

2016/41

Multiplicative Conditional Correlation Models for Realized Covariance Matrices

LUC BAUWENS, MANUELA BRAIONE AND GIUSEPPE STORTI

A large, stylized graphic of the number '50' in a golden-yellow color, with the '0' being a thick, curved band that loops around the text below.

50 YEARS OF
CORE
DISCUSSION PAPERS

CORE

Voie du Roman Pays 34, L1.03.01

Tel (32 10) 47 43 04

Fax (32 10) 47 43 01

Email: immaq-library@uclouvain.be

<http://www.uclouvain.be/en-44508.html>

Multiplicative conditional correlation models for realized covariance matrices

Luc Bauwens^{a,b}, Manuela Braione^{a,*}, Giuseppe Storti^c

^a*Université catholique de Louvain, CORE, B-1348 Louvain-La-Neuve, Belgium*

^b*Université Côte d'Azur, SKEMA, France*

^c*Università di Salerno, Dipartimento di Scienze Economiche e Statistiche (DISES)*

Abstract

We introduce a class of multiplicative dynamic models for realized covariance matrices assumed to be conditionally Wishart distributed. The multiplicative structure enables consistent three-step estimation of the parameters, starting by covariance targeting of a scale matrix. The dynamics of conditional variances and correlations are inspired by specifications akin to the consistent dynamic conditional correlation model of the multivariate GARCH literature, and estimation is performed by quasi maximum likelihood. Simulations show that in finite samples the three-step estimator has smaller bias and root mean squared error than the full estimator when the cross-sectional dimension increases. An empirical application illustrates the flexibility of these models in a low-dimensional setting, and another one illustrates their effectiveness and practical usefulness in high dimensional portfolio allocation strategies.

Keywords: Dynamic conditional correlations; Wishart distribution; Multiplicative models; Realized covariances.

1. Introduction

The availability of high-frequency data to construct measures of daily variances of financial returns, and covariances between these, allows researchers to model time series of realized covariance matrices. One interest of these models is that they can be used for forecasting, which is of use in financial applications such as hedging, option pricing, risk management and portfolio allocation. Another potential interest of models for realized covariance matrices is that they allow researchers to study the macroeconomic and financial determinants of the changes in multivariate volatility. GARCH models can be used for the same purposes - see for example [Engle and Rangel \(2008\)](#) - but since they rely on daily observed returns, in principle they provide less precise estimates and forecasts of variances and covariances than measures based on intraday data.

Modeling a covariance matrix is a challenging task for two reasons. First, since the dimension of the object to be modeled is proportional to the square of the number of assets, the number of parameters easily becomes unmanageable even for a handful of assets. This suggests that models should be parsimoniously parameterized to avoid the parameter proliferation issue. Second, models should be congruent with the property that covariance matrices are positive definite. Several dynamic models for realized covariance matrices use the Wishart distribution, which is a natural choice since the support of the Wishart is the set of positive definite symmetric (PDS) matrices. We also use this distribution as a basis for estimation.

The main contribution of this paper is to propose a class of dynamic models for realized covariance matrices, which are computationally tractable and readily applicable in large dimensional settings. Its main features are a multiplicative structure and a parameterization inspired by the consistent dynamic conditional correlation (cDCC) model of the multivariate GARCH literature (see [Engle \(2002a\)](#) and [Aielli \(2013\)](#)). We

*Corresponding author: Manuela Braione, Université catholique de Louvain, CORE, B-1348 Louvain-La-Neuve, Belgium. *Email address:* manuela.braione@uclouvain.be.

call this type of models realized consistent DCC (Re-cDCC). Jointly with the conditional Wishart assumption, these features lead to two important advantages.

First, the multiplicative structure allows to separate the parameters of the constant long-term covariance matrix from those of the dynamic equations of variances and correlations, such that it is possible to target (i.e. estimate consistently) the constant scale matrix term by the sample average of the observed covariance matrices. After this, the parameters of the dynamic variances and correlations can be estimated in two additional steps, the first one for the realized variances, and the second one for the realized correlation matrix. This extends over similar models making use of a less flexible BEKK-type parameterization, see [Golosnoy et al. \(2012\)](#) and [Noureldin et al. \(2012\)](#).

Second, due to the properties of the Wishart distribution, the estimators of the parameters of the dynamic processes of the variances and correlations at each step have a quasi-maximum likelihood (QML) interpretation.

The stepwise estimation procedure allows us to estimate the models for a relatively large order of the realized covariance matrices, up to fifty in the empirical illustration and one hundred in simulations. Our ability to model matrices of such dimensions contrasts with existing models that also rely on the Wishart assumption: [Gouriéroux et al. \(2009\)](#) work with three assets, [Bonato et al. \(2009\)](#) with four, [Golosnoy et al. \(2012\)](#) and [Jin and Maheu \(2013\)](#) with five, [Noureldin et al. \(2012\)](#) with ten, and [Bonato et al. \(2012\)](#) with twelve. It also surpasses the dimension used in different modeling approaches for realized covariance matrices: [Chiriac and Voev \(2011\)](#) work with six assets and [Bauer and Vorkink \(2011\)](#) with five. The only notable exception is given by [Sheppard and Xu \(2014\)](#) who analyze a portfolio of forty assets. Nevertheless, if the cross-sectional dimension is small, joint estimation of all the Re-cDCC model parameters, including the constant scale matrix, can still be performed by full QML in a single step. Also, in moderately large systems, in order to gain efficiency, it is of interest to consider a variant of the proposed three-step estimator where the short term volatility and correlation parameters are jointly estimated. Focusing on cross-sectional dimensions limited to at most fifteen assets, we perform a simulation study aimed at comparing the statistical properties of step-wise and one-step QMLEs. The results of the analysis show that, as expected, in small systems, the one-step QMLE is slightly outperforming the step-wise estimators although the latter allow for remarkable efficiency gains for cross-sectional dimensions exceeding five assets. Furthermore, letting the system size further increase up to one hundred assets, we find that the efficiency and bias properties of the step-wise estimators quickly improve as both the cross sectional and time series dimension increase.

In order to assess the merits of the Re-cDCC models in practice, we undertake a comprehensive out-of-sample forecasting study using series of daily realized covariance matrices of different cross-sectional dimensions and spanning different time periods. In a first study, we illustrate the models on a dataset composed of two stocks, where the correlation process is parameterized by scalar and non-scalar cDCC processes. We show that the Re-cDCC models perform quite favourably in comparison to benchmarks both in- and out-of-sample, and that the in-sample fit of the scalar model is no worse than the considered alternatives. In a second study, we focus on the scalar specification and evaluate the model forecasting accuracy using two datasets spanning different time periods. An application of the model confidence set (MCS) of [Hansen et al. \(2011\)](#) reveals that the Re-cDCC scalar model leads to statistically significant gains in one-step-ahead predictions, irrespective of the chosen loss function or of the forecasting period considered. Moreover, applying the model to optimal portfolio allocation, we find that during the turbulent period of the 2008 financial crisis it improves the portfolio return and achieves the lowest standard deviation among the considered set of competing models, yet maintaining the smallest concentration. These results demonstrate that the flexibility of the Re-cDCC induced by capturing the dynamics of realized variances and correlations through a cDCC structure, which is missing in similar models featuring BEKK or DECO-type ([Engle and Kelly \(2012\)](#)) specifications, improves the prediction of the conditional covariance matrix of returns.

The structure of the paper is as follows. Section 2 introduces the proposed class of Re-cDCC models for realized covariance matrices. Section 3 covers the three-step QML estimation method applicable to these models. Section 4 presents the results of a simulation study to investigate the potential efficiency loss of the proposed three-step estimator with respect to a one-step estimator (in low dimension) and the bias and efficiency properties of the three-step estimator. Empirical results are provided in Section 5 and concluding remarks are presented in the last section.

2. The multiplicative realized model class

2.1. Multiplicative framework

Let C_t be a $n \times n$ positive-definite symmetric (PDS) unbiased estimator of the latent integrated covariance matrix of the daily logarithmic return vector $r_t = (r_{1,t}, \dots, r_{n,t})'$ of n assets. We assume that conditional on past information \mathfrak{I}_{t-1} consisting of C_τ for $\tau \leq t-1$, C_t follows a n -dimensional central Wishart distribution:

$$C_t | \mathfrak{I}_{t-1} \sim W_n(\nu, S_t / \nu), \quad \forall t = 1, \dots, T, \quad (1)$$

where $\nu (> n - 1)$ is the degrees of freedom parameter and S_t is the PDS conditional expectation of C_t , denoted by $\mathbb{E}(C_t | \mathfrak{I}_{t-1}) := \mathbb{E}_{t-1}(C_t)$. Under the assumption of absence of microstructure noise and other biases (see [Barndorff-Nielsen and Shephard \(2001\)](#)), S_t represents the conditional covariance matrix of returns, so that the i, j -th element of S_t is the conditional covariance between daily returns on assets i and j , $\text{Cov}(r_{i,t}, r_{j,t} | \mathfrak{I}_{t-1})$, for $i, j = 1, \dots, n$.

The unconditional expectation of C_t is assumed to be a constant matrix:

$$\mathbb{E}(S_t) = M = LL', \quad (2)$$

where the Cholesky factorization is used in the second equality, such that L (L') is a $n \times n$ lower (upper) triangular scale matrix of parameters. By applying a multiplicative decomposition, S_t can be rewritten as a product of positive-definite matrices, namely

$$S_t = LH_tL' \quad (3)$$

where H_t must satisfy the property that $\mathbb{E}(H_t) = I_n$, I_n being the $n \times n$ identity matrix. It immediately follows that the conditional covariance S_t can be interpreted as the result of the interaction of two components, namely a mean reverting matrix-valued (short-term) process H_t and a constant long-term level M around which H_t fluctuates from day to day. A similar structure has been proposed by [Golosnoy et al. \(2012\)](#) and, in the context of MGARCH models, by [Bauwens et al. \(2013\)](#). The main difference is that we use a constant matrix L for C_t while they use a time-varying matrix computed either parametrically or as a nonparametric function of time. We discuss the main pros and cons of their approach in the remainder of this section.

2.2. Parameterization of the short-term component

Our idea is to separately model the elements of the matrix H_t in a way that permits some flexibility and does not prevent to estimate the model parameters in large dimensions. Similarly to the approach developed in the Multivariate GARCH literature, we resort to parametric specifications, according to dynamic processes that depend on the available information set.

2.2.1. Modeling volatilities and correlations

There is a substantial literature on how to model the dynamics of the conditional covariance matrix of returns in a stationary framework. This means modeling the fluctuations of S_t around M , which, in our case, translates into modeling the dynamics of the short-term component H_t around its long run level I_n .

Inspired by the dynamic conditional correlation (DCC) model of [Engle \(2002a\)](#), we use the representation of a covariance matrix in terms of the corresponding diagonal matrix of standard deviations and the corresponding correlation matrix (see also [Tse and Tsui \(2002\)](#) and the *consistent* or *corrected* version proposed by [Aielli \(2013\)](#)). Therefore, the H_t component in Eq.(3) is further decomposed as

$$H_t = D_t R_t D_t, \quad (4)$$

where $D_t = \{\text{diag}(H_t)\}^{1/2}$ is a diagonal matrix with i -th diagonal entry equal to $\sqrt{H_{ii,t}}$ and R_t is the corresponding correlation matrix. This type of decomposition enables us to separately specify the dynamic equation of each short-term conditional variance and of the short-term conditional correlation matrix. In the sequel, when the difference between the components is made clear, we shall simply refer to $H_{ii,t}$ as the conditional variance of asset i and to R_t as the conditional correlation matrix of the assets. The specifications used to for their temporal dynamics are explained in the following.

To begin with, we opt for separate univariate specifications for the conditional variances which exclude interaction (spillover) terms, since this simplifies the stepwise estimation approach explained in the next section. We opt for a GARCH-type recursion of the form:

$$H_{ii,t} = (1 - \gamma_i - \delta_i) + \gamma_i C_{ii,t-1}^* + \delta_i H_{ii,t-1}, \quad (5)$$

where $C_{ii,t-1}^*$ is a diagonal element of

$$C_{t-1}^* = L^{-1} C_{t-1} L'^{-1}. \quad (6)$$

The matrix C_t^* is the observed realized covariance matrix transformed to have its unconditional expectation equal to the identity matrix of order n , so that $C_{ii,t-1}^*$ is interpretable as the short-term realized variance of asset i . This is used in Eq.(5) as the factor that drives the volatility dynamics, instead of the less precise squared daily asset return used in GARCH equations, but like in the HEAVY model of Shephard and Sheppard (2010). Restrictions on the parameters are that $\gamma_i > 0, \delta_i > 0$ and $\gamma_i + \delta_i < 1$. Notice that the constant term in Eq.(5) implies that $\mathbb{E}(H_{ii,t}) = 1$, in conformity with the restriction $\mathbb{E}(H_t) = I_n$.

With respect to proposals available in the literature, such as the ARFIMA model used in Andersen et al. (2003) and the HAR model of Corsi (2009), the choice of a GARCH(1,1)-type recursion is convenient as it is parsimonious in parameters and permits to easily impose the identification restriction $\mathbb{E}(H_{ii,t}) = 1$. Richer specifications can be considered, with more lags or nonlinear effects.

The next stage entails defining the dynamic process of the conditional correlation matrix R_t . The most general formulation we consider consists in the following equations:

$$Q_t = (I_n - AA' - BB') + AC_{t-1}^Q A' + BQ_{t-1} B' \quad (7)$$

$$R_t = \{Q_t\}^{-1/2} Q_t \{Q_t\}^{-1/2} \quad (8)$$

where $\{A,B\}$ are $n \times n$ matrices of parameters and Q_t is a PDS matrix representing the correlation driving process at time t , with fixed initial value $Q_0 = I_n$. To shorten notations, $\{Q_t\}^{1/2}$ denotes the diagonal matrix $\text{diag}(Q_{ii,t}^{1/2}, i = 1, 2, \dots, n)$. The relation in Eq.(8) is necessary to transform Q_t into a correlation matrix since its diagonal elements are not necessarily equal to 1. Notice that in order to obtain a consistent (in the sense of Aielli (2013)) DCC-type specification, C_t^Q has to be defined as

$$C_t^Q = \{Q_t\}^{1/2} D_t^{-1} C_t^* D_t^{-1} \{Q_t\}^{1/2}, \quad (9)$$

with C_t^* defined as in Eq.(6) and D_t denoting the diagonal matrix of short-term conditional standard deviations $\sqrt{H_{ii,t}}$. The matrix $D_t^{-1} C_t^* D_t^{-1}$ represents the short-term realized covariance matrix purged from its short-term conditional variances. The pre- and post-multiplication by $\{Q_t\}^{1/2}$ is necessary to achieve the equality between the unconditional expectations of C_t^Q and Q_t :

$$\begin{aligned} \mathbb{E}(C_t^Q) &= E[\{Q_t\}^{1/2} D_t^{-1} L^{-1} \mathbb{E}_{t-1}(C_t) L^{-1} D_t^{-1} \{Q_t\}^{1/2}] \\ &= E[\{Q_t\}^{1/2} D_t^{-1} L^{-1} (L D_t R_t D_t L') L^{-1} D_t^{-1} \{Q_t\}^{1/2}] \\ &= E[\{Q_t\}^{1/2} R_t \{Q_t\}^{1/2}] \\ &= \mathbb{E}(Q_t), \end{aligned}$$

and finally $\mathbb{E}(Q_t) = I_n$ by taking expectations on both sides of Eq.(7).

Equations (4) – (9) define the most general formulation of the Realized consistent DCC (Re-cDCC) model, which we consider as the main representative of the multiplicative realized class. It is probably useful to outline that other members of the same family can be derived just by changing the dynamic process for the correlation matrix R_t . For example, in the spirit of the Dynamic Equicorrelation (DECO) model of Engle and Kelly (2012), the Re-cDECO model is obtained by averaging all the pairwise DCC correlations across the day in order to get a new, equicorrelated matrix R_t^{DECO} , according to the following equations

$$\begin{aligned} R_t^{DECO} &= (1 - \rho_t) I_n + \rho_t J_n, \\ \rho_t &= \frac{1}{n(n-1)} (\iota R_t^{DCC} \iota'), \end{aligned} \quad (10)$$

with $\rho_t > 0$ the time t equicorrelation parameter, I_n the identity matrix of order n , J the n -dimensional matrix on ones and ι a row vector of ones.

While sharing a similar structure, the Re-cDECO is nonnested in the Re-cDCC just like the DECO is nonnested in the DCC. The models receive the same information but use it in very different ways to get a final estimate of the daily covariance matrix of returns. Specifically, the Re-cDECO correlations between any pair of assets are related to past realizations of *all* pairs, so that the model exploits an information pooling mechanism that is lacking in the Re-cDCC. Moreover, the equicorrelation assumption keeps estimation tractable even for very large matrices as these do not have to be inverted at each time step of the procedure. As these features can be beneficial to handle vast dimensional systems, we use the Re-cDECO as a valuable competitor of the Re-cDCC in the large sample portfolio application presented in the empirical section.

2.2.2. Feasible parameterizations

By imposing parametric restrictions on the dynamic matrix process Q_t described in Eq.(7), different conditional correlation models can be obtained. Herein, three readily applicable parameterizations are illustrated: scalar, diagonal and Hadamard; for each of them we define the constraints to be imposed for the correlation process to be stationary and give advises on their application depending on the cross-sectional dimension of the dataset to be modeled. For further insights about usage and performance of more general non-scalar DCC models we refer to the recent paper by [Bauwens et al. \(2015a\)](#).

Scalar specification. The scalar specification is the most used in practical applications when the number of assets in the system is large and parameter proliferation needs to be contained. It obtains by setting $A = \alpha^{1/2}I_n$ and $B = \beta^{1/2}I_n$, thus restricting the dynamics of all the elements of Q_t to be driven by the same parameters. The scalar version of Eq.(7) is written

$$Q_t = (1 - \alpha - \beta) + \alpha C_{t-1}^Q + \beta Q_{t-1}, \quad (11)$$

and the constraints to be imposed on the parameters are $\alpha > 0$, $\beta \geq 0$ and $\alpha + \beta < 1$. The last constraint ensures the existence of $\mathbb{E}(Q_t)$. As the number of parameters implied by Eq.(11) is equal to two independently of the dimension of the cross-section, the scalar specification is worth using in the modeling of realized covariance matrices in dimensions up to fifty, as we show later on, and perhaps more.

Diagonal specification. The diagonal specification permits asset-specific parameter heterogeneity in the dynamic structure. It implies that the parameter matrices $\{A, B\}$ are diagonal, *i.e.* with diagonal elements equal to $\alpha_{ii}^{1/2}$ and $\beta_{ii}^{1/2}$. We further impose that $\alpha_{ii}^{1/2} > 0$ and $\beta_{ii}^{1/2} \geq 0$. Denoting $\tilde{A} = \text{diag}(\alpha_{ii}^{1/2})$ and $\tilde{B} = \text{diag}(\beta_{ii}^{1/2})$ for $i = 1, \dots, n$, the diagonal model is given by

$$Q_t = (I_n - \tilde{A}\tilde{A}' - \tilde{B}\tilde{B}') + \tilde{A}C_{t-1}^Q\tilde{A}' + \tilde{B}Q_{t-1}\tilde{B}', \quad (12)$$

implying a number of correlation parameters to be estimated equal to $2n$, which can be already prohibitive in moderately large systems.

The restriction to be imposed for the existence of $\mathbb{E}(Q_t)$ is that the eigenvalues of the diagonal matrix $A \otimes A + B \otimes B$ be less than 1 in modulus. Necessary and sufficient condition for this to hold are $\alpha_{ii} + \beta_{ii} < 1$, which also assures that $I_n - \tilde{A}\tilde{A}' - \tilde{B}\tilde{B}'$ is positive definite.

Hadamard specification. Define by $\text{math}(\cdot)$ the operator that transforms a $n^* = n(n+1)/2$ dimensional vector into a $n \times n$ symmetric matrix of parameters. Then, given α and β $n^* \times 1$ vectors of parameters, we can rewrite $\{A, B\}$ as $A := \text{math}(\alpha)$ and $B := \text{math}(\beta)$, such that the Hadamard representation of Eq.(7) is expressed as

$$Q_t = (\mathbf{1}'\mathbf{1} - \text{math}(\alpha) - \text{math}(\beta)) + \text{math}(\alpha) \odot C_{t-1}^Q + \text{math}(\beta) \odot Q_{t-1}, \quad (13)$$

where $\mathbf{1}$ is the unit vector of dimension $1 \times n$ and \odot denotes element by element product. We only consider the case where $\{A, B\}$ are full rank matrices. Sufficient condition on the parameters α and β ensuring that $\mathbb{E}(Q_t)$ is finite are $|\alpha_i + \beta_i| < 1$ for $i = 1, \dots, n$ while positivity of Q_t is assured by constraining $\mathbf{1}'\mathbf{1} - \text{math}(\alpha) - \text{math}(\beta)$ to be positive definite. Despite the advantage of having a correlation dynamics that differs for all pairs of assets, this specification exhibits a quadratic dependence on the dimension n of the process with $n(n+1)$ parameters to be estimated. Therefore, to avoid parameter proliferation, its applicability is limited to low dimensional systems.

2.3. Some remarks

Given the recent contributions to the literature listed below, we find opportune to briefly discuss the possibility to include some extra features in our models. At the same time, we also outline pros and cons of pursuing such approaches.

2.3.1. Dynamic attenuation approach

According to the theoretical developments in [Barndorff-Nielsen and Shephard \(2004\)](#) and [Barndorff-Nielsen et al. \(2009\)](#), estimated realized covariance matrices can be decomposed into the sum of two components: the latent true covariance matrix and a measurement error term. The latter quantity is usually present in finite samples and may be more sizeable on some days and less on others, with a direct consequence on the accuracy of the realized covariance measure. Correctly estimating the magnitude of the noise is not a trivial task especially in the multivariate framework, where the error term is not a scalar and knowledge of the exact form of its variance-covariance matrix is required. [Barndorff-Nielsen and Shephard \(2004\)](#) show that for the Realized Covariance estimator, as well as for some other consistent estimators (e.g. the Multivariate Kernel of [Barndorff-Nielsen et al. \(2009\)](#)), the covariance matrix of the error term, let us call it Π_t , is time-varying and proportional to the *integrated quarticity*, which can be estimated consistently using the same high-frequency data that are used for estimating the realized covariance measures.

Recently, [Bollerslev et al. \(2016a\)](#) introduced a simple and intuitive way to account for the heteroskedasticity of the errors when modeling series of realized volatilities. The intuition behind their approach is that, on days when the variance of the measurement error is small, the daily realized variance estimator provides a stronger signal for next day's volatility than on days when the variance is large. By explicitly incorporating this feature into the parameterization of the model it is possible to dynamically attenuate the parameters, and in turn improve the forecasts. A similar reasoning holds in the multivariate setting.

The Re-cDCC model does not account for the estimation error in the series of realized covariance matrices. An improved formulation could be achieved by including those effects in the volatilities and correlation equations, where in the latter case an appropriate transformation of Π_t , e.g. $\Gamma_t = \{\Pi_t\}^{-1/2}\Pi_t\{\Pi_t\}^{-1/2}$, should be implemented to account for the measurement errors in the correlation dynamics. Assuming the simplified scalar case, Equations (5) and (7) are changed as follows:

$$H_{ii,t} = (1 - \gamma_i - \delta_{i,t}) + \gamma_i C_{ii,t-1}^* + \delta_{i,t} H_{ii,t-1}, \quad (14)$$

$$\delta_{i,t} = \delta_i + \delta_{i,Q} \Pi_{ii,t-1}, \quad (15)$$

$$Q_t = (1 - \alpha - \beta_t) + \alpha C_{t-1}^Q + \beta_t \odot Q_{t-1}, \quad (16)$$

$$\beta_t = \beta J_n + \beta_Q \Gamma_{t-1}, \quad (17)$$

with $\delta_{i,Q}$ and β_Q scalar parameters, J_n the n -dimensional matrix of ones and \odot the Hadamard product. Equations (15) and (17) allow the dynamics in each conditional variance series and each of the individual elements in the pseudo-correlation matrix Q_t to depend on their own measurement error variances. Notice that negative values of $\delta_{i,Q}$ and β_Q assign lower weight to those that are measured with greater error.

Clearly, applying this *measurement error-correction* is simpler with a handful of assets but may become difficult in larger dimensions, not only because it requires a higher number of parameters to be estimated (one for each univariate series plus one for the correlation matrix as a whole), but also because the properties of the dynamic evolution of the measurement error for the correlations are not well known in this case.

2.3.2. A time-varying long-run representation

One assumption underlying the Re-cDCC model is that conditional covariances mean-revert to constant quantities. This can be overcome by allowing the long-run trend $M = LL'$ to evolve across time as function of changing economic conditions. This idea has been used for long in the GARCH literature: since [Engle and Lee \(1999\)](#) introduced a GARCH model with additive long and short-run dynamic components, other ideas have been proposed for modeling long-run changes in volatilities and correlations, see e.g. [Engle and Rangel \(2008\)](#), [Engle et al. \(2008a\)](#), [Colacito et al. \(2011\)](#), [Hafner and Linton \(2010\)](#) and [Bauwens et al. \(2013\)](#). Nonparametric as well as parametric specifications of the trend have been experimented, but which one to prefer on a practical ground remains the practitioner's own choice.

The first time-varying long-run model built for realized covariance matrices is the MIDAS-Conditionally Autoregressive Wishart (CAW) of [Golosnoy et al. \(2012\)](#), designed to capture long-run fluctuations in the

levels of covariances by means of a parametric Mixed Data Sampling (MIDAS) filter of past realized covariance matrices. Later, the BEKK-type structure accommodated in the MIDAS-CAW has been extended to more flexible specifications by [Bauwens et al. \(In Pressb\)](#) and [Bauwens et al. \(In Pressa\)](#), who perform an extensive out-of-sample forecasting exercise to evaluate the usefulness of these models for financial applications.

While these dynamic long-run representations involve richer structures than the Re-cDCC, there are still valuable reasons why we believe our model may be useful to practitioners. First, it is extremely easy to programme and fast to estimate, especially in its scalar form, even when the cross-sectional dimension is of several tens of assets. The same is not necessarily true when a time-varying intercept is present in the covariance specification, where, depending on the chosen approach, more parameters have to be estimated.

Second, while targeting can be readily applied in the Re-cDCC model, when a dynamic intercept is present things are not so simple. In fact, for the targeting to work, the user needs to have knowledge of the analytical relationship between the mean of the targeted process and the other model parameters. In many cases, due to the complex dependence structure encountered in the dynamic behavior of covariance matrices, this is far from being obvious (see also [Aielli \(2013\)](#) and the discussion in [Bauwens et al. \(In Pressa\)](#)).

Last, the empirical findings in [Bauwens et al. \(In Pressa\)](#) show that a dynamic long-term component is particularly useful in periods of high market volatility, when the assets experience fast and abrupt changes in their unconditional levels due to non-normal market conditions.¹ However, in periods of calm, it is noticeably more complicated to disentangle the different volatility components, as the long-term trend tends to be flat over time. In such circumstances highly parameterized sophisticated models suffer from additional parameter uncertainty. This suggests that accurate predictions can be obtained by employing models that do not necessarily account for time-varying long run levels, and such models remain suitable benchmarks in empirical applications.

3. Multistep estimation approach

For a Re-cDCC model in the class defined in Section 2, the vector of parameters to be estimated is denoted by ϕ . It can be partitioned into $(\nu, \mathbf{n}'_M, \phi'_v, \phi'_c)'$, where ν is the degrees of freedom parameter of the Wishart distribution, \mathbf{n}_M is the vector of unique parameters included in the constant long-term matrix M (with $n(n+1)/2$ elements), and ϕ_v, ϕ_c are the vectors of conditional variance and correlation parameters, respectively. The dimension of ϕ_v is $2n$ if the GARCH-type formulation of equation Eq.(5) is adopted, while that of ϕ_c depends on the specification of the Q_t process. Estimation of all the components of ϕ , if the cross-sectional dimension is relatively small, can be easily performed by full QML, *i.e.* by maximizing the sample log-likelihood function with respect to all the parameters in one step. In larger settings, this can become prohibitive, thus requiring an *ad hoc* approach. Ours is fully described in the following subsections.

3.1. Covariance targeting

The possibility of covariance targeting is important for the estimation of the proposed model. To understand its usefulness, consider that the Re-cDCC model, even in its scalar specification, requires the estimation of \mathbf{n}_M , which has a dimension proportional to the square of the number of assets. This feature can render the estimation of all the parameters, in a maximum likelihood framework, prohibitive for large values of n . Hence, it is useful to be able to estimate \mathbf{n}_M in a simple and quick way, before estimating the remaining parameters in a maximum likelihood framework. For this, we resort to *targeting*, which consists in estimating the constant matrix M by a method of moment estimator, without having to estimate the other parameters. After this first step, the estimate of M is substituted for the corresponding parameter matrix in the likelihood function to be maximized in the next steps. Since the dynamic equations of variances and correlations both depend on M through C_t^* , see Eq.(6) and Eq.(9), it is desirable that the targeting is consistent, even if it is inefficient. In other words, since the QML estimators of the remaining parameters depend on the targeting estimator, they cannot be consistent if the targeting estimator is not consistent; see [Engle et al. \(2008b\)](#) and [Pedersen and Rahbek \(2014\)](#) among others.

¹Examples are the Dot-Com Bubble and the more recent financial crisis.

For the Re-cDCC model, under conditions that ensure the existence of second-order moments of C_t , consistent targeting of the matrix M is straightforwardly achieved by computing

$$\hat{M} = \frac{1}{T} \sum_{t=1}^T C_t, \quad (18)$$

which amounts to estimate the unconditional covariance matrix of the daily returns by the sample average of the realized covariance matrices up to time T . It is worth to mention that [Ledoit et al. \(2003\)](#) and [Fan et al. \(2012\)](#), among others, noticed that when the number of assets n is a significant share of T , the sample covariance matrix, subject to large estimation errors, becomes a poor estimator of M as well as ill-conditioned. In these cases, they advice using some improved estimators built under varying assumptions that are all suitable for financial applications. They are developed in a different framework (i.e. for MGARCH models), but it could be worth trying to rationalize their use also in the realized covariance framework. For future work, it would be interesting to compare these different approaches empirically.

Overall, applying the targeting has three important implications. Firstly, \hat{L} , such that $\hat{L}\hat{L}' = \hat{M}$, can be substituted for L in the computation of the likelihood function without impairing the possibility of consistent estimation of the parameters of the dynamic equations of conditional variances and correlations. Secondly, along with parsimonious parameterizations of the volatility and correlation processes, it makes the model applicable for large cross-sections. Thirdly, it enables a three-step estimation approach that crucially reduces the computational burden and speeds up estimation.

In order to assess the impact of using targeting along with the three-step approach for scalar Re-cDCC models, in Section 4 we perform a simulation study to assess the bias and efficiency properties of the estimator of the parameters. These properties are compared to those achieved by one-step estimation, namely by maximizing the likelihood function with respect to all the parameters jointly.

3.2. QML estimation

In $\boldsymbol{\phi} = (\nu, \mathbf{n}'_M, \boldsymbol{\phi}'_v, \boldsymbol{\phi}'_c)'$, we also partition $\boldsymbol{\phi}_v$ into $\boldsymbol{\phi}_v = (\boldsymbol{\phi}_v^{(1)'}, \boldsymbol{\phi}_v^{(2)'}, \dots, \boldsymbol{\phi}_v^{(n)'})'$, where $\boldsymbol{\phi}_v^{(i)}$ is the parameter vector of the conditional variance equation specific to asset i . Using the expression of a Wishart density function, and of S_t in Eq.(3), the log-likelihood contribution of observation t , $\ell(C_t; \boldsymbol{\phi} | \mathfrak{F}_{t-1}) := \ell_t(\boldsymbol{\phi})$, is given by

$$\begin{aligned} \ell_t(\boldsymbol{\phi}) &= \frac{\nu n}{2} \log \frac{\nu}{2} + \frac{\nu - n - 1}{2} \log |C_t| - \sum_{i=1}^n \log \Gamma[(\nu + 1 - i)/2] \\ &\quad - \frac{\nu}{2} \log |LH_t(\boldsymbol{\phi})L'| - \frac{\nu}{2} \text{tr}\{(LH_t(\boldsymbol{\phi})L')^{-1}C_t\}, \end{aligned} \quad (19)$$

where the dependence of H_t on $\boldsymbol{\phi}$ is made explicit. When targeting is used, meaning that $M = LL'$ is estimated consistently by $\hat{M} = 1/T \sum_{t=1}^T C_t$, then \hat{L} is substituted for L in Eq.(19), otherwise, L is an unknown set of parameters that is estimated jointly with the remaining parameters (the non-zero elements of L form a set of parameters which is equivalent to \mathbf{n}_M).

Given our interest in applying the model to high-dimensional systems, we consider targeting as an essential part of the estimation procedure, as otherwise estimation would be not feasible. For ease of notation, in the sequel we use L in the expressions of the log-likelihood function instead of \hat{L} . The next proposition (proven in [Appendix A](#)) provides the basis of a three-step estimation procedure.

Proposition 1. *The likelihood contribution $\ell_t(\boldsymbol{\phi})$ in Eq.(19) can be written as*

$$\ell_t(\boldsymbol{\phi}) = \ell_{0,t}(\nu) + \ell_{\nu,t}(\nu, \boldsymbol{\phi}_v; \hat{\mathbf{n}}_M) + \ell_{c,t}(\nu, \boldsymbol{\phi}_c; \hat{\boldsymbol{\phi}}_v, \hat{\mathbf{n}}_M), \quad (20)$$

where

$$\ell_{0,t}(\nu) = \frac{\nu n}{2} \log \frac{\nu}{2} + \frac{\nu - n - 1}{2} \log |C_t| - \sum_{i=1}^n \log \Gamma[(\nu + 1 - i)/2] \quad (21)$$

$$\ell_{\nu,t}(\nu, \boldsymbol{\phi}_v; \hat{\mathbf{n}}_M) = -\nu \log |D_t| - \frac{\nu}{2} \text{tr}\{D_t^{-1}L^{-1}C_tL'^{-1}D_t^{-1}\}, \quad (22)$$

$$\ell_{c,t}(\nu, \boldsymbol{\phi}_c; \hat{\boldsymbol{\phi}}_v, \hat{\mathbf{n}}_M) = -\frac{\nu}{2} \log |R_t| - \nu \log |L| - \frac{\nu}{2} \text{tr}\{(R_t^{-1} - I_n)D_t^{-1}L^{-1}C_tL'^{-1}D_t^{-1}\} \quad (23)$$

If the univariate equations for the conditional variances do not include spillover terms other than their own lagged conditional variance and their parameters are variation-free, $\ell_{v,t}$ can be written as the sum of n univariate functions:

$$\ell_{v,t}(\nu, \boldsymbol{\phi}_v; \hat{\boldsymbol{n}}_M) = \frac{\nu}{2} \left[- \sum_{i=1}^n \log H_{ii,t} - \sum_{i=1}^n H_{ii,t}^{-1} C_{ii,t}^* \right] = \frac{\nu}{2} \sum_{i=1}^n \ell_{v,t}^{(i)}(\boldsymbol{\phi}_v^{(i)}), \quad (24)$$

with $C_{ii,t}^*$ denoting the i -th diagonal element of the matrix $L^{-1}C_tL'^{-1}$.

The above proposition inspires the following comments:

1. The $\ell_{v,t}$ part of the log-likelihood is proportional to the shape parameter ν . This implies that it can be maximized with respect to the elements of $\boldsymbol{\phi}_v$ without estimating ν .
2. Each function $\ell_{v,t}^{(i)}$, defined as the terms between square brackets in Eq.(24), only depends on the parameters $\boldsymbol{\phi}_v^{(i)}$ specific to the conditional variance equation of asset i . It follows that maximization of $\ell_{v,t}$ can be achieved through n separate optimizations (under the assumptions stated in Proposition 1), which simplifies estimation. Notice that $\ell_{v,t}^{(i)}$ corresponds to the log-likelihood of an exponential distribution.
3. The $\ell_{c,t}$ part of the log-likelihood depends on the whole set of parameters in $\boldsymbol{\phi}$. Since it is linear in ν , it can be maximized with respect to these parameters without estimating ν .
4. As our primary interest is in the estimation of the parameters of the conditional variances and correlation processes, we consider the degree of freedom parameter ν as a nuisance parameter, which is moreover justified by the fact that the other parts of the log-likelihood depend linearly on ν . Consequently, the $\ell_{0,t}(\nu)$ part of the likelihood can be ignored.

The main advantage of these results is that we can adopt a three-step procedure to estimate the model parameters:

1. In step 1, the vector of unique parameters \boldsymbol{n}_M of the constant matrix M are consistently estimated by a method of moment estimator and the resulting estimator $\hat{\boldsymbol{n}}_M$ is substituted for the corresponding parameter in the likelihood function.
2. In step 2, conditional on $\hat{\boldsymbol{n}}_M$, the conditional variance parameters are estimated as

$$\hat{\boldsymbol{\phi}}_v^{(i)} = \underset{\boldsymbol{\phi}_v^{(i)}}{\operatorname{argmax}} \sum_{t=1}^T \ell_{v,t}^{(i)}, \text{ for } i = 1, \dots, n.$$

3. In step 3, the parameters of the correlation equation are estimated as

$$\hat{\boldsymbol{\phi}}_c = \underset{\boldsymbol{\phi}_c}{\operatorname{argmax}} \sum_{t=1}^T \ell_{c,t}$$

conditional on $\hat{\boldsymbol{n}}_M$ and $\hat{\boldsymbol{\phi}}_v$ obtained in the previous steps.

In order to provide inferential properties for the three-step procedure, we observe that it can be represented as a three-step moment-estimator. In the following proposition, proven in [Appendix A](#), we derive analytical expressions for the underlying moment conditions and show that their expectation is zero at the true parameter values.

Proposition 2. *The vector of moment conditions for the t -th observation is $\tilde{\boldsymbol{s}}_t = (\tilde{\boldsymbol{s}}'_{n,t}, \tilde{\boldsymbol{s}}'_{v,t}, \tilde{\boldsymbol{s}}'_{c,t})'$, with components:*

$$\tilde{\boldsymbol{s}}_{n,t} = \operatorname{vech}(C_t) - \boldsymbol{n}_M \quad (25)$$

$$\tilde{\boldsymbol{s}}_{v,t} = \frac{\partial \ell_{v,t}(\nu, \boldsymbol{\phi}_v; \hat{\boldsymbol{n}}_M)}{\partial \boldsymbol{\phi}_v} = -\frac{\nu}{2} \frac{\partial \log |H_t|}{\partial \boldsymbol{\phi}_v} - \frac{\nu}{2} \frac{\partial \operatorname{tr}(L^{-1}C_tL'^{-1}H_t^{-1})}{\partial \boldsymbol{\phi}_v} \quad (26)$$

$$\tilde{\boldsymbol{s}}_{c,t} = \frac{\partial \ell_{c,t}(\nu, \boldsymbol{\phi}_c; \hat{\boldsymbol{n}}_M, \hat{\boldsymbol{\phi}}_v)}{\partial \boldsymbol{\phi}_c} = -\frac{\nu}{2} \frac{\partial \log |R_t|}{\partial \boldsymbol{\phi}_c} - \frac{\nu}{2} \frac{\partial \operatorname{tr}(R_t^{-1}D_t^{-1}L^{-1}C_tL'^{-1}D_t^{-1})}{\partial \boldsymbol{\phi}_c}. \quad (27)$$

At the true parameter values $\boldsymbol{\phi}_0 = \{\boldsymbol{n}_{M,0}, \boldsymbol{\phi}_{v,0}, \boldsymbol{\phi}_{c,0}\}$, the expected value of $\tilde{\boldsymbol{s}}_t$ is equal to $\mathbf{0}$.

Under the stated result, the theory developed in Newey and McFadden (1994) can be applied to prove, at the price of the additional regularity conditions stated in their paper, the consistency and asymptotic normality of the three-step estimator. These hold even in cases in which the Wishart assumption is not satisfied, making our approach robust to misspecification of the distribution of the realized covariance matrices. It is worth noting that, if targeting is not used, the resulting estimator can be represented as a two step QML estimator (QMLE) with score given by $(\tilde{s}'_{v,t}, \tilde{s}'_{c,t})'$. In this case the estimation approach is similar to that applied, in different settings, by Engle and Russell (1998), Engle (2002b), and Engle and Gallo (2006) in the estimation of ACD and multiplicative error models with a gamma conditional distribution.

Concerning the asymptotic distribution of the estimator $\hat{\phi}$, under standard regularity conditions, it holds that

$$\sqrt{T}(\hat{\phi} - \phi_0) \xrightarrow{d} N(\mathbf{0}, I^{-1} \Xi (I^{-1})')$$

where $\Xi = \text{var} \left[\frac{1}{\sqrt{T}} \sum_{t=1}^T \tilde{s}_t \right]$ is estimated using a HAC estimator (e.g. Newey and West (1987)) and

$$l = -\mathbb{E} \left[\frac{\partial \tilde{s}_t}{\partial \phi} \right] = -\mathbb{E} \begin{bmatrix} -I_{n^*} & \frac{\partial^2 \tilde{s}_v}{\partial n_M \partial \phi'_v} & \frac{\partial^2 \tilde{s}_c}{\partial n_M \partial \phi'_c} \\ 0 & \frac{\partial^2 \tilde{s}_v}{\partial \phi_v \partial \phi'_v} & \frac{\partial^2 \tilde{s}_c}{\partial \phi_v \partial \phi'_c} \\ 0 & 0 & \frac{\partial^2 \tilde{s}_c}{\partial \phi_c \partial \phi'_c} \end{bmatrix}, \quad (28)$$

with the Hessian having a block upper triangular structure. As targeting is used, the covariance matrix of the estimated parameters in ϕ_c has to include an adjustment to account for the efficiency loss due to the preliminary estimation of the parameters in ϕ_v and of the parameters in the constant term.

3.3. Third stage composite likelihood estimation

A possibly interesting option, which we outline here but do not pursue empirically, is to estimate ϕ_c by composite likelihood (CL). This approach, first introduced by Lindsay (1988) and later on popularized by Engle and Sheppard (2001) and Pakel et al. (2014), consists, roughly speaking, in using an approximation of the joint marginal density based on lower dimensional marginal densities. The estimator that obtains leads to a loss in efficiency compared to a full quasi-likelihood approach but is shown to be potentially more robust to misspecification issues arising with high dimensional models. For this reason, the CL procedure has been extensively applied in the MGARCH literature for the scalar versions² of the BEKK, DCC and cDCC models. In the latter case, Pakel et al. (2014) find that the CL allows for the estimation of models in dimension of many hundreds, being easy to implement and not effected as much as the full QML estimator by the incidental parameter problem.

In the realized co-volatility literature, Bollerslev et al. (2016b) use the CL approach with the aim to simplify estimation and reduce computational burden in large dimensional systems where multistep solutions are not feasible. Nevertheless, none of these papers carries out a formal investigation of the properties of realized covariance models fitted by CL. In our setting, where the likelihood function is already split in blocks, the CL could be applied to the third estimation stage in order to avoid inverting many times matrices of large dimensions. In fact, when many hundreds of assets are involved, $\ell_{c,t}(v, \phi_c; \hat{\phi}_v, \hat{n}_M)$ can easily result in numerically unstable R_t^{-1} matrices.

Luckily, it turns out that for the scalar Re-cDCC – and Re-cDECO – model, the Wishart assumption also enables us to use the CL method. This can be briefly described as follows.

For any square matrix C_t of order n , let $C_{AA,t}$ denote a square matrix of order n_A , extracted from C_t , which has its main diagonal elements on the main diagonal of C_t . Namely, if A stands for a subset of n_A different indexes of $\{1, 2, \dots, n\}$, $C_{AA,t}$ is the matrix that consists of the intersection of the rows and columns of C_t corresponding to the selection of indexes denoted by A .

The method relies on the following three results:

R1: If $C_t \sim W_n(v, S_t/v)$, $C_{AA,t} \sim W_{n_A}(v, S_{AA,t}/v)$ for any selection of n_A indexes.

²There has not been, so far, formal evidence of the theoretical validity of applying CL estimation to non-scalar models. A first attempt in this direction has recently been made by Bauwens et al. (2015b).

R2: If $H_t = D_t R_t D_t$, $H_{AA,t} = D_{AA,t} R_{AA,t} D_{AA,t}$.

R3: If $Q_t = (1 - \alpha - \beta) + \alpha C_{t-1}^Q + \beta Q_{t-1}$, then $Q_{AA,t} = (1 - \alpha - \beta) + \alpha C_{AA,t-1}^Q + \beta Q_{AA,t-1}$.

Result 1 is a property of the Wishart distribution, see [Lutkepohl \(1996\)](#). Notice that if applied with $n_A = 1$, it follows that the marginal distribution of a diagonal element of a Wishart matrix is a gamma, a result that is used to form the first step log-likelihood function $\ell_{v,t}^{(i)}$. Results 2 and 3, given that D_t is diagonal and L is lower triangular, are obvious algebraic derivations. A CL third step estimator of the parameters α and β is then defined as the maximizer of the sum of a number of Wishart marginal log-likelihoods of n_A dimensional sub-matrices corresponding to different choices of indices A . The most obvious choice is to select all unique sub-matrices of order 2, *i.e.* all the $n(n-1)/2$ correlation coefficients or pairs of assets. In this bivariate Wishart log-likelihoods, only matrices of order 2 must be inverted, and this can be efficiently programmed. Such a CL for the contribution of observation t is formed by averaging across all pairs in the sample:

$$CL_t(v, \phi_c; \hat{\phi}_v, \hat{\mathbf{n}}_M) = \frac{1}{n} \sum_{h=1}^n \sum_{k < h} \ell_{hk,t}(v, \phi_c; \hat{\phi}_v^{(h)}, \hat{\phi}_v^{(k)}, \hat{\mathbf{n}}_M^{(hk)}).$$

In this case the CL costs $O(n^2)$ calculations, while other choices, such as when subset of all (or some) contiguous pairs are used, can reduce the cost to $O(n(n-1)/2)$. Hence, depending on the particular sampling strategy adopted, computational gains can be rather substantial (see e.g. [Bauwens and Storti \(2013\)](#)).

4. Simulation study

The three-step estimation approach exposed in Section 3 is easy to implement and provides consistent parameter estimates, under appropriate conditions. However, not much is known about its behavior in finite samples, nor how it compares to the full QML estimator, hereafter denoted 1QML. We do not include v , considered again as a nuisance parameter.

Table 1: **Simulation design for the scalar Re-cDCC process**

n		2, 5, 7, 10, 15, 25, 50, 100
M	in Eq. (2)	$M_{i,i} = 0.1$, $M_{i,j} = 0.02$ for $i \neq j$
γ_i	in Eq. (5)	$\gamma_i = 0.05$, for $i = 1, \dots, n$
δ_i	in Eq. (5)	$\delta_i = 0.9$, for $i = 1, \dots, n$
α	in Eq. (11)	0.1
β	in Eq. (11)	0.8
v	in Eq. (1)	$2n$
T		500, 1000, 2000 (+ 500 burn-in)

A clear-cut advantage of the three-step estimator (3QMLE) over 1QMLE is that it reduces the computational burden arising in the optimization process by splitting it in several steps. However, in medium-low cross-sectional dimensions up to ten-fifteen assets, the 1QMLE is feasible and replacing it with 3QMLE may incur a loss of efficiency. Moreover, it is also possible to define a two-step estimator (2QMLE) that adds to the targeting step a second, single, step estimator of (ϕ_c, ϕ_v) obtained by maximizing the sample average of $\ell_{v,t}(v, \phi_v; \hat{\mathbf{n}}_M) + \ell_{c,t}(v, \phi_c; \phi_v, \hat{\mathbf{n}}_M)$. The 2QMLE does not separate the estimation of the variance and correlation parameters and due to this it may incur a smaller efficiency loss than 3QMLE with respect to 1QMLE, the difference between 2QMLE and 1QMLE being due solely to the targeting step.

In order to compare the three estimators, we report the results of a Monte Carlo experiment, designed to compare the finite-sample bias and precision of 3QMLE, 2QMLE and 1QMLE in dimensions up to fifteen (since 1QMLE is too costly in larger dimensions). In a nutshell, we find that 3QMLE has as good properties as the other two estimators in very low dimensions ($n \leq 5$), and has better properties as soon as n exceeds 7. We also report results for the 3QMLE in dimensions up to $n = 100$.

Specifically, we assume as data generating process (DGP) a scalar Re-cDCC process, from which we simulate $K = 500$ series according to the design summarized in Table 1. To cover a variety of cases, eight different values of the cross-sectional dimension and three time series lengths are considered. In all cases, the estimated model corresponds to the model class to which the DGP belongs, hence it is correctly

specified. The statistical properties of the estimators are evaluated in terms of relative bias (RB) and root mean squared error (RMSE) defined as

$$RB(\hat{\theta}) = \frac{1}{K} \sum_{k=1}^K \frac{(\hat{\theta}_k - \theta_0)}{\theta_0}, \quad RMSE(\hat{\theta}) = \sqrt{\frac{1}{K} \sum_{k=1}^K (\hat{\theta}_k - \theta_0)^2},$$

for the estimator $\hat{\theta}$ of the parameter θ (an element of ϕ or M) having the true value θ_0 and estimated by $\hat{\theta}_k$ for the k -th simulated series. To synthesize results, the statistics for the elements of the matrix M are averaged across parameters: namely, at each k , RB and RMSE are averaged across diagonal (variance) and off diagonal (covariance) elements. Similarly, the results for the parameters of the univariate variance equations are reported as averages across assets.

4.1. Finite sample properties of 3QMLE

Let first comment the finite-sample properties of the 3QMLE, reported in Table 2 and of the targeting estimator, reported in Table 3. A few conclusions can be drawn in terms of bias:

1. The average parameters of the conditional variance equations, i.e. $\bar{\gamma}$ and $\bar{\delta}$, achieve the highest biases across the panels. The bias on $\bar{\gamma}$ is always positive and ranges from the highest value of 13.1% (for $n = 5, T = 500$) to the lowest value of 0.4% (for $n = 10, T = 2000$). The bias on $\bar{\delta}$ is negative and ranges between -14.3% (for $n = 5, T = 500$) and -1% (for $n = 10, T = 2000$). The relative bias of $\bar{\gamma}$ and $\bar{\delta}$ decreases when T increases, while it remains stable with respect to n . However, it is worth to stress that the conditional variance series are estimated equation by equation and the reported estimates are averaged across a number of parameters that increases with n . The average stability of these estimates across n is an encouraging indication that, despite the increasing cross-sectional size, the first step moment targeting is not negatively affecting second stage estimation.
2. For α , the bias is negative, except in two cases, and anyway small, being under one percent in absolute value with only one exception occurring at $n = 2, T = 500$, where $RB(\alpha) = 3\%$. The bias shrinks toward zero as T increases. This is more visible for $n \geq 7$ than for $n \leq 7$, probably because there is more information for large n about a parameter that is common to all correlation processes. There is evidence of a slight decreasing trend in the bias as n increases from $n = 2$ ($n = 5$ for $T = 500$), though the bias becomes essentially stable for $n \geq 25$.
3. The bias for β is always negative and in the majority of cases very small, ranging between the minimum of -0.3% for $n = 100, T = 2000$ and the maximum of -5.3% for $n = 2, T = 500$, the latter caused by a few outlying estimates (the lowest achieved value is almost zero). As a function of T , for each cross-sectional dimension, the relative bias is divided by about two when T is doubled. As a function of n , the relative bias decreases strongly between $n = 2$ and $n = 5$, then very little, and is stable for $n \geq 15$.
4. For the parameters of the matrix M , the average relative bias for the variances is less than 0.1% in absolute value, and less than 1% for the covariances. They are decreasing with n , this being visible more quickly for $T = 1000$ and 2000 than for $T = 500$.

Table 2 shows that the RMSEs of the GARCH and cDCC equations decrease when n and T increase. This is not surprising as an increase in n implies that there is more information to use, while a larger number of observations leads to more efficient estimators. The values of the RMSE are larger for $\bar{\delta}$ and β than for $\bar{\gamma}$ and α , reflecting a larger dispersion of the estimates of the former through the Monte Carlo replications. The same downward trends are observed in the RMSEs reported in Table 3, for the variance and covariance terms of the matrix M .

Table 2: Simulation results for the 3-step QMLE of the GARCH and cDCC parameters

The table reports the relative bias (RB), root mean squared error (RMSE), and standard deviation (MCS) of the 3-step QML estimators of the GARCH and cDCC parameters for T={500, 1000, 2000} and n={2, 5, 7, 10, 15, 25, 50, 100}. The univariate GARCH parameters $\tilde{\gamma}$ and $\tilde{\delta}$ are averaged across n . DGP parameter values are listed in the top line of the table. Reported values are averaged across 500 simulations.

	T = 500				T = 1000				T = 2000			
	$\tilde{\gamma} = 0.05$	$\tilde{\delta} = 0.9$	$\alpha = 0.1$	$\beta = 0.8$	$\tilde{\gamma} = 0.05$	$\tilde{\delta} = 0.9$	$\alpha = 0.1$	$\beta = 0.8$	$\tilde{\gamma} = 0.10$	$\tilde{\delta} = 0.8$	$\alpha = 0.1$	$\beta = 0.8$
	n=2				n=2				n=2			
RB	0.130	-0.114	0.005	-0.055	0.032	-0.040	0.005	-0.023	0.013	-0.011	-0.002	-0.008
RMSE	0.019	0.182	0.032	0.138	0.013	0.099	0.024	0.082	0.009	0.028	0.018	0.043
	n=5				n=5				n=5			
RB	0.131	-0.143	-0.008	-0.016	0.052	-0.048	-0.003	-0.008	0.014	-0.013	-0.001	-0.004
RMSE	0.014	0.171	0.010	0.028	0.009	0.073	0.008	0.019	0.006	0.023	0.005	0.012
	n=7				n=7				n=7			
RB	0.125	-0.135	-0.004	-0.015	0.051	-0.047	-0.004	-0.007	0.013	-0.012	-0.001	-0.004
RMSE	0.012	0.151	0.007	0.021	0.007	0.067	0.005	0.013	0.005	0.020	0.003	0.009
	n=10				n=10				n=10			
RB	0.112	-0.138	-0.009	-0.014	0.042	-0.047	-0.000	-0.007	0.004	-0.010	-0.001	-0.003
RMSE	0.010	0.148	0.005	0.016	0.006	0.060	0.004	0.011	0.004	0.016	0.002	0.006
	n=15				n=15				n=15			
RB	0.109	-0.123	-0.004	-0.014	0.049	-0.047	-0.003	-0.006	0.012	-0.013	-0.002	-0.003
RMSE	0.009	0.124	0.003	0.014	0.005	0.056	0.002	0.007	0.003	0.018	0.001	0.004
	n=25				n=25				n=25			
RB	0.108	-0.122	-0.007	-0.013	0.044	-0.046	-0.002	-0.006	0.011	-0.013	-0.001	-0.003
RMSE	0.007	0.119	0.002	0.011	0.004	0.049	0.001	0.006	0.002	0.016	0.000	0.003
	n=50				n=50				n=50			
RB	0.108	-0.136	-0.007	-0.012	0.044	-0.045	-0.002	-0.006	0.012	-0.013	-0.001	-0.003
RMSE	0.007	0.127	0.002	0.010	0.003	0.045	0.000	0.005	0.052	0.870	0.100	0.803
	n=100				n=100				n=100			
RB	0.126	-0.128	-0.006	-0.013	0.044	-0.047	-0.002	-0.006	0.012	-0.012	-0.001	-0.003
RMSE	0.006	0.106	0.000	0.010	0.002	0.045	0.000	0.005	0.001	0.013	0.000	0.002

Table 3: Simulation results for the targeted scale matrix M

The table reports the relative bias (RB) and root mean squared error (RMSE) of the estimated targeted scale matrix M for T=500, 1000, 2000 and n=2, 5, 7, 10, 15, 25, 50, 100. $RB_{(i,i)}$ and $RMSE_{(i,i)}$ denote averaged values over diagonal terms; $RB_{(i,j)}$ and $RMSE_{(i,j)}$ denote averages over off diagonal terms. The DGP values are such that $M_{(i,i)}=0.1$ and $M_{(i,j)}=0.02$. Reported values are averaged across 500 simulations.

	T=500	T=1000	T=2000	T=500	T=1000	T=2000
	n=2			n=15		
$RB_{(i,i)}$	0.0006	-0.0007	-0.0005	-0.0002	-0.0001	1.20e-05
$RB_{(i,j)}$	0.0038	-0.0039	-0.0016	0.0009	-0.0022	-0.0007
$RMS E_{(i,i)}$	0.0060	0.0045	0.0031	0.0023	0.0016	0.0012
$RMS E_{(i,j)}$	0.0042	0.0033	0.0023	0.0017	0.0012	0.0008
	n=5			n=25		
$RB_{(i,i)}$	0.0008	-0.0005	-0.0002	0.0001	4.50e-05	-5.97e-05
$RB_{(i,j)}$	0.0069	0.0036	0.0005	0.0003	-0.0006	-0.0002
$RMS E_{(i,i)}$	0.0040	0.0029	0.0020	0.0018	0.0013	0.0008
$RMS E_{(i,j)}$	0.0029	0.0020	0.0015	0.0013	0.0009	0.0006
	n=7			n=50		
$RB_{(i,i)}$	-0.0002	-0.0005	-0.0002	-4.88e-05	-2.76e-05	-4.48e-05
$RB_{(i,j)}$	-0.0020	-0.0027	0.0016	-0.0001	0.0002	-8.63e-05
$RMS E_{(i,i)}$	0.0034	0.0024	0.0017	0.0012	0.0008	0.0006
$RMS E_{(i,j)}$	0.0024	0.0017	0.0012	0.0009	0.0006	0.0004
	n=10			n=100		
$RB_{(i,i)}$	0.0002	-0.0004	-0.0002	-7.16e-05	-8.71e-05	-6.13e-05
$RB_{(i,j)}$	0.0014	-0.0020	-0.0003	-0.0002	-0.0002	-0.0002
$RMS E_{(i,i)}$	0.0028	0.0020	0.0014	0.0008	0.0006	0.0004
$RMS E_{(i,j)}$	0.0020	0.0014	0.0010	0.0006	0.0004	0.0003

4.2. Comparison of QML estimators in 15-dimensional model

Of interest is to analyze the relative performance of the 3QMLE against the other two approaches in medium-low dimensional settings. Table 4 presents Monte Carlo results up to dimension 15, where entries are pairwise ratios of the RB (in absolute value) and RMSE statistics achieved by each estimator. Values smaller than unity indicate that the estimator at the numerator is less biased or more precise. Conceivably, we expect the differences between the three estimators to be minimal when only two assets are considered, and to increase with the number of series in the cross-section. The use of the 2QMLE, which can be considered as an intermediate approach between the other two, serves the purpose to disentangle the effects of targeting and multi-step estimation. Of course, 2QMLE and 3QMLE do not differ in the estimation of

Table 4: Monte Carlo comparison results: GARCH and cDCC parameters

The table reports ratios of RB and RMSE values for 1QML, 2QML, 3QML estimators. RB and RMSE are computed as described in Table 2.

		$T = 500$				$T = 1000$				$T = 2000$			
		$\bar{\gamma}$	δ	α	β	$\bar{\gamma}$	δ	α	β	$\bar{\gamma}$	δ	α	β
		n=2											
RB	2QML/1QML	1.076	1.061	0.721	1.001	1.131	1.118	0.696	0.967	1.043	1.022	1.496	0.991
	3QML/1QML	1.028	1.034	0.433	0.982	1.048	1.096	0.687	1.038	0.936	1.019	2.240	0.966
	3QML/2QML	0.956	0.975	0.600	0.981	0.927	0.980	0.988	1.073	0.898	0.997	1.497	0.975
RMSE	2QML/1QML	1.014	1.048	0.992	1.000	0.994	1.084	0.996	0.989	0.997	1.005	0.999	1.000
	3QML/1QML	1.007	1.040	0.994	0.999	1.007	1.067	0.998	1.107	1.009	1.020	0.998	0.998
	3QML/2QML	0.993	0.992	1.002	1.000	1.013	0.984	1.002	1.119	1.012	1.015	0.999	0.998
		n=5											
RB	2QML/1QML	0.997	0.985	1.436	0.980	1.012	1.004	1.688	0.965	0.961	0.980	2.435	0.966
	3QML/1QML	0.926	0.964	1.654	0.961	0.973	1.018	1.986	0.947	1.005	1.049	3.794	0.951
	3QML/2QML	0.929	0.979	1.152	0.980	0.962	1.014	1.177	0.981	1.046	1.071	1.558	0.985
RMSE	2QML/1QML	0.996	0.988	0.998	0.994	1.004	1.019	1.000	0.996	0.994	0.975	0.999	0.997
	3QML/1QML	0.980	0.996	0.998	0.989	1.006	1.023	1.000	0.993	1.015	1.042	0.999	0.996
	3QML/2QML	0.985	1.008	1.000	0.995	1.002	1.004	1.000	0.998	1.021	1.069	1.000	0.999
		n=7											
RB	2QML/1QML	0.247	0.475	0.315	0.787	0.142	0.324	0.357	0.756	0.048	0.174	0.015	0.673
	3QML/1QML	0.231	0.443	0.380	0.779	0.153	0.345	0.420	0.748	0.042	0.173	0.013	0.665
	3QML/2QML	0.936	0.934	1.206	0.989	1.077	1.062	1.176	0.989	0.885	0.995	0.894	0.988
RMSE	2QML/1QML	0.400	0.547	0.944	0.881	0.397	0.464	0.969	0.912	0.348	0.311	0.968	0.919
	3QML/1QML	0.401	0.514	0.945	0.878	0.405	0.482	0.969	0.910	0.357	0.304	0.968	0.918
	3QML/2QML	1.002	0.940	1.002	0.997	1.019	1.038	1.000	0.998	1.025	0.980	1.000	0.999
		n=10											
RB	2QML/1QML	0.146	0.491	0.690	0.746	0.076	0.280	0.008	0.720	0.005	0.078	0.089	0.655
	3QML/1QML	0.128	0.441	0.727	0.738	0.071	0.285	0.070	0.712	0.090	0.090	0.111	0.650
	3QML/2QML	0.876	0.899	1.054	0.989	0.924	1.017	0.777	0.989	1.512	1.149	1.246	0.991
RMSE	2QML/1QML	0.257	0.565	0.940	0.849	0.219	0.405	0.916	0.874	0.150	0.140	0.874	0.870
	3QML/1QML	0.243	0.521	0.940	0.844	0.212	0.394	0.915	0.870	0.155	0.158	0.874	0.869
	3QML/2QML	0.944	0.921	1.000	0.994	0.968	0.973	0.999	0.995	1.037	1.130	1.000	0.998
		n=15											
RB	2QML/1QML	0.174	1.148	0.419	0.864	0.080	0.463	0.303	0.886	0.023	0.129	0.032	0.780
	3QML/1QML	0.124	0.857	0.451	0.853	0.055	0.400	0.339	0.872	0.014	0.118	0.051	0.771
	3QML/2QML	0.709	0.789	1.076	0.987	0.691	0.864	1.119	0.984	0.629	0.912	1.591	0.989
RMSE	2QML/1QML	0.233	1.262	0.617	0.609	0.143	0.599	0.910	0.934	0.072	0.194	0.880	0.899
	3QML/1QML	0.198	1.027	0.617	0.603	0.119	0.523	0.911	0.926	0.071	0.180	0.881	0.895
	3QML/2QML	0.850	0.814	1.001	0.991	0.830	0.873	1.001	0.992	0.989	0.930	1.000	0.996

Table 5: Monte Carlo comparison results: M matrix

The table reports ratios of Frobenius norms of the 1QML estimator of M and of the targeting estimator of M (used for 2QML and 3QML estimations).

Frobenius norm defined as $FN(\hat{M}) = \sqrt{\sum_{i,j=1}^n |M_{i,j} - \hat{M}_{i,j}|^2}$, where $M_{i,j}$ denotes the true value (see DGP in Table 1) and $\hat{M}_{i,j}$ is the element of the estimated M matrix, obtained either by 1-step estimation (1QML) or by the targeting approach. Values below (above) 1 are in favour of 1QMLE (targeting). Results based on 500 simulations.

n	$T = 500$	$T = 1000$	$T = 2000$
2	0.988	0.946	0.986
5	0.992	0.971	0.967
7	0.999	0.990	0.993
10	1.198	1.072	1.025
15	1.787	1.408	1.249

the constant matrix M which is done by targeting in the same way for both estimators.

The most noticeable finding emerges from the comparison of the biases of the estimators of the univariate GARCH parameters γ and δ . The table reveals that the 2QMLE is generally very close to the 3QMLE, while the 1QMLE estimates appear much more biased as n increases. For example, when $n=15$ and $T=2000$, the average biases on γ and δ estimated via 3QML are respectively only 1.4% and 11.8% of those achieved via 1QML, and 63% and 91% of those obtained via 2QML. Similarly, the RMSE ratios show that the performance of 1QMLE compared to the others deteriorates when n increases. This is spectacular as soon as n exceeds 5. The effect of this distortion is also visible for the correlation parameters, but much lessened.

Our explanation of these results is that the 1QMLEs of the GARCH and cDCC parameter are affected by the increasing number of model parameters due to the dimension of the matrix M : for example, when $n = 15$ the 1QML optimizes the likelihood function over 152 parameters (120 for M , 32 for the rest), while 2QML optimizes over 32 parameters, and 3QML achieves this by steps involving two parameters at a time after targeting. As the ratios reported in Table 4 between the 2QMLE and 3QMLE are close to unity, it is clear that the main source of burden for 1QMLE comes from the estimation of the parameters of the constant matrix M . By looking at Table 5, it is easy to realize that while in the lowest dimensions (up to $n = 7$) the

1QMLE delivers slightly more accurate estimates of M than the targeting estimator used in 3QMLE and 2QMLE, from $n > 7$ there is a clear countertrend in favor of the targeting estimator.

What is more surprising from Tables 4 is that, for fixed n , the relative discrepancy does not decrease with larger T . By inspection of Table B.11 in Appendix B, it appears that RB and RMSE for 1QMLE decrease with T , but at a slower rate than for the competing estimators (at $n=15$, $RB(\bar{\gamma})$ implied by the 1QMLE is reduced by 5.8 percent from $T=500$ (0.933) to $T=1000$ (0.879) and increases slightly from $T=1000$ to $T=2000$ (0.883), as opposed to the $RB(\bar{\gamma})$ of the 3QMLE which is reduced by 61 percent in the first case (from 0.127 to 0.049) and by 73 percent in the second (from 0.049 to 0.013).

This suggests that some improvements in the 1QMLE relative performance could be achieved by increasing further the number of observations in the time series; this is in accordance with the strand of literature claiming that a multistep estimator can not be asymptotically more efficient than the standard one-step estimator; see Pedersen and Rahbek (2014) for the BEKK-GARCH model, Francq et al. (2014) in the CCC-GARCH case and Francq et al. (2011) for an overview in the univariate GARCH framework.

To summarize, the simulation exercise strongly supports the use of the proposed three-step estimator in practical applications, especially of large dimension. The general message is that, under the assumption of a scalar Re-cDCC DGP, the estimator has good finite sample properties, it works no worse than the two-step approach and it favourably compares to the standard one-step QMLE, even when both are feasible.

5. Empirical analysis

This section illustrates the results of two empirical applications on real data. First, we analyse the main features of the Re-cDCC model on a bivariate dataset including two stocks. Next, we apply the model to more heterogeneous high dimensional data and investigate its usefulness for one-step-ahead forecasting and minimum variance portfolio allocation.

In both exercises, we consider the two following competing models:

- the Conditional Autoregressive Wishart (CAW) model of Golosnoy et al. (2012), where the daily covariance process is driven by a BEKK-type recursion. Assuming a full parametrization and only one lag it is defined as Eq.(1) with the conditional expectation of the Wishart distribution parameterized as $S_t = CC' + AC_{t-1}A' + BS_{t-1}B'$, with $\{C, A, B\}$ $n \times n$ matrices of parameters, C being lower triangular;
- the Rotated DCC (RDCC) model of Noureldin et al. (2014) which belongs to the rotated ARCH class of models where parameters of interest are obtained by fitting a DCC model of the time varying covariance to rotated daily returns having as long-run covariance an identity matrix.

In the second application, we also consider the Re-cDECO model previously illustrated and an Exponentially Weighted Moving Average (EWMA) model is applied to the series of realized covariance matrices, *i.e.* $S_t = (1 - \lambda)C_{t-1} + \lambda S_{t-1}$, with the parameter λ parameter is set equal to 0.94.

5.1. Bivariate dataset AA-XOM

The dataset comprises the series of daily realized covariance matrices of the Alcoa (AA) and Exxon Mobil (XOM) stocks, traded on the NYSE, estimated from five minute intraday returns over the period February 1, 2001 to December 31, 2009.³ Descriptive statistics are shown in Table 6. For fitting the RDCC model, we use the associated series of open-to-close daily returns, as provided by Noureldin et al. (2012).⁴

For the models featuring a DCC component we adopt the three-stage estimation approach explained in Section 3 and fit scalar (S), diagonal (D) and Hadamard (H) specifications; in the case of the CAW, which is based on a BEKK-type recursion, estimation is done in two steps (covariance targeting is applied in the first step and then the remaining parameters are estimated via QML) and a full (F) specification is fitted in place of the Hadamard. In all cases, estimation is performed while imposing the stationarity constraints. For the F(ull)-CAW, where the matrices $\{A, B\}$ are unconstrained, *i.e.* not necessarily symmetric, covariance stationarity is achieved by imposing that the eigenvalues of the matrix $\{A + B\}$ are less than unity in modulus and A_{11}, B_{11} are restricted to be positive.

³The same dataset has been used in the paper by Noureldin et al. (2014).

⁴Dataset downloaded from <http://realized.oxford-man.ox.ac.uk>.

Table 6: Descriptive statistics for the Alcoa (AA) and Exxon Mobil (XOM) realized variances

The table reports descriptive statistics of daily realized (co)variances and correlation for the AA and XOM stocks over the period February 1, 2001 to December 31, 2009 (2242 observations).

Stock		Mean	Max.	Min.	Std.dev.	Skewness	Kurtosis
AA	(var)	4.94	160.24	0.29	8.93	7.63	92.124
XOM	(var)	2.07	115.37	0.12	4.15	13.29	288.06
AA-XOM	(cov)	1.36	71.88	-0.36	3.69	9.89	140.85
AA-XOM	(corr)	0.33	0.84	-0.26	0.17	-0.05	2.71

Parameter estimates and robust standard errors over the full sample period from the CAW, Re-cDCC and RDCC models are reported in Table 7 along with the corresponding maximized loglikelihood and AIC values. For the RDCC, the reported likelihood values refer to the unrotated returns evaluated at the estimated parameter values.

The upper panel of the table provides the parameter estimates of the univariate GARCH(1,1) processes and the correlation (or covariance) process of the scalar specifications. At first, we notice an important difference between the GARCH estimates of the S-RDCC and the S-Re-cDCC models: the γ_i estimates are much larger in the S-Re-cDCC model, and the δ_i estimates are much lower, than in the S-RDCC model. Thus, the realized conditional variances are more responsive to their own shocks than the GARCH variances based on daily returns, while in the S-RDCC they are smoother. This is a consequence of the different information used by the models. Hansen et al. (2012) document a similar effect when comparing estimates from standard GARCH and Realized GARCH models. They attribute the discrepancy in the value of the shock response coefficient, the so called ARCH term, to the highly noisy nature of the squared returns as an estimator of the latent volatility which prevents from correctly identifying the signal coming from the latent volatility measure. In both cases, the sum of the γ_i and δ_i estimated parameters is always very close to one, with the S-RDCC model being the most persistent.

This type of difference is also found for the α and β parameters of the correlation process⁵, but with a slightly lower level of persistence. Note that the estimates obtained for the S-CAW model drive the dynamics of all elements of the conditional covariance matrix and hence they are much more comparable to the estimates of the univariate GARCH processes than to those of the DCC correlation processes.

Since we perform QML estimation of the models relying on the Wishart assumption, likelihood values cannot be compared with those of the RDCC (which use a Gaussian QL), but only with models belonging to the same class. The S-Re-cDCC delivers the best fit in terms of log-likelihood value, but the AIC criterion leads to select the S-CAW as the preferred model, revealing that the four additional parameters of the former are not relevant for this dataset.

The diagonal and full specifications do not much improve the fit with respect to the scalar versions. The H-RDCC slightly improves the corresponding scalar and diagonal types, whereas for the other models there seems to be no additional benefit from enriching the correlation (or covariance) dynamics. This may be a consequence of the minimal cross-sectional dimension of the dataset, which only demands one conditional correlation (or covariance) to be fitted.

Finally, the Panel C of Table 7 shows the p -values of Ljung-Box Q(10) tests for the null hypothesis of absence of serial correlation of order ten in the standardized residuals, computed as $\epsilon_t^* = S_t^{-1/2} C_t (S_t^{-1/2})'$, for $t = 1, \dots, T$. These results reveal that the scalar and diagonal specifications of the Re-cDCC model clearly dominate the CAW models by successfully capturing the dynamics in the variances of the stocks, and their covariance.

In order to compare the forecasting performance of the models, we perform a one-step-ahead prediction exercise using the diagonal specification of each model. We recursively estimate the parameters on the first 1542 daily observations of the sample (that is using the data from February 1, 2001 until March 26, 2007) and use the last 700 as the forecast sample. We compare the forecasting accuracy of the models using the QLIK loss function, which is often used in the presence of noisy proxies; see Patton and Sheppard (2009) and Patton (2011). Specifically, if S_t^a denotes the time- t conditional forecast from model a , the corresponding QLIK loss function is defined as

$$\ell_t^a = \log|S_t^a| + (S_t^a)^{-1} \text{tr}(C_t), \quad (29)$$

⁵Analogously to what Noureldin et al. (2014) report, the α parameter in the S-RDCC is not significantly different from zero. Noureldin et al. (2014) test for constant conditional correlation in the scalar case imposing $\alpha = \beta = 0$ and find that the null cannot be rejected.

Table 7: Full sample estimates for the the Alcoa and Exxon Mobil dataset

Panels A and B show the parameter estimates, except constant terms, computed with the full sample: 1/2/2001 – 31/12/2009, of the scalar (S), diagonal (D), Hadamard (H) Re-cDCC and RDCC models, and the S, D, and F(ull) CAW. Robust standard errors are in brackets in Panel A. Panel B (where the superscript * denotes significance at the 1% level) reports estimated parameter matrices for the non-scalar specifications. Panel C provides the p -values of the Ljung-Box Q(10) test statistics for serial correlation of order ten in the standardized residuals.

PANEL A: SCALAR PARAMETER ESTIMATES									
	CAW			Re-cDCC			RDCC		
	S	D	F	S	D	H	S	D	H
γ_1				0.376 (0.032)	0.376 (0.032)	0.376 (0.032)	0.084 (0.017)	0.084 (0.017)	0.084 (0.017)
γ_2				0.298 (0.031)	0.298 (0.031)	0.298 (0.031)	0.054 (0.012)	0.054 (0.012)	0.054 (0.012)
δ_1				0.603 (0.034)	0.603 (0.034)	0.603 (0.034)	0.902 (0.015)	0.902 (0.015)	0.902 (0.015)
δ_2				0.684 (0.035)	0.684 (0.035)	0.684 (0.035)	0.942 (0.012)	0.942 (0.012)	0.942 (0.012)
α	0.314 (0.022)			0.170 (0.033)			0.023 (0.014)		
β	0.669 (0.024)			0.807 (0.043)			0.945 (0.044)		
Total LL	-3812	-3811	-3811	-3809	-3809	-3809	-7844	-7843	-7842
AIC	7627	7630	7636	7630	7634	7639	15704	15707	15708
Num. of parameters	2	4	8	6	8	10	8	10	12

PANEL B: ESTIMATED PARAMETER MATRICES						
	D-CAW		D-Re-cDCC		D-RDCC	
	A	0.343*	0.276*	0.119*	0.171*	0.008
B	0.639*	0.710*	0.871*	0.829*	0.956*	0.932*

	F-CAW		H-Re-cDCC		H-RDCC	
	A	0.830*	0.060*	0.181*	0.139*	0.011
	-0.150*	0.519*	0.139*	0.163	0.027	0.043
B	0.028*	0.067*	0.818*	0.860*	0.986*	0.970*
	-0.328*	-0.703*	0.860*	0.265	0.970*	0.949*

PANEL C: P-VALUES FOR THE LJUNG-BOX TEST FOR THE RESIDUALS, 10 LAGS			
	ϵ_{11}^*	ϵ_{22}^*	ϵ_{12}^*
S-CAW	0.01	0.28	0.00
D-CAW	0.08	0.14	0.00
F-CAW	0.00	0.02	0.00
S-Re-cDCC	0.25	0.28	0.22
D-Re-cDCC	0.25	0.30	0.16
H-Re-cDCC	0.02	0.34	0.06

where C_t is the the 5-minute realized covariance estimator chosen as proxy for the latent covariance matrix.⁶ To compare two models $\{a, b\}$ over the out-of-sample period τ , we look at the average QLIK loss difference

$$L_{a,b} = \frac{1}{\tau} \sum_{t=h+1}^{\tau} \ell_{a,b,t}, \quad \ell_{a,b,t} = \ell_t^a - \ell_t^b \quad (30)$$

and then we test the null hypothesis of equal predictive accuracy via the Diebold and Mariano (2012) test, where the Newey-West covariance estimator is used to compute the long run variance of $L_{a,b}$. The alternative is one-sided, so that positive (negative) values of $L_{a,b}$ correspond to superior performance of model b (a).

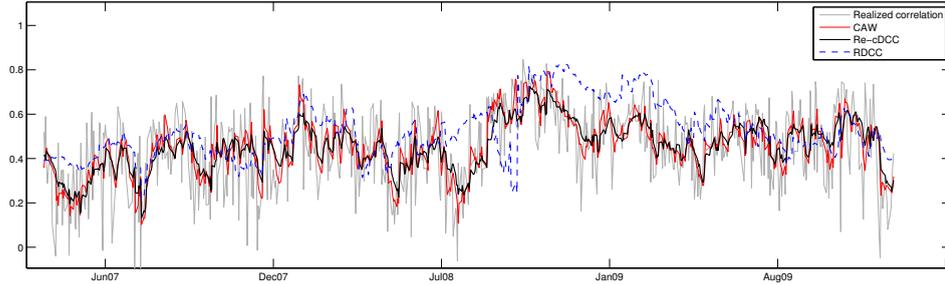
Comparing the D-Re-cDCC model (taken as model a) to the D-CAW and the D-RDCC, we get t -statistics of -2.86 and -6.68 , respectively, which are statistically significant at 1% and indicate that the D-Re-cDCC model strongly outperforms both competitors. Despite its higher number of parameters to be estimated than in the D-CAW model, it seems that using the D-Re-cDCC one for out-of-sample forecasting pays off by providing superior one-step-ahead predictions.

Figure 1 plots the predicted conditional correlations from the diagonal models along with the selected proxy. Not surprisingly, the assets are highly correlated across the whole period, with the highest peaks

⁶Any estimator that fulfils the unbiasedness condition can be used as proxy.

reached during the more volatile months of the 2008 financial crisis. The Re-cDCC and CAW correlations generally exhibit a similar pattern that fits the realized correlation much better than the D-RDCC, which is not really surprising since the latter are based on daily returns only. The D-RDCC predicted correlations also appear to be more persistent than the predicted correlations obtained from the models that use the realized measures.

Figure 1: Predicted conditional correlations of XOM-AA from diagonal CAW, Re-cDCC and RDCC models (27/03/2007-31/12/2009).



5.2. High dimensional datasets

In this subsection we push the analysis to a higher dimension. In particular, we are interested in the model forecasting ability and consequently, in its potential usefulness for portfolio allocation strategies. We conduct the same empirical application on two large-scale datasets, so as to investigate and compare the responsiveness of the model to different realized covariance estimators across different time periods. The main characteristics of the two datasets, hereafter referred to as Dataset A and Dataset B, are summarized below.

- Dataset A pertains to 46 stocks traded in the NYSE and NASDAQ. Their tickers and issue names are shown in the upper panel of Table B.12. The sample runs from January 5, 1999 to December 31, 2008, for a total of 2483 trading days. The dataset has been cleaned from weekends, holidays and early closing days. Also, days with several consecutive missing values or constant prices have been removed and rare missing values linearly interpolated. The realized covariance matrices are constructed using the *Realized Covariance* estimator (see [Barndorff-Nielsen and Shephard \(2004\)](#)) based on intraday returns computed from six-minute intervals last mid-quotes. Since the trading period of the NYSE and NASDAQ is 6.5 hours per day, 65 intraday observations are available per day. Relying on the arguments of [Andersen et al. \(2010\)](#), we use open-to-close realized covariances.
- Dataset B pertains to 50 stocks among the largest U.S. financial institutions, with three assets (AXP, BAC, WFC) in common with Dataset A. Tickers are given in the lower panel of Table B.12. The trading period runs from January 3, 2007 till December 13, 2012 for a total of 1499 daily observations. Data have been cleaned using the step-by-step cleaning procedure of [Barndorff-Nielsen et al. \(2009\)](#) whereby overnight returns have been excluded to favor estimation accuracy. The daily realized covariance matrices have been computed using the *Cholcov* estimator of [Boudt et al. \(2014\)](#), which is guaranteed to be positive semidefinite in the presence of asynchronicity and illiquidity.

The upper panel of Figure 2 plots the series of daily returns of the AXP stock over the two sample periods. The financial crisis is included in both datasets but located at the opposite extremes of the samples, namely at the very end of Dataset A and at the beginning of Dataset B. This allows us to use a different scenario in our forecasting exercises and therefore constitutes an important test for assessing the robustness of the models to varying market conditions.

Given the large number of parameters to be estimated for non-scalar models in the considered dimensions ($n = 46$ and 50), we use scalar specifications. Full sample parameter estimates for the scalar CAW, Re-cDCC, Re-cDECO and RDCC models are reported in Table 8. The main difference with respect to the estimates in Table 7 occurs for the estimated values of the α parameter across models: it is found to be much smaller than in the case of two assets. While in the MGARCH literature, as mentioned, composite likelihood has been found to alleviate this issue, whether this applies as well to realized models is still an open question. Nevertheless, in the latter case the effect seems somehow attenuated.

Figure 2: Daily returns (upper panel) and predicted conditional volatilities (lower panel) of the AXP stock for Dataset A (left column) and Dataset B (right column) over the two sample periods analyzed.

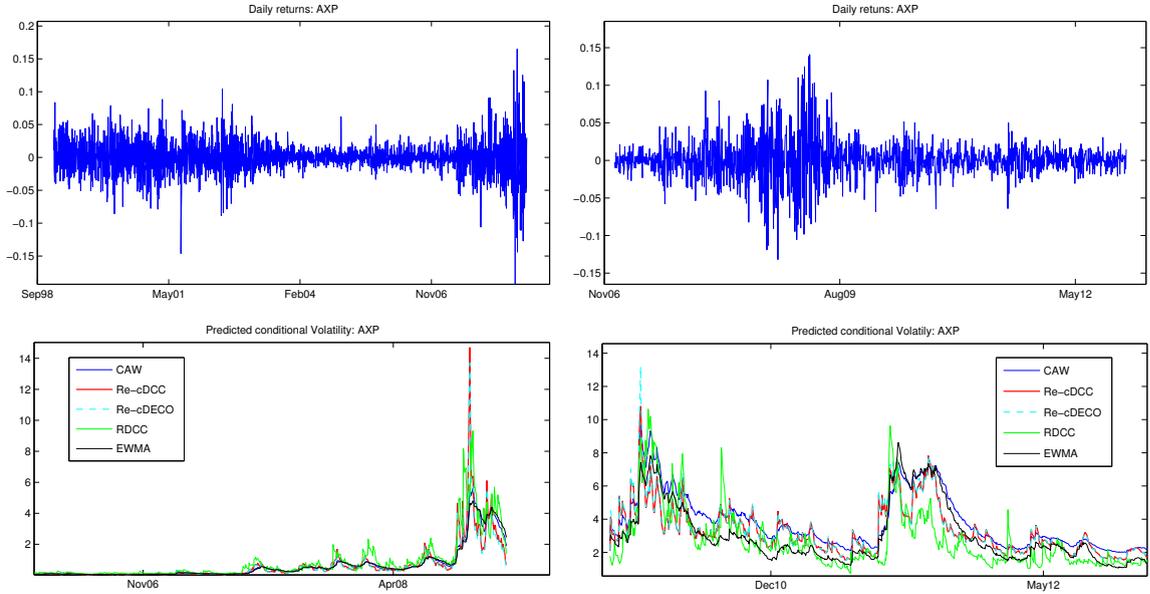


Table 8: Full sample parameter estimates

The table shows full sample estimates (except constant terms) and robust standard errors in brackets for each dataset; $\bar{\gamma}$ and $\bar{\delta}$ are the means of the estimates of the γ_i and δ_i parameters for the different stocks (with the corresponding averages of the standard errors in brackets). Dataset A contains 46 equities from the NYSE and NASDAQ while Dataset B contains stocks of 50 among US largest institutions.

	S-CAW	S-Re-cDCC	S-Re-cDECO	S-RDCC
DATASET A: JANUARY 5, 1999 – DECEMBER 31, 2008				
$\bar{\gamma}$	-	0.348 (0.050)	0.348 (0.050)	0.110 (0.025)
$\bar{\delta}$	-	0.635 (0.054)	0.635 (0.054)	0.868 (0.028)
α	0.091 (0.005)	0.015 (0.000)	0.116 (0.028)	0.003 (0.000)
β	0.904 (0.006)	0.982 (0.001)	0.864 (0.034)	0.979 (0.004)
Total LL	433866	435747	434744	292223
AIC	-867728	-871307	-869300	-584166
Num. of parameters	2	94	94	140
DATASET B: JANUARY 3, 2007 – DECEMBER 13, 2012				
$\bar{\gamma}$	-	0.194 (0.053)	0.194 (0.053)	0.124 (0.023)
$\bar{\delta}$	-	0.794 (0.059)	0.794 (0.059)	0.871 (0.021)
α	0.055 (0.004)	0.010 (0.000)	0.041 (0.012)	0.004 (0.001)
β	0.940 (0.005)	0.978 (0.005)	0.936 (0.023)	0.924 (0.013)
Total LL	-81846	-78718	-78724	-128500
AIC	163695	157639	157652	257310
Num. of parameters	2	102	102	152

As a general finding, for both datasets, the S-Re-cDCC model better fits the data than its competitors using the same data, followed by the S-Re-cDECO model. Not surprisingly, the S-CAW is too restrictive for a large number of assets, since all variance processes are constrained to share the same parameters.

To determine whether the better in-sample performance of the S-Re-cDCC goes along with forecasting gains, we perform an out-of-sample one-step-ahead prediction exercise for each dataset. The parameters of each model are estimated on a fixed rolling window and kept fixed to produce forecasts over the following ten days, which usually corresponds to two weeks of trading. Then the estimation window is shifted forward by ten observations and the models are re-estimated. This procedure is repeated until all data have been

Table 9: MCS results

The table reports 90% and 75% MCS p -values for the two datasets based on 700 one-step-ahead conditional covariance matrix predictions. Models included in the MCS have their p -values highlighted in bold.

	S-CAW	S-Re-cDCC	S-Re-cDECO	S-RDCC	EWMA
DATASET A: CRISIS PERIOD (7/2/2006–31/12/2008)					
Panel A: 90% MCS					
sfrob	0.530	0.333	0.530	0.024	1.000
stein	0.000	1.000	0.000	0.000	0.000
Qlik	0.000	1.000	0.000	0.000	0.000
VND	1.000	0.110	0.000	0.000	0.000
Panel B: 75% MCS					
sfrob	0.537	0.343	0.537	0.025	1.000
stein	0.000	1.000	0.000	0.000	0.000
Qlik	0.000	1.000	0.000	0.000	0.000
VND	0.000	1.000	0.000	0.106	0.000
DATASET B: POST-CRISIS PERIOD (8/3/2010–13/12/2012)					
Panel A: 90% MCS					
sfrob	0.000	0.772	1.000	0.000	0.772
stein	0.000	1.000	0.000	0.000	0.000
Qlik	0.000	1.000	0.000	0.000	0.000
VND	0.000	1.000	0.935	0.000	0.000
Panel B: 75% MCS					
sfrob	0.000	0.780	1.000	0.000	0.780
stein	0.000	1.000	0.000	0.000	0.000
Qlik	0.000	1.000	0.000	0.000	0.000
VND	0.000	1.000	0.934	0.001	0.000

used and a set of 700 predicted covariance matrices is obtained. Predicted volatilities for the AXP stock are reproduced in the lower panels of Figure 2.

The first comparison of the different models is done using the Model Confidence Set (MCS) procedure of Hansen et al. (2011). This method allows us to compare the forecasts of the models based on a loss function without having to choose a reference model and leads to a subset of models that cannot be discriminated at a selected confidence level. Given that five models are in competition, this approach is preferable to performing pairwise DM tests of equal predictive accuracy. In order to make the results robust to the choice of the loss functions, in addition to the QLIK, we use three additional consistent loss functions (formulas are given in Table B.13 of Appendix B) and set the rejection levels at 10 and 25%, such that models are removed until the final model set includes the best model(s) with confidence levels of 90 and 75%. The resulting MCS p -values are given in Table 9. The results are similar for both datasets: the S-Re-cDCC model is included in all MCS, and is the single one at both levels in each dataset for the Stein and QLIK loss functions. For VND, a second model is in the MCS in three cases, and for the sfrob, only the S-RDCC is excluded. Actually, the latter is excluded from all MCS.

In order to compare the forecasting performances of the five models in an economically meaningful way, we also apply them to the problem of optimal portfolio allocation. We analyze the portfolio gains obtained by allocating the portfolios based on the S-Re-cDCC covariance forecasts instead of the forecasts from the chosen benchmark models. The advantage of studying global minimum variance (GMV) portfolios (in contrast to minimum variance portfolios for a given expected return) is that the changes in the ex ante optimal portfolio weights are solely determined by the forecasts of the conditional covariance matrix over the given investment horizon. The GMV portfolio weights w_t are the solution to the problem

$$\min_w \quad \frac{1}{2} w_t' S_t w_t, \quad s.t. \quad \sum_{i=1}^n w_{t,i} = 1$$

where w_t is the vector of portfolio weights for time t chosen at time $t - 1$, S_t denotes the conditional covariance forecast from a generic model, and the only requirement is on the vector of weights to add up to unity. We further impose the portfolio weights to be nonnegative, thus ruling out short-selling operations.

We measure the portfolio performance in two ways. First, we compute the out-of-sample portfolio gross returns, defined as $r_t^P = \sum_{i=1}^n w_{t,i} r_{t,i}$, with $w_{t,i}$ and $r_{t,i}$ respectively denoting the weight and the return on asset i on that day. From the return series, we report the corresponding mean and standard deviation over the forecast period. Furthermore, we quantify the average portfolio concentration of the resulting GMV

Table 10: **GMV portfolio results**

The table reports GMV portfolio summary statistics for the two datasets based on 700 one-step-ahead conditional covariance matrix predictions; μ_p stands for the mean and σ_p for the standard deviations of the gross returns of the portfolio. The best model is accompanied by * or ** if its standard deviation is different than the one of the next best model at the 5% or 1% (two-sided) level, respectively. CO is defined in (31), Turnover in (32).

	S-CAW	S-Re-cDCC	S-Re-cDECO	S-RDCC	EWMA
DATASET A: CRISIS PERIOD (7/2/2006–31/12/2008)					
<i>Gross return performance</i>					
$\mu_p(e^{-03})$	-0.0774	-0.0322	-0.0650	-0.1309	-0.0772
σ_p	0.0067	0.0061*	0.0063	0.0068	0.0065
CO	0.3612	0.3338	0.3594	0.3591	0.3737
<i>Turnover</i>					
Mean	0.1203	0.1550	0.1678	0.2732	0.0984
Median	0.0907	0.1273	0.1379	0.2395	0.0703
Standard deviation	0.0862	0.0919	0.0957	0.1571	0.0826
Kurtosis	24.2441	8.1248	9.3996	10.8337	29.9466
DATASET B : POST-CRISIS PERIOD (8/3/2010–13/12/2012)					
<i>Gross return performance</i>					
μ_p	0.0323	0.0329	0.0361	0.0133	0.0315
σ_p	0.0012	0.0012	0.0013	0.0017	0.0010**
CO	0.2147	0.2118	0.2242	0.5334	0.2471
<i>Turnover</i>					
Mean	0.2321	0.2038	0.3268	0.3033	0.2887
Median	0.2011	0.1871	0.2549	0.2356	0.2648
Standard deviation	0.1280	0.0981	0.2296	0.2174	0.1293
Kurtosis	9.2625	6.1505	4.6816	8.0879	8.5586

portfolio weights, as large values of this measure can imply too little diversification. Concentration (CO) is measured by the Euclidean norm of the vector of portfolio weights, and averaged over the forecast period:

$$CO = \frac{1}{T} \sum_{t=1}^T \|w_t\|_2 = \frac{1}{T} \sum_{t=1}^T \left(\sum_{i=1}^n w_{t,i}^2 \right)^{1/2} \quad (31)$$

At each date, the measure is minimized for an equally weighted portfolio, i.e. for $w_t = (1/n)\mathbf{1}$.

Second, following de Pooter et al. (2008) and Boudt et al. (2013), we evaluate the portfolio turnover rates (for the forecast period), which can be used to assess the stability of the covariance matrix forecasts. In the moment just before rebalancing to the next period, denoted as t^- , the actual weight on each asset in the portfolio changes to $w_{i,t^-} = w_{i,t-1} \frac{1+r_{i,t}}{1+r_t^p}$, such that the total change in weights required at time t is equal to

$$TO_t = \sum_{i=1}^n |w_{t,i} - w_{i,t^-}|. \quad (32)$$

In order to compare the turnover rates generated from the different models, we report the mean, standard deviation and kurtosis of the resulting TO_t series. The results are collected in Table 10.

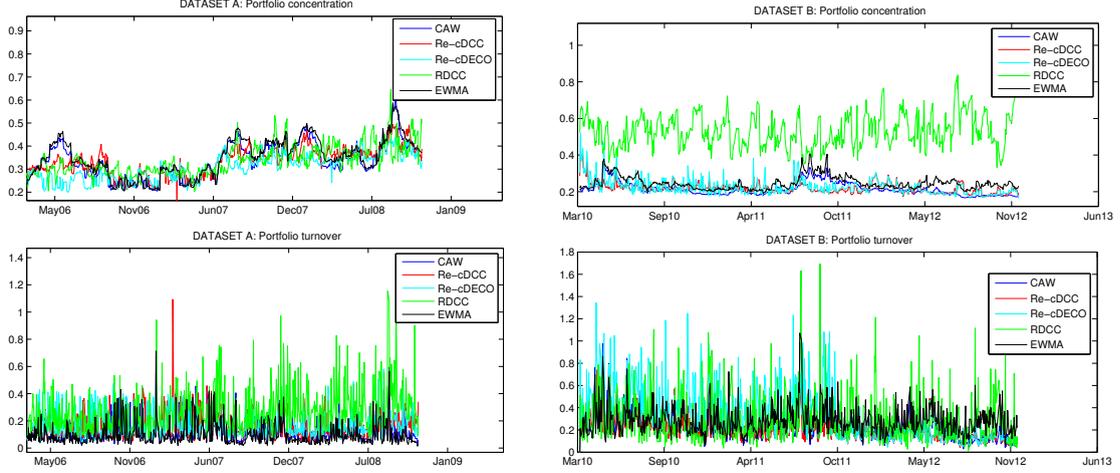
The differences between the two time periods strongly emerge by comparing the portfolio gross return performances achieved by the models. Indeed, the impact of the 2008 financial crisis is evident in the negative average portfolio returns and higher standard deviations in the upper panel of the table, in contrast to the positive returns predicted for the calmer period of 2010-2012, which is also characterized by lower average values of standard deviations.

In terms of gross return performance, the calm period shows no clear evidence of dominance of one specific model over the other, except that the S-RDCC is generating a lower mean return, a higher standard deviation and a much higher concentration than the other models, as also evident from the upper right panel of Figure 3. As for the turnover rates, the S-Re-cDCC compares favorably to the competitors by delivering on average more stable portfolio weights and thus lower portfolio churn, which is desirable if the transaction cost level is sizeable.

The benefits of the S-Re-cDCC allocation are more visible during the turbulent crisis period, where it generates the largest mean return and the lowest standard deviation yet maintaining the smallest concentration. The S-Re-cDECO model is less rewarding in terms of mean return but comparable on the other two measures, while the S-RDCC model performs quite poorly in terms of mean return. On the other hand, the allocations delivered by the S-CAW and the EWMA models increase volatility but lower portfolio turnover.

However, kurtosis values are largely in favor of the S-Re-cDCC, with percentage differences ranging between 16 and 268%. This denotes more stability in the S-Re-cDCC conditional covariance matrices, and overall, a higher responsiveness to the changing economic environment.

Figure 3: Daily portfolio concentration and turnover for Dataset A (left side) and Dataset B (right side).



6. Conclusions

This paper proposes a dynamic class of models for realized covariance matrices which combine flexibility and computational tractability that make them applicable to a large number of assets. Within the different specifications nested in the proposed class, we focused on the so-called Re-cDCC model and explained in details its several advantages. First and foremost, it can be specified and estimated in steps, thanks to the joint use of a consistent DCC specification and of a central Wishart distribution, with each step having a QML interpretation. Then, due to the possibility of applying a preliminary covariance targeting step, parameter proliferation is kept under control so that the model can be easily fit to matrices of large dimension, where large in this context means of the order of fifty. Moreover, computational tractability and speed of implementation are preserved even in the largest case. This is already a significant progress relative to existing models relying on less flexible parameterizations.

As the implemented procedure is expected to be less efficient than a one-step QML estimation, the cost of introducing targeting along with the multistep approach is investigated via a Monte Carlo study, revealing that the proposed three-step estimator is more precise than the standard one-step QMLE when the cross-sectional dimension is larger than ten and the time series length is up to two thousand. This favorably supports our conjecture that as the sample dimension increases, the usual one-step estimator, even if feasible, leads to a considerable distortion in the estimated parameters especially if the time-series is not sufficiently long. Our multistep approach proves useful to considerably reduce the bias in this situation.

Finally, the empirical forecasting performance of the model against some benchmarks is illustrated in low and large dimensional settings, in the latter case using different realized covariance estimators, with qualitatively similar results. The Re-cDCC model is found to perform best in the cases under study, thus providing evidence to its potential usefulness for all kind of financial applications beyond the one of portfolio allocation we reported.

Appendix A. Proofs

Proof of Proposition 1: Let $\mathbf{n}_M = \text{vech}(M)$ denote the model's static parameters, and let $\boldsymbol{\phi}_s$ and $\boldsymbol{\phi}_c$ respectively denote the vectors of dynamic parameters indexing $\{\boldsymbol{\gamma}, \boldsymbol{\delta}\}$ in Eq.(5) and $\{A, B\}$ in Eq.(7). The full parameter vector is $\boldsymbol{\phi} = (\mathbf{v}, \mathbf{n}'_M, \boldsymbol{\phi}'_s, \boldsymbol{\phi}'_c)$, and the true parameter vector is denoted by $\boldsymbol{\phi}_0$. Recall from Eq.(19) that the log-likelihood contribution of observation t , ignoring constant terms and those not depending on H_t , is given by

$$\ell_t(\boldsymbol{\phi}) = -\frac{\nu}{2} \log |S_t| - \frac{\nu}{2} \text{tr}(S_t^{-1} C_t),$$

where H_t and C_t are $n \times n$ PDS matrices and L is lower triangular. The structure of the log-likelihood allows for a three-step estimation strategy to estimate ϕ , where the first step is represented by a method of moment estimator: $\hat{M} = \hat{L}\hat{L}' = 1/T \sum_{t=1}^T C_t$, which implies \hat{n}_M . Note that for notational ease, we use L instead of \hat{L} . The following two steps are obtained by decomposing the log-likelihood function as in Eq.(20). First, by Eq.(3), we have:

$$\ell_t(\phi) = -\frac{\nu}{2} \left[\log|LH_tL'| + \text{tr}(L'^{-1}H_t^{-1}L^{-1}C_t) \right],$$

and by Eq.(4), *i.e.* $H_t = D_tR_tD_t$,

$$\ell_t(\phi) = -\frac{\nu}{2} \left[\log|LD_tR_tD_tL'| + \text{tr}(L'^{-1}D_t^{-1}R_t^{-1}D_t^{-1}L^{-1}C_t) \right].$$

By noticing that $\log|LD_tR_tD_tL'| = 2 \log|D_t| + \log|R_t| + 2 \log|L|$ and using the trace property of invariance under cyclical rotation, we can write

$$\begin{aligned} \ell_t(\phi) &= -\frac{\nu}{2} [2 \log|D_t| + \log|R_t| + 2 \log|L|] \\ &\quad - \frac{\nu}{2} \left[\text{tr} \left((R_t^{-1} - I_n) D_t^{-1} L^{-1} C_t L'^{-1} D_t^{-1} + D_t^{-1} L^{-1} C_t L'^{-1} D_t^{-1} \right) \right], \end{aligned}$$

where we added and subtracted the term $D_t^{-1} L^{-1} C_t L'^{-1} D_t^{-1}$ in the second term on the right hand side. Finally, reallocating elements we get

$$\begin{aligned} \ell_{v,t}(\nu, \phi_v, \hat{n}_M) &= -\nu \log|D_t| - \frac{\nu}{2} \text{tr} \left(D_t^{-1} L^{-1} C_t L'^{-1} D_t^{-1} \right), \\ \ell_{c,t}(\nu, \phi_c, \hat{n}_M, \hat{\phi}_v) &= -\frac{\nu}{2} \left[\log|R_t| + \text{tr} \left((R_t^{-1} - I_n) D_t^{-1} L^{-1} C_t L'^{-1} D_t^{-1} \right) \right] - \nu \log|L|, \end{aligned}$$

which correspond to Equations Eq.(22)–Eq.(23).

Proof of Proposition 2: Most steps of the following proofs rely on the matrix derivatives rules stated in Lutkepohl (1996). We start deriving the $(1 \times m_v)$ score vector $\partial \ell_{v,t}(\nu, \phi_v, \hat{n}_M) / \partial \phi_v$ and prove that its expected value in $\phi_{v,0}$ is zero. From the $\ell_{v,t}$ log-likelihood function we have:

$$\begin{aligned} \frac{\partial \ell_{v,t}(\nu, \phi_v, \hat{n}_M)}{\partial \phi_v} &= -\nu \frac{\partial \log|D_t|}{\partial \phi_v} - \frac{\nu}{2} \frac{\partial \text{tr} \left(D_t^{-1} L^{-1} C_t L'^{-1} D_t^{-1} \right)}{\partial \phi_v} \\ &= -\frac{\nu}{2} \frac{\partial \log|H_t|}{\partial \phi_v} - \frac{\nu}{2} \frac{\partial \text{tr} \left(L^{-1} C_t L'^{-1} H_t^{-1} \right)}{\partial \phi_v} \\ &= \frac{\nu}{2} \left[-\sum_{i=1}^n \frac{\partial \log H_{ii,t}}{\partial \phi_v} - \sum_{i=1}^n L_{ii}^{-1} C_{ii,t} L'_{ii}^{-1} \left(\frac{\partial H_{ii,t}^{-1}}{\partial \phi_v} \right) \right] \\ &= \frac{\nu}{2} \left[-\sum_{i=1}^n \frac{1}{H_{ii,t}} + \sum_{i=1}^n L_{ii}^{-1} C_{ii,t} L'_{ii}^{-1} H_{ii,t}^{-2} \right], \end{aligned}$$

where we used $2 \log|D_t| = \log|H_t|$ and $D_t^{-2} = H_t^{-1}$. Using the law of iterated expectation, we obtain

$$\begin{aligned} \mathbb{E} \left[\frac{\partial \ell_{v,t}(\nu, \phi_v, \hat{n}_M)}{\partial \phi_v} \right] &= \frac{\nu}{2} \mathbb{E} \left\{ \mathbb{E}_{t-1} \left[-\sum_{i=1}^n \frac{1}{H_{ii,t}} + \sum_{i=1}^n L_{ii}^{-1} C_{ii,t} L'_{ii}^{-1} H_{ii,t}^{-2} \right] \right\} \\ &= \frac{\nu}{2} \mathbb{E} \left\{ -\sum_{i=1}^n \frac{1}{H_{ii,t}} + \sum_{i=1}^n \mathbb{E}_{t-1} \left(L_{ii}^{-1} C_{ii,t} L'_{ii}^{-1} \right) H_{ii,t}^{-2} \right\}. \end{aligned}$$

The expected second step score for observation t is equal to zero as, by the conditional moment assumption stated in Eq.(3), it holds $\mathbb{E}_{t-1} \left(L_{ii}^{-1} C_{ii,t} L'_{ii}^{-1} \right) = H_{ii,t}$ and hence $\mathbb{E} \left[\frac{\partial \ell_{v,t}(\nu, \phi_v, \hat{n}_M)}{\partial \phi_v} \right] = 0$.

The derivation of the $(1 \times m_c)$ score vector $\partial \ell_{c,t}(\nu, \phi_c, \hat{n}_M, \hat{\phi}_\nu) / \partial \phi_c$ follows a similar procedure. From the $\ell_{c,t}$ likelihood function with n_M and ϕ_ν fixed at the estimates obtained in the previous steps, we get

$$\begin{aligned} \frac{\partial \ell_{c,t}(\nu, \phi_c, \hat{n}_M, \hat{\phi}_\nu)}{\partial \phi_c} &= -\frac{\nu}{2} \frac{\partial \log |R_t|}{\partial \phi_c} - \frac{\nu}{2} \frac{\partial \text{tr} \left(R_t^{-1} D_t^{-1} L^{-1} C_t L'^{-1} D_t^{-1} \right)}{\partial \phi_c} \\ &= -\frac{\nu}{2} \text{tr} \left(R_t^{-1} \frac{\partial R_t}{\partial \phi_c} \right) + \frac{\nu}{2} \text{tr} \left(R_t^{-1} \frac{\partial R_t}{\partial \phi_c} R_t^{-1} D_t^{-1} L^{-1} C_t L'^{-1} D_t^{-1} \right). \end{aligned}$$

Again, by taking expectations on both sides, it gives

$$\begin{aligned} \mathbb{E} \left[\frac{\partial \ell_{c,t}(\nu, \phi_c, \hat{n}_M, \hat{\phi}_\nu)}{\partial \phi_c} \right] &= -\frac{\nu}{2} \mathbb{E} \left\{ \mathbb{E}_{t-1} \left[\text{tr} \left(R_t^{-1} \frac{\partial R_t}{\partial \phi_c} \right) + \frac{\nu}{2} \text{tr} \left(R_t^{-1} \frac{\partial R_t}{\partial \phi_c} R_t^{-1} D_t^{-1} L^{-1} C_t L'^{-1} D_t^{-1} \right) \right] \right\} \\ &= -\frac{\nu}{2} \mathbb{E} \left\{ \text{tr} \left(R_t^{-1} \frac{\partial R_t}{\partial \phi_c} \right) + \frac{\nu}{2} \text{tr} \left(R_t^{-1} \frac{\partial R_t}{\partial \phi_c} R_t^{-1} D_t^{-1} L^{-1} L \mathbb{E}_{t-1}(C_t) L' L^{-1} D_t^{-1} \right) \right\} \end{aligned}$$

where at the true parameter value $\phi_c = \phi_{c,0}$, it holds $\mathbb{E}_{t-1}(C_t) = S_t = L D_t R_t D_t L'$ and hence $\mathbb{E} \left[\frac{\partial \ell_{c,t}(\nu, \phi_c, \hat{n}_M, \hat{\phi}_\nu)}{\partial \phi_c} \right] = 0$.

Appendix B. Tables

Table B.12: **Data**

The table reports tickers and issue names for the stocks used in the empirical applications. Dataset A pertains to 46 equities from the S&P 500 traded over the period 5/1/1999–31/12/2008 while Dataset B pertains to 50 stocks among the largest U.S. financial institutions for the period 02/01/2007–31/12/2012. The assets in common (AXP, BAC, WFC) are only reported in the upper panel.

Stock	Issue Name	Stock	Issue Name
Dataset A: 46 selected constituents of the NYSE and NASDAQ on January 1999			
AAPL	Apple	JNJ	Johnson & Johnson
ABT	Abbott Laboratories	JPM	J P Morgan Chase & Co
AXP	American Express Company	KO	Coca Cola
BA	The Boeing Company	ETR	Entergy Corporation
BAC	Bank of America	LLY	Eli Lilly and Company
BMY	Bristol-Myers Squibb Company	MCD	McDonald's Corporation
BP	BP p.l.c.	MMM	3M Company
C	Citigroup	MRK	Merck & Company, Inc.
CAT	Caterpillar	MS	Morgan Stanley
CL	Colgate-Palmolive Co.	MSFT	Microsoft Corporation
CSCO	Cisco Systems	SLB	Schlumberger N.V.
CVX	Chevron Corporation	T	AT&T
DELL	Dell Inc.	TWX	Time Warner
DIS	Walt Disney Company	UN	Unilever NV
EXC	Exelon Corporation	VZ	Verizon Communications
F	Ford Motor Co.	PEP	Pepsico, Inc.
FDX	FedEx Corporation	PFE	Pfizer
GE	General Electric	PG	Procter & Gamble Company
HD	The Home Depot	QCOM	QUALCOMM Incorporated
HNZ	HNZ Group	WFC	Wells Fargo & Company
HON	Honeywell International	XOM	Exxon Mobil Corporation
INTC	Intel Corporation	XRX	Xerox Corporation
Dataset B: 50 selected US financial institutions on January 2007			
ACAS	American Capital	GS	The Goldman Sachs Group
AET	Aetna	HBAN	Huntington Bancshares Incorporated
AFL	Aflac Incorporated	HIG	Hartford Financial Services Group
AIG	American International Group	HNT	Health Net Inc.
AIZ	Assurant	ICE	Intercontinental Exchange Inc.
ALL	The Allstate Corporation	JNS	Janus Capital Group, Inc
AMP	Ameriprise Financial	KEY	KeyCorp
BBT	BB&T Corporation	NTRS	Northern Trust Corporation
BEN	Franklin Resources, Inc.	NYX	NYX Gaming Group Limited
BK	Bank Of New York Mellon Corporation	PFG	Principal Financial Group Inc
BLK	BlackRock, Inc	PGR	Progressive Corporation
BRKB	Berkshire Hathaway	PNC	PNC Financial Services Group
CB	Chubb Limited	PRU	Prudential Financial
CBG	CBRE Group	RF	Regions Financial Corporation
CINF	Cincinnati Financial Corporation	SEIC	SEI Investments Company
CMA	Comerica Incorporated	SNV	Synovus Financial Corp.
COF	Capital One Financial Corporation	STI	SunTrust Banks
CVH	CV Holdings	STT	State Street Corporation
EV	Eaton Vance Corporation	TMK	Torchmark Corporation
FITB	Fifth Third Bancorp	TROW	T.Rowe Price Group
FNF	Fidelity National Financial	UNH	UnitedHealth Group
GNW	Genworth Financial Inc	UNM	Unun Group
MET	MetLife, Inc.	USB	U.S.Bancorp
MTB	M&T Bank Corporation	WU	The Western Union Company
ZION	Zions Bancorporation		

Table B.11: Monte Carlo study results: comparison of estimators in low dimensions

The table reports Monte Carlo results for the one-step QMLE (1QMLE) and the two-step QMLE (2QMLE) across $T=500, 1000, 2000$ and $n=2, 5, 7, 10, 15$. RB and RMSE respectively denote the relative bias and the root mean square error (see Section 4). Results based on 500 simulations.

		$T = 500$		$T = 1000$		$T = 2000$	
		RB	RMSE	RB	RMSE	RB	RMSE
n=2							
$\bar{\gamma}$	1QML	0.127	0.019	0.031	0.013	0.014	0.009
	2QML	0.136	0.020	0.035	0.013	0.015	0.009
$\bar{\delta}$	1QML	-0.111	0.175	-0.037	0.093	-0.011	0.028
	2QML	-0.118	0.184	-0.041	0.101	-0.011	0.028
α	1QML	0.012	0.033	0.008	0.024	-0.001	0.018
	2QML	0.009	0.033	0.005	0.024	-0.001	0.018
β	1QML	-0.057	0.138	-0.022	0.074	-0.008	0.043
	2QML	-0.057	0.138	-0.021	0.074	-0.008	0.043
n=5							
$\bar{\gamma}$	1QML	0.142	0.015	0.053	0.009	0.014	0.006
	2QML	0.141	0.015	0.054	0.009	0.013	0.006
$\bar{\delta}$	1QML	-0.148	0.173	-0.047	0.072	-0.012	0.022
	2QML	-0.146	0.170	-0.047	0.073	-0.012	0.022
α	1QML	-0.005	0.011	-0.002	0.008	0.000	0.005
	2QML	-0.008	0.011	-0.003	0.008	-0.000	0.005
β	1QML	-0.017	0.029	-0.008	0.019	-0.004	0.013
	2QML	-0.016	0.029	-0.008	0.019	-0.004	0.012
n=7							
$\bar{\gamma}$	1QML	0.543	0.030	0.336	0.019	0.243	0.013
	2QML	0.134	0.012	0.048	0.008	0.012	0.005
$\bar{\delta}$	1QML	-0.305	0.294	-0.138	0.143	-0.067	0.065
	2QML	-0.145	0.161	-0.045	0.066	-0.012	0.020
α	1QML	0.010	0.007	0.005	0.005	0.007	0.004
	2QML	-0.003	0.007	-0.002	0.005	0.000	0.004
β	1QML	-0.020	0.025	-0.010	0.015	-0.006	0.010
	2QML	-0.016	0.022	-0.007	0.014	-0.004	0.010
n=10							
$\bar{\gamma}$	1QML	0.877	0.045	0.593	0.030	0.479	0.024
	2QML	0.128	0.011	0.045	0.007	0.003	0.004
$\bar{\delta}$	1QML	-0.314	0.286	-0.166	0.152	-0.112	0.102
	2QML	-0.154	0.162	-0.047	0.061	-0.009	0.014
α	1QML	0.013	0.005	0.010	0.004	0.005	0.003
	2QML	-0.009	0.005	-0.000	0.004	-0.000	0.002
β	1QML	-0.018	0.020	-0.010	0.012	-0.005	0.007
	2QML	-0.014	0.017	-0.007	0.011	-0.003	0.006
n=15							
$\bar{\gamma}$	1QML	0.933	0.047	0.879	0.044	0.883	0.044
	2QML	0.163	0.011	0.070	0.006	0.020	0.003
$\bar{\delta}$	1QML	-0.141	0.127	-0.118	0.106	-0.112	0.100
	2QML	-0.162	0.161	-0.055	0.064	-0.014	0.019
α	1QML	0.015	0.005	0.009	0.003	0.007	0.002
	2QML	-0.006	0.003	-0.003	0.002	-0.000	0.002
β	1QML	-0.017	0.023	-0.007	0.008	-0.005	0.005
	2QML	-0.015	0.014	-0.006	0.007	-0.004	0.005

Table B.13: Implemented Loss functions

The table gives the formulas of the chosen matrix loss functions. S_t is the predicted conditional covariance matrix while C_t is the realized measure; n denotes the number of assets and λ_i are the eigenvalues of $(S_t - C_t)^2$.

Matrix loss function		
sfrob	Squared Frobenius Distance	$\sum_{i=1}^n \lambda_i$
stein	Stein	$\text{tr}(S_t^{-1}C_t) - \log S_t^{-1}C_t - n$
Qlik	QLIK	$\log S_t + \text{tr}(S_t^{-1}C_t)$
VND	von Neumann Divergence	$\text{tr}(C_t \log C_t - C_t \log S_t - C_t + S_t)$

Acknowledgements

The authors would like to thank Sébastien Laurent and Francesco Violante for providing the data used in the paper. Luc Bauwens and Manuela Braione acknowledge support of the “Communauté française de Belgique” through contract “Projet d’Actions de Recherche Concertées 12/17-045”, granted by the “Académie universitaire Louvain”. Giuseppe Storti acknowledges funding from the Italian Ministry of Education, University and Research (MIUR) through PRIN project “Forecasting economic and financial time series: understanding the complexity and modeling structural change” (code 2010J3LZEN).

References

- Aielli, G., 2013. Dynamic conditional correlations: on properties and estimation. *Journal of Business & Economic Statistics* 31 (3), 282–299.
- Andersen, T., Bollerslev, T., Diebold, F., Labys, P., 2003. Modeling and forecasting realized volatility. *Econometrica* 71, 579–625.
- Andersen, T., Bollerslev, T., Frederiksen, P., Nielsen, M., 2010. Continuous-time models, realized volatilities and testable distributional implications for daily stock returns. *Journal of Applied Econometrics* 25, 233–261.
- Barndorff-Nielsen, O., Shephard, N., 2001. Normal modified stable processes. *Theory of Probability and Mathematics Statistics* 65, 1–19.
- Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., Shephard, N., 2009. Realized kernels in practice: Trades and quotes. *The Econometrics Journal* 12 (3), C1–C32.
- Barndorff-Nielsen, O. E., Shephard, N., 2004. Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. *Econometrica* 72 (3), 885–925.
- Bauer, G., Vorkink, K., 2011. Forecasting multivariate realized stock market volatility. *Journal of Econometrics* 160, 93–101.
- Bauwens, L., Braïone, M., Storti, G., In Pressa. A dynamic component model for forecasting high-dimensional realized covariance matrices. *Journal of Econometrics and Statistics*.
- Bauwens, L., Braïone, M., Storti, G., In Pressb. Forecasting comparison of long term component dynamic models for realized covariance matrices. *Annals of Economics and Statistics*.
- Bauwens, L., Grigoryeva, L., Ortega, J.-P., 2015a. Estimation and empirical performance of non-scalar dynamic conditional correlation models. *Computational Statistics & Data Analysis*.
- Bauwens, L., Grigoryeva, L., Ortega, J.-P., 2015b. Non-scalar garch models: Composite likelihood estimation and empirical comparisons.
- Bauwens, L., Hafner, C. M., Pierret, D., 2013. Multivariate volatility modeling of electricity futures. *Journal of Applied Econometrics* 28 (5), 743–761.
- Bauwens, L., Storti, G., 2013. Computationally efficient inference procedures for vast dimensional realized covariance models, . in *Complex Models and Computational Methods in Statistics*. Springer, pp. 37–49, m. Grigoletto, F. Lisi and S. Petrone Eds.
- Bollerslev, T., Patton, A. J., Quaedvlieg, R., 2016a. Exploiting the errors: A simple approach for improved volatility forecasting. *Journal of Econometrics* 192 (1), 1–18.
- Bollerslev, T., Patton, A. J., Quaedvlieg, R., 2016b. Modeling and forecasting (un) reliable realized covariances for more reliable financial decisions. Available at SSRN 2759388.
- Bonato, M., Caporin, M., Ranaldo, A., 2009. Forecasting realized (co)variances with a block structure Wishart autoregressive model, working Papers 2009-03, Swiss National Bank.
- Bonato, M., Caporin, M., Ranaldo, A., 2012. A forecast-based comparison of restricted Wishart autoregressive models for realized covariance matrices. *The European Journal of Finance* 18 (9), 761–774.
- Boudt, K., Danielsson, J., Laurent, S., 2013. Robust forecasting of dynamic conditional correlation GARCH models. *International Journal of Forecasting* 29 (2), 244–257.
- Boudt, K., Laurent, S., Lunde, A., Quaedvlieg, R., 2014. Positive semidefinite integrated covariance estimation, factorizations and asynchronicity. Tech. rep.
- Chiriac, R., Voev, V., 2011. Modelling and forecasting multivariate realized volatility. *Journal of Applied Econometrics* 26, 922–947.
- Colacito, R., Engle, R. F., Ghysels, E., 2011. A component model for dynamic correlations. *Journal of Econometrics* 164 (1), 45–59.
- Corsi, F., 2009. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics* 7, 174–196.
- de Pooter, M., Martens, M., van Dijk, D., 2008. Predicting the daily covariance matrix for s&p 100 stocks using intraday data, but which frequency to use? *Econometric Reviews* 27 (1-3), 199–229.
- Diebold, F. X., Mariano, R. S., 2012. Comparing predictive accuracy. *Journal of Business & economic statistics*.
- Engle, R., 2002a. Dynamic conditional correlation - a simple class of multivariate GARCH models. *Journal of Business and Economic Statistics* 20, 339–350.
- Engle, R., Gallo, G., 2006. A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics* 131, 3–27.
- Engle, R., Kelly, B., 2012. Dynamic equicorrelation. *Journal of Business and Economic Statistics* 30 (2), 212–228.
- Engle, R., Lee, G., 1999. A Permanent and Transitory Component Model of Stock Return Volatility, . *Cointegration, Causality, and Forecasting: A Festschrift in Honor of Clive W.J. Granger*. Oxford University Press - R. Engle and H. White eds., pp. 475–497.
- Engle, R., Rangel, J., 2008. The spline-GARCH model for low-frequency volatility and its global macroeconomic causes. *Review of Financial Studies* 21, 1187–1222.
- Engle, R., Russell, J., 1998. Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica* 66, 1127–1162.
- Engle, R., Sheppard, K., 2001. Theoretical and empirical properties of dynamic conditional correlation multivariate garch, mimeo, UCSD.
- Engle, R. F., 2002b. New frontiers for ARCH models. *Journal of Applied Econometrics* 17, 425–446.
- Engle, R. F., Shephard, N., Sheppard, K., 2008a. Fitting vast dimensional time-varying covariance models.
- Engle, R. F., Shephard, N., Sheppard, K., 2008b. Fitting vast dimensional time-varying covariance models.
- Fan, J., Li, Y., Yu, K., 2012. Vast volatility matrix estimation using high-frequency data for portfolio selection. *Journal of the American Statistical Association* 107 (497), 412–428.
- Franco, C., Horvath, L., Zakoïan, J.-M., 2011. Merits and drawbacks of variance targeting in GARCH models. *Journal of Financial Econometrics* 4, 619–656.
- Franco, C., Horvath, L., Zakoïan, J.-M., 2014. Variance targeting estimation of multivariate GARCH models. *Journal of Financial Econometrics* 14 (2), 353–382.
- Golosnoy, V., Gribisch, B., Liesenfeld, R., 2012. The conditional autoregressive Wishart model for multivariate stock market volatility. *Journal of Econometrics* 167, 211–223.
- Gouriéroux, C., Jasiak, J., Sufana, R., 2009. The Wishart autoregressive process of multivariate stochastic volatility. *Journal of Econometrics* 150, 167–181.
- Hafner, C., Linton, O., 2010. Efficient estimation of a multivariate volatility model. *Journal of Econometrics* 159 (1), 55–73.

- Hansen, P. R., Huang, Z., Shek, H. H., 2012. Realized garch: a joint model for returns and realized measures of volatility. *Journal of Applied Econometrics* 27 (6), 877–906.
 URL <http://dx.doi.org/10.1002/jae.1234>
- Hansen, P. R., Lunde, A., Nason, J. M., 2011. The model confidence set. *Econometrica* 79 (2), 453–497.
- Jin, X., Maheu, J., 2013. Modelling realized covariances and returns. *Journal of Financial Econometrics* 11 (2), 335–369, wP 11-08, The Rimini Center for Economic Analysis.
- Ledoit, O., Santa-Clara, P., Wolf, M., 2003. Flexible multivariate garch modeling with an application to international stock markets. *The Review of Economics and Statistics* 85, 735–747.
- Lindsay, B. G., 1988. Composite likelihood methods. *Contemporary mathematics* 80 (1), 221–39.
- Lutkepohl, H., 1996. *Handbook of matrices*. Wiley.
- Newey, W. K., West, K. D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55 (3), 703–708.
- Noureldin, D., Shephard, N., Sheppard, K., 2012. Multivariate high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics* 27 (6), 907–933.
- Noureldin, D., Shephard, N., Sheppard, K., 2014. Multivariate rotated ARCH models. *Journal of Econometrics* 179 (1), 16–30.
- Pakel, C., Shephard, N., Sheppard, K., Engle, R. F., 2014. Fitting vast dimensional time-varying covariance models. Tech. rep., Working Paper.
- Patton, A. J., 2011. Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics* 160 (1), 246–256.
- Patton, A. J., Sheppard, K., 2009. Evaluating volatility and correlation forecasts. In: *Handbook of financial time series*. Springer, pp. 801–838.
- Pedersen, R. S., Rahbek, A., 2014. Multivariate variance targeting in the BEKK–GARCH model. *The Econometrics Journal* 17 (1), 24–55.
- Shephard, N., Sheppard, K. K., 2010. Realising the future: forecasting with high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics* 25, 197–231.
- Sheppard, K., Xu, W., 2014. Factor high-frequency based volatility (heavy) models, available at SSRN: <http://ssrn.com/abstract=2442230> or <http://dx.doi.org/10.2139/ssrn.2442230>.
- Tse, Y., Tsui, A., 2002. A multivariate GARCH model with time-varying correlations. *Journal of Business and Economic Statistics* 20, 351–362.