsymbol{\epsilon}_t|I_{t-1}) = H_t = D^{1/2}\Lambda\,\text{diag}(\sigma^2_{f_{1t}},...,\sigma^2_{f_{mt}})\Lambda^{'}D^{1/2} \tag{1.48}$$

where H_t depends only on the $3\times m$ parameters $(\omega_i, \alpha_i, \beta_i)$, given that Σ and Λ are estimated on sample data. The O-GARCH models can allow for a strong parameter reduction but they may show identification problems in some cases (see van der Weide [2002]).

1.5 Conditional Correlation models

The models for conditional correlations exploit the decomposition of the conditional covariance matrix into the related conditional standard deviations and correlations which are separately estimated in a two-step procedure. From a theoretical point of view these models can be classified as nonlinear combinations of univariate GARCH's (see Bauwens *et al.* [2006]). Since they can be estimated in two stages, they generally exhibit a greater computational tractability compared to models like VEC or BEKK and they also turn out to be more flexible. This section presents a review of: the Constant Conditional Correlation (CCC-) GARCH by Bollerslev (1990) Subsection 1.5.1); the Varying Correlation (VC-) GARCH introduced in Tse and Tsui (2002) (Subsection 1.5.2); the Dynamic Conditional Correlation (DCC-) GARCH introduced in Engle (2002) (Subsection 1.5.3), with a particular attention to the related likelihood decomposition (Subsection 1.5.4) and to Aielli's critique (Aielli [2013]) (Subsection 1.5.5).

1.5.1 Constant Conditional Correlation model

The ancestor of models for conditional correlations is the CCC-GARCH by Bollerslev (1990). In order to model short-run nominal exchange rates the author suggests to split the conditional covariance into multiplicative components determined by conditional standard deviations, $\sigma_{1t},...,\sigma_{Nt}$, and correlations, Γ:

$$\begin{aligned} H_t &= D_t \Gamma D_t \\ D_t &= \text{diag}(\sigma_{1t}, ..., \sigma_{Nt}). \end{aligned} \tag{1.49}$$

As we can see from the definitions in (1.49), the correlation matrix Γ is assumed to be constant over time. The matrices H_t are almost surely positive definite if and only if each of the N conditional variances are well defined and Γ is in turn positive definite. To estimate the standard deviations in D_t univariate GARCH$(1,1)$ models are used[24]:

$$\sigma_{it}^2 = \omega_i + \alpha_i \epsilon_{i,t-1}^2 + \beta_i \sigma_{i,t-1}^2. \tag{1.50}$$

A very important topic that Bollerslev addresses in his paper is the likelihood decomposition (see also the next Subsection (1.5.4)) which has become a cornerstone in the following literature, especially for computational reasons:

$$\begin{aligned} L(\boldsymbol{\theta}) \propto -\frac{1}{2} \sum_{t=1}^{T} \left(ln|H_t| + \boldsymbol{\epsilon}_t' H_t^{-1} \boldsymbol{\epsilon}_t \right) = \\ = -\sum_{t=1}^{T} ln|D_t| - \frac{1}{2} \sum_{t=1}^{T} \left(ln|\Gamma| + \boldsymbol{\eta}_t' \Gamma^{-1} \boldsymbol{\eta}_t \right) \end{aligned} \tag{1.51}$$

where $\boldsymbol{\eta}_t = D_t^{-1} \boldsymbol{\epsilon}_t$ are the standardized residuals. In Equation (1.51) there is a first part depending on the conditional volatilities alone, while the correlation parameters are included only under the second summation operator: this feature allows to maximize the likelihood in two steps. Moreover, due to the assumed constancy of the correlation matrix, the likelihood optimization needs the inversion of only one matrix, Γ, instead of T matrices, H_t, and this is advantageous in case of very large sets of assets[25]. Hereinafter we are going to discuss this issue again. He and Terasvirta (2004) propose a generalization (Extended CCC-GARCH) of

[24] In general any GARCH(p,q) can be used

[25] Actually Bollerslev further simplifies the likelihood by substituting Γ with its sample counterpart

Bollerslev's model in which interactions between the conditional variances are allowed. In other terms the relationship (1.50) in this case becomes:

$$\sigma_{it}^2 = \omega_i + \sum_{j=1}^{N} \alpha_j \epsilon_{j,t-1}^2 + \sum_{j=1}^{N} \beta_j \sigma_{j,t-1}^2 \tag{1.52}$$

and, hence, the conditional variance of any asset depends on the first lag of squared innovations and conditional variances of all the assets considered in the portfolio.

1.5.2 Tse and Tsui's Varying Correlation GARCH

The hypothesis that the conditional correlations are constant over time may be unrealistic in most empirical cases, therefore generalizations of the CCC model in which the correlation matrix is considered time-dependent have been proposed in the early 2000's: we refer in particular to the works of Tse and Tsui (2002) and Engle (2002), in which the constant correlation matrix, Γ, is replaced by a time-varying matrix, Γ_t. As is known from standard matrix algebra, in order to guarantee the positive definiteness of decomposed conditional covariance matrices, $H_t = D_t \Gamma_t D_t$, it is sufficient that the elements of D_t (that is $h_{ii,t}^{1/2} = \sigma_{it}$) are well defined and the Γ_t's are in turn positive definite. The latter property can be ensured under simple conditions on the parameters. On the other hand, compared to CCC–GARCH, it is apparent that, when considering time-varying conditional correlations Γ_t, the advantage of numerically simple estimation is lost because the correlation matrix has to be inverted for every t during each iteration (*cf.* the log-likelihood function in (1.51)).
In Tse and Tsui (2002) the demeaned returns $\boldsymbol{\epsilon}_t \in \mathbb{R}^N$ are assumed to have conditional covariances of the following shape:

$$H_t = D_t \Gamma_t D_t \tag{1.53}$$

where $D_t = \text{diag}(h_{11t}^{1/2}, ..., h_{NNt}^{1/2})$ with conditional variances governed by GARCH specifications: $h_{iit} = \omega_i + \sum_{h=1}^{p} \alpha_{ih} \epsilon_{i,t-h}^2 + \sum_{h=1}^{q} \alpha_{ih} h_{ii,t-h}$[26] and Γ_t is a time-varying correlation

[26] As a matter of fact, also different more complex GARCH specifications may be used

matrix. If we consider the standardized residuals $\boldsymbol{\eta}_t = D_t^{-1}\boldsymbol{\epsilon}_t$, the dynamics of the conditional correlation matrices are described by the following iterative equation:

$$\begin{aligned} &\Gamma_t = (1-\theta_1-\theta_2)\Gamma + \theta_1\Gamma_{t-1} + \theta_2\Psi(\boldsymbol{\eta}_{t-1},...,\boldsymbol{\eta}_{t-m}) \\ &\theta_1,\theta_2 > 0 \quad , \quad \theta_1+\theta_2 \leq 1 \end{aligned} \tag{1.54}$$

where Ψ is a matrix function of the latest M standardized residuals and Γ is a suitable constant matrix[27]. More precisely, the functional form adopted by Tse and Tsui (2002) for Ψ is such that the generic entry of the matrix is:

$$\psi_{ijt} = \frac{\sum_{l=1}^{M}\eta_{i,t-l}\eta_{j,t-l}}{\left(\sum_{l=1}^{M}\eta_{i,t-l}^2 \sum_{l=1}^{M}\eta_{j,t-l}^2\right)^{1/2}} \tag{1.55}$$

where it must be $M \geq N$ for Ψ_t to be positive definite. Hence, if $M \geq N$ and Γ, Γ_0 and Ψ_0 are positive definite, from (1.54) it derives that the generic Γ_t is positive definite as well. The model proposed by Tse and Tsui (2002) is named Varying Correlation (VC-) GARCH.

1.5.3 Engle's Dynamic Conditional Correlation (DCC) model

Engle (2002) introduces the Dynamic Conditional Correlation (DCC-) GARCH model. The nucleus of this model is expressed by the following dynamic matrix process:

$$Q_t = (\boldsymbol{i}\boldsymbol{i}' - A - B) \odot S + A \odot \boldsymbol{\eta}_t\boldsymbol{\eta}_t' + B \odot Q_{t-1} \tag{1.56}$$

where $\boldsymbol{i}$ is an N-dimensional vector of ones, A and B are symmetric parameter matrices and S, according to the correlation targeting (see Subsection 1.3.7), is the unconditional covariance of $\boldsymbol{\eta}_t$ or, alternatively, the unconditional correlation of $\boldsymbol{\epsilon}_t$[28]. In Ding and Engle

[27] Γ must be positive definite and have unit diagonal

[28] Like in Subsection 1.5.2, the vector $\boldsymbol{\eta}_t$ denotes the standardized returns: $\boldsymbol{\eta}_t = D_t^{-1}\boldsymbol{\epsilon}_t$; the same notation is kept in the next sections

(2001) it is shown that if A, B and $(\boldsymbol{ii}^{'} - A - B)$ are positive semidefinite, then also Q_t will be; further, Q_t is positive definite if any one of the matrices on the right of the equal sign is positive definite. Although the process in (1.56) ensures positive definiteness, it does not generally produce well defined correlation matrices. They can be obtained by rescaling Q_t:

$$R_t = (I \odot Q_t)^{-1/2} Q_t (I \odot Q_t)^{-1/2}. \tag{1.57}$$

The conditional variances are consequently obtained as in Equation (1.53): $H_t = D_t R_t D_t$, where D_t is as usual the diagonal matrix of the conditional standard deviations. However the model in (1.56) needs a large number of parameters, also when the number of assets is only moderately high. For this reason Engle (2002) proposes also a scalar version of the specification in (1.56):

$$Q_t = (1 - \alpha - \beta)S + \alpha \boldsymbol{\eta}_t \boldsymbol{\eta}_t^{'} + \beta Q_{t-1}. \tag{1.58}$$

where $\alpha, \beta \in \mathbb{R}$. The scalar version of DCC clearly simplifies the estimation procedure in case of large N but at the same time usually turns out to be empirically unfounded: indeed, forcing the dynamics of the comovements of many heterogenous assets to be described by only two scalar parameters, α and β, generally causes a great loss of information which translates into inaccurate estimates and poor forecasting performances.

1.5.4 DCC and likelihood decomposition

Another essential issue Engle (2002) deals with is the likelihood decomposition (see also Engle and Sheppard [2001]). Let the parameters in $D_t = \text{diag}(\sigma_{1t}, ..., \sigma_{Nt})$ be denoted by the vector $\boldsymbol{\theta}$ and the additional parameters in R_t be denoted by $\boldsymbol{\varphi}$. We assume that the conditional variances σ_{it}^2 are governed by univariate GARCH models. Then, assuming multivariate normal innovations $\boldsymbol{\epsilon}_t$, the log-likelihood of a DCC model is obtained as follows:

$$L(\boldsymbol{\theta}, \boldsymbol{\varphi}) = -\frac{1}{2}\sum_{t=1}^{T}\left(N\ln 2\pi + \ln|H_t| + \boldsymbol{\epsilon}_t' H_t^{-1}\boldsymbol{\epsilon}_t\right). \tag{1.59}$$

By using some algebra, the log-likelihood in (1.59) can be rearranged in the form:

$$\begin{aligned} L(\boldsymbol{\theta}, \boldsymbol{\varphi}) = &-\frac{1}{2}\sum_{t=1}^{T}\left(N\ln 2\pi + 2\ln|D_t| + \boldsymbol{\epsilon}_t' D_t^{-1} D_t^{-1}\boldsymbol{\epsilon}_t\right) + \\ &-\frac{1}{2}\sum_{t=1}^{T}\left(-\boldsymbol{\eta}_t'\boldsymbol{\eta}_t + \ln|R_t| + \boldsymbol{\eta}_t' R_t^{-1}\boldsymbol{\eta}_t\right) = \\ &= L_1(\boldsymbol{\theta}) + L_2(\boldsymbol{\theta}, \boldsymbol{\varphi}) \end{aligned} \tag{1.60}$$

where the first part of the function (namely L_1, in the first line of (1.60)) depends only on $\boldsymbol{\theta}$. Engle and Sheppard (2001) show that under weak conditions the first-step estimation of $\boldsymbol{\theta}$, obtained by maximizing L_1, and the second-step estimation of $\boldsymbol{\varphi}$, obtained by maximizing L_2 conditional on $\hat{\boldsymbol{\theta}}$, give consistent and asymptotically normal results[29]. This feature of DCC clearly makes it more computationally advantageous than other M-GARCHs for the modeling of conditional covariance matrices (like VEC and BEKK, for istance).

1.5.5 Aielli's critique and cDCC

Due to the possibility of two-step consistent estimation and to the presence of only two scalar parameters α and β, the scalar DCC model is very attractive because of its feasibility also when N is high. Nevertheless, the computational easiness of parameter estimation in Equation (1.58) also presumes the correlation targeting, that is the estimation of S by means of the sample covariance of the estimated standardized returns, $\hat{\boldsymbol{\eta}}_t$:

$$S \equiv \hat{S} = \frac{1}{T}\sum_{i=1}^{T}\hat{\boldsymbol{\eta}}_t\hat{\boldsymbol{\eta}}_t'. \tag{1.61}$$

[29] The authors claim that their proofs closely follow some related results presented in White (1996)

However, as shown in Aielli (2013), taking the unconditional expected value of Equation (1.58) does not return:

$$S = E(\boldsymbol{\eta}_t \boldsymbol{\eta}_t^{'}) \tag{1.62}$$

unless $E\left(Q_t\right) = E(\boldsymbol{\eta}_t \boldsymbol{\eta}_t^{'}) = E\left(R_t\right)$, which is not true in general. Thus, the assumption (1.61) yields unconsistent estimates of Q_t and, hence, of R_t and H_t. To solve this problem Aielli (2013) proposes a *corrected* version (cDCC) of the specification in (1.58):

$$Q_t = (1 - \alpha - \beta)S + \alpha \left(I \odot Q_{t-1}\right)^{1/2} \boldsymbol{\eta}_{t-1} \boldsymbol{\eta}_{t-1}^{'} \left(I \odot Q_{t-1}\right)^{1/2} + \beta Q_{t-1} \tag{1.63}$$

where as usual I is the identity matrix. The specification in (1.63) along with the assumption in (1.61) produces consistent estimates but needs an identifying restriction: $[S]_{ii} = 1$ for every i. Let us note that the correlation targeting may be avoided by using the alternative dynamic equation (see for example Caporin and McAleer [2012]):

$$Q_t = CC^{'} + \alpha \boldsymbol{\eta}_t \boldsymbol{\eta}_t^{'} + \beta Q_{t-1} \, . \tag{1.64}$$

where C is upper triangular. Obviously with this specification the number of parameters increases up to $2 + N(N-1)/2$.

1.6 Some more recent contributions in DCC literature

More recently we have assisted to the development of a new trend in literature that aims to overcome the strict assumptions of the scalar DCC model and to propose richer and more flexible parametrizations. In this section we will examine some of these contributions: the Diagonal (D-) DCC (Cappiello, Engle and Sheppard [2006]) (Subsection 1.6.1); the Generalized (G-) DCC (Hafner and Franses [2009]) (Subsection 1.6.2); the Rotated GARCH (RARCH) and the Rotated DCC (RCC) (Noureldin, Shephard and Sheppard [2014]) (Subsection 1.6.3).

1.6.1 Diagonal DCC

As is well known from empirical observations, volatility typically increases more after a negative shock (for instance, news about a worsening of economic outlook) than after a positive shock of the same strength (for example after a swift drop in oil price); further, the effect of such shocks is different for different types of assets. Cappiello *et al.* (2006) remark that these changes in volatilities cause subsequent changes in conditional covariances and correlations. On this background they propose a diagonal version (previously introduced in a manuscript of Sheppard [2002]) of the DCC model able to take into account the above-mentioned asymmetric effects:

$$\begin{aligned} Q_t = (\overline{Q} - A^{'}\overline{Q}A - B^{'}\overline{Q}B - G^{'}\overline{N}G) + A^{'}\boldsymbol{\eta}_{t-1}\boldsymbol{\eta}_{t-1}^{'}A + \\ + B^{'}Q_{t-1}B + G^{'}\boldsymbol{n}_{t-1}\boldsymbol{n}_{t-1}^{'}G \end{aligned} \tag{1.65}$$

where A, B and G are diagonal parameter matrices. The i-th vector element $n_{i,t-1}$ is set equal to $\eta_{i,t-1}$ when $\eta_{i,t-1} < 0$, while it is assumed null otherwise. The two expectation matrices $\overline{Q}$ and $\overline{N}$ are replaced with their sample counterpart $T^{-1}\sum_t \hat{\boldsymbol{\eta}}_t\hat{\boldsymbol{\eta}}_t^{'}$ and $T^{-1}\sum_t \hat{\boldsymbol{n}}_t\hat{\boldsymbol{n}}_t^{'}$. Note that in Equation (1.65) Q_t is almost surely positive definite if $(\overline{Q} - A^{'}\overline{Q}A - B^{'}\overline{Q}B - G^{'}\overline{N}G)$ is. The empirical application is carried out on 21 national equity and 13 national bond indices. Relative to worldwide bond and stock indices, the authors find evidence of the superiority of specifications with diagonal matrices compared to the scalar ones. The other interesting result of their empirical application is that both stock and bond indices exhibit asymmetry in conditional correlations, while there is strong evidence of asymmetric effects only in the volatility of stock index returns. It is worth noting that, provided that the correlation targeting is used, the specification (1.65) involves the estimation of $3N$ parameters ($2N$ if asymmetric effects are not considered): so the specification is definitely more flexible than the scalar DCC, but the number of parameters increases linearly with the number of assets.

1.6.2 Generalized DCC

Another more flexible version of DCC can be found in Hafner and Franses (2009) where it is applied on 18 German and 25 UK stock returns. The authors name their model Generalized (G-) DCC and find convincing evidence that it improves on the DCC model. In Hafner and Franses' specification the matrices associated with both the lagged standardized innovations and the autoregressive term are built by outer self-product of two N-dimensional vector (where as usual N is the number of assets) so that they turn out to be rank one matrices. More precisely, the proposed dynamic specification is:

$$Q_t = (1 - \overline{\alpha}^2 - \overline{\beta}^2)\overline{Q} + \boldsymbol{\alpha}\boldsymbol{\alpha}^{'} \odot \boldsymbol{\eta}_{t-1}\boldsymbol{\eta}_{t-1}^{'} + \boldsymbol{\beta}\boldsymbol{\beta}^{'} \odot Q_{t-1} \tag{1.66}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are N-dimensional vectors (so that the numbers of parameters raises up to $2N$) and $\overline{\alpha} = \frac{1}{N}\sum_{i=1}^{N}\alpha_i$ and $\overline{\beta} = \frac{1}{N}\sum_{i=1}^{N}\beta_i$. Since in many empirical applications the parameters related to the innovations' matrix are more variable across the assets than the ones associated with the autoregressive term, Hafner and Franses (2009) suggest that, when $2N$ is still too large, the matrix $\boldsymbol{\beta}\boldsymbol{\beta}'$ can be replaced by a scalar. As an alternative, the α_i's and the β_i's may be pooled into sufficiently homogeneous clusters (for example based upon *a priori* economic considerations or previous empirical findings). In Equation (20) the authors use the matrix $(1-\overline{\alpha}^2-\overline{\beta}^2)\overline{Q}$ because $(\boldsymbol{\iota}\boldsymbol{\iota}^{'} - \boldsymbol{\alpha}\boldsymbol{\alpha}^{'} - \boldsymbol{\beta}\boldsymbol{\beta}^{'}) \odot \overline{Q}$ would not be in general positive definite; on the contrary, provided that Q_0 is positive definite, the specification in (20) ensures the positive definiteness of Q_t and, hence, of H_t. Besides, note that in this case, even though we assume $E(Q_t) = E(\boldsymbol{\eta}_t\boldsymbol{\eta}_t^{'}) = R$, we will obtain:

$$\overline{Q} = \frac{(\boldsymbol{\iota}\boldsymbol{\iota}^{'} - \boldsymbol{\alpha}\boldsymbol{\alpha}^{'} - \boldsymbol{\beta}\boldsymbol{\beta}^{'})}{(1 - \overline{\alpha}^2 - \overline{\beta}^2)} \odot E(\boldsymbol{\eta}_t\boldsymbol{\eta}_t^{'}) \tag{1.67}$$

and, therefore, setting $\overline{Q} = T^{-1}\sum_{t=1}^{T}\hat{\boldsymbol{\eta}}_t\hat{\boldsymbol{\eta}}_t^{'}$ leads to a biased and inconsistent estimate of the intercept matrix. Despite this circumstance the authors employ the correlation targeting.

1.6.3 Rotated ARCH and Rotated DCC

The main idea in Noureldin, Shephard and Sheppard (2014) is related to the orthogonal models (see Section 1.4) and consists in transforming the raw returns by *rotations* in order to fit flexible multivariate models using the covariance targeting. However, unlike the O-GARCH, in this paper the transformed returns are not assumed to be conditionally uncorrelated. They propose two different specifications: the RARCH (Rotated ARCH) which is based upon a BEKK-type parametrization; the RCC (Rotated DCC) that entails the application of the DCC framework. One of the crucial elements of the model is that the unconditional variance of the rotated residuals is equal to the identity matrix; indeed, the rotated residuals $\boldsymbol{e}_t$ are obtained as follows:

$$\boldsymbol{e}_t = \overline{H}^{-1/2}\boldsymbol{\epsilon}_t = P\Lambda^{-1/2}P^{'}\boldsymbol{\epsilon}_t \tag{1.68}$$

where $\overline{H} = \text{var}(\boldsymbol{\epsilon}_t)$ with spectral decomposition $\overline{H} = P\Lambda P^{'}$, where P is the orthogonal matrix of the eigenvectors and Λ is the diagonal one of the eigenvalues. That being said, a BEKK-type specification is applied:

$$G_t = (I - AA^{'} - BB^{'}) + A\boldsymbol{e}_{t-1}\boldsymbol{e}_{t-1}^{'}A^{'} + BG_{t-1}B^{'} \tag{1.69}$$

where $G_t = \text{var}(\boldsymbol{e}_t|I_{t-1})$ is the rotated residuals' conditional variance, $E(G_t) = I$[30] and A and B are two conformable parameter matrices such that $(I - AA^{'} - BB^{'})$ is positive semidefinite. With these assumptions the positive definiteness of G_t is easy to be ensured. The model in (1.69) is covariance stationary if all the eigenvalues of $(A \otimes A + B \otimes B)$ are less than 1 in modulus (see Engle and Kroner [1995] and Subsection 1.3.6). In the article in discussion some special cases are taken into account. The first one is a straightforward scalar specification with $A = \alpha^{1/2}I$ and $B = \beta^{1/2}I$. In the second A and B are assumed

[30] *Cf.* Equation 1.68

to be diagonal with elements $\alpha_{ij}^{1/2}, \beta_{ij}^{1/2} \neq 0$ only when $i = j$. In the context of this second special case the so-called persistence parameters are introduced[31]:

$$\lambda_{ij} = \alpha_{ii}^{1/2}\alpha_{jj}^{1/2} + \beta_{ii}^{1/2}\beta_{jj}^{1/2}. \tag{1.70}$$

On the other hand, the authors define as smoothness parameters the $\beta_{ii}^{1/2}\beta_{jj}^{1/2}$'s alone. Setting all the persistence parameters equal (that is $\lambda_{ij} = \lambda$ for each couple i, j) we have a third special case (the common persistence or CP specification):

$$G_t = (1 - \lambda)I + A(\boldsymbol{e}_{t-1}\boldsymbol{e}_{t-1}^{'} - G_{t-1})A^{'} + \lambda G_{t-1} \tag{1.71}$$

with A diagonal and $0 < \lambda < 1$. The specification (1.71) is motivated by the empirical observation that the persistences λ_{ij} are not so much heterogenous across assets while the heterogeneity of the α_{ii}'s and that of the β_{ii}'s seems to be inversely correlated[32]. The CP specification is rather parsimonious as it entails only $N + 1$ parameters: $\{\lambda, \alpha_{11}, ..., \alpha_{NN}\}$. Once estimated the parameters of the RARCH specification, the conditional variance of the unrotated returns $\boldsymbol{\epsilon}_t$ is coherently obtained as $H_t = \overline{H}^{1/2} G_t \overline{H}^{1/2}$.
In the case of the RCC model the same procedure applies to the standardized residuals $\boldsymbol{\eta}_t = D_t^{-1}\boldsymbol{\epsilon}_t$, where as usual $D_t = \text{diag}(h_{11t}^{1/2}, ..., h_{NNt}^{1/2})$ and $h_{iit} \sim GARCH(\boldsymbol{\theta}_i)$[33]. Also in this case Noureldin *et al.* (2014) decompose the unconditional variance of $\boldsymbol{\eta}_t$, Π, to obtain the rotated devolatilized residuals:

$$\begin{aligned} \Pi &= P\Lambda P' \\ \tilde{\boldsymbol{e}}_t &= P\Lambda^{-1/2}P'\boldsymbol{\eta}_t \\ \text{var}(\tilde{\boldsymbol{e}}_t) &= I. \end{aligned} \tag{1.72}$$

[31] In order to realize why these parameters are defined that way, one just needs to notice that $g_{iit} = (1 - \alpha_{ii} - \beta_{ii}) + \alpha_{ii}e_{i,t-1}^2 + \beta_{ii}g_{ii,t-1}$ and $g_{ijt} = \alpha_{ii}^{1/2}\alpha_{jj}^{1/2}e_{i,t-1}e_{j,t-1} + \beta_{ii}^{1/2}\beta_{jj}^{1/2}g_{ij,t-1}$

[32] In other terms, it is observed empirically that, while the $\alpha_{ii}^{1/2}\alpha_{jj}^{1/2}$'s and $\beta_{ii}^{1/2}\beta_{jj}^{1/2}$'s tend to vary across assets, at the same time their sum tends to be somewhat costant

[33] The structure and the dimension of the parameter vector $\boldsymbol{\theta}_i$ will obviously depend on the type of thechosen GARCH specification

Hence, the dynamics of the conditional correlation matrices of the rotated standardized residuals are modeled as follows:

$$Q_t^* = (I - AA^{'} - BB^{'}) + A\tilde{\boldsymbol{e}}_{t-1}\tilde{\boldsymbol{e}}_{t-1}^{'}A^{'} + BQ_{t-1}^*B^{'} \quad , \quad Q_0^* = I \tag{1.73}$$

where the Q_t^*'s are the *quasi* conditional correlation matrices, according to the classic Engle's DCC symbology (see Equation (1.56)). Then, the correlation matrices of the unrotated standardized residuals is reconstructed *via* the following two steps:

$$\begin{aligned} Q_t =& P\Lambda^{1/2}P'Q_t^*P\Lambda^{1/2}P' \\ R_t =& (I \odot Q_t)^{-1/2}Q_t(I \odot Q_t)^{-1/2}. \end{aligned} \tag{1.74}$$

Applying the dynamic equation to the rotated residuals also avoids the inconsistence of the correlation targeting because the intercept $(I - AA^{'} - BB^{'})$ needs not be estimated by a sample estimator. The paper ends with empirical applications that show the superior predictive capability of RARCH and RCC compared to O-GARCH and GO-GARCH. Note that the empirical applications are carried out on 10 assets.

Chapter 2

Models and methods for the estimation and regularization of high dimensional volatility matrices

2.1 Introduction

In the previous Chapter 1 the problem of the curse of dimensionality has been reminded several times and has been recognized as the main hurdle in conveniently adapting the M-GARCH scheme to large portfolios. As seen, many simplified versions of the original model by Bollerslev, Engle and Wooldridge (1988) have been introduced, but they have often turned out to be not completely satisfying: in many cases the imposed restrictions are too tight with respect to empirical observations and economic notions, while sometimes even these simplified specifications keep not being applicable when the number of assets exceed a certain limit. For these reasons in the last years a specific line of research has been expressly dedicated to find new methods of estimation which are feasible in case of large portfolios without causing too big losses of information. In this chapter, without any presumption of completeness, we want to offer a synthetic overlook of this recent current of literature along with a review

of some contributions on a closely related issue which is the regularization of ill-defined covariance and correlation matrices. Our attention to the latter topic is due to the fact that, in Chapter 3, some of the regularization techniques that are here covered will represent the key ingredients for the derivation of new flexible prediction methods for DCC models, which is the main methodological contribution of this thesis. Consequently, this chapter is split in two parts: the first one concerns some models and methods for the estimation of high dimensional volatility matrices in the DCC framework (Section 2.2), while the second one is about a few methods of regularization of covariance and correlation matrices and their application to financial econometrics (Section 2.3).

2.2 Models and methods for the estimation of high dimensional volatility matrices

This section deals with models and methods for the estimation of high dimensional volatility matrices in the DCC framework (Engle [2002]) (see Section 1.5) and, in more detail, it contains: a brief discussion of the so-called MacGyver estimator introduced by Engle (2009) (Subsection 2.2.1); an illustration of the composite likelihood (CL) method of estimation used in Engle, Shephard and Sheppard (2008) (Subsection 2.2.2); a presentation of Engle and Kelly (2012) who introduce the dynamic equi-correlation model or, more simply, DECO (Subsection 2.2.3); an assessment of composite-likelihood (CL-) DCC and DECO carried out by Clements, Scott and Silvennoinen (2012) in which the two models are compared to some simpler method for estimation and forecasting of large covariance matrices (Subsection 2.2.4); a presentation of a recent contribution by Bauwens, Grigoryeva and Ortega (2014) in which the empirical performances of some parsimonious parametrizations of the DCC model are assessed and where the estimation of the proposed models is carried out relying on efficient constrained optimization techniques borrowed from the machine learning literature (Subsection 2.2.5).

2.2.1 MacGyver estimator

In Chapter 1 it has been seen that in the DCC framework the likelihood function can be maximized in two steps by decomposing it in a first part relative to the estimation of volatility parameters and a second part relative to conditional correlation ones. The segment involved in the estimation of the dynamic parameters of the DCC model (see Equation (1.60)) is:

$$L_2 = -\frac{1}{2}\sum_{t=1}^{T}\left(\ln|R_t| + \boldsymbol{\eta}_t' R_t^{-1}\boldsymbol{\eta}_t\right) \tag{2.1}$$

whose maximization in case of large sets needs the computation of the determinant and, especially, the inversion of T big correlation matrices. We know that this may cause numerical problems which are often insolvable. Besides, Engle and Sheppard (2001) point out that in the DCC dynamic equation:

$$Q_t = (1-\alpha-\beta)S + \alpha\boldsymbol{\eta}_t\boldsymbol{\eta}_t' + \beta Q_{t-1} \tag{2.2}$$

the parameters α and β regularly suffer from bias in finite sample when the problem size N is large and the two-step likelihood maximization is used. In Engle (2009) the so-called MacGyver estimation method is introduced in order to face this typical large-N issues. The MacGyver is meant to be used in the scalar DCC framework and is built upon bivariate DCC estimations of conditional correlations ρ_{ijt}:

$$\begin{aligned}\rho_{ijt} &= \frac{q_{ijt}}{(q_{iit}q_{jjt})^{1/2}}\\ q_{ijt} &= (1-\alpha-\beta)\bar{q}_{ij} + \alpha\eta_{it}\eta_{jt} + \beta q_{ij,t-1}.\end{aligned} \tag{2.3}$$

Therefore, the explicit formulation of the pairwise likelihood is:

$$L_{2ij} = -\frac{1}{2}\sum_{t=1}^{T}\left(\ln(1-\rho_{ijt}^2) + \frac{\eta_{it}^2 + \eta_{jt}^2 - 2\rho_{ijt}\eta_{it}\eta_{jt}}{1-\rho_{ijt}^2}\right) \tag{2.4}$$

which is clearly easy to be numerically maximized. The numerical optimization of the likelihood in (2.4) for each pair of assets yields $N(N-1)/2$ couples of estimated parameters $(\hat{\alpha}_{ij}, \hat{\beta}_{ij})$; these couples are then blended together by means of a suitable function. More precisely, in the cited paper the author uses and compares three blending functions: mean, 5%-trimmed mean and median, and two types of likelihood: the ordinary unrestricted likelihood and a constrained version in which the parameters are forced to be $\in]0,1[$ based on the following modified specification:

$$q_{ijt} = \overline{q}_{ij} + \frac{e^{\tilde{\alpha}}}{1+e^{\tilde{\alpha}}}(\eta_{it}\eta_{jt} - \overline{q}_{ij}) + \frac{e^{\tilde{\beta}}}{1+e^{\tilde{\beta}}}(q_{ij,t-1} - \overline{q}_{ij}). \tag{2.5}$$

The finite sample properties of the MacGyver method are assessed by means of a Monte Carlo experiment with different problem size whose net result is that the smallest root mean square errors and biases are generally achieved by using the median. The consistency of the MacGyver estimation is easily established when $T \to \infty$, given that the bivariate estimates are consistent themselves under standard regularity assumptions.

2.2.2 Composite likelihood estimation and DCC

The composite (quasi) likelihood estimator for the parameters of DCC models has been introduced in Engle, Shephard and Sheppard (2008). The method has been proposed in order to address the bias problem caused by the two-step quasi likelihood estimation (see Equation (1.60)) and to make the likelihood decomposition plausible for large portfolios. Let us denote the vectors of return residuals by $\boldsymbol{\epsilon}_1, ..., \boldsymbol{\epsilon}_T$, with T the time length of the considered sample and $\boldsymbol{\epsilon}_t \in \mathbb{R}^N$. The standard inference method is based on the maximization of a Gaussian likelihood function:

$$\ln L(\boldsymbol{\varphi}; \boldsymbol{\epsilon}_1, ..., \boldsymbol{\epsilon}_T) \propto \sum_{t=1}^{T} \left(-\frac{1}{2}|H_t| - \frac{1}{2}\boldsymbol{\epsilon}_t' H_t^{-1} \boldsymbol{\epsilon}_t \right) \tag{2.6}$$

where $\boldsymbol{\varphi}$ is the vector of parameters varying in the parameter space Φ. The authors point out that every matrix inversion in (2.6) has a computation complexity of order $O(N^3)$ for each likelihood evaluation. Furthermore, the parameter space Φ is typically large, so the numerical difficulties get harder. As is known, the covariance (or correlation) targeting consists of a first step in which a moment estimator is used for the nuisance parameters[1], say $\boldsymbol{\lambda}$, and of a second one in which the dynamic parameters of interest, say $\boldsymbol{\theta}$, are estimated; the second-stage maximization problem consequently becomes: $\max_{\boldsymbol{\theta}} \ln L(\boldsymbol{\theta}; \hat{\boldsymbol{\lambda}}, \boldsymbol{\epsilon}_1, ..., \boldsymbol{\epsilon}_T)$. However, when N and the dimension of $\boldsymbol{\lambda}$ are large, the two-step procedure may yield an estimated $\boldsymbol{\theta}$ far from its true value: this is a general statistical problem, the so-called incidental parameter problem, that was discussed for the first time in Neyman and Scott (1948). Engle *et al.* (2008) carry out an empirical application on a set of $N \leq 96$ assets and $T = 2516$ observations by fitting three different scalar models[2] and they show that the innovations' coefficient dramatically decreases as N gets higher, disclosing an apparent bias in the estimation based on the marginal likelihood. The composite likelihood method, instead, can overcome both the bias and the numerical problem. It is based on summing up the quasi likelihood functions of the data subsets built with every distinct pair of the considered assets:

$$\begin{aligned} &Y_t = \{\boldsymbol{y}_{1t}, ..., \boldsymbol{y}_{Kt}\} \quad , \quad K = \frac{N(N-1)}{2} \\ &\boldsymbol{y}_{1t} = (\epsilon_{1t}, \epsilon_{2t})' \\ &\boldsymbol{y}_{2t} = (\epsilon_{1t}, \epsilon_{3t})' \\ &..... \\ &\boldsymbol{y}_{Kt} = (\epsilon_{N-1,t}, \epsilon_{N,t})' \end{aligned} \tag{2.7}$$

[1] In the paper in discussion the nuisance parameters refer to the entries of the long-run covariance (or correlation) included in the intercept of the dynamic equation of the model

[2] The used specification are a scalar version of BEKK, the EWMA and the scalar DCC

or, alternatively, with all the contiguous overlapping pairs[3]:

$$
\begin{aligned}
&Y_t = \{\boldsymbol{y}_{1t}, ..., \boldsymbol{y}_{Kt}\} \quad , \quad K = N-1 \\
&\boldsymbol{y}_{1t} = (\epsilon_{1t}, \epsilon_{2t})' \\
&\boldsymbol{y}_{2t} = (\epsilon_{2t}, \epsilon_{3t})' \\
&..... \\
&\boldsymbol{y}_{K-1,t} = (\epsilon_{N-2,t}, \epsilon_{N-1,t})' \\
&\boldsymbol{y}_{Kt} = (\epsilon_{N-1,t}, \epsilon_{Nt})'.
\end{aligned}
\tag{2.8}
$$

The composite likelihood function is, for each period t:

$$
\begin{aligned}
&C_t(\boldsymbol{\varphi}) = \frac{1}{K}\sum_{i=1}^{K} l_{it}(\boldsymbol{\varphi}) \\
&l_{it}(\boldsymbol{\varphi}) \propto -\frac{1}{2}\ln|H_{it}| - \frac{1}{2}\boldsymbol{y}_{it}' H_{it}^{-1}\boldsymbol{y}_{it} \\
&H_{it} = \mathrm{var}(\boldsymbol{y}_{it}|I_{t-1}).
\end{aligned}
\tag{2.9}
$$

The composite likelihood estimator has the advantage that there is no need to invert large covariance matrices[4] because it involves the inversion of a 2×2 matrix at a time; besides, it does not yield the incidental parameter problem. The underlying assumption is obviously that each pair's composite likelihood is a function only of that pair's nuisance parameters, $l_{it}(\boldsymbol{\varphi}) = l_{it}(\boldsymbol{\theta}, \boldsymbol{\lambda}_i)$: this allows to estimate $\boldsymbol{\lambda}_i$ basing upon $\boldsymbol{y}_{i1}, ..., \boldsymbol{y}_{iT}$ which causes a loss of efficiency but not bias. On the other hand, a related issue is the overlapping of some components of vectors $\boldsymbol{\lambda}_1, ..., \boldsymbol{\lambda}_K$: this pitfall can be eluded by choosing separated subsets, $\{(\epsilon_{1t}, \epsilon_{2t}), (\epsilon_{3t}, \epsilon_{4t}), ..., (\epsilon_{N-1,1t}, \epsilon_{Nt})\}$, (with an additional computational simplification but with a certain loss of information about the target parameters) or by imposing equality

[3] In theory the composite likelihood may be built with each kind of subsets, separated or not and of different dimension, but we refer to the choice made in Engle *et al.* (2008) for their empirical exercise

[4] The computational complexity order of the estimator is $O(N^2)$ in the case of scheme (2.7) and $O(N)$ in the case of scheme (2.8)

constraints across the $\boldsymbol{\lambda}_i$. However, in the cited paper the pairwise nuisance parameters are assumed to be variation free (*i.e.* no cross constraint in the case of overlapping subsets). In a general formulation the problem looks as follows[5]:

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \frac{1}{K} \sum_{t=1}^{T} \sum_{i=1}^{K} l_{it}(\boldsymbol{\theta}, \hat{\boldsymbol{\lambda}}_i) \\ \text{s.t. } \hat{\boldsymbol{\lambda}}_i &: \sum_{t=1}^{T} g_{it}(\boldsymbol{\theta}, \hat{\boldsymbol{\lambda}}_i) = 0 \end{aligned} \tag{2.10}$$

where g_{it} is a suitable moment function. The statistical performances of the estimator are evaluated with a Monte Carlo experiment on a DCC scalar model in which it is shown that, as N increases, the composite likelihood method outperforms the classical quasi likelihood as regards the bias of the innovations' coefficient but not that of the autoregressive parameter; the performance of the composite likelihood built with each distinct pair of assets and the one built with the contiguous ones show analogous results, with a dramatic reduction of the estimation time in the latter case. The paper ends with an application of the composite likelihood estimator to a set of $N \leq 96$ assets (95 stocks plus a market index) and $T = 2516$ observations by fitting three different scalar models (scalar BEKK, EWMA and scalar DCC); the comparisons are made *via* an out-of-sample hedging exercise on the last one third of the sample. The application shows that, compared to the classical quasi likelihood estimator, the composite likelihood guarantees stability in the estimation of the parameters when N gets higher, so overcoming the bias problem, while the results of the Giacomini-White test (see Giacomini and White [2006]) ensure that the difference between the forecasts are significative for more than a half of the stocks. Furthermore, it is worth noting that the conditional correlations estimated with the composite likelihood are much more time-variable than the ones estimated with the classical method, which are almost constant throughout the period.

[5] In the second line of (2.10) the symbol 0 may represent a scalar, a vector or a matrix according to the structure of g_{it}

2.2.3 Dynamic Equicorrelation model

Factor models (see Section 1.4) can be a good solution to overcome the curse of dimensionality because they can model conditional correlations of large sets by using a very small number of factors. Further, they are generally efficient under the hypothesis of correct specification. Nevertheless, they impose stringent structures and this may cause misspecification; besides it is not always clear which and how many factors have to be employed. On the other hand, the composite likelihood (CL) method applied to the estimation of DCC specifications allows for more generality because different parameters are estimated for each pair of assets (see Section 2.2.2), but yields less efficient estimators given that little information is included in each likelihood function. Engle and Kelly (2012) propose a new estimator which explicitly bodes to be a solution to this tradeoff between the generality of CL-DCC and the efficiency of factor models: the dynamic equicorrelation (DECO) model. The crucial hypothesis is that all pairwise conditional correlations are equal in each period of time. This is an assumption that is sometimes pragmatically made in many areas of finance and economics. Equicorrelated matrices have an easy analytic formulation of their determinants and inverses, so likelihood computation and maximization is extremely simplified. Also, a bit of structure may be introduced into the system by blocking together (Block DECO) similar assets (for example those belonging to the same industry) and then obtaining determinants and inverses of the *blocked* correlations or, rather, applying to them the composite likelihood method. An interesting feature of DECO is that it can provide consistent estimates even when the equicorrelation assumption is irrealistic: more precisely the authors prove that, if DCC is consistent, DECO and Block DECO are consistent as well. The equicorrelation matrix is defined as:

$$R_t = (1 - \rho_t)I + \rho_t J \tag{2.11}$$

where I usually denotes the $N \times N$ identity matrix, while J is an $N \times N$ matrix of ones. It can be shown that:

$$
\begin{aligned}
R_t^{-1} &= \frac{1}{1-\rho_t} I - \frac{\rho_t}{(1-\rho_t)(1-(N-1)\rho_t)} J \\
|R_t| &= (1-\rho_t)^{N-1}(1-(N-1)\rho_t)).
\end{aligned}
\tag{2.12}
$$

R_t is positive definite if and only if $\rho_t \in]-\frac{1}{N-1}, 1[$, in which interval the inverse surely exists. A random vector of return residuals $\boldsymbol{\epsilon}_t$ follows a DECO model when $\text{var}(\boldsymbol{\epsilon}_t | I_{t-1}) = D_t R_t D_t$, where, as usual, D_t is the diagonal matrix of conditional standard deviations (modeled by GARCH specifications) and R_t is defined by Equation (2.11). Under weak conditions of regularity the usual two-step quasi maximum likelihood estimator is consistent. Also, Engle and Kelly (2012) prove the asymptotic normality of the estimator as a conjecture deriving from a White's theorem (White [1996]), pointing out that a rigorous analysis of asymptotic theory for M-GARCH is still a broadly unanswered question. The starting point in developing the DECO model is the consistent DCC specification proposed by Aielli (2013)[6]:

$$
\begin{aligned}
Q_t &= (1-\alpha-\beta)S + \alpha \left(I \odot Q_{t-1}\right)^{1/2} \boldsymbol{\eta}_{t-1} \boldsymbol{\eta}_{t-1}^{'} \left(I \odot Q_{t-1}\right)^{1/2} + \beta Q_{t-1} \\
R_t^{DCC} &= \left(I \odot Q_{t-1}\right)^{-1/2} Q_t \left(I \odot Q_{t-1}\right)^{-1/2} \\
\boldsymbol{\eta}_t &= D_t^{-1} \boldsymbol{\epsilon}_t;
\end{aligned}
\tag{2.13}
$$

then the model is fulfilled by the following system of dynamic equations:

$$
\begin{aligned}
R_t^{DECO} &= (1-\rho_t) I + \rho_t J \\
\rho_t &= \frac{1}{N(N-1)} \left(\boldsymbol{i}^{'} R_t^{DCC} \boldsymbol{i} - N \right)
\end{aligned}
\tag{2.14}
$$

where $\boldsymbol{i}$ denotes a N-dimensional vector of ones[7]. Stationarity and positive definiteness are ensured by the usual assumptions: S is positive definite; $\alpha + \beta < 1$; $\alpha, \beta > 0$. It can be easily

[6] See also Subsection 1.5.5

[7] Note that the last equation in (2.14) is nothing but the grand mean of all the scalar DCC conditional correlations

shown that the second-step likelihood segment takes the following shape:

$$L_2^{DECO} = -\frac{1}{T}\sum_{t=1}^{T}\ln\left([1-\rho_t]^{N-1}[1-(N-1)\rho_t]\right) + \\ -\frac{1}{T}\sum_{t=1}^{T}\frac{1}{1-\rho_t}\left(\sum_{i=1}^{N}\hat{\eta}_{it}^2 - \frac{\rho_t}{1+(N-1)\rho_t}\left(\sum_{i=1}^{N}\hat{\eta}_{it}\right)^2\right) \tag{2.15}$$

that does not need any numerical computation of determinants and inverses. One of the differences between DCC and DECO that the authors stress is the ability of the DECO model to exploit pooled information as the conditional equicorrelation absorbs all pairs of assets' history: $\rho_t = \frac{2}{N(N-1)}\sum_{i>j}\frac{q_{ijt}}{(q_{iit}q_{jjt})^{1/2}}$[8]. According to the authors, this *pooling ability* is the reason behind the globally better estimation performance of DECO compared to DCC, proved by their numerical experiments. Furthermore, in the paper in discussion it is asserted that, under high-level regularity conditions, DECO provides consistent estimates of the DCC parameters when DCC is the true model, even if the equicorrelation assumption is violated: this is DECO's winning card when the portfolio dimension exceeds certain orders of magnitude, so making the DCC model infeasible. The flexibility of DECO is particularly significant when the model takes a block structure, for example when assets belonging to the same industry are gathered together. In this case the equicorrelation matrix becomes:

$$R_t = \begin{bmatrix} (1-\rho_{11t})I_{n_1} & \cdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \cdots & (1-\rho_{kkt})I_{n_k} \end{bmatrix} + \begin{bmatrix} \rho_{11t}J_{n_1} & \cdots & \rho_{1kt}J_{n_1\times nk} \\ \cdots & \cdots & \cdots \\ \rho_{1kt}J_{n_1\times n_k} & \cdots & \rho_{kkt}J_{n_k} \end{bmatrix} \tag{2.16}$$

where k is the number of blocks and $n_1 + ... + n_k = N$. Nevertheless, when the number of blocks is greater than two the analytic calculation of inverses and determinants becomes untractable and so the structure needs to be forced to be diagonal: $\rho_{mlt} = 0\,, \forall l \neq m$. On

[8]This expression of ρ_t is equivalent to the formulation in (2.14)

the other hand, an interesting solution to the non-diagonal case is the composite likelihood[9] applied to the subsets consisting of all the pairs of blocks. Indeed, in case of two-block models analytic expressions of determinants and inverses remain quite simple[10] and this further simplifies the use of the CL procedure. As regards numerical experiments, Monte Carlo simulations show that the DCC performance[11] is poor when the data generating process is of DECO-type (equicorrelation); when the data are instead generated by DCC the DECO model seems to be consistent as $T \to \infty$, so confirming the theoretical claims of the authors.

2.2.4 An empirical evaluation of CL-DCC and DECO

In Clements, Scott and Silvennoinen (2012) an assessment of composite likelihood (CL-) DCC (see Subsection (2.2.2)) and DECO (see Subsection 2.2.3) is carried out. The two models are compared to some simpler methods for the estimation and forecasting of large covariance matrices and the tool utilized for evaluation is the out-of-sample volatility of minimum variance portfolios. The econometric exercise is applied to a selection of 200 stocks contained in the Standard&Poor1500 index, daily observed from 1994 to 2009 for 4029 global observations: the first 2000 days are used as estimation sample, whereas the last 2029 constitute the out-of-sample interval. A minimum variance portfolio is built for each out-of-sample period by solving the classic portfolio problem $\min_{\boldsymbol{w}_t} \boldsymbol{w}_t' H_{l,t} \boldsymbol{w}_t$ under the constraint $\boldsymbol{w}_t' \boldsymbol{i} = 1$, where $\boldsymbol{i}$ is a vector of ones and $H_{l,t}$ is the conditional covariance matrix generated from a generic model l (see Engle and Colacito [2006]). As is known from basic portfolio theory, the algebraic solution is $\boldsymbol{w}_t^* = \frac{H_{l,t}^{-1}\boldsymbol{i}}{\boldsymbol{i}' H_{l,t}^{-1}\boldsymbol{i}}$. Before weight computation, univariate volatilities are modeled with GJR-GARCH(1,1) specifications (see Glosten, Jagannathan and Runkle [1993]):

$$h_{it} = \omega_i + \alpha_i \epsilon_{i,t-1}^2 + \delta_i \, \mathrm{I}_{i,t-1} \, \epsilon_{i,t-1}^2 + \beta_i h_{i,t-1} \tag{2.17}$$

[9] See again Subsection 2.2.2

[10] For the sake of conciseness these expressions are not reported; however, they can be obviously found in the original article

[11] The estimation of the DCC specification is carried out with the composite likelihood

where $\epsilon_{i,t}$ denotes the i-th series of demeaned returns, with $\omega_i \geq 0$, $\alpha_i + \delta_i/2 \geq 0$, $\beta_i \geq 0$ and $\alpha_i + \delta_i/2 + \beta_i < 1$. The variable $\mathrm{I}_{i,t-1}$ is an indicator function which is equal to 1 when $\epsilon_{i,t-1} < 0$ and 0 otherwise[12]. The signficance of δ_i is tested and the parameter is omitted when its significance is rejected (at the 5% level). The volatility series obtained from (2.17) are as usual employed to standardize the returns: $\eta_{it} = \epsilon_{it}/h_{it}^{1/2}$. The second step of the evaluation exercise in Clements *et al.* (2012) is the estimation of CL-DCC and DECO in order to compute one-step-ahead forecasts of covariance matrices on the out-of-sample period[13]. Two estimation strategies are employed: the first one involves a *one-shot* parameter estimation only on the first 2000 data (fixed-window approach), while the second is carried out by re-estimating the parameters at each period $2000 + t$ ($t = 1, ..., 2029$) on all the previous observations (expanding-window approach). The predicted covariances are then utilized to construct the series of minimum variance portfolios according to the above reminded analytic solution. Moreover, alternative predicted covariance matrices are derived from simpler methods, namely a simple moving average (SMA)[14]:

$$Q_t^{sma} = \frac{1}{\tau} \sum_{j=1}^{\tau} \boldsymbol{\eta}_{t-j} \boldsymbol{\eta}_{t-j}^{'}; \tag{2.18}$$

an exponentially-weighted moving average (EWMA) from Fleming, Kirby and Ostdiek (2001) (see also Subsection 1.3.1):

$$Q_t^{ewma} = \mathrm{e}^{-\alpha}\, Q_{t-1}^{ewma} + \alpha\, \mathrm{e}^{-\alpha}\, \boldsymbol{\eta}_{t-1} \boldsymbol{\eta}_{t-1}^{'}; \tag{2.19}$$

a mixed data sampling specification (MIDAS) introduced by Ghysels, Santa-Clara and Valkanov (2006)[15]:

[12] The model (2.17) has been proposed by Glosten *et al.* (1993) in order to model the empirical finding of higher impacts of negativity shocks on volatility

[13] In both cases a cDCC model (see Equation (1.63)) with only two scalar parameters has been used; the CL-DCC estimation has been carried out on all the unique pairs of stocks

[14] In Equation (2.18) the $\boldsymbol{\eta}_t$'s are the vectors built with the components η_{it}, that is the standardized demeaned returns as previously mentioned

[15] $\overline{Q}$ is the unconditional covariance matrix of vectors $\boldsymbol{\eta}_t$

$$Q_t^{midas} = \overline{Q} + \varphi \sum_{j=1}^{\tau} \theta_j \boldsymbol{\eta}_{t-j} \boldsymbol{\eta}_{t-j}^{'}. \tag{2.20}$$

In the previous model equations the authors set $\tau = 252$ and $\alpha = 2/(\tau + 1)$; the weights θ_j are computed using the Beta function with parameters $(1, 0.98)$[16] and then normalized, while φ is estimated. After having estimated the series Q_t^{sma}, Q_t^{ewma} and Q_t^{midas} the covariance matrices are obtained in the usual way used for dynamic conditional correlation models[17]:

$$\begin{aligned} R_t^j &= \left(I \odot Q_t^j\right)^{-1/2} Q_t^j \left(I \odot Q_t^j\right)^{-1/2} \\ H_t^j &= D_t R_t^j D_t \\ D_t &= \text{diag}(h_{1t}^{1/2}, ..., h_{Nt}^{1/2}). \end{aligned} \tag{2.21}$$

The different models are assessed in terms of minimum portfolio variances for various problem size: $N = 5, 10, 50, 100, 200$. The evaluation is enriched by applying the model confidence set (MCS) procedure by Hansen, Lunde and Nason (2011) with loss function $L(H_t^j) = \boldsymbol{w}_t^{j*'} H_t^j \boldsymbol{w}_t^{j*}$. The MCS is an iterative procedure of elimination of the significantly worst models from the set and is based upon the null $H_0 : \text{E}(L(H_t^j) - L(H_t^k)) = \text{E}(d_{jkt}) = 0$. The procedure entails the comparison of all the pairs of models in the set and continues until failure to reject H_0. The test statistic is computed from the sample values and has a standard normal distribution:

$$t_{jk} = \frac{\overline{d}_{jk}}{\hat{\text{se}}(d_{jk})} \tag{2.22}$$

where $\overline{d}_{jk}$ and $\hat{\text{se}}(d_{jk})$ are the sample mean and sample standard error of d_{jkt}. The empirical results show a global superiority of the simpler method of estimation and forecasting, also when the out-of-sample interval is split into high and low volatility subperiods and basically for every size N, although the difference between performances tends to increase for high

[16] See, among others, Andrews, Askey and Roy (1999, chapter 1) about the features of the Beta function

[17] In equation (2.21), and on in the subsection, the superscripts j and k stand for two generic types of model

N (100 and 200); further, DECO globally outperforms CL-DCC and there is no substantial difference between the fixed-window and the expanding-window estimation strategies. Nevertheless, it is worth noting that the weights computed based on the simpler models are more variable than those based on DECO and, especially, on CL-DCC and this is an important practical shortcoming of SMA, EWMA and MIDAS because of the economic costs that portfolio reallocation has in practice.

2.2.5 Estimation of richly parameterized DCC models

In a recent paper Bauwens, Grigoryeva and Ortega (2014) have introduced a new method to estimate richly parameterized DCC models. The method consists in an algorithm based upon a procedure involving Bregman matrix divergences (see Bregman [1967]); it can solve likelihood maximization problems handling a high number of parameters and, simultaneously, non-linear stationarity and positivity constraints. The Bregman matrix divergence is a measure defined as follows[18,19]:

$$d_{\varphi}(X,Y) = \varphi(X) - \varphi(Y) + \text{trace}[(\nabla\varphi(Y))^{'}(X-Y)] \tag{2.23}$$

where X and Y are two $N \times N$ symmetric matrices, $\varphi : \mathbb{R}^{N\times N} \to \mathbb{R}$ is a convex and differentiable function and ∇ denotes the gradient operator. The Bregman measure includes different type of metrics according to the choice of φ. In the cited paper the authors use a special case called Stein's loss (see James and Stein [1961]):

$$d_{stein}(X,Y) = \text{trace}(XY^{-1}) - \ln|XY^{-1}| - N. \tag{2.24}$$

As is shown in the article, Bregman divergences are particularly appropriate in optimization when dealing with positive (semi)definiteness constraints. The econometric specification

[18] *Cf.* the equivalent formulation in Equation (3.5)

[19] The trace of a square matrix A is the sum of its diagonal elements: $\sum_{i=1}^{N}[A]_{ii}$ (see Lutkpohl [1996], p.41)

which the procedure is applied to is the most general version of Engle's DCC (see Equation (1.56)):

$$Q_t = (\boldsymbol{ii}^{'} - A - B) \odot \overline{Q} + A \odot \boldsymbol{\eta}_{t-1}\boldsymbol{\eta}_{t-1}^{'} + B \odot Q_{t-1}. \tag{2.25}$$

The intercept $\overline{Q}$ is obtaneid *via* the correlation targeting, while the parameter matrices take on different structure: A and B are alternatively assumed to be full rank, rank two and rank one[20]. In case of full rank parameter matrices, the model corresponds with the more general specification introduced in Ding and Engle (2001) and Engle (2002) in which the overall number of parameters is $N(N+1)$. It has been already stressed in Chapter 1 that a sufficient condition for such a model to yield positive definite covariance matrices is that A and B are positive semidefinite and $(\boldsymbol{ii}^{'} - A - B) \odot \overline{Q}$ is positive definite. Further, a sufficient condition for stationarity, that is:

$$E(Q_t) = Q \quad \text{where every entry of Q is constant and finite,} \tag{2.26}$$

is $|a_{ij} + b_{ij}| < 1\,, \forall i, j \in \{1, ..., N\}$. The full rank version of the model is the more flexible form but it is practically unfeasible when the dimension of portfolio N is large, as long as ordinary numeric optimization method are used. On the other hand, the algorithm proposed in Bauwens *et al.* (2014) can overcome this complication and reach convergence in the constrainted maximization of the likelihood function. A computationally feasible alternative is to work with reduced rank matrices. The rank one matrices are built as outer self-products of two vectors, $A = \boldsymbol{aa}^{'}$ and $B = \boldsymbol{bb}^{'}$: in one case $\boldsymbol{a}$ and $\boldsymbol{b}$ are two $\mathbb{R}^N$-vectors having independent components, for a total number of $2N$ parameters; instead, in the other case, they are defined by using an Almon function (see Almon [1965]), that is their generic i-th element is of type $v_1 + \exp(v_2\dot{i} + v_3\dot{i}^2)$: in other words, the Almon operator ($: \mathbb{R}^3 \to \mathbb{R}^N$) shifts a 3-component vector into an N-component one. Therefore, when using the Almon function there will be only 6 parameters to be estimated, say $(\alpha_1, \alpha_2, \alpha_3)$ and $(\beta_1, \beta_2, \beta_3)$, and

[20] as regards the rank one specification *cf.* Equation describing the G-DCC by Hafner and Franses (2009) (Subsection 1.6.2)

for this reason the so-called Almon DCC turns out to be very parsimonious and a natural competitor of the scalar specification. It is also noteworthy that the Almon specification is a novel contribution introduced by Bauwens *et al.* (2014) in the DCC literature. In case of rank one parameter matrices with $A = \boldsymbol{a}\boldsymbol{a}^{'}$ and $B = \boldsymbol{b}\boldsymbol{b}^{'}$ the positive definiteness constraints reduce to the positive definiteness of $(\boldsymbol{i}\boldsymbol{i}^{'} - A - B) \odot \overline{Q}$, while the stationarity is ensured by $|a_i a_j + b_i b_j| < 1\ ,\ \forall i, j \in \{1, ..., N\}$. The authors also use two (necessary) identification constraints: $a_1 > 0$, $b_1 > 0$. All these constraints remain virtually valid also in case of use of the Almon function, with the only difference that in this case each one of the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ is built with only 3 parameters. The rank two parameter matrices are obtained by the self-product of two matrices, $A = \tilde{A}\tilde{A}^{'}$ and $B = \tilde{B}\tilde{B}^{'}$, where the matrices denoted by the *tilde* belong to the space $\mathbb{R}^{N\times 2}$ and are of the following type[21]:

$$\begin{bmatrix} v_1 & 0 \\ v_2 & v_{N+1} \\ \vdots & \vdots \\ v_N & v_{2N-1} \end{bmatrix}. \tag{2.27}$$

Therefore any parameter matrix of the rank two specification is defined by a vector with $2N-1$ components, so the model needs a total amount of $4N-2$ parameters. The conditions for stationarity are $\sum_{k=1}^{2} |\tilde{a}_{ik}\tilde{a}_{jk} + \tilde{b}_{ik}\tilde{b}_{jk}| < 1\ ,\ \forall i, j \in \{1, ..., N\}$, while the condition for the positive definiteness is guaranteed by the positive definiteness of $(\boldsymbol{i}\boldsymbol{i}^{'} - \tilde{A}\tilde{A}^{'} - \tilde{B}\tilde{B}^{'}) \odot \overline{Q}$ provided that the starting pseudo-correlation Q_0 is positive semidefinite. Furthermore, the identification constraints are in this case $a_1 > 0$, $a_{N+1} > 0$, $b_1 > 0$ and $b_{N+1} > 0$. The different parametrizations are tested in an empirical study involving daily closing prices of the thirty components included in the Dow Jones Industrial Average Index (DJIA) belonging to two consecutive intervals, each of 3750 days (the first 3000 observations are used for estimating the parameters, the residual 750 for an out-of-sample study). It is worth noting that the residuals are obtained by preprocessing the raw data according to the capital asset

[21] In the discussed paper definitions and conditions for generic rank r parameter matrices are reported, but here only the rank two case is presented as it is the only actually used in the empirical application along with the rank one, the Almon specification and the full rank

pricing model (CAPM) (*i.e.* a one-common-factor regression model with intercept). The goodness of the various specifications is evaluated by using the model confidence set by Hansen *et al.* (2011). In this extensive empirical study the authors find substantial evidence that some of the proposed parametrizations are advantageous (more precisely the Almon and the rank one models) compared to the classical scalar DCC.

2.3 Methods for the regularization of covariance and correlation matrices

This section presents two techniques for regularizing sample covariance and correlation matrices in case they are ill-conditioned or even singular. In more detail, the section contains: an illustration of the shrinkage technique of estimation and regularization of sample covariance matrices and of its applications in finance, based on the work of Ledoit and Wolf (2004a) (Subsection 2.3.1); a presentation of two modified versions of the above mentioned shrinkage technique introduced in Ledoit and Wolf (2003) and Ledoit and Wolf (2004b) in which the choice of the so-called target matrix is different from that in Ledoit and Wolf (2004a) (Subsection 2.3.2); the discussion on an empirical application of shrinkage estimation techniques to covariance targeting carried out by Hafner and Reznikova (2012) (Subsection 2.3.3); an overview of an alternative technique of regularization, borrowed from the random matrix theory (RMT), known as eigenvalue cleaning, introduced in econometrics by Laloux, Cizeau, Bouchaud and Potters (1999) (Subsection 2.3.4); a summary of two applications in which the eigenvalue cleaning is employed (Tola, Lillo, Gallegati and Mantegna [2008]; Hautsch, Kyj and Oomen [2009]) (Subsection 2.3.5).

2.3.1 Shrinkage regularization of large covariance matrices

It is known from matrix algebra and mathematical statistics that, when the number of variables is not negligible compared to the number of observations, sample covariance matrices

are typically not well-conditioned[22]. Further, when the sample length is not greater than the number of variables the covariance matrix is not even invertible. Indeed, consider a random vector $\boldsymbol{x}_t \in \mathbb{R}^N$ and the matrix X made up of T observations of $\boldsymbol{x}_t$; its mean and covariance matrix are:

$$\begin{aligned} \boldsymbol{\mu} &= \frac{1}{T} X \boldsymbol{i} \\ S &= \frac{1}{T} X (I - \frac{1}{T} \boldsymbol{i}\boldsymbol{i}^{'}) X^{'} \end{aligned} \tag{2.28}$$

with $\boldsymbol{i}$ an N-dimensional vector of ones and I an $N \times N$ identity matrix. From matrix algebra it is known that $\text{rank}(AB) \leq \min[\text{rank}(A), \text{rank}(B)]$ and that a square matrix is invertible only if it is full rank (see for example Magnus and Neudecker [2007], p.8)). In the matrix $(I - \frac{1}{T}\boldsymbol{i}\boldsymbol{i}^{'})$ each of the columns can be obtained as a sum of the remaining ones multiplied by -1 and so, when $N \geq T$, it is $\text{rank}(S) \leq \text{rank}(I - \frac{1}{T}\boldsymbol{i}\boldsymbol{i}^{'}) = T - 1$ and the $N \times N$ covariance matrix S is singular. Ill-conditioned or even singular covariance matrices are often found in practical applications. Ledoit and Wolf (2004a) introduce in econometrics the so-called shrinkage, a method of regularization of ill-conditioned or singular sample covariances which does not only yield well-conditioned but even more accurate estimated covariances than the ordinary moment estimator does. The method is not new in mathematical statistics and dates back to the classical work of James and Stein (1961). In its standard version, the shrinkage estimator consists of a weighted average of the sample covariance and the most simple structured covariance estimator[23]: a rescaled identity matrix. The difficulty the authors have to face is that the true optimal weight depends on the unobservable true covariance, but this issue is solved by finding a consistent estimator of the weight (the so-called shrinkage intensity) which makes the estimated shrinkage asymptotically equivalent to the *true* one. The asymptotical results on the shrinkage and weight estimator are derived under what the authors define general asymptotics, that is with both the number of variables, N, and the number of sample observations, T, going to infinity, plus the additional constraint

[22] A square matrix is ill-conditioned when its condition number (that is the ratio of the largest to smallest singular value in the singular value decomposition) is too large: this means a determinant next to 0 and numerical problems in inversion (see Gentle [2007], Chapter 6)

[23] The structured estimator used to *shrink* the matrix to be regularized is usually called target

that the ratio N/T must remain bounded[24]. The *true* shrinkage estimator is given by the minimization of the squared Frobenius norm[25], $\| \ \|^2$, of the difference between the true covariance, Σ , and the shrinkage itself, Σ^*:

$$\begin{aligned} &\min_{\mu,\alpha} E(\|\Sigma^* - \Sigma\|^2) \\ &\text{s.t. } \Sigma^* = \alpha\mu I + (1-\alpha)\, S \end{aligned} \tag{2.29}$$

whose optimal solutions are:

$$\begin{aligned} \mu^* &= \text{trace}(\Sigma)/N \\ \alpha^* &= \frac{E(\|S - \Sigma\|^2)}{E(\|S - \mu^* I\|^2)}. \end{aligned} \tag{2.30}$$

The shrinkage must be interpreted as the optimal tradeoff between bias and variance in the mean square error (MSE):

$$E(\|\Sigma^* - \Sigma\|^2) = E(\|\Sigma^* - E(\Sigma^*)\|^2) + \|E(\Sigma^*) - \Sigma\|^2 \tag{2.31}$$

where the MSE is all bias when $\Sigma^* = I$, while it is all variance when $\Sigma^* = S$ (provided that $E(S) = \Sigma$). As previously mentioned (see Equations (2.30)), the true optimal weight depends on the true latent covariance Σ which is unknown and not observable. However, Ledoit and Wolf (2004a) show under some general-asymptotics conditions that the solutions can be consistently estimated as:

$$\begin{aligned} \hat{\mu} &= \text{trace}(S)/N \\ \hat{\alpha} &= \frac{T^{-2} \sum_{i=1}^{T} \|\boldsymbol{x}_{.i}\boldsymbol{x}_{.i}' - S\|^2}{\|S - \hat{\mu} I\|^2} \end{aligned} \tag{2.32}$$

[24] *General* is counterposed to *standard* asymptotics that is characterized by $T \to \infty$ and N fixed: under standard asymptotics the sample covariance is asymptotically well defined

[25] The Frobenius norm of a real-valued matrix A is equal to $\sqrt{\text{trace}(A'A)}$ (see Lutkepohl [1996], p.103)

where $\boldsymbol{x}_{\cdot i}$ denotes the i-th column of the data matrix X and, then, the estimated shrinkage $\hat{\Sigma}(\hat{\mu}, \hat{\alpha})$ is a consistent (in quadratic mean and norm) estimator of the true shrinkage Σ^* and has the same asymptotic expected loss, that is for $T \to \infty$[26]:

$$\begin{aligned} \|\hat{\Sigma} - \Sigma^*\| &\xrightarrow{qm} 0 \\ E\|\hat{\Sigma} - \Sigma\|^2 - E\|\Sigma^* - \Sigma\|^2 &\to 0. \end{aligned} \tag{2.33}$$

Through Monte Carlo simulations Ledoit and Wolf's shrinkage is compared to other estimators from finite sample decision theory (namely: the Haff estimator; the Stein-Haff estimator; the Dey-Srinivasan-Stein estimator) and it turns out to be the one with the bigger percentage gain in expected loss compared to that of the sample covariance under fixed values of N/T, NT and eigenvalue dispersion of the true covariance. Moreover, the shrinkage estimator shows the best performances also when the mentioned parameters (N/T and NT) are let to vary[27]. The asymptotic theory developed by the authors approximates finite sample behaviour well as soon as N and T become of the order of 20. Surprisingly, the Ledoit-and-Wolf shrinkage is not only more accurate (that is with a smaller expected loss) than the sample covariance but is even better conditioned than the true covariance matrix: its condition number has the smaller growth when either the the ratio N/T or the eigenvalue dispersion increases and converges quickly (to a smaller constant) when the product NT grows. It is worth noting that the estimated *shrunk* covariance remains well defined and invertible even when $N > T$.

2.3.2 Shrinkage with alternative targets

The same idea discussed in Subsection 2.3.1 is applied in Ledoit and Wolf (2003), where this time the target is not the identity matrix but the covariance deriving from a single-factor model:

[26] A series of random variables X_n converges in quadratic mean to a random variable X ($X_n - X \xrightarrow{qm} 0$) if $E(X^2) < +\infty$ and $\lim_{n\to\infty} E(X_n - X) = 0$ (see Rohatgi and Ehsanes Saleh [2001], p.263)

[27] In more detail, N/T is allowed to change from $4/200$ up to $200/4$ while NT from 3×6 up to 40×80

$$\begin{aligned} \boldsymbol{x}_t &= \boldsymbol{\alpha} + \boldsymbol{\beta} f_t + \boldsymbol{\epsilon}_t \\ \text{var}(\boldsymbol{x}_t) &= \Psi = \sigma_f^2 \boldsymbol{\beta}\boldsymbol{\beta}' + \Delta \end{aligned} \tag{2.34}$$

where f_t is the factor (a broad-based market index) with variance σ_f^2 and $\Delta = \text{var}(\boldsymbol{\epsilon}_t) = \text{diag}(\delta_1, ..., \delta_N)$ (residuals assumed to be uncorrelated). The model defined by Equations (2.34) can be estimated *via* a linear regression and σ_f^2 by its sample counterpart. Also in this case the basic intuition is that the sample covariance matrix S has a lot of estimation error though it is consistent, whereas the covariance Ψ (and hence its least squares estimator $\hat{\Psi}$) has a lower estimation error but it is biased due to a stringent structure assumption: in the wake of Stein (1956) and following literature, Ledoit and Wolf (2003) assert that there will be an optimal point between these two extremes and this validates the use of a suitable linear combination of them. In this case the asymptotic framework is standard: $T \to \infty$ and N fixed. The optimal shrinkage intensity is obtained by minimizing the expected loss $E[\|\alpha\hat{\Psi} + (1-\alpha)S - \Sigma\|^2]$, where Σ denotes the true latent covariance matrix. In the paper in discussion it is shown that the true optimal shrinkage intensity asymptotically is :

$$\begin{aligned} \alpha^* &= \frac{1}{T}\frac{\pi - \rho}{\gamma} + O(1/T^2) \\ \pi &= \sum_i \sum_j \text{avar}(\sqrt{T} s_{ij}) \\ \rho &= \sum_i \sum_j \text{acov}(\sqrt{T} s_{ij}, \sqrt{T}\hat{\psi}_{ij}) \\ \gamma &= \sum_i \sum_j (\psi_{ij} - \sigma_{ij}) \end{aligned} \tag{2.35}$$

with $O(1/T^2)$ negligible infinitesimal and where avar and acov denote the asymptotical variance and covariance. The shrinkage intensity α^* in (2.35) asymptotically vanishes as $1/T$ and depends on unobservables (namely on the couples (ψ_{ij}, σ_{ij}))[28], but Ledoit and Wolf

[28] According to the definition used by the authors, it is not a *bona fide* estimator

(2003) obtain a consistent estimator of it[29]. The shrinkage-to-market estimator is tested on a problem of portfolio selection (Markowitz [1952]):

$$\begin{aligned} &\min_{\boldsymbol{w}} \boldsymbol{w}^{'}\Sigma\boldsymbol{w} \\ &\text{s.t. } (\boldsymbol{w}^{'}\boldsymbol{i}=1)\cap(\boldsymbol{w}^{'}\boldsymbol{\mu}=q) \end{aligned} \tag{2.36}$$

where as usual $\boldsymbol{\mu}$ and Σ are the mean and the covariance of a random vector of returns, $\boldsymbol{i}$ is a vector of ones and q is a scalar constant. The solution to the portfolio problem depends on the inverse of the (estimated) covariance Σ, hence it needs to be invertible and well conditioned. The application is carried out on monthly returns of a big basket of stocks ($N \simeq 1000$)[30] assuming re-estimation intervals much smaller than N ($T = 120$) and portfolio updates on every August. The shrunk-to-market covariance is checked on its out-of-sample performance and is compared to some competing estimators: identity matrix, constant correlation, pseudo-inverse, Sharpe's model, k-factor model, principal component analysis, shrinkage to identity (see Equation (2.29)). In the application the shrinkage-to-market estimator is the one that performs better in terms of out-of-sample portfolio standard deviation in the case of both unconstrained and constrained global return (q in the problem (2.36)).

Another application of the shrinkage approach to portfolio optimization is carrried out in Ledoit and Wolf (2004b). In this case the shrinkage is obtained by linearly combining the sample covariance matrix with a constant-correlation covariance. The analytic formulation of the problem in this case is:

$$\begin{aligned} &\min_{\boldsymbol{x}} \boldsymbol{x}^{'}\Sigma\boldsymbol{x} \\ &\text{s.t. } \boldsymbol{x}^{'}\boldsymbol{\alpha}\geq g \\ &\quad \boldsymbol{x}^{'}\boldsymbol{i}=0 \\ &\quad -\boldsymbol{w}_B\leq\boldsymbol{x}\leq c\boldsymbol{i}-\boldsymbol{w}_B \end{aligned} \tag{2.37}$$

[29] See the article for the analytic expression

[30] The number of considered stocks changes throughout the period of observation 1972 – 1994

where $\boldsymbol{w}_B$ are the weights of a chosen benchmark portfolio while $\boldsymbol{x}$ represents the deviations of the weights of the chosen allocation relative to the benchmark; further, $\boldsymbol{\alpha}$ is the vector of the expected extra-returns (compared to those of the benchmark), g is the minimum performance level and c is the biggest allowed position in each stock. The two quantities to be estimated are α and Σ[31] . This time the shrinkage is obtained as a linear combination of the sample covariance matrix $S = [s_{ij}]$ and a constant-correlation covariance $F = [f_{ij}]$ which is defined as follows:

$$f_{ij} = \begin{cases} s_{ij} & \text{if } i = j \\ \bar{r}(s_{ii}s_{jj})^{1/2} & \text{if } i \neq j \end{cases} \tag{2.38}$$

where:

$$\bar{r} = \frac{2}{N(N-1)} \sum_{i<j}^{1,\ldots,N} r_{ij} \tag{2.39}$$

and where the r_{ij}'s are the sample correlations. Hence, the shrinkage estimator is $\hat{\Sigma} = \delta F + (1-\delta)S$. The optimal shrinkage intensity δ is computed on the same theoretical results obtained in Ledoit and Wolf (2003) and its analytic expression is provided in the appendix of their paper[32]. Also in this case an application on monthly data (taking into account different cross-section sizes) shows the superiority of portfolio management based on the shrinkage rather than the widely used sample covariance.

2.3.3 Shrinkage regularization in covariance targeting

In Hafner and Reznikova (2012) the bias affecting the quasi maximum likelihood estimator in case of application to DCC and large portfolios (see Subsection 2.2.2) is ascribed to the ill-conditioning of the sample covariance matrix used in the covariance targeting. As is known, there are two main problems related to the classic DCC estimation. The first is

[31] The estimation of the extra-returns is explicitly kept out from the purposes of the article
[32] We refer to the article for the technical details

the inconsistency due to the original formulation of the dynamic equation based upon the incorrect assumption: $S = E(\boldsymbol{\eta}_t \boldsymbol{\eta}_t^{'})$. As we have seen in Chapter 1, Aielli (2013) addresses this problem by correcting Engle's original dynamic specification as follows (cDCC):

$$\begin{aligned} Q_t &= (1-\alpha-\beta)S + \alpha \left(I \odot Q_{t-1}\right)^{1/2} \boldsymbol{\eta}_{t-1}\boldsymbol{\eta}_{t-1}^{'} \left(I \odot Q_{t-1}\right)^{1/2} + \beta Q_{t-1} \\ S &= \frac{1}{T}\sum_{t=1}^{T} \left(I \odot Q_t\right)^{1/2} \boldsymbol{\eta}_t \boldsymbol{\eta}_t^{'} \left(I \odot Q_t\right)^{1/2}. \end{aligned} \tag{2.40}$$

The second problem is the bias associated with the estimates of the dynamic parameters α and β when the cross-section size is large and the likelihood decomposition is used in the estimation (see Engle and Sheppard [2001]). This second issue is solved in Engle *et al.* (2008) with the introduction of the composite likelihood (see Subsection 2.2.2). Hence, combining the composite likelihood method and the corrected DCC of Aielli one can obtain consistent estimates of both the intercept and the dynamic parameters. Nevertheless, the composite likelihood is not efficient because of the partial use of information due to subsampling. In the article in discussion Hafner and Reznikova (2012) claim that the bias of the dynamic parameters is caused by ill-conditioned sample matrices employed in covariance targeting. Consequently they propose to use the shrinkage technique (Ledoit and Wolf [2003], [2004a], [2004b]) in order to numerically regularize such sample covariances (see Subections 2.3.1 and 2.3.2). The estimated shrinkage intensity associated to the target matrix (Equations (2.32) and (2.35)) decreases as $1/T$ with each one of the three versions of the shrinkage estimators, so it practically vanishes for large T and Ledoit and Wolf's estimators become asymptotically equivalent to the efficient ordinary MLE. The authors run a numerical experiment with differently sized portfolios ($N = 5, 10, 100, 200$) and time spans ($T = 100, 250, 500, 1000, 2000$). The data are simulated on the basis of N univariate GARCH models ($\omega_i = 0.01, \alpha_i = 0.05, \beta_i = 0.9$) plus, alternatively, a DCC and an Aielli's cDCC with $\alpha = 0.05, \beta = 0.93$. Then, eight parametric models are used in the estimation: plain DCC and cDCC; DCC and cDCC with shrinkage respectively to identity matrix, to single-factor model and to equicorrelated covariance. Further, the no-shrinkage versions of DCC and cDCC are estimated with both the complete and the composite likelihood. In summary,

the Monte Carlo experiment results show that the parameter bias of ordinary DCC generally grows as N increases; on the other hand the shrinkage-adjusted DCCs reach apparent improvement in terms of bias, particularly for the α coefficient. As regards the root mean square errors the shrinkage DCCs are substantially equivalent to the classic ones, while gains compared to the composite likelihood DCC are observed when N is small. There is no significant global difference between Engle's and Aielli's specification; however, it is worth noting that the improvements shown by the shrinkage-adjusted DCCs are less relevant when the ordinary DCC is estimated with the composite likelihood. Furthermore, the numerical simulations illustrate how shrinkage intensities decreases toward zero as the sample dimension T grows. The article ends with an empirical example carried out on stocks belonging to the *S&P500* observed throughout a (short) sample ($T = 521$), with different size of cross-sections ($N = 5, 10, 25, 50, 75, 100$). The returns are previously cleaned by employing AR specifications and then the various models seen above are applied to them. The ordinary DCC and cDCC specifications show the typical bias problems[33] which are instead overcome by the composite likelihood estimator and, to a minor extent, by the various shrinkage-adjusted DCCs. Like in the simulation, the shrinkage intensity decreases as the cross-sectional dimension grows. The conditional correlations are almost constant when estimated with the classic DCC, while the more variable are the ones derived from the composite likelihood DCC; the shrinkage DCCs show a halfway situation.

2.3.4 Eigenvalue cleaning

In Sections 2.3.1, 2.3.2 and 2.3.3 the shrinkage regularization technique and its use in financial econometrics have been discussed. However, other regularization procedures exist in mathematical statistics and algebra. One of these alternative techniques has been developed starting from random matrix theory (RMT) and has been applied in financial econometrics by Laloux, Cizeau, Bouchaud and Potters (1999): it is called eigenvalue cleaning. Empirical correlation matrices usually contain measurement noise which generally increases as the ratio of observations to number of time series, T/N, gets smaller. On the other hand, in

[33] More precisily, in this context the bias is indirectly deduced from the strong decreasing tendency of α and β when N gets higher

econometric applications and asset management it is crucial to distinguish *noise* from *signal* (*i.e.* true information). In order to do that, Laloux *et al.* (1999) suggest to compare the considered empirical correlation matrix to its pure-noise counterpart, that is the one obtained from finite strictly independent time series: deviations from this random matrix case can detect the presence of true information. In the article the empirical correlation matrix C is assumed to be constructed from the demeaned time series of returns:

$$C = \frac{1}{T} M M' \tag{2.41}$$

where M is the $N \times T$ data matrix. In case of independent and identically distributed matrix entries, M becomes a so-called random Wishart matrix whose eigenvalue distribution, for $N \to \infty$ and $T \to \infty$ with $T/N = constant \geq 1$, is (see for example Mehta [1995])[34]:

$$\begin{aligned} \rho(\lambda) &= \frac{1}{N}\frac{dn(\lambda)}{d\lambda} = \frac{T\sqrt{(\lambda - \lambda_{max})(\lambda - \lambda_{min})}}{2\pi\sigma^2 N\lambda} \\ \lambda_{min}^{max} &= \sigma^2(1 + \frac{N}{T} \pm 2\sqrt{\frac{N}{T}}) \end{aligned} \tag{2.42}$$

where $dn(\lambda)$ naturally denotes the number of eigenvalues less than λ, λ_{min} and λ_{max} are respectively the minimum and maximum eigenvalue and σ^2 is the global variance of the $N \times T$ data matrix[35]. A sample correlation will show more or less wide deviations from the distribution in (2.42) and, as mentioned, these deviations from the random matrix case should suggest the presence of true information (signal hypothesis). Basing on this theoretical background, the authors carry out an empirical application on 1309 daily observations of 406 price series of *S&P500* stocks ($T/N = 3.22$). They find that the empirical maximum eigenvalue, $\hat{\lambda}_{max}$, is 25 times bigger than the theoretical λ_{max}, showing that the pure-noise hypothesis is incongruous. The empirical maximum eigenvalue is interpreted as the market

[34] The distribution shown in (2.42) is known as Marchenko-Pastur distribution

[35] σ^2 can be set equal to 1 when considering the theoretical correlation matrix (rescaled data with zero mean and unit volatility)

signal and it is related to the common behaviour of the considered assets. Then, in order to obtain a more reliable theoretical distribution, the contribution of $\hat{\lambda}_{max}$ to the overall variance is subtracted: $\sigma^2 = 1 - \frac{\hat{\lambda}_{max}}{N}$. As many eigenvalues keep being larger than the theoretical λ_{max}, this subtraction procedure is repeated until the best fitting is reached; in other terms, σ^2 is treated as a tuning parameter. Practically, the discussed procedure is used to split the empirical correlation matrix in a signal component, described by the eigenvalues laying outside the adjusted theoretical distribution, and a purely random component, relative to the eigenvalues well described by the adjusted theoretical distribution. The adjusted correlation matrix is obtained by replacing the noise component of the empirical correlation with a rescaled identity matrix, such that the trace is conserved, and is utilized in the empirical application consisting in a classic Markowitz portfolio problem. The results show that the estimate of future risks on optimized portfolios is substantially improved with the *cleaned* correlation matrix and, on the other hand, that the plain sample correlation underestimate the real risk.

2.3.5 Other applications of eigenvalue cleaning

An application of the theory described in Laloux *et al.* (1999) (Subsection 2.3.4) can be found in Tola, Lillo, Gallegati and Mantegna (2008). The authors point out that the eigenvalues of a sample correlation matrix of stock time series almost always can be divided into three classes: the biggest one that is totally incompatible with the theoretical distribution in (2.42) and that describes the common behaviour of all stocks (the so-called market signal); a certain number of eigenvalues between the largest one and the theoretical λ_{max}, which probably describe economic information different from the market signal; the remaining large part of the eigenvalues between the theoretical λ_{min} and λ_{max} that are congruous with the theoretical distribution in (2.42) and, therefore, come out from measurement noise. That being said, the authors suggest to use the following algorithm in order to clean the correlation matrix in the framework of global minimum variance portfolio: compute the sample correlation and its eigenvalues $\lambda_1 < ... < \lambda_N$; set $\sigma^2 = 1 - \frac{\lambda_N}{N}$ in equation (2.42); find λ_{min} and λ_{max} by using the (2.42); build the filtered eigenvalue matrix by setting to zero all the eigenvalues smaller than λ_{max}; reconstruct the related filtered correlation by pre- and post-multiplying

the filtered eigenvalue matrix by the original eigenvector one, forcing the diagonal elements to be 1. The adjusted correlation will preserve only the information recognized as signal. In the paper an empirical application is shown which presents a superior predictive ability compared to the Markowitz basic estimation.

Another application of the eigenvalue cleaning is carried out in Hautsch, Kyj and Oomen (2009). Working on high frequency data, they use the kernel estimator proposed in Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008) to estimate large block-wise realized covariance matrices, where the assets are blocked together with respect to their trading frequencies. These realized covariances are constructed upon a series of asset clusters whose prices are observed with different sampling frequencies (the so-called refresh time samplings) so that the global number of observations taken into account is larger for the more liquid set of assets and gradually smaller for the more illiquid ones; by doing so, the authors claim gains in efficiency (that is minimization of data loss) and, simultaneously, an adequate control of the market microstructure noise (which tends to become larger as the sampling frequency increases). Nevertheless, a built-by-blocks covariance may not be in general well defined, so they exploit the random matrix theory findings presented in Laloux *et al.* (1999): hence, in this case the regularization technique is not employed to filter out the noise but to obtain instead positive definite and well conditioned matrices. The regularization is carried out on the correlation matrices and the first algorithm step is the computation of the eigenvalue decomposition of each realized correlation: $\hat{R} = \hat{Q}\hat{\Lambda}\hat{Q}'$, where $\hat{\Lambda}$ is the diagonal matrix of eigenvalues and $\hat{Q}$ the relative eigenvector matrix. The larger eigenvalue is then employed to filter out the market signal and to adjust the distribution in Equations (2.42), in more detail: $\tilde{\lambda}_{max} = (1 - \frac{\hat{\lambda}_{max}}{N})(1 + \frac{N}{T} + 2\sqrt{\frac{N}{T}})$. Afterwards, the original eigenvalue structure $\hat{\Lambda}$ is modified into $\tilde{\Lambda}$ as follows:

$$\tilde{\lambda}_i = \begin{cases} \hat{\lambda}_i & \text{if } \hat{\lambda}_i > \hat{\lambda}_{max} \\ \delta & \text{if } \hat{\lambda}_i \leq \hat{\lambda}_{max} \end{cases} \tag{2.43}$$

where δ is the following quantity:

$$\delta = \frac{\text{trace}\,\hat{R}_+ - \sum_{\hat{\lambda}_i > \hat{\lambda}_{max}} \hat{\lambda}_i}{N - \#(\hat{\lambda}_i > \hat{\lambda}_{max})}. \tag{2.44}$$

The symbol $\hat{R}_+$ stands for the positive semidefinite projection of matrix $\hat{R}$, that is the matrix $\hat{Q}\hat{\Lambda}^0\hat{Q}'$ where in $\hat{\Lambda}^0$ all the $\hat{\lambda}_i < 0$ are set equal to zero. The transformation in Equation (2.44) ensures the positive definiteness of the recontructed $\tilde{\Lambda}$. The article in discussion is integrated with an extensive numerical study devised to simulate the empirical features of the *S&P1500* universe, in which the block-wise realized kernel estimator shows significant improvements compared to the plain realized kernel by Barndorff-Nielsen *et al.* (2008), especially in case of high dimensions and heterogeneous observation frequencies of assets. Furthermore, an empirical application is carrried out to yield out-of-sample forecasts of the volatility of portfolios made up of *S&P500* components and, also in this case, the proposed estimator shows significant gains over the one-block realized kernel estimator.

Chapter 3

A new flexible estimation method for DCC models

3.1 Introduction

In the previous chapters the issue of the tradeoff between flexibility and feasibility of M-GARCH models has been repeatedly raised and discussed and a synthetic overview of literature devolopment in this field of econometrics has been presesented. As mentioned, after the introduction of Engle's DCC, many methods have been proposed in the area of models for conditional correlations in order to allow more parametrized specifications to be estimated without extremely overloading computations; in most of the cases, these approaches are based upon some typical assumptions that we shortly recall for reader's convenience (see Chapter 1 for further details):

- the parameter matrices which define the most general DCC specification have some simple *a priori* structure (*e.g.* diagonal, block-diagonal, rank deficient);
- the returns (hence their covariance structure, too) can be decomposed according to a factor model.

Obviously both of these strategies may implicate in general a certain loss of information because of parameter pooling or other imposed simplifications. The goal of this chapter is to propose in the DCC framework a new method able to exploit more information without imposing any *a priori* structure and at the same time without incurring overwhelming calculations. In more detail, this chapter contains: a recall of the general DCC framework and of its limits and the outline of a possible solution (Section 3.2); the introduction of a shrinkage-type predictor inspired by the shrinkage estimator of Ledoit and Wolf (2004a) (Section 3.3); the development of a novel method for obtaining more accurate conditional correlation forecasts (Section 3.4); an application to real data aimed to provide an empirical assessment of the performances of the proposed predictors (Section 3.5) and, finally, a comparison between the proposed predictors and those implied by some richly parametrized DCC models (Section 3.6).

3.2 Some limitations of the general DCC framework for high-dimensional portfolios

As it has been seen in Chapter 1, given a vector time series of returns $\boldsymbol{r}_t$, with conditional expectation $E(\boldsymbol{r}_t|I_{t-1})$ and cross-sectional dimension N, the most general version of the DCC model (Engle [2002]) is completely described by the following equations:

$$
\begin{aligned}
&\mathbf{r}_t - \boldsymbol{\mu}_t = \boldsymbol{\epsilon}_t = H_t^{1/2}\mathbf{z}_t \quad , \quad \mathbf{z}_t \sim \text{i.i.d.}(\mathbf{0}, I) \\
&D_t = \text{diag}(h_{11,t}^{1/2}, ..., h_{NN,t}^{1/2}) \quad , \quad h_{ii,t} = \text{var}(\epsilon_{it}|I_{t-1}) \\
&\boldsymbol{\eta}_t = D_t^{-1}\boldsymbol{\epsilon}_t \\
&\overline{Q} = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\eta}_t\boldsymbol{\eta}_t^{'} \\
&Q_t = \left(\boldsymbol{ii}^{'} - A - B\right) \odot \overline{Q} + A \odot \boldsymbol{\eta}_{t-1}\boldsymbol{\eta}_{t-1}^{'} + B \odot Q_{t-1} \\
&R_t = \left(Q_t \odot I_t\right)^{-1/2} Q_t \left(Q_t \odot I_t\right)^{-1/2} \\
&H_t = D_t R_t D_t
\end{aligned}
\tag{3.1}
$$

where the parametric symmetric matrices A and B have dimension equal to $N \times N$ and are full rank, for a global amount of $N(N+1)$ parameters (because of the symmetry of the two matrices). Due to the computational infeasibility of this model even when N is moderately large, Engle (2002) proposes as an alternative a scalar version in which A and B collapse into scalars:

$$Q_t = (1 - \alpha - \beta)\,\overline{Q} + \alpha \boldsymbol{\eta}_{t-1}\boldsymbol{\eta}_{t-1}^{'} + \beta Q_{t-1}. \tag{3.2}$$

It is well known that the scalar DCC is particularly advantageous thanks to: the separate estimation of each univariate conditional variance and, then, of the correlation matrices (likelihood decomposition); the possibility of separately estimating the intercept term $\overline{Q}$ by a sample estimator (correlation targeting); the computational feasibility (only two scalar parameters to be estimated: α and β). However, in case of large portfolios, it relies on a too restrictive homogeneity assumption given that the correlation dynamics is ruled by the same two parameters for all the pairs of assets[1]. In addition to this, the two-step estimation based upon the likelihood decomposition may yield massively biased estimates of the two parameters α and β due to the so-called incidental parameter problem (see Engle, Shephard and Sheppard [2008]) and this flaw generally gets more and more serious as the number of assets increases. Given these DCC's features, a natural solution can be conceived whose essentials can be summarized in three steps:

- fit separately a scalar DCC to each pair of assets;
- estimate the pairwise correlation dynamics;
- reconstruct the multivariate high dimensional correlation matrices by using the pairwise estimated correlations (obtaining something that may be called *merged* correlation matrices).

[1] in Engle and Sheppard (2001) it is shown, for example, that this leads to sub-optimal portfolio selection in case of many assets

This solution would allow to keep exploiting the simplicity of the scalar DCC, without removing crucial heterogeneity information and succumbing to overwhelming or infeasible computations. On the other hand, it is known that well defined correlation (and, in general, covariance) matrices must be positive (semi)definite[2] (see Equation (1.13)); nevertheless, we expect that reconstructing a correlation matrix by putting together unrelatedly estimated entries may not yield well defined matrices, so some kind of regularization is needed. In the next sections the features of the proposed method and the related issues of estimation and regularization are discussed in detail.

3.3 A new forecasting approach based on shrinkage estimation

As already said in Section 3.1, the goal of this chapter is to propose a more flexible and possibly accurate approach to modeling and, especially, to forecasting conditional correlations in case of large portfolios. In Section 3.2 the solution outline has been sketched in three steps which should allow to exploit the easiness of estimation of scalar DCC while introducing more information without incurring overwhelming or *de facto* impossible computations. The gain in information is achieved by computing the *merged* conditional correlations matrices or, in other words, the conditional correlations built by gathering the pairwise estimates relative to each couple of assets belonging to a considered portfolio. Such matrices are expected to be more informative about the real dynamics of conditional correlations because they take into account the heterogeneity that the scalar DCC can hide, but it has been already stressed that they will not be in general well defined (positive definite or semidefinite, in particular) because obtained from unrelatedly estimated entries. In Ledoit and Wolf (2004a) a similar problem is faced. Their work (see Section 2.3.1) is concerned with the regularization of ill-conditioned or even singular sample covariance matrices through the shrinkage method (James and Stein [1961]). Through numerical experiments the authors find that does not only the shrinkage provide an estimator better conditioned than the sample covariance but

[2]Sample correlations are surely positive definite if the number of observations T is greater than the number of assets N (see Equations 2.28 and the related proof)

also generates more accurate estimates of the true covariance. We want to borrow from Ledoit and Wolf (2004a) and transfer their intuition into our context. The idea consists in linearly combining the correlation matrix forecasts derived from a full-sized scalar DCC with those obtained from the estimation of all the pairwise scalar DCC models, applied to each one of the $N(N-1)/2$ pairs of assets. More precisely, in order to build what we have defined *merged* correlation matrices, the generic pairwise DCC forecast, say $r^*_{ij,t}$, is used as the (i,j) entry of the *merged* matrix at time t; then, a suitable convex linear combination is computed between the full-sized DCC predicted correlation, say $r_{ij,t}$, and the pairwise one: $\alpha r^*_{ij,t} + (1-\alpha) r_{ij,t}$, where α is the same for every couple (i,j). In an analogical comparison with the shrinkage *à la* Ledoit and Wolf, in this context:

- the correlation matrices estimated by the full-sized scalar DCC play the role of the shrinkage target
- the reconstructed *merged* correlations play the role of the matrices to shrink

and, like in the cited paper, we expect the resulting estimator/predictor to be well defined and also more accurate according to some chosen loss function. Further, the proposed solution is computationally feasible because it involves the estimation of only two parameters at a time. A consequent issue is the estimation of the optimal weight in the linear combination. In Ledoit and Wolf (2004a) the problem is solved by minimizing the squared Frobenius norm of the difference between the combined estimator and the latent covariance, which is obviously unknown: to overcome this problem they derive some asympotical results that allow them to obtain consistent estimates without observing the true covariance. However, Ledoit and Wolf's general results cannot be useful in the DCC framework as they have been derived for sample estimators. That being said, in the optimization problem we replace the true correlations with consistent and unbiased proxies: adequately computed realized correlation matrices[3]. This kind of regularization scheme can be defined *a posteriori* because it applies *after* having built the linear combination and along with the weight estimation. In more detail, the general analytic formulation of the *a posteriori* regularization problem is the following:

[3]For definitions and properties of realized covariance matrices we refer to Andersen, Bollerslev and Diebold (2009)

$$\begin{aligned} \alpha_{opt} &= \underset{0\leq\alpha\leq 1}{\operatorname{argmin}} \sum_{t=1}^{T} \left\| \alpha R_t^{merged} + (1-\alpha)\, R_t^{sDCC} - R_t^{real} \right\|^2 \\ &\text{s.t. } \alpha R_t^{merged} + (1-\alpha)\, R_t^{sDCC} \text{ is p.d. , } \forall t = \tau_1, ..., \tau_s \end{aligned} \tag{3.3}$$

where R_t^{real} is the realized correlation matrix and $\| \ \|$ denotes the Frobenius norm. Hence, the regularization is achieved by the introduction of a numerical constraint which forces the shrinkage estimator/predictor to be positive definite, where it is assumed that $\tau_1 < ... < \tau_s$ and that, in general, the intersection $[\tau_1, ...\tau_s] \cap [1, ...T]$ may be $= \emptyset$ or $\neq \emptyset$. The constraint is obtained by using a Matlab function[4] related to the Cholesky decomposition of the matrices to be regularized. This function applies to each symmetric (in general Hermitian) matrix, say A, and produces two main outputs: an upper triangular matrix, say U, such that $A = U'U$ and a real number, say p, such that $p = 0$ if and only if A is positive definite. The latter output, p, can be therefore employed to write a suitable function to bind the minimization problem in (3.3). More precisily, the constraining algorithm employed to bind the minimization problem[5] is the following function loss_cons:

$$\begin{aligned} &\text{function}[c, ceq] = \text{loss\_cons}(\alpha, R^{merged}, R^{sDCC}, \tau_1, ..., \tau_s) \\ &psum = 0; \\ &\text{for } t = \tau_1 : \tau_s \\ &\quad [U, p] = \text{chol}(\alpha, R_t^{merged} + (1-\alpha) R_t^{sDCC}); \\ &\quad psum = psum + p; \\ &\text{end} \\ &c = [\,]; \\ &ceq = psum; \\ &\text{end} \end{aligned} \tag{3.4}$$

[4] All of the computations relative to the empirical applications of this thesis have been carried out by using the software Matlab

[5] The minimization has been carried out by using of the common Matlab function for the constrained problem: fmincon

where the operator chol denotes the Matlab function for the Cholesky decomposition[6] and R^{merged} and R^{sDCC} denote the 3-dimensional arrays built with conditional correlation matrices R_t^{merged} and R_t^{sDCC}. That being said, there is no regularization of the reconstructed *merged* correlations through this technique: what is regularized is the entire combined estimator, according to the original Ledoit and Wolf's intuition.

As it can be seen from (3.3), relying on the results in Laurent, Rombouts and Violante (2013) here the (sum of) squared Frobenius norms, $\sum \| \ \|^2$, are minimized rather than the (sum of) Frobenius norms, $\sum \sqrt{\| \ \|^2}$. Indeed, when different forecasting models or estimators are used, their performance can be evaluated by using a loss function depending on the true variable. However, as previously stressed, true covariances and correlations are not observable, so they are generally substituted with a proxy (for example, the realized correlations in our case). Nevertheless, when the model evaluation is based upon a target observed with error (a proxy), the resulting model ranking may be different from that resulting from using the latent variable (see Hansen and Lunde [2006]): for these reasons the choice of the loss function is not irrelevant. In the cited article the authors introduce a generalized class of functional forms which guarantee consistency of the ranking under certain conditions. This class of forms correspond to a Bregman-type distance function (see Bregman [1967]; see also Subection 2.2.5):

$$L(\Sigma_t^*, H_t) = f(H_t) - f(\Sigma_t^*) + \nabla f(H_t)' \operatorname{vech}(\Sigma_t^* - H_t). \tag{3.5}$$

In Equation (3.5) $f : \mathbb{R}^{N\times N} \to \mathbb{R}$ is a scalar-valued matrix function and it is assumed to be three times continuously differentiable, while Σ_t^* and H_t are respectively the covariance proxy and its forecast at time t; the term $\nabla f(H_t)$ is the gradient of f with respect to the $N(N+1)/2$ unique elements of H_t. When one sets $f(A) = -\operatorname{tr}(A^2)$ the definition in (3.5) becomes the squared Frobenius distance: $L(\Sigma_t^*, H_t) = \operatorname{tr}[(\Sigma_t^* - H_t)(\Sigma_t^* - H_t)']$[7]. Hence, the squared Frobenius distance is a special case of the class of loss functions which preserve the

[6] For details about the Cholesky decomposition see Lutkepohl [1996], p.86

[7] The operator tr denotes the trace of a square $K \times K$ matrix: $\operatorname{tr}(M) = \sum_{i=1}^{K} m_{ii}$ (see see Lutkepohl [1996], p.41)

ordering of different forecast models when the true covariance is substituted with a proxy: this validates our choice.

3.4 Nearest correlation problem

In the previous section the shrinkage of the *merged* conditional correlations has been defined as an *a posteriori* regularization because it applies *after* having built the linear combination and along with the estimation of the optimal weight. However, in the statistical and mathematical literature there exist methods for regularizing ill-defined covariance and correlation matrices directly. In our context, in opposition to the *a posteriori* regularization, this type of methods may be defined for simplicity *a priori*, as they can be used on the matrices to be regularized alone and before a possible combination of estimates or forecasts. In Higham (2002) a regularization algorithm has been introduced with the aim to solve the so-called nearest correlation problem[8]. This problem consists in finding the nearest (according to a certain norm function) symmetric positive (semi)definite unit-diagonal matrix to a given symmetric matrix. From an analytic point of view the problem is the following:

$$\begin{aligned} &R_{opt} = \underset{R}{\operatorname{argmin}} \|A - R\| \\ &\text{s.t. } \{R \text{ is positive semidefinite}\} \cap \{[R]_{ii} = 1\} \end{aligned} \tag{3.6}$$

where A is the given generic symmetric matrix, $\| \ \|$ denotes the Frobenius norm and R is the matrix variable. If we define the two sets:

$$\begin{aligned} S &= \{X = X' \in \mathbb{R}^{N \times N} \ : \text{ X is positive semidefinite}\} \\ U &= \{X = X' \in \mathbb{R}^{N \times N} \ : \ [X]_{ii} = 1, \text{ for } i = 1, ...N\} \end{aligned} \tag{3.7}$$

[8] Another common method of direct regularization is the eigenvalue cleaning (see Subection 2.3.4). The eigenvalue cleaning is a regularization technique based on the adjustment of the largest and/or smallest eigenvalues. Nevertheless, this method is useless for our problem because it can apply to sample covariance matrices only

the nearest correlation problem is equivalent to finding the nearest matrix to the given A, belonging to the intersection $S \cap U$. This intersection is a convex set as S and U are both convex as well, so the solution to the problem (3.6) exists and is unique. The problem can be solved with the so-called alternating projections method (see Algorithm 3.3 in Higham[2002]) which consists in iteratively projecting the resulting matrix on the set U (*i.e.* forcing the diagonal element to be equal to 1) and then on the set S (*i.e.* forcing the negative eigenvalues to be equal to 0). The convergence is determined by comparing convenient relative norms with a chosen tolerance[9]: more precisely, the default tolerance is set equal to $N \cdot 2.2204 \cdot 10^{-16}$, while the convergence is achieved as soon as one of the three following relative Frobenius norms is lesser than the tolerance:

$$\frac{\|s_i - s_{i-1}\|}{\|s_{i-1}\|}, \frac{\|u_i - u_{i-1}\|}{\|u_{i-1}\|}, \frac{\|s_i - u_i\|}{\|s_i\|} \tag{3.8}$$

where $s_i \in S$, $u_i \in U$ and i denotes the algorithm iteration number. Although the alternating projection method is reliable, in our empirical applications we have used a more complex Newton method introduced in Qi and Sun (2006) which converges more quickly and whose software implementation has been publicly shared by the authors[10]. However, the two algorithms give practically the same results.

3.5 Empirical application

The main empirical application has been carried out on a data set of 50 liquid stocks included in the index *S&P500* with prices observed every minute from 9.30AM to 4.00PM in the interval 07/27/1997 through 07/18/2008 (2549 daily observations). The data are part of the One-Minute Equity Data (OMED) database provided by Tick Data[11]. In Table 3.1 names and related abbreviations of the stocks included in the data set are reported. The data

[9] The Matlab code (nearcorr.m) is available at the internet page https://nickhigham.wordpress.com/2013/02/13/the-nearest-correlation-matrix/

[10] The Matlab code (CorrelationMatrix.m) is available at the internet page http://www.math.nus.edu.sg/~matsundf/CorrelationMatrixNewton.m

[11] www.tickdata.com

have been converted in log-returns and demeaned: they are represented in Figure 3.1 along with their squared values, from which it can be seen that the first part of the considered sample is characterized by a higher overall volatility with a new increasing tendency in the very last part of the interval (at the beginning of the latest world-wide recession). The related five-minute computed realized covariances have been used as a proxy for the latent covariances: consequently the proxy series for the realized correlations has been computed by simply standardizing the realized covariance matrices (Figure 3.2).

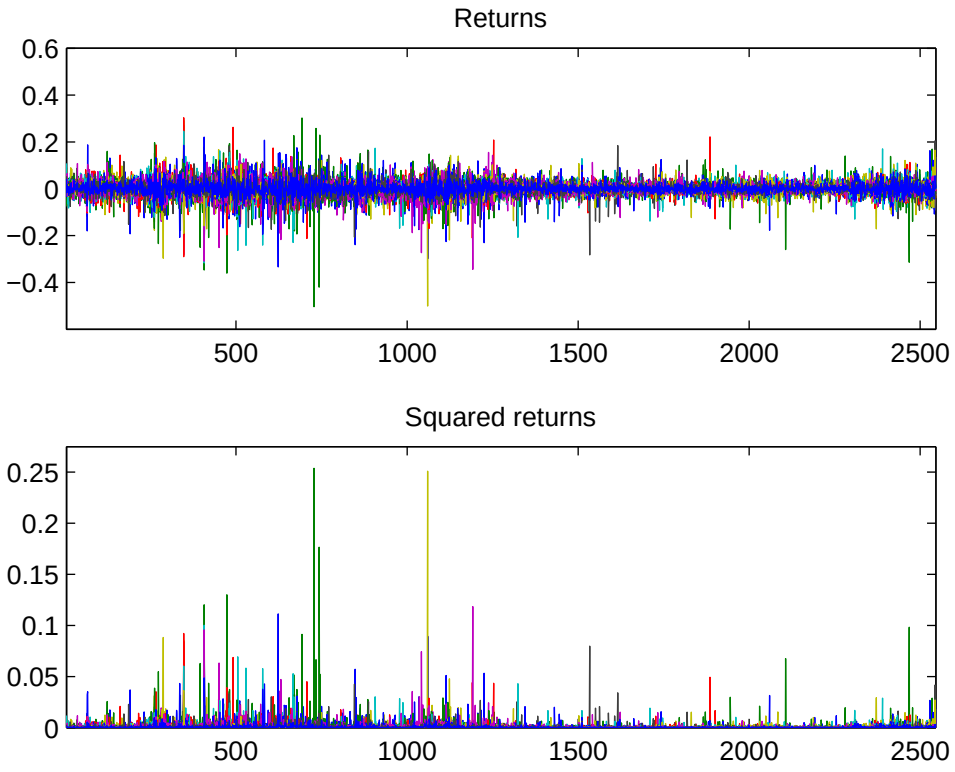

FIGURE 3.1: Returns relative to the fifty price time series used for the empirical application (upper box) and related squared values (lower box)

After that, the standardized residuals (upper box of Figure 3.3) have been obtained by applying a univariate GARCH(1,1) model to each of the 50 series of demeaned returns ϵ_{it}:

Stock	Issue Name	Stock	Issue Name
	Dataset: October 1997 – July 2008		
AA	Alcoa	COST	Costco Wholesale Corporation
ABT	Abbott Laboratories	CPB	Campbell Soup
ADI	Analog Devices	CSC	Computer Sciences Corporation
AFL	Aflac Incorporated	CTAS	Cintas Corporation
AIG	American International Group	CTL	CenturyLink
ALL	Allstate	DOV	Dover Corporation
APD	Air Products & Chemicals	DOW	The Dow chemical Company
APOL	Apollo Education Group	DTE	DTE Energy Company
AVY	Avery Dennison Corporation	EIX	Edison International
AXP	American Express Company	EMN	Eastman Chemical Co.
AZO	AutoZone	ETR	Entergy Corporation
BAC	Bank of America	F	Ford Motor Co.
BAX	Baxter Corporation	FDO	Family Dollar Stores
BBBY	Bed Bath & Beyond	FISV	Fiserv
BDX	Becton, Dickinson and Company	GCI	Gannett Co.
BHI	Braveheart Resources	GE	General Electric
BMY	Bristol-Myers Squibb Company	GIS	General Mills
C	Citigroup	GPC	Genuine Parts Company
CAG	ConAgra Foods	GPS	The Gap
CAH	Cardinal Health	HD	The Home Depot
CL	Colgate-Palmolive Co.	HNZ	HNZ Grou
CLX	Clorox	HON	Honeywell International
CMA	Comerica Incorporated	HPQ	Hewlett-Packard Company
CMS	CMS Energy Corp.	KO	Coca Cola
COF	Capital One Financial Corporation	T	AT&T

TABLE 3.1: Stocks included in the data set used for the empirical application

$h_{ii,t} = \omega_i + \alpha_i \epsilon^2_{i,t-1} + \beta h_{ii,t-1}$, with $i = 1, ..., 50$. The related conditional variances are depicted in the lower box of Figure 3.3.

The standardized residuals' sample has been then split into nb blocks ($nb = 26\,,\, 109$) by employing a rolling window of $2000 + s$ observations ($s = 21\,,\, 5$); for each of the nb data

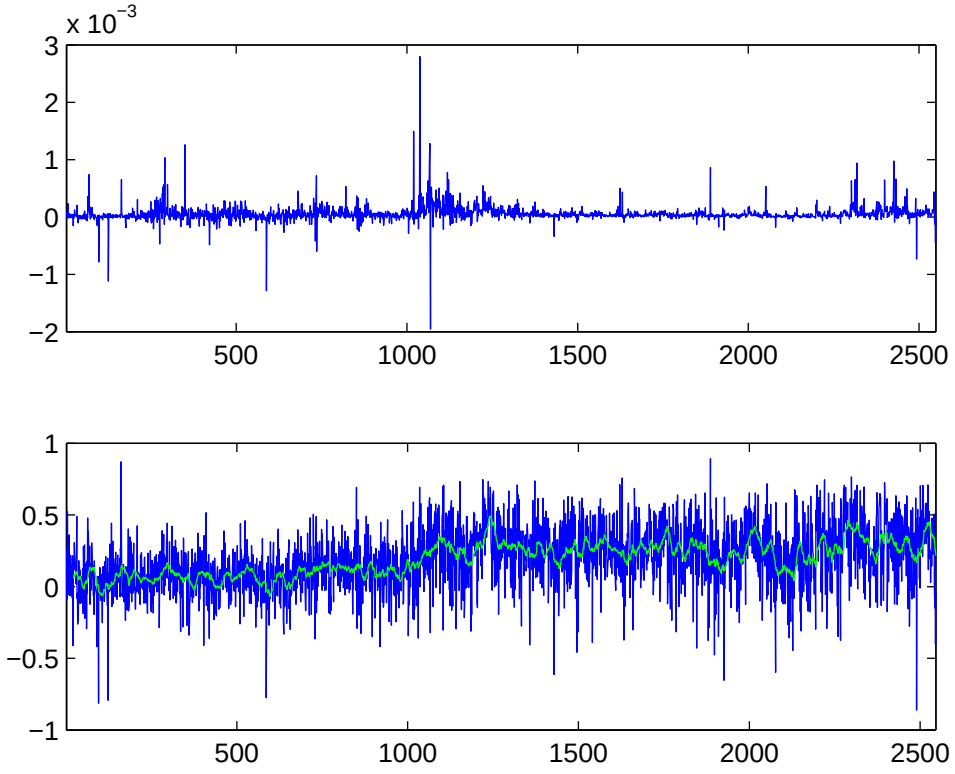

FIGURE 3.2: Realized covariances (upper box) and correlations (lower box) relative to the first couple of assets (Alcoa and Abbott Laboratories) of the portfolio used for the empirical application (the green line represents a 21-lag moving average)

blocks, a full-sized scalar DCC and $N(N-1)/2$ bivariate scalar DCCs have been estimated (according to the solution outlined in Sections 3.2 and 3.3) on the first 2000 data points; then the estimates have been used to obtain two series (one from the full-size scalar DCC and one from the $N(N-1)/2$ bivariate scalar DCCs) of $s \cdot nb$ predicted conditional correlation matrices. The choice of a rolling window scheme is based on a greater reliability of forecasts resulting from limited memory estimators compared to those obtained from expanding window estimators. The issue is discussed in detail in Giacomini and White (2006) where the authors claim the superiority of rolling window estimators and predictors, especially in case of inadequately modeled heterogeneity or misspecification: a typical example could be the

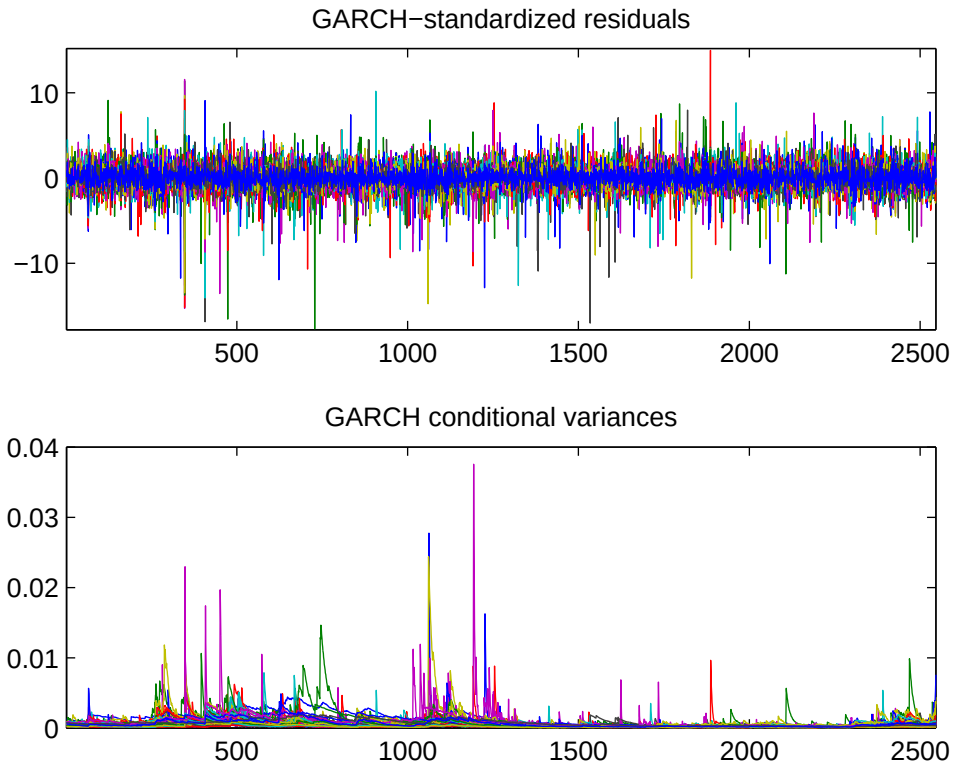

FIGURE 3.3: Upper box: Standardized GARCH(1,1) residuals. Lower box: Related conditional variances

presence of different regimes, which is often the case in long samples. To be more precise, the approach here adopted cannot strictly be defined as a pure rolling window scheme but it is rather a *mixed rolling window* approach where the forecasting model is re-estimated each s observations. The two selected re-estimation intervals ($s = 21\,,\,5$) coincide respectively with the typical trading month and week. We focus on one-step ahead forecasts which in practical terms means that in the dynamic equation:

$$Q_t = (1 - \alpha - \beta)\,\overline{Q} + \alpha \hat{\boldsymbol{\eta}}_{t-1} \hat{\boldsymbol{\eta}}_{t-1}^{'} + \beta Q_{t-1} \tag{3.9}$$

the standardized residuals $\hat{\boldsymbol{\eta}}_t$, with $t = 2001, ..., 2000 + s$, are computed assuming the knowledge of the returns relative to the same time points. In Figure 3.4 we can see the series of the predicted correlations ($t = 2001, ..., 2546$) for the first couple (namely Alcoa and Abbott Laboratories) of the considered stocks in case of 26 blocks and consequent re-estimation interval of 21 periods: the forecasts resulting from the pairwise DCC model are clearly more variable than the ones obtained with the scalar DCC applied to the entire portfolio.

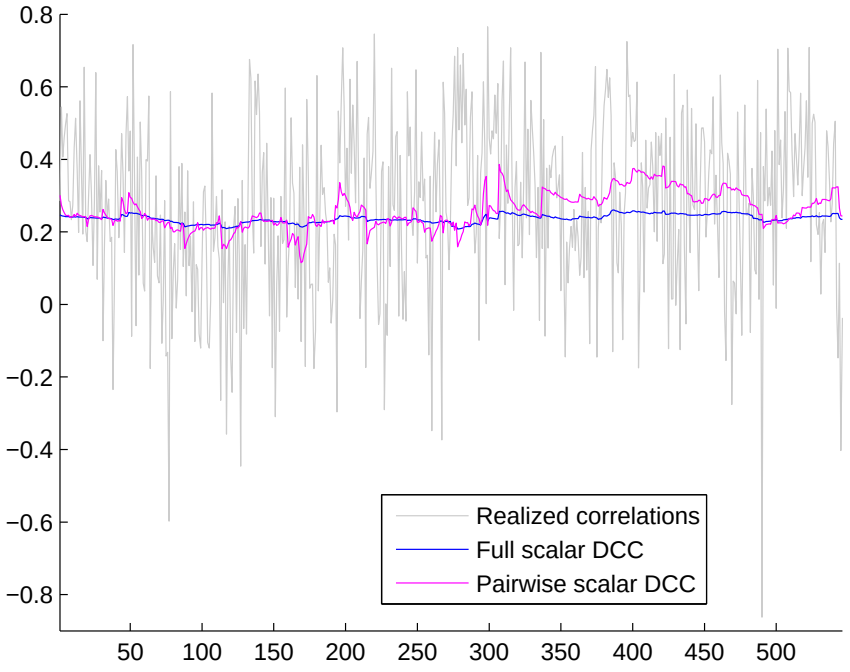

FIGURE 3.4: Predicted full scalar DCC and pairwise scalar DCC correlations for the first couple of stocks (Alcoa and Abbott Laboratories) of the considered sample with $nb = 26$ and $s = 21$ (number of predictions $= nb \cdot s = 546$)

As it has been repeatedly stressed, we expect the *merged* correlation matrices not to be well-conditioned. In Figure 3.5 the eigenvalue numerical distributions of the *merged* predicted correlation matrices are represented in case of $nb = 26$, while the related time series of each of the 50 eigenvalues can be seen in Figure 3.6; the two graphs regard the *merged* correlations before any regularization. Figure 3.5 shows that many of the smallest eigenvalues are less than

zero, whereas the biggest three have simple averages respectively equal to 1.978, 2.617 and 14.880 (see Figure 3.6). In order to obtain well conditioned predicted conditional correlations, the two techniques described in Section 3.3 and Section 3.4 have been alternatively adopted in order, so obtaining three different sets of forecasts that hereafter we will define: constrained shrinkage (CS); simple regularization (SR); shrinkage after regularization (SAR).

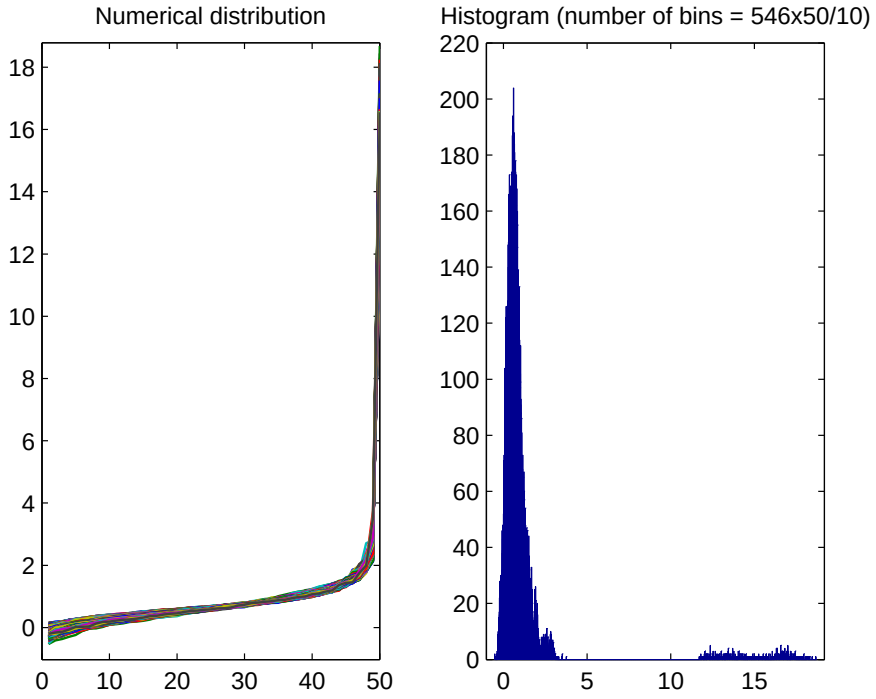

FIGURE 3.5: Eigenvalue distribution for the 546 predicted *merged* correlation matrices in case of 26 data blocks (s=21)

3.5.1 First case: constrained shrinkage (CS)

In this case the regularization has been achieved by imposing the positive definiteness on the predicted correlation matrices through a numerical constraint (see Section 3.3), that is:

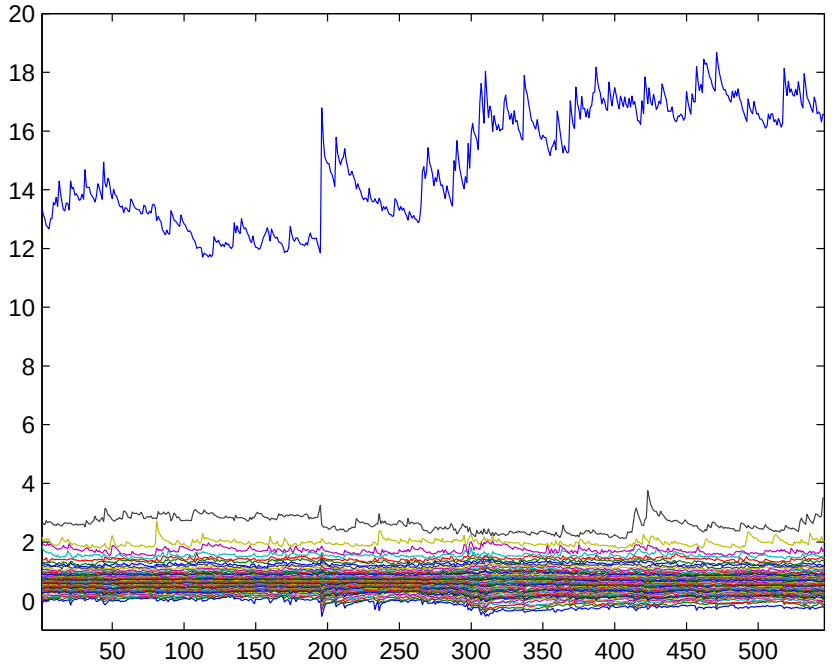

FIGURE 3.6: Time series of the eigenvalues of the 546 predicted *merged* correlation matrices in case of 26 data blocks (s=21)

$$\begin{aligned} \alpha_{j,opt} &= \underset{0\leq\alpha_j\leq 1}{\operatorname{argmin}} \sum_{t=1}^{2000} \left\| \alpha_j R_{j,t}^{merged} + (1-\alpha_j) R_{j,t}^{sDCC} - R_{j,t}^{real} \right\|^2 \\ &\text{s.t. } \alpha_j R_{j,t}^{merged} + (1-\alpha_j) R_{j,t}^{sDCC} \text{ is p.d. } , \quad t = 2001, ..., 2000+s \end{aligned} \tag{3.10}$$

with $j = 1, ..., nb$. The weights are represented (in blue) in Figure 3.7, in the case of a re-estimation interval equal to 21 (upper box) and 5 (lower box) respectively. In both cases an increasing trend in the weights is apparent: higher weights correspond with a greater importance of the pairwise estimates in the linear combination and, consequently, they mean an increase of heterogeneity across the conditional correlations in the sample. In Figure 3.8

the predicted conditional correlations obtained *via* the constrained shrinkage are represented (in blue) for the first pair of stocks in the portfolio (namely Alcoa and Abbott Laboratories).

3.5.2 Second case: simple regularization (SR)

In this case the regularization has been applied through the Qi and Sun algorithm (see Section 3.4) to the entire series of estimated *merged* conditional correlation matrices in order to obtain an entire set of regularized matrices: $\tilde{R}_{j,t}^{merged}$ with $t = 1, ..., 2000 + s$ and $j = 1, ..., nb$. In Figure 3.8 the predicted *merged* conditional correlations regularized with the Qi and Sun algorithm are shown (in green) for the first pair of stocks in the portfolio (Alcoa and Abbott Laboratories).

3.5.3 Third case: shrinkage after regularization (SAR)

The *merged* matrices regularized with the Qi and Sun procedure, $\tilde{R}_{j,t}^{merged}$, have been used to build a combined predictor by solving the problem:

$$\alpha_{j,opt} = \underset{0 \leq \alpha_j \leq 1}{\operatorname{argmin}} \sum_{t=1}^{2000} \left\| \alpha_j \tilde{R}_{j,t}^{merged} + (1 - \alpha_j) R_{j,t}^{sDCC} - R_{j,t}^{real} \right\|^2 \tag{3.11}$$

where $j = 1, ..., nb$. This combined estimator has been built to check whether a shrinkage-type predictor with an *a priori* regularization may outperform the constrained shrinkage (CS). While in the *a posteriori* (CS) case the combined forecasts, $\alpha_j R_{j,2000+t}^{merged} + (1 - \alpha_j) R_{j,2000+t}^{sDCC}$ with $t = 1, ..., s$, must be numerically regularized along with the computation of the optimal weights α_j, in the *a priori* one they are already well defined because of the prior application of the Qi and Sun algorithm to the pairwise-predicted correlation matrices. Hence, in the latter case the α_j's can be calculated analytically. For a generic data block of length T, let $r_{lm,t}^* = [\tilde{R}_t^{merged}]_{lm}$, $r_{lm,t} = [R_t^{sDCC}]_{lm}$ and $\rho_{lm,t} = [R_t^{real}]_{lm}$. If we denote $\delta_{lm,t}^* = r_{lm,t}^* - r_{lm,t}$ and $\delta_{lm,t} = \rho_{lm,t} - r_{lm,t}$, the optimization problem (3.3) can be written as:

$$\alpha_{opt} = \underset{0\leq\alpha\leq 1}{\operatorname{argmin}} \sum_{t=1}^{T} \left\| \begin{bmatrix} 0 & \alpha\delta^*_{12,t} - \delta_{12,t} & \cdots & \alpha\delta^*_{1N,t} - \delta_{1N,t} \\ \alpha\delta^*_{12,t} - \delta_{12,t} & 0 & \cdots & \alpha\delta^*_{2N,t} - \delta_{2N,t} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \delta^*_{1N,t} - \delta_{1N,t} & \alpha\delta^*_{2N,t} - \delta_{2N,t} & \cdots & 0 \end{bmatrix} \right\|^2 . \tag{3.12}$$

where as usual the notation $\| \ \|$ denotes the Frobenius norm. On the other hand, for the symmetry of the matrices under the summation in (3.12), the problem can be reformulated as follows:

$$\alpha_{opt} = \underset{0\leq\alpha\leq 1}{\operatorname{argmin}} \sum_{t=1}^{T} \sum_{l=1}^{N-1} \sum_{m=l+1}^{N} (2\alpha^2\delta^{*2}_{lm,t} + 2\delta^2_{lm,t} - 4\alpha\delta^*_{lm,t}\delta_{lm,t}) \tag{3.13}$$

and, taking into account that $\frac{\partial}{\partial\alpha}\sum_t\sum_l\sum_m = \sum_t\sum_l\sum_m\frac{\partial}{\partial\alpha}$, the analytical solution for the optimal weight is obtained by deriving with respect to α the function in (3.13) and setting the derivative equal to zero:

$$\alpha_{opt} = \frac{\sum_{t=1}^{T}\sum_{l=1}^{N-1}\sum_{m=l+1}^{N}(r^*_{lm,t} - r_{lm,t})(\rho_{lm,t} - r_{lm,t})}{\sum_{t=1}^{T}\sum_{l=1}^{N-1}\sum_{m=i+1}^{N}(r^*_{lm,t} - r_{lm,t})^2}. \tag{3.14}$$

As the second-order derivative of the function in (3.13) is equal to $\sum_{t=1}^{T}\sum_{l=1}^{N-1}\sum_{m=l+1}^{N}(r^*_{lm,t} - r_{lm,t})^2$, which is the sum of $TN(N-1)/2$ squared terms, the second-order conditions for the minimization problem in (3.13) are always satisfied: so one can be sure that the value determined by the condition (3.14) uniquely minimizes the problem (3.12). Moreover, since we are interested only in convex combinations, we set:

$$\alpha_{opt} = \begin{cases} 0 \quad \text{if} \quad \frac{\sum_{t=1}^{T}\sum_{l=1}^{N-1}\sum_{m=l+1}^{N}(r^*_{lm,t}-r_{lm,t})(\rho_{lm,t}-r_{lm,t})}{\sum_{t=1}^{T}\sum_{l=1}^{N-1}\sum_{m=l+1}^{N}(r^*_{lm,t}-r_{lm,t})^2} \leq 0 \\ \\ 1 \quad \text{if} \quad \frac{\sum_{t=1}^{T}\sum_{l=1}^{N-1}\sum_{m=l+1}^{N}(r^*_{lm,t}-r_{lm,t})(\rho_{lm,t}-r_{lm,t})}{\sum_{t=1}^{T}\sum_{l=1}^{N-1}\sum_{m=l+1}^{N}(r^*_{lm,t}-r_{lm,t})^2} \geq 0. \end{cases} \tag{3.15}$$

The formula (3.14) has been used to compute the optimal weights which solve the problem (3.11). The weights are shown (in red) in Figure 3.7, for $s = 21$ (upper box) and $s = 5$ (lower box) respectively: like for the solutions to the problem (3.10), also in this case the estimated weights exhibit an increasing trend, probably because (as already considered) the volatility heterogeneity of the data raises as the rolling window moves ahead. In Figure 3.8 the predicted correlations for the first couple of the considered assets (namely Alcoa and Abbott Laboratories) are displayed (in red): it is apparent that the two combined estimators yield very close forecasts and that they obviously constitute a *middle way* between the full scalar DCC forecasts and the more variable pairwise ones.

3.5.4 Evaluation of forecasting performances

The performances of the three estimators discussed in the previous subsections and that of the full scalar DCC have been evaluated and compared by computing the squared Frobenius norms of the differences between the predicted and the corresponding realized correlations. In order to stress the dependence of the performances of the proposed estimators on problem dimension, the forecasting evaluation has been carried out for a variable number of stocks equal to 10, 20, 30, 40 and 50 respectively. The values of the squared Frobenius norms along with the percentage gains (in parentheses) relative to the full scalar DCC performances are reported in Tables 3.2 to 3.6. Three main facts come to light:

- the combined predictors (CS, SAR) show similar results and outperform the full-size scalar DCC as soon as $N \geq 20$

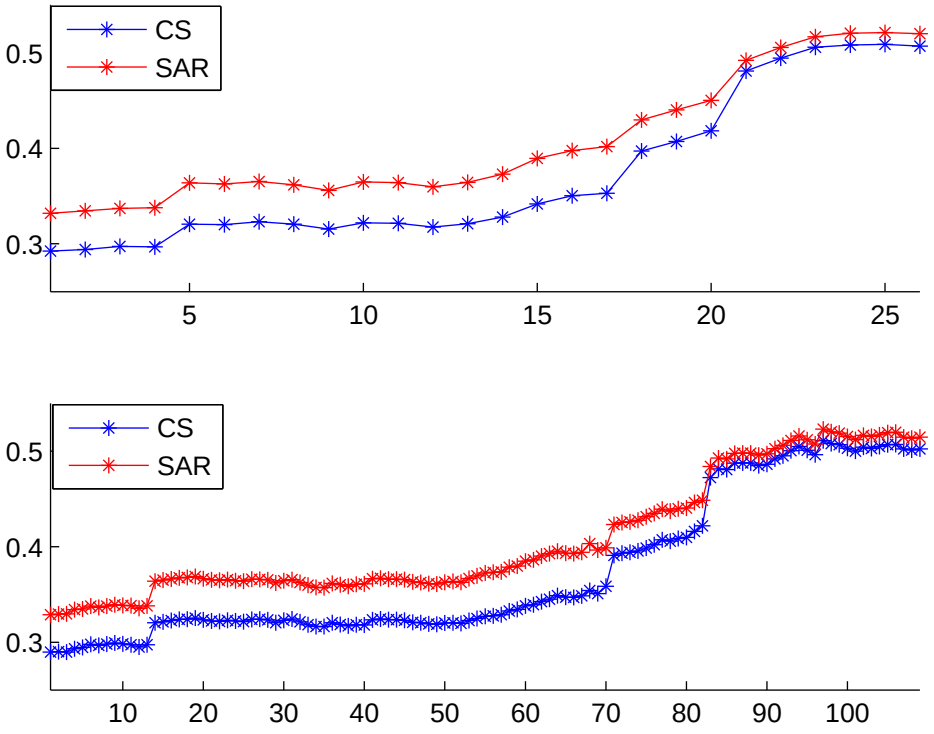

FIGURE 3.7: Optimal weights in case of 26 (upper box) and 109 data bolcks (lower box)

- the pairwise-estimated regularized forecasts (SR) perform worse than the full scalar model up to $N = 40$

- the estimators CS, SR and SAR show increasing performances as N grows.

The latter observation suggest that the three proposed predictors would probably be good candidates in the context of very large portfolios ($N \geq 100$); we currently leave the empirical investigation of this issue to future research.

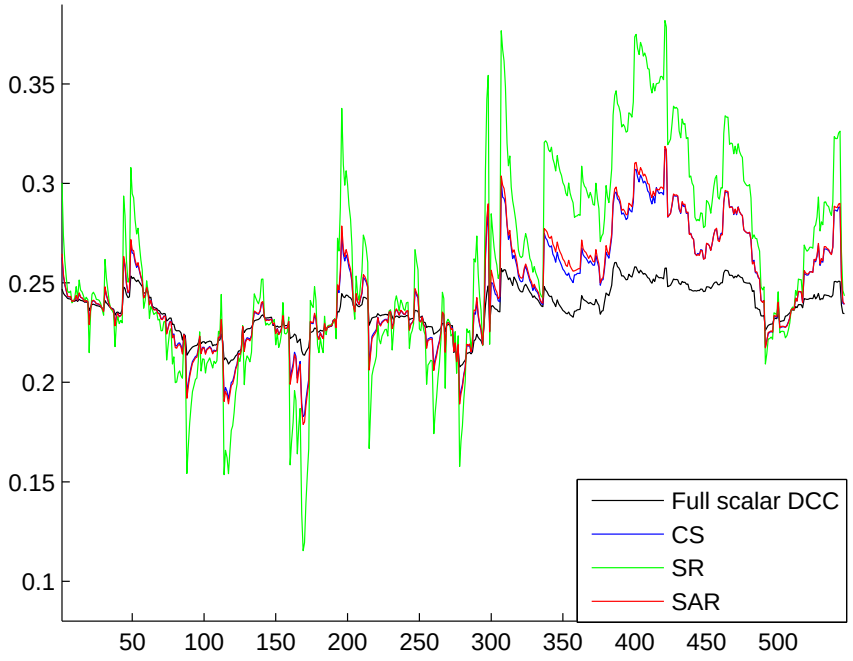

FIGURE 3.8: Predicted conditional correlations for the first couple (Alcoa and Abbott Laboratories) of the considered fifty assets in case of 26 data blocks ($s = 21$)

Re-estimation interval	**Scalar DCC**	**Constrained shrinkage**	**Simple regularization**	**Shrinkage after regularization**
s = 5	$2.6085 \cdot 10^3$ (−)	$2.6091 \cdot 10^3$ (0.023%)	$2.7314 \cdot 10^3$ (4.709%)	$2.6091 \cdot 10^3$ (0.023%)
s = 21	$2.6112 \cdot 10^3$ (−)	$2.6112 \cdot 10^3$ (0.002%)	$2.7259 \cdot 10^3$ (4.395%)	$2.6112 \cdot 10^3$ (0.001%)

TABLE 3.2: Frobenius norms of the differences between the employed predictors and the realized correlation matrices in case of $N = 10$

Re-estimation interval	**Scalar DCC**	**Constrained shrinkage**	**Simple regularization**	**Shrinkage after regularization**
s = 5	$1.1452 \cdot 10^4$ (−)	$1.1326 \cdot 10^4$ (-1.101%)	$1.1605 \cdot 10^4$ (1.336%)	$1.1325 \cdot 10^4$ (-1.116%)
s = 21	$1.1460 \cdot 10^4$ (−)	$1.1333 \cdot 10^4$ (-1.115%)	$1.1598 \cdot 10^4$ (1.200%)	$1.1331 \cdot 10^4$ (-1.130%)

TABLE 3.3: Frobenius norms of the differences between the employed predictors and the realized correlation matrices in case of $N = 20$

Re-estimation interval	**Scalar DCC**	**Constrained shrinkage**	**Simple regularization**	**Shrinkage after regularization**
s = 5	$2.5344 \cdot 10^4$ (−)	$2.4939 \cdot 10^4$ (-1.598%)	$2.5433 \cdot 10^4$ (0.351%)	$2.4925 \cdot 10^4$ (-1.652%)
s = 21	$2.5356 \cdot 10^4$ (−)	$2.4948 \cdot 10^4$ (-1.609%)	$2.5413 \cdot 10^4$ (0.227%)	$2.4935 \cdot 10^4$ (-1.662%)

TABLE 3.4: Frobenius norms of the differences between the employed predictors and the realized correlation matrices in case of $N = 30$

Re-estimation interval	**Scalar DCC**	**Constrained shrinkage**	**Simple regularization**	**Shrinkage after regularization**
s = 5	$4.5027 \cdot 10^4$ (−)	$4.4125 \cdot 10^4$ (-2.002%)	$4.5064 \cdot 10^4$ (0.083%)	$4.4074 \cdot 10^4$ (-2.116%)
s = 21	$4.5057 \cdot 10^4$ (−)	$4.4158 \cdot 10^4$ (-1.995%)	$4.5058 \cdot 10^4$ (0.002%)	$4.4109 \cdot 10^4$ (-2.103%)

TABLE 3.5: Frobenius norms of the differences between the employed predictors and the realized correlation matrices in case of $N = 40$

Re-estimation interval	**Scalar DCC**	**Constrained shrinkage**	**Simple regularization**	**Shrinkage after regularization**
s = 5	$7.0797 \cdot 10^4$ (−)	$6.9186 \cdot 10^4$ (-2.276%)	$7.0465 \cdot 10^4$ (-0.469%)	$6.9081 \cdot 10^4$ (-2.425%)
s = 21	$7.0857 \cdot 10^4$ (−)	$6.9260 \cdot 10^4$ (-2.254%)	$7.0475 \cdot 10^4$ (-0.539%)	$6.9157 \cdot 10^4$ (-2.399%)

TABLE 3.6: Frobenius norms of the differences between the employed predictors and the realized correlation matrices in case of $N = 50$

3.5.5 Comparison between forecasting performances

In Subsection 3.5.4 the results in terms of squared Frobenius norms obtained by the proposed estimators have been shown and these results might be used to compare and rank them. Nevertheless, it is well known that every econometric model always shows some degree of misspecification mainly due to unmodeled dynamics, incorrect functional forms or finite sample distortions: in these cases estimators and predictors resulting from the model may produce partially unreliable forecasts whose comparisons do not make sense. In Giacomini and White (2006) two predictive ability tests are introduced that allow to compare different forecasting models taking into account these features. The tests are in a way a generalization of those proposed in the previous literature (in particular Diebold and Mariano [1995] and West [1996]) and are especially (but not exclusively) conceived for forecasts based on limited memory estimators (like rolling window predictors). In the cited paper the limited memory estimators are particularly considered because they can overcome one the main shortfall of the economic forecasting models which is the heterogeneity of the data generating process[12]. The first test concerns the conditional predictive ability, while the second is introduced as a special case of the first one and pertains the unconditional predictive ability. An important feature of Giacomini-White tests is that their validity is independent on the chosen loss

[12]See also the introduction of Section 3.5

function and, besides, they allow for a unified treatment of nested and non-nested models[13]. While the tipical null hypothesis in classic tests is typically:

$$H_0 : E[L_{t+\tau}(X_{t+\tau}, f_t(\boldsymbol{\phi})) - L_{t+\tau}(X_{t+\tau}, g_t(\boldsymbol{\gamma}))] = 0 \tag{3.16}$$

where L is a suitable loss function, τ is the forecast period, X is the considered variable, f and g are two competing models and $\boldsymbol{\phi}$ and $\boldsymbol{\gamma}$ are the probability limits of the related parameters, the null in Giacomini-White tests takes into account the available information, G_t, and the sample-dependent estimated parameters, $\hat{\boldsymbol{\phi}}_t$ and $\hat{\boldsymbol{\gamma}}_t$:

$$\begin{aligned} H_0 : E[L_{t+\tau}(X_{t+\tau}, f_t(\hat{\boldsymbol{\phi}}_t)) - L_{t+\tau}(X_{t+\tau}, g_t(\hat{\boldsymbol{\gamma}}_t)) \,|\, G_t] = \\ = E[\Delta L_{t+\tau} \,|\, G_t] = 0. \end{aligned} \tag{3.17}$$

Basing on this background, these tests do not operate assuming that in-sample estimated parameters converge to their population values, instead they consider the sample behaviour (evaluated on the basis of the information G_t) of the competing estimators. The authors derive the conditional and unconditional tests by assuming respectively $G_t = I_t$[14] and $G_t = \{\emptyset, \Omega\}$[15]. The multi-step test statistic for the conditional predictive test is the following:

$$T_{m,n,\tau} = n \left(n^{-1} \sum_{t=m}^{T-\tau} \boldsymbol{h}_t \Delta L_{m,t+\tau} \right)' \tilde{\Sigma}_n^{-1} \left(n^{-1} \sum_{t=m}^{T-\tau} \boldsymbol{h}_t \Delta L_{m,t+\tau} \right). \tag{3.18}$$

In Equation (3.18) $\boldsymbol{h}_t \in \mathbb{R}^q$ is a suitable test function[16], m is the length of the estimation sample and $n = T - m - \tau$; the central term is the consistently $q \times q$ estimated covariance of $\boldsymbol{h}_t \Delta L_{m,t+\tau}$:

[13] In our case the simple scalar DCC predictor can be regarded as a forecasting model nested in the shrinkage-type estimator when $\alpha_j = 0$

[14] I_t stands for the common time-t information set

[15] $G_t = \{\emptyset, \Omega\}$ is the largest σ-field on Ω, where as usual Ω denotes the set of all possible outcomes (see Casella and Berger [2002], p.6)

[16] See Giacomini and White (2006) for the technical details about the properties of $\boldsymbol{h}_t$ and $\boldsymbol{h}_t \Delta L_{m,t+\tau}$

$$\tilde{\Sigma}_n = n^{-1} \sum_{t=m}^{T-\tau} (\boldsymbol{h}_t \Delta L_{m,t+\tau})(\boldsymbol{h}_t \Delta L_{m,t+\tau})'. \tag{3.19}$$

Giacomini and White (2006) show that $T_{m,n,\tau} \sim \chi^2_q$ under the null. The test for the unconditional predictivity ability is similarly obtained as a test on the sample average of the loss function difference $n^{-1} \sum_{t=m}^{T-\tau} \Delta L_{m,t+\tau}$ and its statistic has a standardized Gaussian distribution.

The conditional Giacomini-White test has been used to verify the statistical significance of the differences in norm between the proposed estimators' performances. The test results are displayed in Table 3.8 and Table 3.7 for $N = 50$: the differences are always significant except for that between the full scalar DCC and the SR forecasts. The signs $+$ and $-$ in parentheses state that the sample mean of the differences in the null hypothesis (3.17) are respectively > 0 and < 0 , where the predictor f in the formula (3.17) is here the corresponding model in the first column of the table, while g denotes the corresponding model in the first row. Hence, these results can be used to rank the employed predictors and the resulting ordering can be schematized as follows:

$$\begin{aligned} &\text{Shrinkage after regularization} \succ \text{Constrained shrinkage} \succ \ ... \\ &\qquad\qquad ... \ \succ \ \text{Simple regularization} \ \succeq \ \text{Full scalar DCC} \end{aligned} \tag{3.20}$$

where, in this context, the symbol $\succeq$ means that the positive difference between the norms associated to the full scalar DCC and the SR predictors is not statistical significant and so the latter cannot be surely preferred over the former.

3.6 Comparison with richly parametrized DCC models

In this section the proposed predictors are compared to the approach by Bauwens, Grigoryeva and Ortega (2014) (BGO hereafter) which has been previously discussed in Subsection 2.2.5.

s = 21	**Scalar DCC**	**Constrained shrinkage**	**Simple regularization**	**Shrinkage after regularization**
Scalar DCC	- -	**29.54**(+) (*0.000*)	3.48(+) (*0.176*)	**29.86**(+) (*0.000*)
Constrained shrinkage	- -	- -	**5.99**(−) (*0.015*)	**22.64**(+) (*0.000*)
Simple regularization	- -	- -	- -	**31.55**(+) (*0.000*)
Shrinkage after regularization	- -	- -	- -	- -

TABLE 3.7: Results of the conditional Giacomini-White test on the differences between the loss functions relative to the considered predictors in the case of a re-estimation interval of 21 periods and $N = 50$ (χ^2_2 test statistics, p-values in parantheses)

s = 5	**Scalar DCC**	**Constrained shrinkage**	**Simple regularization**	**Shrinkage after regularization**
Scalar DCC	- -	**43.14**(+) (*0.000*)	2.57(+) (*0.276*)	**42.81**(+) (*0.000*)
Constrained shrinkage	- -	- -	**14.76**(−) (*0.001*)	**26.42**(+) (*0.000*)
Simple regularization	- -	- -	- -	**109.89**(+) (*0.000*)
Shrinkage after regularization	- -	- -	- -	- -

TABLE 3.8: Results of the conditional Giacomini-White test on the differences between the loss functions relative to the considered predictors in the case of a re-estimation interval of 5 periods and $N = 50$ (χ^2_2 test statistics, p-values in parantheses)

The authors introduce a new (for econometric applications) method which can solve likelihood maximization problems handling a high number of parameters and, simultaneously, non-linear stationarity and positivity constraints. The method consists in an algorithm based upon a procedure involving Bregman matrix divergences (see Bregman [1967]). In Bauwens *et al.* (2014) this algorithm is used to estimate on real data different parametric specifications of the general DCC model (see Equation (2.25)): a full rank DCC (called Hadamard DCC); a rank two specification; a rank one version; an additional, more parsimonious rank one specification obtained by using an Almon function (called Almon DCC). The number of parameters needed by the listed specifications are respectively: $N(N+1)$, $4N-2$, $2N$ and 6.

In our first empirical comparison the rank two, the rank one and the Almon DCC have been estimated by using the BGO method and are compared to the ones obtained by the methods described in Sections 3.3, 3.4 and 3.5 (CS, SR and SAR) on the entire data set ($N = 50$) already considered for the empirical application in Section 3.5; nevertheless, in this case the rolling window scheme has been abandoned because of the computational cost needed by the BGO method with $N = 50$. Also in this case the standardized residuals (Figure 3.3) have been obtained by applying a univariate GARCH(1,1) model to each one of the 50 series of the demeaned returns. Then the data set has been split in two parts: the first 2000 observations have been used for the parameter estimation, while the second subsequent block of 545 data points has been utilized for forecast evaluations. In this first application we have turned down estimating the Hadamard DCC because it would have involved the estimation of 2450 parameters which exceed the number of observation, 2000, included in the estimation data block. After having estimated the parameters on the first 2000 data[17], one step-ahead forecasts of the correlation matrices have been computed (see Equation (3.9) and related considerations). Regularization and construction of the two combined estimator have been carried out according to the outlines in Section 3.5: in this case the estimated weights are equal to 0.3219 for the constrained shrinkage (CS) and 0.3640 for the shrinkage after regularization (SAR). Finally, like in the empirical application in Subection 3.5.4, the evaluation has been done by computing the squared Frobenius norms of the differences

[17]The estimation of the specifications proposed in Bauwens *et al.* (2014) has been carried out by applying some software routines kindly shared by the authors: https://github.com/lgrigoryeva/DCC-nonScalar_estimation

between the predicted and the corresponding realized correlations. The norms associated with the different employed predictors are reported in Table 3.9: gains with respect to the scalar DCC are relevant in the case of the two predictors CS and SAR; on the other hand, among the specifications estimated with the BGO method, only the rank one DCC performs better than the scalar version. These results basically match those shown in the discussion of empirical application and in the conclusions of Bauwens *et al.* (2014), where it is asserted that (only) some non-scalar DCC parameterizations (in particular the rank one and the Almon shuffle, here not considered[18]), are worth being used in modeling the volatility of portfolios of dimension up to thirty (see again Section 2.2.5).
In order to obtain a reliable ranking of the different predictors, the conditional Giacomini-White test has been carried out, setting the forecast horizon equal to one. The results are summarized in Table 3.11: according to the test, all the differences between the norms are statistically significant and, taking into account the signs of sample means of the norm differences, the predictors can be ordered as follows:

$$\begin{aligned} \text{Shrinkage after regularization} &\succ \text{Constrained shrinkage} \succ \ldots \\ \ldots &\succ \text{Simple regularization} \succ \text{Rank one DCC} \ \ldots \\ \ldots &\succ \text{Scalar DCC} \succ \text{Almon DCC} \succ \text{Rank two DCC.} \end{aligned} \tag{3.21}$$

In Figure 3.9, for the first couple of stocks in the sample (namely Alcoa and Abbott Laboratories), the predicted conditional correlations derived from some of the estimated models are graphically compared to the realized correlations; in order to better highlight the evolutions of the different forecast series, a 21-lag moving average of the realized correlations has been included in the graph: it is apparent that the more variable series is the one coming out from the SR predictor; however, the SAR forecasts appear to *stay nearer* the realized correlations' trend *on average*, particularly after sudden changes; further, it can be seen that the forecasts

[18]The Almon shuffle DCC is a modified version of the Almon DCC in which the different asset time series included in the portfolio are reordered so that the vector entries are as monotonous as possible in order to enhance the performance of the Almon parameterization

resulting from the rank one DCC do not drift much apart from those obtained by the scalar model[19].

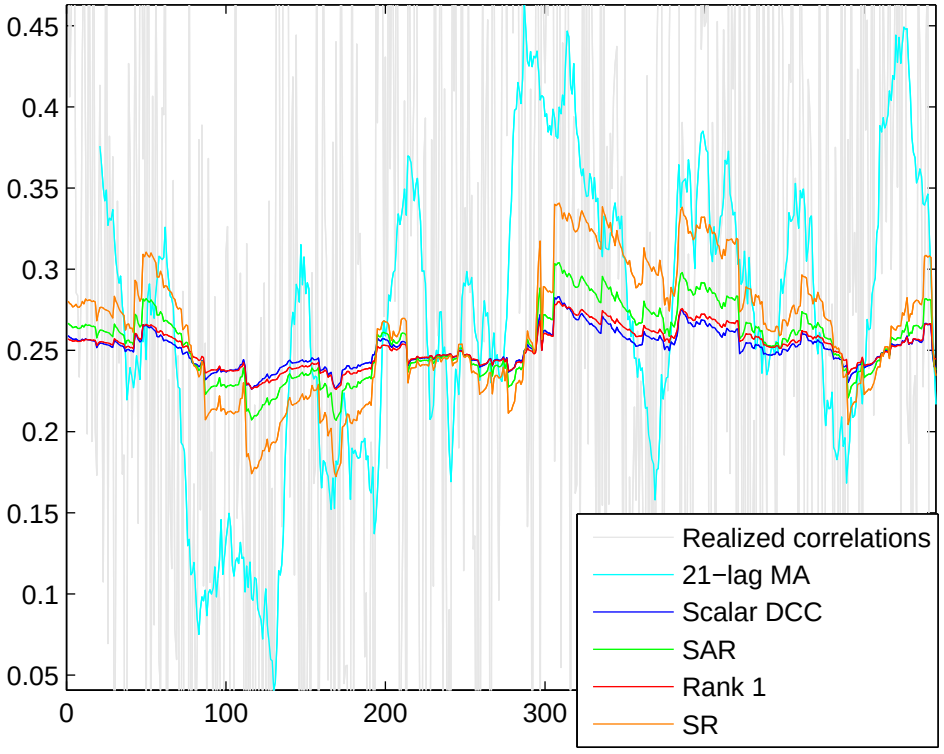

FIGURE 3.9: Conditional correlations between the first couple of stocks (Alcoa and Abbott Laboratories) in the sample predicted by: scalar DCC; SAR; rank one DCC; SR

Finally, in order to compare the three proposed predictors CS, SR and SAR also with the full rank (Hadamard) DCC, the estimation and forecasting experiment has been repeated using the same data intervals employed in the previous comparison (that is, 2000 data points for the estimation, 545 for the one-step ahead forecasting) but on a restricted number of assets: $N = 25$. In this case the parameters to be estimated relative to the Hadamard specification are 650 on a sample estimation interval of length 2000. The results in terms of square Frobenius norms are presented in Table 3.10. Apart from the CS and SAR[20]

[19]These features can be observed generally with each pair of assets

[20]The estimated shrinkage weights are 0.3066, in the case of CS, and 0.3301, in the case of SAR

predictors which even in this case achieve the best relative performances, this time the rank one specification turns out be the best alternative predictor with respect to the full scalar DCC; it is noteworthy that the full rank (Hadamard) parametrization performs poorly. Also in this case the results in terms of norms have been tested with the Giacomini-White test, whose outcomes are displayed in Table 3.12: the differences between the norms turn out to be statistically significant in each comparison. The resulting ranking can be schematized as follows:

$$\begin{aligned} \text{Shrinkage after regularization} &\succ \text{Constrained shrinkage} \succ \text{Rank one DCC} \succ \ldots \\ \ldots &\succ \text{Scalar DCC} \succ \text{Almon DCC} \succ \text{Rank two DCC} \succ \ldots \\ \ldots &\succ \text{Simple regularization} \succ \text{Hadamard DCC.} \end{aligned} \tag{3.22}$$

Model	Frobenius norm	Percentage gain
Scalar DCC	$7.0807 \cdot 10^4$	-
Constrained shrinkage	$6.9262 \cdot 10^4$	-2.181%
Simple regularization	$6.9865 \cdot 10^4$	-1.330%
Shrinkage after regularization	$6.9135 \cdot 10^4$	-2.362%
Rank two	$7.0940 \cdot 10^4$	$+0.188\%$
Rank one	$7.0638 \cdot 10^4$	-0.238%
Almon DCC	$7.0807 \cdot 10^4$	$\approx 0\%$

TABLE 3.9: Frobenius norm values and percentage gains (with respect to the scalar DCC) of the differences between the employed predictors and the realized correlation matrices in case of $N = 50$

Model	Frobenius norm	Percentage gain
Scalar DCC	$1.7966 \cdot 10^4$	-
Constrained shrinkage	$1.7758 \cdot 10^4$	-1.161%
Simple regularization	$1.8010 \cdot 10^4$	$+0.240\%$
Shrinkage after regularization	$1.7750 \cdot 10^4$	-1.206%
Hadamard DCC	$1.8388 \cdot 10^4$	$+2.345\%$
Rank two	$1.7992 \cdot 10^4$	$+0.145\%$
Rank one	$1.7924 \cdot 10^4$	-0.237%
Almon DCC	$1.7967 \cdot 10^4$	$+0.002\%$

TABLE 3.10: Frobenius norm values and percentage gains (with respect to the scalar DCC) of the differences between the employed predictors and the realized correlation matrices in case of $N = 25$

	Scalar	**CS**	**SR**	**SAR**	**Rank 2**	**Rank 1**	**Almon**
Scalar	-	**119.64**(+)	**15.82**(+)	**112.80**(+)	**59.19**(−)	**59.03**(+)	**160.58**(−)
	-	(*0.001*)	(*0.000*)	(*0.000*)	(*0.000*)	(*0.000*)	(*0.000*)
CS	-	-	**14.21**(−)	**62.75**(+)	**155.31**(−)	**125.06**(−)	**119.65**(−)
	-	-	(*0.000*)	(*0.000*)	(*0.000*)	(*0.000*)	(*0.000*)
SR	-	-	-	**17.24**(+)	**18.76**(−)	**14.47**(−)	**15.83**(−)
	-	-	-	(*0.000*)	(*0.000*)	(*0.001*)	(*0.000*)
SAR	-	-	-	-	**144.83**(−)	**116.66**(−)	**112.81**(−)
	-	-	-	-	(*0.000*)	(*0.000*)	(*0.000*)
Rank 2	-	-	-	-	-	**82.80**(+)	**59.20**(+)
	-	-	-	-	-	(*0.000*)	(*0.000*)
Rank 1	-	-	-	-	-	-	**59.13**(−)
	-	-	-	-	-	-	(*0.000*)
Almon	-	-	-	-	-	-	-
	-	-	-	-	-	-	-

TABLE 3.11: Results of the conditional Giacomini-White test on the differences between the loss functions relative to the considered predictors with forecast horizon $s = 1$ and $N = 50$ (χ^2_2 test statistics, p-values in parantheses)

	Scalar DCC	**CS**	**SR**	**SAR**	**Hadamard DCC**	**Rank 2 DCC**	**Rank 1 DCC**	**Almon DCC**
Scalar	-	**73.87**(+)	**21.74**(−)	**70.33**(+)	**104.91**(−)	**30.17**(−)	**39.63**(+)	**117.72**(−)
	-	(*0.000*)	(*0.000*)	(*0.000*)	(*0.000*)	(*0.000*)	(*0.000*)	(*0.000*)
CS	-	-	**39.04**(−)	**34.19**(+)	**129.65**(−)	**92.90**(−)	**63.92**(−)	**73.93**(−)
	-	-	(*0.000*)	(*0.000*)	(*0.000*)	(*0.000*)	(*0.000*)	(*0.000*)
SR	-	-	-	**41.33**(+)	**38.14**(−)	**22.86**(+)	**23.95**(+)	**21.73**(+)
	-	-	-	(*0.000*)	(*0.000*)	(*0.000*)	(*0.000*)	(*0.000*)
SAR	-	-	-	-	**127.84**(−)	**89.60**(−)	**62.38**(−)	**70.39**(−)
	-	-	-	-	(*0.000*)	(*0.000*)	(*0.000*)	(*0.000*)
Hadamard	-	-	-	-	-	**124.70**(+)	**115.55**(+)	**104.83**(+)
	-	-	-	-	-	(*0.000*)	(*0.000*)	(*0.000*)
Rank 2	-	-	-	-	-	-	**71.75**(+)	**30.20**(+)
	-	-	-	-	-	-	(*0.000*)	(*0.000*)
Rank 1	-	-	-	-	-	-	-	**39.81**(−)
	-	-	-	-	-	-	-	(*0.000*)
Almon	-	-	-	-	-	-	-	-
	-	-	-	-	-	-	-	-

TABLE 3.12: Results of the conditional Giacomini-White test on the differences between the loss functions relative to the considered predictors with forecast horizon $s = 1$ and $N = 25$ (χ^2_2 test statistics, p-values in parantheses)

Bibliography

[1] RiskMetricsTM - Technical Document (1996), J.P.Morgan/ Reuters

[2] Aielli G.P. (2013). Dynamic Conditional Correlation: on Properties and Estimation, Journal of Business & Economic Statistics, 31(3): 282-299

[3] Alexander C. (2001). Mastering Risk Vol.2, Financial Times/Prentice Hall

[4] Almon S. (1965). The distributed lag between capital appropriations and expenditures, Econometrica 33(1): 178-196

[5] AndersenT., Bollerslev T. and Diebold F.X. (2009). Parametric and Nonparametric Volatility Measurement, in Handbook of Financial Econometrics Vol.1, Ait-Sahalia Y. and Hansen L.P. Editors, North Holland.

[6] Andrews G.E., Askey R. and Roy R. (1999). Special functions, Cambridge University Press

[7] Attanasio O.P. (1991). Risk, Time-Varying Second Moments and Market Efficiency, Review of Economic Studies 58: 479-494

[8] Barndorff-Nielsen O., Hansen P., Lunde A. and Shephard N. (2008). Multivariate Realized Kernels: Consistent Positive Semi-Definite Estimators of the Covariation of Equity Prices with Noise and Non-synchronous Trading, Working Paper, University of Oxford,

[9] Bauwens L., Grigoryeva L. and Ortega J.P. (2014). Estimation and Empirical Performance of Non-Scalar Dynamic Conditional Correlation Models, CORE Discussion Paper 12

[10] Bauwens L., Hafner C. and Laurent S. (2011). Volatility Models, CORE Discussion Paper 58

[11] Bauwens L., Laurent S. and Rombouts J.V.K. (2006). Multivariate GARCH Models: a Survey, Journal of Applied Econometrics 21: 79–109

[12] Bauwens L. and Otranto E. (2013). Modeling the Dependence of Conditional Correlations on Volatility, CORE Discussion Paper 14

[13] Billio M. and Caporin M. (2005). Multivariate Markov Switching Dynamic Conditional Correlation GARCH Representations for Contagion Analysis, Statistical Methods & Applications 14: 145-161

[14] Billio M., Caporin M. and Gobbo M. (2003). Block Dynamic Conditional Correlation Multivariate GARCH Models, GRETA Working Paper 03.03

[15] Billio M., Caporin M. and Gobbo M. (2006). Flexible Dynamic Conditional Correlation Multivariate GARCH Models for Asset Allocation, Applied Financial Economics Letters 2(2): 123-130

[16] Bollerslev T. (1990). Modelling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized Arch Model, The Review of Economics and Statistics 72(3): 498-505

[17] Bollerslev T., Engle R.F. and Wooldridge J.M. (1988). A Capital Asset Pricing Model with Time-Varying Covariances, Journal of Political Economy 96: 116-131

[18] Boussama F., Fuchs F. and Stelzer R. (2011). Stationarity and Geometric Ergodicity of BEKK Multivariate GARCH Models, Stochastic Processes and their Applications 121(10): 2331-2360

[19] Bregman L. (1967). The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming, Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki 7(3): 620-631

[20] Caporin M. and McAleer M. (2012). Do we really need both BEKK and DCC? A tale of two multivariate GARCH models, Journal of Economic Surveys 26(4): 736-751

[21] Cappiello L., Engle R.F. and Sheppard K. (2006). Asymmetric Dynamics in the Correlations of Global Equity and Bond Returns, Journal of Financial Economics 4(4): 537-572

[22] Casella G. and Berger R.L. (2002). Statistical inference, 2nd Edition, Duxbury Thomson Learning

[23] Clements A., Scott A. and Silvennoinen A. (2012). Forecasting multivariate volatility in larger dimensions: some practical issues, NCER Working Paper Series 80

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

[25] Comte F. and Lieberman O. (2003). Asymptotic theory for multivariate GARCH processes, Journal of Multivariate Analysis 84: 61-84.

[26] Diebold F.X. and Mariano R.S. (1995). Comparing Predictive Accuracy, Journal of Business & Economic Statistics 13: 253–263

[27] Ding Z. and Engle R.F. (2001). Large Scale Conditional Covariance Matrix Modeling, Estimation and Testing, NYU Working Paper S-DRP-01-07

[28] Engle R.F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, Econometrica 50(4): 987-1007.

[29] Engle R.F. (2002). Dynamic conditional correlation - A simple class of multivariate GARCH models, Journal of Business & Economic Statistics 20: 339-350

[30] Engle R.F. (2009). The Methodology and Practice of Econometrics - Chapter 5, Oxford Univerity Press

[31] Engle R.F. and Colacito R. (2006). Testing and valuing dynamic correlations for asset allocation, Journal of Business & Economic Statistics 24: 238–253

[32] Engle R.F. and Kelly B. (2012). Dynamic Equicorrelation, Journal of Business & Economic Statistics, 30(2): 212-228

[33] Engle R.F. and Kroner K.F. (1995). Multivariate Simultaneous Generalized ARCH, Econometric Theory, Cambridge University Press 11(01): 122-150

[34] Engle R.F., Ng V.K. and Rothschild M. (1990), Asset Pricing with a Factor-ARCH Covariance Structure - Empirical Estimates for Treasury Bills, Journal of Econometrics 45: 213-237

[35] Engle R.F., Shephard N. and Sheppard K. (2008). Fitting and testing vast dimensional time-varying covariance models, NYU Working Paper FIN-07-046

[36] Engle R.F. and Sheppard K. (2001). Theoretical and Empirical properties of Dynamic Conditional Correlation Multivariate GARCH, NBER Working Paper 8554

[37] Fernandes M., de Sa Mota B. and Rocha G. (2005). A Multivariate Conditional Autoregressive Range Model, Economics Letters 86: 435-440

[38] Fleming J., Kirby C. and Ostdiek B. (2001). The Economic Value of Volatility Timing, The Journal of Finance 56(1): 329-352

[39] Gentle J.E. (2007). Matrix Algebra - Theory, Computations and Applications in Statistics, Springer Texts in Statistics, Springer

[40] Ghysels E., Santa-Clara P. and Valkanov R. (2006). Predicting volatility: getting the most out of return data sampled at different frequencies, Journal of Econometrics 131: 59-95

[41] Giacomini R. and White H. (2006). Tests of Conditional Predictive Ability, Econometrica 74(6): 1545-1578

[42] Glosten L.W., Jagannathan R. and Runkle D.E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks, Journal of Finance 48: 1779-1801

[43] Gourieroux C. (1997). ARCH Models and Financial Applications, Springer Texts in Statistics, Springer

[44] Hafner C.M. and Franses P.H. (2009). A generalized Dynamic Conditional Correlation model: simulation and application to many assets, Econometric Reviews 28(6): 612-631

[45] Hafner C.M. and Preminger A. (2009a). Asymptotic theory for a factor GARCH model, Econometric Theory 25: 336-363.

[46] Hafner C.M. and Preminger A. (2009b). On asymptotic theory for multivariate GARCH models, Journal of Multivariate Analysis 100: 2044-2054

[47] Hafner C.M. and Reznikova O. (2012). On the Estimation of Dynamic Conditional Correlation Models, Computational Statistics & Data Analysis 56(11): 3533-3545

[48] Hansen P.R. and Lunde A. (2006). Consistent ranking of volatility models, Journal of Econometrics 131: 97-121

[49] Hansen P.R., Lunde A. and Nason J.M. (2011). The model confidence set, Econometrica 79(2): 453-497

[50] Hansson B. and Hordahl P. (1998). Testing the conditional CAPM using multivariate GARCH-M, Applied Financial Economics 8: 377-388

[51] Hautsch N., Kyj L.M. and Oomen R.C.A. (2009). A Blocking and Regularization Approach to High Dimensional Realized Covariance Estimation, Journal of Applied Econometrics 27(4): 625-645

[52] He C. and Terasvirta T. (2004). An Extended Constant Correlation GARCH Model and Its Fourth-Moment Structure, Econometric Theory 20: 904-926

[53] Higham N.J. (2002). Computing the Nearest Correlation Matrix - A Problem from Finance, Journal of Numerical Analysis 22: 329-343

[54] Horn R.A. and Johnson C.R. (1999). Matrix analysis, 2nd Edition, Cambridge University Press

[55] James W. and Stein C. (1961). Estimation with quadratic loss, Proceedings of 4th Berkeley Symposium of Mathematical Statistics and Probability 1: 361-379, University of California Press

[56] Jeantheau T. (1998). Strong consistency for multivariate of ARCH models, Econometric Theory 14: 70-86.

[57] Kawakatsu H. (2003). Cholesky Factor GARCH, mimeo

[58] Kroner K.F. and Ng V.K. (1998). Modeling asymmetric comovements of asset returns, The Review of Financial Studies 11: 817-844

[59] Laloux L., Cizeau P., Bouchaud J.P. and Potters M. (1999). Noise Dressing of Financial Correlation Matrices, Physical Review Letters 83

[60] Lanne M. and Saikkonen P. (2007). A Multivariate Generalized Orthogonal Factor GARCH Model, Journal of Business & Economic Statistics 25: 61-75

[61] Laurent S., Rombouts J.V.K and Violante F. (2013). On Loss Functions and Ranking Forecasting Performances of Multivariate Volatility Models, Journal of Econometrics 173: 1-10

[62] Ledoit O., Santa-Clara P. and Wolf M. (2003). Flexible Multivariate GARCH Modeling with an Application to International Stock Markets, The Review of Economics and Statistics 85(3): 735-747

[63] Ledoit O. and Wolf M. (2003). Improved Estimation of the Covariance Matrix of Stock Returns With an Application to Portfolio Selection, Journal of Empirical Finance 10(5): 365-411

[64] Ledoit O. and Wolf M. (2004a). A Well-Conditioned Estimator for Large-Dimensional Covariance Matrices, Journal of Multivariate Analysis 88: 365-411

[65] Ledoit O. and Wolf M. (2004b). Honey, I Shrunk the Sample Covariance Matrix, Journal of Portfolio Management 30(4): 110-119

[66] Ling S. and McAleer M. (2003). Asymptotic theory for a vector ARMA-GARCH model, Econometric Theory 19: 280-310

[67] Lutkepohl H. (1996). Handbook of Matrices, John Wiley & Sons

[68] Magnus J. and Neudecker H. (2007). Matrix Differential Calculus with Applications in Statistics and Econometrics, 3rd Edition, John Wiley & Sons

[69] Marcus M. and Minc H. (1964). Inequalities for general matrix functions, Bulletin of the American Mathematical Society 70(2): 308-313

[70] Markowitz H. (1952). Portfolio selection, Journal of Finance 7: 77-91

[71] Mehta M. (1995). Random Matrices, Academic Press New York

[72] Meitz M. and Saikkonen P. (2008). Ergodicity, mixing and existence of moments of a class of Markov models with applications, Econometric Theory 24: 1291-1320.

[73] Meyn S. and Tweedie R.L. (2009). Markov Chains and Stochastic Stability, 2nd Edition, Cambridge University Press

[74] Neyman J. and Scott E.L. (1948). Consistent estimates based on partially consistent observations, Econometrica 16: 1-16

[75] Ng L. (1991). Tests of the CAPM with time-varying covariances: a multivariate GARCH approach, The Journal of Finance 46: 1507-1521

[76] Noureldin D., Shephard N. and Sheppard K. (2014). Multivariate Rotated ARCH models, Journal of Econometrics 179(1): 16-30

[77] Pedersen R.S. and Rahbek A. (2013). Multivariate Variance Targeting in the BEKK-GARCH Model, The Econometrics Journal 17(1): 24-55

[78] Pelletier D. (2006). Regime Switching for Dynamic Correlations, Journal of Econometrics 131(1-2): 445-473

[79] Qi H. and Sun D. (2006). A Quadratically Convergent Newton Method for Computing the Nearest Correlation Matrix, Journal of Matrix Analysis and Applications 28(2): 360-385

[80] Rohatgi V.K. and Ehsanes Saleh A.K.Md. (2001). An Introduction to Probability and Statistics, 2nd Edition, John Wiley & Sons

[81] Ross S.A. (1976). The Arbitrage Theory of Capital Asset Pricing, Journal of Economic Theory 13: 341-360

[82] Sheppard K. (2002). Understanding the Dynamics of Equity Covariance, mimeo

[83] Shlens J. (2014). A tutorial on principal component analysis, preprint on arXiv.org

[84] Silvennoinen A. and Terasvirta T. (2008). Multivariate GARCH models, SSE/EFI Working Paper Series in Economics and Finance 669

[85] Stein C. (1956). Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution, Proceeding of 3rd Berkeley Symposium on Mathematical Statististics and Probability 1, University of California Press

[86] Tola V., Lillo F., Gallegati M. and Mantegna R.N. (2008). Cluster Analysis for Portfolio Optimization, Journal of Economic Dynamics & Control 32: 235-258

[87] Tsay R.S. (2010). Analysis of Financial Time Series, 3rd Edition, John Wiley & Sons

[88] Tse Y.K. and Tsui A.K.C. (2002). A Multivariate Generalized Autoregressive Conditional Heteroscedasticity Model with Time-Varying Correlations, Journal of Business & Economic Statistics 20(3): 351-362

[89] Vrontos I.D., Dellaportas P. and Politis D.N. (2003). A full-factor multivariate GARCH model, Econometrics Journal 6: 311-333

[90] van der Weide R. (2002). Generalized Orthogonal GARCH - A Multivariate GARCH Model, Journal of Applied Econometrics 17(5): 549-564

[91] West K.D. (1996). Asymptotic Inference about Predictive Ability, Econometrica 64: 1067-1084

[92] White H. (1996). Estimation, Inference and Specification Analysis, Cambridge University Press