DYNAMIC CONDITIONAL CORRELATION MODELS FOR REALIZED COVARIANCE MATRICES

(Preliminary and incomplete version)

Luc Bauwens[1], Giuseppe Storti[2] and Francesco Violante[3]

August 24, 2012

Abstract

New dynamic models for realized covariance matrices are proposed. The expected value of the realized covariance matrix is specified in two steps: a model for each realized variance, and a model for the realized correlation matrix. The realized correlation model is a dynamic conditional correlation model. Estimation can be done in two steps as well, and a QML interpretation is given to each step, by assuming a Wishart conditional distribution. Moreover, the model is applicable to large matrices since estimation can be done by the composite likelihood method.

Keywords: Realized covariance; dynamic conditional correlations; covariance targeting; Wishart distribution; composite likelihood.

JEL Classification: C32, C58.

[1]Université catholique de Louvain, CORE, B-1348 Louvain-La-Neuve.
E-mail: luc.bauwens@uclouvain.be

[2]Department of Economics and Statistics, Universiti Salerno, Fisciano, Italy.
E-mail: storti@unisa.it

[3]Department of Quantitative Economics, Maastricht University, The Netherlands.
E-mail: f.violante@maastrichtuniversity.nl

L. Bauwens thanks the University of Sassari for its hospitality during a visit while working on this paper. Research supported by the contract "Projet d'Actions de Recherche Concertées" 07/12-002 of the "Communauté francaise de Belgique", granted by the "Académie universitaire Louvain". The scientific responsibility is assumed by the authors.

1 Introduction

The availability of measures of daily variances of financial returns, and covariances between these, allows researchers to model time series of covariance matrices. One interest of these models is that they can be used for forecasting future values, which is typically 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.

Models have firstly been proposed for realized variances alone, such as ARFIMA models, see e.g. Andersen, Bollerslev, Diebold, and Labys (2003), and the HAR model of Corsi (2009). Empirical studies show that probability distributions of the log of realized variances are very close to being Gaussian, and from a time series perspective, the data exhibit a long memory feature, see Andersen, Bollerslev, Diebold, and Ebens (2001). The development of multivariate models has come afterwards. Modeling a covariance matrix is challenging since the dimension of the object to be modeled is proportional to the square of the number of assets, and thus the number of parameters is likely to be large even for a handful of assets. Another challenge is that the model should be congruent with the property that covariance matrices are positive definite. Thus vectorizing the covariance matrix and assuming a vector autoregressive moving average model (VARMA) for its non-redundant elements is not a solution. Chiriac and Voev (2010) model the elements of the Cholesky decomposition of the realized covariance matrix by a VARFIMA process or by a VAR version of the univariate HAR model of Corsi. Using the Cholesky decomposition ensures that forecasts of the covariance matrix are positive definite and avoids imposing complex parameter restrictions to ensure positivity. It also allows to include exogenous variables easily. A drawback of this approach is that the model specification is not invariant with

respect to a permutation of the rows and columns of the covariance matrices from which the Cholesky decompositions are computed. Another drawback is that a VARFIMA model is heavily parameterized, but this problem can be attenuated by imposing restrictions on the parameter space. In the their empirical application, Chiriac and Voev work with covariance matrices of order six, implying a VARFIMA model for six variances and fifteeen covariances. Another transformation of a covariance matrix that ensures positivity is the matrix-log function. This transformation is used by Bauer and Vorkink (2006) for realized covariance matrices. If C_t denotes a realized covariance matrix and $Q_t\Lambda_t Q_t$ its spectral decomposition, then $Q_t \log(\Lambda_t) Q_t$ is the matrix-log transformation of C_t, where the log function is applied to each eigenvalue. Then the vectorized matrix-log can be modeled directly as a VARFIMA process, and by taking the matrix exponential transformation of the forecast implied by the model, one obtains a positive-definite symmetric (PDS) forecast. Their application is to matrices of order five.

Several dynamic models for realized covariance matrices use the Wishart distribution. This is a natural choice since the Wishart has been initially derived as the sampling distribution of the sample covariance matrix of a Gaussian process. Several authors use the Wishart assumption. Gouriéroux, Jasiak, and Sufana (2009) were probably the first to use the idea of a Wishart process for realized covariance matrices, with the WAR(p) (Wishart autoregressive) process, where p is a lag order parameter. It is too heavily parameterized for a large number of assets since it uses a number of parameters equal to $3n^2/2 + n/2 + 1$ (for the one lag case), where n is the order of C_t (i.e. the number of assets), and it does not fit the long memory feature of the data. Bonato, Caporin, and Ranaldo (2009) propose a block structure of the WAR model, which reduces the number of parameters but keeps it of the order of n^2. Jin and Maheu (2010), and Noureldin, Shephard, and Sheppard (2011) propose a joint dynamic model for a return vector and its realized covariance matrix. In both papers, the realized covariance part of the model specifies the conditional distribution of C_t as a Wishart distribution, whose expected value is proportional to the scale matrix of the Wishart. In Jin and Maheu's paper, that scale matrix is a function of a few lags of itself, in the spirit of stochastic volatility models. In Noureldin, Shephard, and

Sheppard (2011), it is specified as a BEKK process taken from the multivariate GARCH literature. Golosnoy, Gribisch, and Liesenfeld (2010) also use the idea of a time-varying Wishart distribution and a scale matrix specified as a BEKK process. In these papers, the number of parameters of the dynamic equation of the scale matrix is thus proportional to n^2, excepting the case of a scalar BEKK process also considered by Noureldin, Shephard, and Sheppard (2011).

Our contribution is to replace the BEKK specification by DCC-type specifications inspired by Tse and Tsui (2002) and by Engle (2002a) as modified by Aielli (2008). This has four important advantages: i) The model can be specified in steps, one for each realized variance, and one for the realized correlation matrix, without departing from the Wishart assumption; ii) Correspondingly, maximum likelihood (ML) estimation can be split in two steps, one for the parameters of each realized variance dynamic process and one for those of the realized correlation process; iii) The estimators at each step have a quasi-ML interpretation; iv) Due to the properties of the Wishart distribution and of one of the DCC specifications (the scalar model) we propose, the estimation of the second step can be done by a composite likelihood (CL) approach, which makes it possible to estimate the model for a relatively large order of the realized covariance matrices (up to one hundred in this paper). This is in contrast to the models mentioned above: Gouriéroux, Jasiak, and Sufana (2009) work on three assets, Bonato, Caporin, and Ranaldo (2009) four, Golosnoy, Gribisch, and Liesenfeld (2010) six, Jin and Maheu (2010) five, and Noureldin, Shephard, and Sheppard (2011) ten. Some of our specifications allow us to use correlation targeting of the matrix constant term of the dynamic equation for the correlation matrix, i.e. preliminary estimation of this matrix by a method of moment estimator. The number of remaining parameters in the dynamic correlation equation is then either fixed (in our scalar model) or linear in the number of assets (in our diagonal model).

The paper is structured as follows. Realized dynamic conditional correlation (RDCC) models are presented in Section 2. In Section 3 we present the ML estimation procedure and in Section 4 we discuss the QML interpretation and present the asymptotic properties of the estimators. Correlation targeting is discussed in Section 5, and estimation by the

composite maximum likelihood (CML) method in Section 6. In Section 7 we report the results of a simulation study comparing the ML and CML estimators. In Section 8, we apply the methods to real data sets, and we offer some conclusions in the last section.

2 Model specifications

Let C_t be a sequence of PDS realized covariance matrices of order n, for $t = 1, \ldots, T$. We assume that conditional on past information I_{t-1} consisting of C_τ for $\tau \leq t-1$, and for all t, C_t follows a n-dimensional central Wishart distribution and denote this assumption by

$$C_t | I_{t-1} \sim W_n(\nu, S_t/\nu), \tag{1}$$

where ν ($> n-1$) is the degrees of freedom parameter and S_t/ν is a PDS scale matrix of order n. Equation (1) defines a generic conditional autoregressive Wishart (CAW) model, as proposed by Golosnoy, Gribisch, and Liesenfeld (2010). From the properties of the Wishart distribution - see e.g. Anderson (1984)- it follows that

$$E(C_t | I_{t-1}) := E_{t-1}(C_t) = S_t, \tag{2}$$

so that the i,j-th element of S_t is defined as the conditional covariance between returns on assets i and j, $cov(r_{i,t}, r_{j,t}|I_{t-1})$, for $i, j = 1, \ldots, n$, $r_{i,t}$ denoting the logarithmic return on asset i between the ends of periods $t-1$ and t.

Several choices are available for specifying the dynamics of S_t. Golosnoy, Gribisch, and Liesenfeld (2010) use the BEKK formulation of the multivariate GARCH literature. Assuming only one lag, this corresponds to

$$S_t = GG' + AC_{t-1}A' + BS_{t-1}B' \tag{3}$$

where A and B are square matrices of order n, and G is a lower triangular matrix such that GG' is PDS. This choice ensures that S_t is PDS for all t if S_0 is itself PDS. For large n, this choice renders the estimation infeasible due to the high number of parameters. Golosnoy, Gribisch, and Liesenfeld (2010) are able to estimate the model in (3) for five assets and two

lags, for a total of one hundred and sixteen parameters. This is a remarkable performance that is probably difficult to be improved on unless imposing strong parameter restrictions, as for instance, common dynamics for all the elements of S_t. However, they do not consider covariance targeting. This is easy to implement, since the unconditional expectation of C_t and of S_t is known analytically, see Corollary 1 in their paper. Nevertheless the number of parameters in (3) remains of order n^2. This holds even with covariance targeting, unless the matrices A and B are restricted to be diagonal, to have their rank equal to one, or are replaced by scalars.

The scalar RBEKK (R for realized) version implies that the conditional variances and covariances all follow the same dynamic pattern, which may be too restrictive. To avoid this restriction, and to enrich the class of possible models, we use the following decomposition of the covariance matrix in terms of the corresponding diagonal matrix of standard deviations and correlation matrix. Thus, we express S_t in equation (1) as

$$S_t \;=\; D_t R_t D_t, \tag{4}$$

where R_t is the conditional correlation matrix of the return vector $r_t = (r_{1,t}, \ldots, r_{n,t})'$ and $D_t = \{\text{diag}(S_t)\}^{1/2}$ is the diagonal matrix whose i-th diagonal entry is given by the conditional standard deviation $\sqrt{S_{ii,t}}$ of asset i. For any matrix A_t, the notation $A_{ij,t}$ indicates the (i,j)-th element of A_t.

This decomposition, introduced in a similar context by Engle (2002a) and Tse and Tsui (2002), enables us to specify separately the dynamic equation of each conditional variance and of the conditional correlation matrix R_t. For the conditional variances, we can choose among available univariate specifications, such as a GARCH-type equation, the HAR equation of Corsi (2009), an ARFIMA model as in Andersen, Bollerslev, Diebold, and Labys (2003), or any other suitable model. Each univariate model for $C_{ii,t}$ depends on lags of $C_{ii,t}$ and in some cases of $S_{ii,t}$, but cannot depend on lags of other realized variances or conditional realized variances (spillover effects) to allow for the two-step estimation developed in Section 3. More details are provided in Section 8 in the context of empirical illustrations.

In the following subsections we propose and discuss several suitable specifications for R_t, starting with the most parsimonious one.

2.1 Common correlation dynamics: scalar models

The most parsimonious specification that we propose imposes a scalar dynamic equation on the conditional correlation matrix. A possible dynamic updating equation for R_t, inspired by that of Tse and Tsui (2002) for multivariate GARCH models, is given by

$$R_t = (1 - \alpha - \beta)\bar{R} + \alpha P_{t-1} + \beta R_{t-1}, \tag{5}$$

where

$$P_t = \{\text{diag}(C_t)\}^{-1/2} C_t \{\text{diag}(C_t)\}^{-1/2} \tag{6}$$

is the realized correlation matrix at time t. The parameters α and β, and their sum, are constrained to lie between zero and one. The matrix $\bar{R}$ is a parameter that must satisfy the constraints of a correlation matrix, i.e. positive definite symmetric with unit diagonal elements. Since P_t has unit diagonal elements, R_t is a well defined correlation matrix for all t if the initial matrix R_0 is a correlation matrix. The matrix $\bar{R}$ can be parameterized by using the representation $\bar{R} = \{\text{diag}(CC')\}^{-1/2} CC' \{\text{diag}(CC')\}^{-1/2}$ where C is a lower triangular matrix of parameters. Notice that, although C is uniquely identifiable only if its diagonal elements are constrained to be positive (i.e. if is a Cholesky type decomposition) and CC' is identifiable only up to a multiplicative constant, $\bar{R}$ remains uniquely identifiable. A drawback of this specification is that it does not imply that $\bar{R}$ is the unconditional expectation of P_t and of R_t, which has some consequences discussed in Section 5 We label this model 'scalar RDCC'.

A different specification is in spirit close to the cDCC model of Aielli (2008), which is itself a modification of Engle (2002a). Thus, the representation in (4) is complemented by the following dynamic equation:

$$R_t = \{\text{diag}(Q_t)\}^{-1/2} Q_t \{\text{diag}(Q_t)\}^{-1/2}. \tag{7}$$

The correlation driving process Q_t is then defined by

$$Q_t = (1 - \alpha - \beta)\bar{Q} + \alpha P^*_{t-1} + \beta Q_{t-1}, \tag{8}$$

where

$$P^*_t = \{\text{diag}(Q_t)\}^{1/2} D_t^{-1} C_t D_t^{-1} \{\text{diag}(Q_t)\}^{1/2}. \tag{9}$$

We label this model 'scalar cRDCC' (c for consistent).

By taking expectations on both sides of (8), one obtains, assuming $\alpha + \beta < 1$, that $E(Q_t) = \bar{Q}$ if $E(P^*_t) = E(Q_t)$. The latter results holds using (2), (4) and (7) since

$$\begin{aligned} E\big(P^*_t\big) &= E\big[\{\text{diag}(Q_t)\}^{1/2} D_t^{-1} E_{t-1}\big(C_t\big) D_t^{-1} \{\text{diag}(Q_t)\}^{1/2}\big] \\ &= E\big[\{\text{diag}(Q_t)\}^{1/2} D_t^{-1} D_t R_t D_t D_t^{-1} \{\text{diag}(Q_t)\}^{1/2}\big] \\ &= E\big[\{\text{diag}(Q_t)\}^{1/2} R_t \{\text{diag}(Q_t)\}^{1/2}\big] = E(Q_t). \end{aligned} \tag{10}$$

In the above scalar models, the number of parameters is $O(n^2)$ due to the matrix $\bar{R}$ or $\bar{Q}$. We discuss targeting, i.e. estimation of these matrices before ML estimation of the remaining parameters, in Section 5. Targeting thus enables us to use the models for large dimensions. In Section 5 we show that an advantage of (5) is that the targeting does not depend on unknown parameters and thus is robust to specification errors in the variance equations.

2.2 Idiosyncratic correlation dynamics: diagonal models

The assumption of common dynamics, while crucial for applications to very large dimensional systems, may be relaxed for models of medium size. In such cases we can afford a more flexible specification allowing for idiosyncratic dynamics for each conditional correlation. By extension of (5) and (8), the recursions of R_t and Q_t can be specified respectively as the diagonal RDCC:

$$R_t = (\bar{R} - A_d \bar{R} A_d - B_d \bar{R} B_d) + A_d P_{t-1} A_d + B_d R_{t-1} B_d, \tag{11}$$

and the diagonal cRDCC:

$$Q_t = (\bar{Q} - A_d \bar{Q} A_d - B_d \bar{Q} B_d) + A_d P^*_{t-1} A_d + B_d Q_{t-1} B_d, \tag{12}$$

where A_d and B_d are diagonal parameter matrices. Let $A_d = \text{diag}(\text{a}')$ where $a' = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ and $B_d = \text{diag}(\text{b}')$, where $b' = (\beta_1, \beta_2, \ldots, \beta_n)$. Notice that the elements of A_d and B_d are identifiable only for $n > 2$. In the case of $n = 2$, the models boils down to the scalar versions.

The condition in the scalar models, that α, β, and their sum be smaller than unity is replaced by $\alpha_i\alpha_j$, $\beta_i\beta_j$, and their sum to be smaller than unity for all i and j. If we impose moreover that α_i and β_i are positive for all i, these restrictions ensure that R_t in (11) is a correlation matrix provided that R_0 is a correlation matrix and $\bar{R} - A_d\bar{R}A_d - B_d\bar{R}B_d$ is positive semidefinite (PSD). Notice that $\bar{R} - A_d\bar{R}A_d - B_d\bar{R}B_d$ is equal to $(U_n - aa' - bb') \odot \bar{R}$, where U_n is has all its elements equal to one. Hence the constant matrix of (11) is PSD if $U_n - aa' - bb'$ is itself PSD (given that $\bar{R}$ is a positive definite correlation matrix) which is a restriction that may not be easily imposed. For Q_t we need not impose that α_i and β_i be positive for all i.

Without targeting, the number of parameters of these diagonal models is $O(n^2)$, though with targeting it is $O(n)$.

REMARK: (12) can also be written

$$Q_t = (U_n - aa' - bb') \odot \bar{Q} + aa' \odot P^*_{t-1} + bb' \odot Q_{t-1} \tag{13}$$

and can be made more flexible by replacing aa' and bb' by full rank square matrices $A = (a_{ij})$ and $B = (b_{ij})$, respectively. The expressions for $Q_{t,ij}$, the (i,j)-element of Q_t, corresponding to (12) and (16) are

$$Q_{t,ij} = (1 - \alpha_i\alpha_j - \beta_i\beta_j)\bar{Q}_{ij} + \alpha_i\alpha_j P^*_{t-1,ij} + \beta_i\beta_j Q_{t-1,ij} \tag{14}$$

$$Q_{t,ij} = (1 - \alpha_{ij} - \beta_{ij})\bar{Q}_{ij} + \alpha_{ij} P^*_{t-1,ij} + \beta_{ij} Q_{t-1,ij}. \tag{15}$$

In matrix format, the more general specification is written

$$Q_t = (U_n - A - B) \odot \bar{Q} + A \odot P^*_{t-1} + B \odot Q_{t-1}. \tag{16}$$

Restrictions ensuring that Q_t in (16) is positive definite (PD) are that A is PD while B, Q_0 and $U_n - A - B$ are PSD, or that B and Q_0 are PD while A and $U_n - A - B$ are PSD, or that $U_n - A - B$ is PD while A, B and Q_0 are PSD.

2.3 Correlation spillovers: the full model

The models in (11) and (12), although more flexible than the corresponding scalar versions, still neglect the effect known as correlation spillovers, which is relevant in some empirical applications. In order to include this feature we need to abandon dynamic equations for R_t, since, although P_t has unit diagonal elements, simply allowing the matrices of parameters in (11) to have all non-zero elements (which would generate the correlation spillovers) does not ensure R_t to have unit diagonal elements. Thus, the specification in (12) can be replaced by the BEKK-type model

$$Q_t = CC^{'} + AP^{*}_{t-1}A^{'} + BQ_{t-1}B^{'}, \tag{17}$$

where A and B are parameter square matrices of order n, C is lower triangular such that CC' is PDS (but not necessarily with unit diagonal elements). In this model, each element of Q_t and R_t is a function of all the elements of P^{*}_{t-1} and of Q_{t-1}. Similarly to the RBEKK model in (3), these models are practicable only in small dimensional problems ($n \leq 5$) since the number of parameters is of order n^2. Further investigations on the full version of the model are left for the future.

2.4 Summary

In Table 1, we summarize the main models presented in this section, assuming they are "(1,1)" models, i.e. they have only one lag of the observed covariance or correlation matrix, and one lag of the corresponding conditionally expected matrix. In the table we have added a diagonal BEKK model, which is defined by equation (3) where A and B are diagonal, rather than full, matrices. The numbers in the first column of the table are purely indicative.

3 Likelihood function and estimation

In this section we focus on the estimation by the ML method of all the parameters of the RDCC models defined in the previous section. We do not consider in this section the

Table 1: Summary of (1,1) models

Dimension	Name (eq. numbers)	Targeting	Number of parameters[(a)]
$n \leq 5$	full RBEKK (3)	yes	$2n^2$
	full rank RDCC (4)-(17)	no	$0.5n(n+1) + 2n^2 + p_U$
$n \leq 25$	diagonal RBEKK[(b)] (3)	yes	$2n$
	diagonal RDCC (4)-(11)	yes, (34)	$2n + p_U$
	diagonal cRDCC (4)-(12)	no	$0.5n(n+1) + 2n + p_U$
$n \leq 100$	scalar RBEKK[(c)] (3)	yes	2
	scalar RDCC (4)-(5)	yes, (34)	$2 + p_U$
	scalar cRDCC (4)-(8)	no	$0.5n(n+1) + 2 + p_U$

p_U is the number of parameters of the univariate models for the n realized variances.
[(a)] not including the degrees of freedom parameter ν.
[(b)] model as in equation (3) where A and B are diagonal matrices.
[(c)] model as in equation (3) where A and B are scalars times the identity matrix of order n.

targeting of the constant matrix term (such as $\bar{R}$). We show that, based on the Wishart distribution assumption of equation (1), the estimation can be split in two steps, one for the parameters of the univariate realized variance equations if they do not include spillover terms, and one for the parameters of the realized correlation equation and the degrees of freedom parameter. We also show that the estimators of the variance and correlation equations have in each step a quasi-ML (QML) interpretation and can be obtained without estimating the degrees of freedom. This result is interesting especially for the estimation of the realized variance parameters, since it has been found empirically that the distribution of realized variances is often very close to being lognormal, see e.g. Andersen, Bollerslev, Diebold, and Ebens (2001) and Andersen, Bollerslev, Diebold, and Labys (2001). The Wishart assumption implies that the distribution of a realized variance is marginally Gamma, but since the estimation method we propose has a QML interpretation, the estimator is consistent even if the true distribution is lognormal and the conditional mean is correctly specified.

Our estimation method is different from that of Golosnoy, Gribisch, and Liesenfeld (2010), who deal with the BEKK formulation and extensions of it. They use ML estimation, but do not give a QML interpretation to it. They cannot separate estimation in

two steps, due to the structure of the BEKK model, and hence they cannot apply their model to more than a few assets. Our results differ also partly from those of Noureldin, Shephard, and Sheppard (2011), who deal with a model that has a realized covariance matrix equation in addition to a modified multivariate GARCH equation. They do provide a QML interpretation to ML estimators based on the Wishart assumption in the context of BEKK formulations, but the latter also prevent two step estimation.

The vector of unknown parameters to be estimated,, denoted by θ, can be partitioned into three components:

$$\theta = (\nu, \theta_c', \theta_v')' \tag{18}$$

where θ_v and θ_v are the vectors containing the conditional variance and correlation parameters, respectively. Since no targeting is considered, θ_c includes the parameters of the matrix $\bar{R}$ of (5) or (11), and $\bar{Q}$ of (8) or (12). We also partition θ_v as

$$\theta_v = (\theta_v^{(1)'}, \theta_v^{(2)'}, \ldots, \theta_v^{(n)'})', \tag{19}$$

where $\theta_v^{(i)}$ is the vector containing all the parameters of the conditional realized variance equation specific to asset i.

Using the expression of a Wishart density function, and of S_t in (4), we obtain the log-likelihood contribution $\ell(C_t; \theta | I_{t-1})$ of observation t, denoted by $\ell_t(\theta)$:

$$\begin{aligned} \ell_t(\theta) &= \frac{\nu n}{2} \log\left(\frac{\nu}{2}\right) + \frac{\nu - n - 1}{2} \log|C_t| - \sum_{i=1}^{n} \log\Gamma[(\nu + 1 - i)/2] \\ &\quad - \frac{\nu}{2} \log|D_t R_t D_t| - \frac{\nu}{2} \mathrm{tr}\{(D_t R_t D_t)^{-1} C_t\}. \end{aligned} \tag{20}$$

Proposition 1. *The likelihood contribution $\ell_t(\theta)$ in (20) can be written as*

$$\ell_t(\theta) = \ell_{1t}(\nu, \theta_v) + \ell_{2t}(\nu, \theta_c, \theta_v), \tag{21}$$

where

$$\ell_{1t}(\nu, \theta_v) = -\nu \log(D_t) - \frac{\nu}{2} tr\{D_t^{-1} C_t D_t^{-1}\} \tag{22}$$

and

$$\begin{aligned}\ell_{2t}(\nu,\theta_c,\theta_v) &= -\frac{\nu}{2}\log|R_t| + \frac{\nu n}{2}\log\frac{\nu}{2} + \frac{\nu-n-1}{2}\log|C_t| \\ &- \sum_{i=1}^{n}\log[\Gamma((\nu+1-i)/2] - \frac{\nu}{2}tr\{(R_t^{-1}-I_n)D_t^{-1}C_tD_t^{-1}\}. \end{aligned} \tag{23}$$

Moreover, assuming that the univariate equations for the conditional realized variances do not include spillover terms and that their parameters are variation-free, ℓ_{1t} can be written as the sum of n univariate functions:

$$\ell_{1t}(\nu,\theta_v) = \frac{\nu}{2}\left[-\sum_{i=1}^{n}\log S_{ii,t} - \sum_{i=1}^{n}S_{ii,t}^{-1}C_{ii,t}\right] = \frac{\nu}{2}\sum_{i=1}^{n}\ell_{1t}^{(i)} \propto \sum_{i=1}^{n}\ell_{1t}^{(i)}(\nu,\theta_v^{(i)}). \tag{24}$$

Proof. The proof relies on the following results:

(i) $\log|D_tR_tD_t| = 2\log|D_t| + \log|R_t|$.

(ii) $\mathrm{tr}\{(D_tR_tD_t)^{-1}C_t\} = \mathrm{tr}\{R_t^{-1}D_t^{-1}C_tD_t^{-1}\} = \mathrm{tr}\{(R_t^{-1}-I_n)D_t^{-1}C_tD_t^{-1}\} + \mathrm{tr}\{D_t^{-1}C_tD_t^{-1}\}$.

(iii) $2\log|D_t| = \sum_{i=1}^{n}\log S_{ii,t}$.

(iv) $\mathrm{tr}\{D_t^{-1}C_tD_t^{-1}\} = \sum_{i=1}^{n}S_{ii,t}^{-1}C_{ii,t}$. □

At this stage there are three important considerations to make:

1. The ℓ_{1t} 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 θ_v independently of the value of ν which is not affecting the first order conditions for θ_v.

2. Each function $\ell_{1t}^{(i)}$, defined as the terms between square brackets in (24), only depends on the parameters $\theta_v^{(i)}$ specific to the conditional variance dynamics of asset i. It follows that maximization of ℓ_{1t} can be achieved through n separate optimizations (under the assumptions stated in the proposition). Notice that $\ell_{1t}^{(i)}$ corresponds to the log-likelihood of an exponential distribution.

3. The second part ℓ_{2t} of the log-likelihood depends on the whole set of parameters θ. It depends on the parameters θ_v and θ_c through the first and last terms of (23), which are linear in ν. This implies that it can be maximized with respect to these parameters independently of the value of ν.

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

1. In step 1, the conditional realized variance parameters are estimated by maximizing $\ell_{1t}^{(i)}$ with respect to $\theta_v^{(i)}$ for $i = 1, \ldots, n$.

2. In step 2, the degrees of freedom and the correlation equation parameters are estimated by maximizing the full likelihood ℓ_t with respect to (ν, θ_c), after fixing θ_v to the estimate provided by step 1. The reason for maximizing ℓ_t and not ℓ_{t2} is the presence of the degrees of freedom parameters in ℓ_{1t}.

4 Consistency and asymptotic normality

We show the consistency and asymptotic normality of the two step estimator of the previous section, and we show that it has a QML interpretation. The consistency of the first step estimator of θ_v is easily proven. By equation (24), we can just consider the estimation of the parameter $\theta_v^{(i)}$ for a given i. The first step score vector for observation t is given by

$$m_{1t}(\theta_v^{(i)}) \;=\; \partial \ell_{1t}^{(i)} / \partial \theta_v^{(i)} = (-1 + C_{ii,t}/S_{ii,t})(1/S_{ii,t}) \partial S_{ii,t} / \partial \theta_v^{(i)}. \tag{25}$$

Using the law of iterated expectation it is easy to show that at the true parameter value $\theta_{v,0}^{(i)}$, the expected first step score is equal to 0:

$$E[m_{1t}(\theta_v^{(i)})] = E\left\{ E_{t-1}\left[(-1 + C_{ii,t}/S_{ii,t})(1/S_{ii,t}) \partial S_{ii,t} / \partial \theta_v^{(i)} \right] \right\} = 0 \tag{26}$$

since $E_{t-1}(C_{ii,t}) = S_{ii,t}$, see equation (2). This implies that (24) is a quasi-likelihood function and its maximizer $\hat{\theta}_v$ is a QML estimator (QMLE). Hence, by the results in Bollerslev and Wooldridge (1992), under the stated regularity conditions, consistency and asymptotic normality hold. The result in (26), and the implied interpretation of $\hat{\theta}_v$ as a QMLE, makes our approach robust to misspecification of the distribution of univariate realized volatilities. Even if the Gamma assumption (implied by the Wishart) is not satisfied, we are still able to obtain consistent estimates of the elements of θ_v. In the

literature, a similar estimation problem has been considered 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.

Consistency of the first step estimator $\hat{\theta}_v$ implies consistency of the second step estimators of ν an θ_c obtained as

$$(\hat{\nu}, \hat{\theta}_c')' = \underset{(\nu,\theta_c')}{\operatorname{argmax}} \sum_{t=1}^{T} \ell_t(\nu, \theta_c, \hat{\theta}_v).$$

This result directly follows from the application of Theorem 3.10 in White (1994), under the general regularity conditions that are stated there. Furthermore, consistency of the estimator of the second step parameters still holds if the first step parameters in θ_v are consistently estimated by optimizing an objective function different from $\sum_{t=1}^{T} \ell_{1t}$. So, in principle, different estimation methods could be used without invalidating the properties of second step estimators. For example, following the mainstream literature on univariate modeling of realized variances, we could adopt a maximum likelihood estimator based on the maximization of a lognormal likelihood.

In some applications of multivariate volatility modeling, such as optimal portfolio choice and hedging, to mention some notable examples, the interest of the modeler is in the estimation of the conditional covariance matrix rather than in its distributional properties. As a consequence, the parameters of interest would be the vectors θ_v and θ_c rather than the shape parameter ν.

In this respect, it is worth noting that, similarly to what we already observed for the first step estimation of θ_v, the first order conditions for the maximization of the full likelihood ℓ_t with respect to θ_c are proportional to ν, since the score for observation t is given by

$$\frac{\partial \ell_t}{\partial \theta_c} = \frac{\partial \ell_{2t}}{\partial \theta_c} = -\frac{\nu}{2} \left\{ \frac{\partial \log(|R_t|)}{\partial \theta_c} + \frac{\partial \operatorname{tr}(R_t^{-1} D_t^{-1} C_t D_t^{-1})}{\partial \theta_c} \right\}. \tag{27}$$

This implies that the value of ν is not affecting the estimation of θ_c which can be consistently estimated independently of the value of ν. Furthermore, by standard ML theory results, the gradient in (27), which is the component of the overall score vector related to θ_c, has expected value equal to 0 when $\theta_c = \theta_{c,0}$, where $\theta_{c,0}$ is the value of θ_c in the data generating

process. Namely, by applying standard results on the differentiation of matrix functions, we have

$$\begin{aligned}\partial \ell_t(\theta)/\partial\theta_c &= \frac{\nu}{2}\left\{\mathrm{tr}\left(R_t^{-1}\frac{\partial R_t}{\partial\theta_c}\right)+\mathrm{tr}\left(D_t^{-1}C_tD_t^{-1}\frac{\partial R_t^{-1}}{\partial\theta_c}\right)\right\}\\ &= \frac{\nu}{2}\left\{\mathrm{tr}\left(R_t^{-1}\frac{\partial R_t}{\partial\theta_c}\right)-\mathrm{tr}\left(D_t^{-1}C_tD_t^{-1}R_t^{-2}\frac{\partial R_t}{\partial\theta_c}\right)\right\}. \qquad (28)\end{aligned}$$

Taking expectations conditional on past information I_{t-1}, we obtain

$$E_{t-1}\left(\partial \ell_t(\theta)/\partial\theta_c\right) = \frac{\nu}{2}\left\{\mathrm{tr}\left(R_t^{-1}\frac{\partial R_t}{\partial\theta_c}\right)-\mathrm{tr}\left(D_t^{-1}E_{t-1}(C_t)D_t^{-1}R_t^{-2}\frac{\partial R_t}{\partial\theta_c}\right)\right\}. \qquad (29)$$

At the true parameter value $\theta_c=\theta_{c,0}$, $E_{t-1}(C_t)=D_tR_tD_t$ by equation (2). By substituting this expression in (29) we obtain

$$E_{t-1}\left(\partial \ell_t(\theta)/\partial\theta_c\right)_{\theta_c=\theta_{c,0}} = \frac{\nu}{2}\left\{\mathrm{tr}\left(R_t^{-1}\frac{\partial R_t}{\partial\theta_c}\right)-\mathrm{tr}\left(R_t^{-1}\frac{\partial R_t}{\partial\theta_c}\right)\right\}=0. \qquad (30)$$

Equation (30) has great practical relevance since it implies that, under the usual regularity conditions - see e.g. Newey and McFadden (1994), Wooldridge (1994) - any GMM estimator based on the moment conditions in (28) is a consistent estimator of θ_c. In other words, even if the 'true' distribution of C_t is not Wishart, we can still consider $\hat{\theta}_c$, which is the maximizer of $\sum_{t=1}^{T}\ell_{2t}$ with respect to θ_c, as a QMLE.

Next we focus on the derivation of the asymptotic distribution of the second step estimator $\hat{\gamma}=(\hat{\nu},\hat{\theta}_c')'$ of $\gamma=(\nu,\theta_c')'$. Similarly to Engle (2002a), we can represent the overall estimation problem as a two step GMM estimator where the first and second step moment conditions are given by the first and second step score, respectively:

$$M_{1T}(\theta_v)=\frac{1}{T}\frac{2}{\nu}\sum_{t=1}^{T}\frac{\partial \ell_{1t}(\nu,\theta_v)}{\partial\theta_v}=\frac{1}{T}\sum_{t=1}^{T}m_{1t}(\theta_v) \qquad (31)$$

where $m_{1t}(\theta_v)=\left(m_{1t}(\theta_v^{(1)})',m_{1t}(\theta_v^{(2)})',\ldots,m_{1t}(\theta_v^{(n)})'\right)'$, and

$$M_{2T}(\gamma,\hat{\theta}_v)=\frac{1}{T}\sum_{t=1}^{T}\frac{\partial \ell_t(\gamma,\hat{\theta}_v)}{\partial\gamma}=\frac{1}{T}\sum_{t=1}^{T}m_t(\gamma,\hat{\theta}_v). \qquad (32)$$

We can then apply Theorem 6.1 in Newey and McFadden (1994) to derive the asymptotic distribution of $\hat{\gamma}$. In particular, it follows from these results that

$$\sqrt{T}(\hat{\gamma}-\gamma_{c,0})\underset{d}{\longrightarrow}N(0,V), \qquad (33)$$

where $\gamma_0 = (\nu_0, \theta_{c,0}^{'})^{'}$ is the 'true' value, and

$$V = G_\gamma^{-1} E\left\{ \left[m_t(\gamma_0, \theta_{v,0}) + G_v \psi_t(\theta_{v,0}) \right] \left[m_t(\gamma_0, \theta_{v,0}) + G_v \psi_t(\theta_{v,0}) \right]^{'} \right\} (G_\gamma^{'})^{-1}$$

with

$$\begin{aligned} G_\gamma &= E\{\nabla_\gamma m_t(\gamma_0, \theta_{v,0})\} \\ G_v &= E\{\nabla_{\theta_v} m_t(\gamma_0, \theta_{v,0})\} \\ \psi_t(\theta_{v,0}) &= -F_v^{-1} m_{1t}(\theta_{v,0}) \\ F_v &= E\{\nabla_{\theta_v} m_{1t}(\theta_{v,0})\}. \end{aligned}$$

5 Correlation targeting

In the scalar and diagonal RDCC models presented in Section 2, the dynamic equations depend on constant matrices $\bar{R}$ or $\bar{Q}$, see (5), (8), (11), and (12). To avoid having a large number of parameters (of order n^2) in the numerical maximization of the quasi-likelihood function of the second step of the estimation of the models, which renders the computations impossible in practice for large values of n, we can use 'targeting'. This means a preliminary estimation of these constant matrices by a method of moment estimator. If this estimator is substituted for the corresponding parameter matrix in the quasi-likelihood function (of the second step), the numerical burden is much reduced since the number of parameters is either independent of n (in the scalar models) or linear in n (in the diagonal models). It is desirable that the targeting estimator of a parameter is consistent, even if inefficient. Indeed, since the QML estimators of the remaining parameters depend on the targeting estimator, they cannot be consistent if the targeting estimator is not consistent.

5.1 RDCC

For the RDCC models, we have mentioned in Section 2.1 that their specification does not imply that $\bar{R}$ is the unconditional expectation of P_t and of R_t. Indeed, although $E_{t-1}(C_t) = S_t$, by assumption, see (2), and thus $E(C_t) = E(S_t)$, this does not imply that $E(P_t)$ is equal

to the unconditional correlation matrix $\{\text{diag}[E(S_t)]\}^{-1/2}E(S_t)\{\text{diag}[E(S_t)]\}^{-1/2}$ deduced from the unconditional covariance $E(S_t)$, due to the non-linearity of the transformation from covariances to correlations. Thus a consistent estimator of the unconditional correlation matrix, given by

$$\bar{P}_T = \sum_{t=1}^{T} P_t/T \tag{34}$$

is not consistent for $\bar{R}$ (because $\bar{R}$ is not the unconditional correlation matrix), and targeting $\bar{R}$ by $\bar{P}_T$ is inconsistent. However the finite sample bias of doing that may not be important if C_t is constructed consistently from a large enough number (H) of high-frequency independent returns $r_{t,h}$ assumed to be independently distributed and $N_n(0, S_t/H)$ for $h = 1, 2, \ldots, H$. Indeed this assumption implies that $C_t := C_{t,H} = \sum_{h=1}^{H} r_{t,h}r'_{t,h} \sim W_n(H, S_t/H)$. Then $C_{t,H} \xrightarrow{p} S_t$ as $H \to \infty$, and

$$P_{t,H} = \{\text{diag}(C_{t,H})\}^{-1/2}C_{t,H}\{\text{diag}(C_{t,H})\}^{-1/2} \xrightarrow{p} R_t = \{\text{diag}(S_t)\}^{-1/2}S_t\{\text{diag}(S_t)\}^{-1/2}.$$

Thus for large H, $E_{t-1}(P_{t,H})$ should be close to R_t, hence $E(P_{t,H})$ should be close to $E(R_t)$. Then $\bar{P}_T$ is estimating $E(R_t)$ consistently. If $E(P_t)$ were equal to $E(R_t)$, then $\bar{R}$ would be equal to $E(R_t)$. Since this holds approximately for large enough H, we expect that targeting $\bar{R}$ by $\bar{P}_T$ should not lead to a strong bias if the observed matrices are obtained form high frequency data and free from contamination by microstructure noise.

5.2 cRDCC

For the cRDCC models, we have shown in Section 2.1 that $\bar{Q} = E(Q_t) = E(P_t^*)$. Hence,

$$\bar{P}_T^* = \sum_{t=1}^{T} P_t^*/T \xrightarrow{p} \bar{Q}. \tag{35}$$

In practice, this estimator is not feasible since it depends on the unknown parameters of the conditional variance equations through D_t. We have explained in Section 3 that the parameters of the variance equations can be estimated consistently in the first step of estimation. If D_t is thus replaced by a consistent estimator $\hat{D}_t$ and $\hat{P}_t^*$ stands for (9) with

$\hat{D}_t$ replacingf D_t, then

$$\hat{\bar{P}}_T^* = \sum_{t=1}^{T} \hat{P}_t^* / T \xrightarrow{p} \bar{Q}. \tag{36}$$

Nevertheless $\hat{\bar{P}}_T^*$ cannot be used for targeting $\bar{Q}$ since it depends also on the parameters α and β of (8) through the diagonal elements of Q_t itself. We can obviously estimate the cDCC model by maximizing a quasi-likelihood function with respect to $\bar{Q}$ in addition to α and β. This approach limits the use of the model to small dimensions since the number of parameters is then $O(n^2)$.

To circumvent this problem, we can substitute $\hat{\bar{P}}_T^*(\alpha, \beta)$ (making the dependence of $\hat{\bar{P}}_T^*$ on α and β explicit) for $\bar{Q}$ in (8) in evaluating the quasi-likelihood function for any value of α and β. If $\hat{\alpha}$ and $\hat{\beta}$ are the values that maximize the quasi-likelihood function, we finally estimate consistently $\bar{Q}$ by $\hat{\bar{P}}_T^*(\hat{\alpha}, \hat{\beta})$. This procedure makes it possible to estimate the scalar cRDCC model for a large number of assets.

6 Composite likelihood estimation

The second step of the estimation method presented in Section 3 may not be practicable for very large dimensions. This is due to the need to invert the matrix R_t appearing in the log-likelihood function for each observation. This operation is time consuming for the sample sizes of typical empirical applications. The same issue arises in the estimation of the DCC version of a multivariate GARCH model and has motivated Engle, Shephard, and Sheppard (2008) to use the composite likelihood (CL) method based on the conditional normal distribution for the return vector. It turns out that for the scalar RDCC and RBEKK models, the Wishart assumption also enables us to use the CL method explained below. We present the CL estimator that applies to the second step of the estimation procedure of Section 3 for the scalar RDCC specification (5) using targeting of $\bar{R}$ by $\bar{P}_T$ defined in (34).

The method is based on three results for which we need the following notations. For any square matrix M_t of order n, we denote by $M_{AA,t}$ a square matrix of order n_A extracted

from M_t, which has its main diagonal elements on the main diagonal of M_t. Namely, if A stands for a subset of n_A different indices of $\{1, 2, \ldots, n\}$, $M_{AA,t}$ is the matrix that consists of the intersection of the rows and columns of M_t corresponding to the selection of indices denoted by A. The three results are:

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

R2: If $S_t = D_t R_t D_t$, $S_{AA,t} = D_{AA,t} R_{AA,t} D_{AA,t}$.

R3: If $R_t = (1-\alpha-\beta)\bar{R} + \alpha P_{t-1} + \beta R_{t-1}$, $R_{AA,t} = (1-\alpha-\beta)\bar{R}_{AA} + \alpha P_{AA,t-1} + \beta R_{AA,t-1}$.

Result 1 is a property of the Wishart distribution already mentioned at the end of Section 3. Notice that applied with $n_A = 1$, it corresponds to the result that the margin of a diagonal element of a Wishart matrix is a Gamma distribution, a result that is used in Section 3 to form the log-likelihood for the first step of the estimation procedure. Results 2, given that D_t is diagonal, and 3 are obvious algebraic results.

A CL second 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 for sub-matrices $P_{AA,t}$ corresponding to different choices of indices A. The most obvious choice is to select all the log-likelihoods corresponding to sub-matrices of order 2, i.e. to all the $n(n-1)/2$ correlation coefficients or pairs of assets. In each bivariate Wishart term, the parameters of the conditional variances are fixed at the estimates of the first step, and the matrix $\bar{R}_{AA}$ is set to the corresponding matrix extracted from $\bar{P}_T$. Notice that with these bivariate Wishart log-likelihoods, only matrices of order 2 must be inverted, which can be efficiently programmed. Such a CL is denoted $CL2_t$ for the contribution of observation t. Formally,

$$CL2_t(\nu, \alpha, \beta, \hat{\bar{R}}, \hat{\theta}_v) = \sum_{h=2}^{n} \sum_{k<h} \ell_{hk,t}(\nu, \alpha, \beta, \bar{P}_T^{(hk)}, \hat{\theta}_v^{(h)}, \hat{\theta}_v^{(k)}) \tag{37}$$

with

$$\begin{aligned} \ell_{hk,t}(.) &= \nu \log\left(\frac{\nu}{2}\right) + \frac{\nu-3}{2} \log|C_t^{(hk)}| - \sum_{i=1}^{2} \log\Gamma[(\nu+1-i)/2] \\ &\quad -\frac{\nu}{2} \log|D_t^{(hk)} R_t^{(hk)} D_t^{(hk)}| - \frac{\nu}{2} \mathrm{tr}\{(D_t^{(hk)} R_t^{(hk)} D_t^{(hk)})^{-1} C_t^{(hk)}\}, \end{aligned} \tag{38}$$

where for any matrix M_t, $M_t^{(hk)}$ is the matrix of order 2 extracted at the intersection of rows h and k of M_t. One can use less terms (e.g. consecutive pairs) than the $n(n-1)/2$

terms in (37) especially if the number of terms is very large. One can also use marginal log-likelihoods of sub-matrices of higher dimension, e.g. all or a subset of triplets of indices of $\{1, 2, \ldots, n\}$, to form a CL of order three, denoted $CL3$.

We conjecture that the CL estimators can be shown to be consistent and asymptotically normal if a consistent estimator of $\bar{R}$ is used, but we leave the proof for further work. As anyway we do not have a consistent estimator of $\bar{R}$, this issue is not much relevant and we rely on a simulation study to get insights on the finite sample bias of some CL estimators, and their efficiency with respect to the ML estimator. The simulation study will also give information on the finite sample properties of the estimators for the cRDCC model.

7 Simulation study

This section presents the results of a Monte Carlo simulation study aimed at comparing the finite sample properties of the maximum likelihood (L) and composite maximum likelihood estimators of the parameters of the conditional correlation process. For CL, we use estimators based on pairs (CL2) and on triplets (CL3). We are interested in the bias of the estimators, especially for the RDCC model with targeting since we know that the targeting we use is not consistent, and in their relative efficiencies.

7.1 Simulation design

We consider as data generating process (DGP) a scalar RDCC model and a scalar cRDCC model, where $\alpha = \alpha_0$ and $\beta = \beta_0$, α_0 and β_0 being positive scalars such that $\alpha_0 + \beta_0 < 1$. Both $\bar{R}$ and $\bar{Q}$ are equicorrelated matrices, i.e. matrices having diagonal elements equal to one and off-diagonal elements equal to ρ. The value of ρ is fixed to 0.6 because this is in the range of plausible values for stock markets. Non reported results for different values of ρ show that the specific value of ρ does not change the conclusions drawn from the simulations.

In the simulations for each of the two DGP considered, we have generated 1000 time

series of length $T = 1000$ and $T = 2500$ with three different choices of α_0 and β_0 and six different values of the cross sectional dimension n (5, 15, 25, 50, 75, 100). In all the cases, the value of the degrees of freedom parameter (ν) has been set equal to $3n$. The DGP for the realized variances associated to the RDCC and cRDCC correlation models are GARCH-type recursions defined by

$$S_{ii,t} = (1 - \gamma_i - \delta_i) + \gamma_i C_{ii,t-1} + \delta_i S_{ii,t-1} \qquad i = 1, \ldots, n. \tag{39}$$

In order to allow for some variation in the volatility dynamics, for each $i = 1, \ldots, n$, we draw γ_i and δ_i from dependent uniform distributions defined as

$$\gamma_i \sim U(\gamma_0 - 0.02, \gamma_0 + 0.02), \qquad \delta_i|\gamma_i \sim U(2\delta_0 + \gamma_i - 1 + \epsilon, 1 - \gamma_i - \epsilon),$$

with γ_0=0.05 and δ_0=0.90. This ensures that $E(\gamma_i) = \gamma_0$ (set to 0.05), $E(\delta_i) = \delta_0$ (set to 0.90) and $\gamma_i + \delta_i < 1 - \epsilon$ ($= 0.99$).

In both cases (RDCC and cRDCC), the estimated model corresponds to the model class to which the DGP belongs, so that the estimated model is correctly specified. Estimation is performed in two steps (by each method – L, CL2, and CL3) with correlation targeting for RDCC, and without it for cRDCC. For the latter we use the approach described in Section 2.1 for the estimation of $\bar{Q}$. The first step of the estimation of the likelihood and composite likelihood methods being identical, we do not report the corresponding simulation results.

For CL2, we use all pairs of assets, and for CL3 we use all of them for $n \leq 25$, while we use 5000 randomly selected triplets for $n > 25$ since the number of triplets is then so large that the Monte Carlo study would require too much time.

In order to assess the statistical properties of the estimates we have computed from the simulated values the percentage relative bias (RB) and root mean squared error (RMSE):

$$RB(\theta) = 100 \times \frac{1}{1000} \sum_{i=1}^{1000} \frac{(\hat{\theta}_i - \theta)}{\theta},$$

$$RMSE(\theta) = 100 \times \sqrt{\frac{1}{1000} \sum_{i=1}^{1000} (\hat{\theta}_i - \theta)^2},$$

where $\hat{\theta}_i$ is either $\hat{\alpha}_i$ (when θ is α_0) or $\hat{\beta}_i$ (when θ is β_0), with $(\hat{\alpha}_i, \hat{\beta}_i)$ denoting the estimated parameter values for the i-th simulated series.

7.2 Bias results

The simulation results for the scalar RDCC and cRDCC processes are reported in Tables 2 and 3, respectively. A few conclusions arise from these results:

1) The biases for β_0 are negative and in most cases very small, being smaller than one per cent in absolute value, except for the cRDCC model, $n = 5$, when β_0 is equal to 0.90 and 0.85, where the largest bias is 1.22 per cent.

2) For α_0, the biases are positive (except for $\alpha_0 = 0.03$ by CL2 and CL3). The largest biases occur for $n = 5$ and the RDCC model, with the maximum being 6.35 per cent (4.61 for the cRDCC). They decrease as n increases, being smaller than one per cent for $n \geq 50$ for both models. This decrease is due to the increase of information brought by a larger cross-sectional dimension, given the scalar nature of the models. This effect (bias decrease) is not visible for β_0 (except comparing $n = 5$ and 15) since the biases are very small.

3) The biases are smaller in the cRDCC model than in the RDCC, but the differences are far from impressive. Thus, in the RDCC case, the targeting of the constant matrix of the correlation process by a (presumably hardly) inconsistent estimator does not seem to create a bias problem in the estimation of the dynamic parameters.

4) The biases tend to decrease when T increases from 1000 to 2500; exceptions happen only for α, and the increases are minor.

7.3 Efficiency results

The simulation results for the scalar RDCC and cRDCC processes are reported in Tables 4 and 5, respectively. What is reported in the tables is the ratio of the RMSE of each CL estimator and the L estimator, and of CL2 with respect to CL3.

Several conclusions emerge from these results:

1) As expected, the efficiency of the CL2 and CL3 estimators is smaller than that of the L

Table 2: Relative biases of estimators of α and β, scalar RDCC model

n	L	CL2	CL3	L	CL2	CL3	L	CL2	CL3
T = 1000		α_0=0.03			α_0=0.05			α_0=0.10	
5	5.39	4.65	4.75	6.35	6.30	6.27	5.99	5.98	5.97
15	2.66	1.50	3.65	2.43	2.72	3.66	2.42	2.61	3.17
25	1.06	-0.74	-0.28	1.12	1.13	1.10	1.36	1.53	1.48
50	0.52	-0.99	-0.66	0.45	0.57	0.63	0.72	0.98	0.98
75	0.31	-0.92	-0.75	0.17	0.53	0.43	0.46	0.80	0.76
100	0.26	-0.86	-0.59	0.09	0.54	0.35	0.38	0.79	0.67
T = 2500		α_0=0.03			α_0=0.05			α_0=0.10	
5	6.32	5.85	6.08	6.47	6.19	6.28	5.92	5.74	5.81
15	2.07	1.43	1.59	2.14	2.42	2.31	2.22	2.28	2.58
25	1.35	0.72	0.87	1.30	1.35	1.31	1.32	1.39	1.36
50	0.70	-0.11	0.10	0.58	0.54	0.56	0.66	0.71	0.72
75	0.48	-0.23	-0.11	0.37	0.45	0.50	0.44	0.57	0.49
100	0.42	-0.09	-0.15	0.28	0.50	0.26	0.35	0.56	0.41
T = 1000		β_0=0.95			β_0=0.90			β_0=0.85	
5	-0.49	-0.52	-0.49	-0.74	-0.91	-0.81	-0.54	-0.65	-0.59
15	-0.52	-0.50	-0.52	-0.64	-0.84	-0.72	-0.46	-0.63	-0.56
25	-0.52	-0.41	-0.43	-0.67	-0.81	-0.76	-0.47	-0.63	-0.58
50	-0.53	-0.41	-0.43	-0.66	-0.78	-0.75	-0.46	-0.61	-0.59
75	-0.54	-0.42	-0.43	-0.65	-0.78	-0.73	-0.45	-0.60	-0.57
100	-0.54	-0.42	-0.43	-0.65	-0.76	-0.71	-0.46	-0.59	-0.55
T = 2500		β_0=0.95			β_0=0.90			β_0=0.85	
5	-0.20	-0.18	-0.19	-0.29	-0.29	-0.28	-0.22	-0.21	-0.21
15	-0.18	-0.13	-0.14	-0.26	-0.33	-0.30	-0.19	-0.25	-0.25
25	-0.20	-0.15	-0.16	-0.26	-0.31	-0.29	-0.19	-0.24	-0.22
50	-0.20	-0.14	-0.15	-0.25	-0.29	-0.28	-0.18	-0.23	-0.22
75	-0.20	-0.15	-0.14	-0.25	-0.31	-0.28	-0.18	-0.24	-0.20
100	-0.20	-0.15	-0.15	-0.25	-0.31	-0.27	-0.18	-0.25	-0.22

Table 3: Relative biases of estimators of α and β, scalar cRDCC model

n	L	CL2	CL3	L	CL2	CL3	L	CL2	CL3
T = 1000	α_0=0.03			α_0=0.05			α_0=0.10		
5	2.97	2.46	2.50	3.52	3.59	3.57	4.51	4.59	4.61
15	1.32	0.83	1.00	1.16	1.91	1.74	1.68	2.17	2.09
25	0.58	-1.19	-0.71	0.52	0.64	0.61	0.99	1.25	1.22
50	0.23	-1.33	-0.93	0.06	0.15	0.20	0.46	0.76	0.74
75	0.23	-1.33	-0.93	-0.10	0.53	0.41	0.28	0.81	0.74
100	0.08	-1.35	-1.01	-0.13	0.18	0.07	0.22	0.60	0.53
T=2500	$\alpha_0 = 0.03$			$\alpha_0 = 0.05$			$\alpha_0 = 0.10$		
5	4.16	3.69	3.91	3.79	3.57	3.68	4.54	4.39	4.50
15	1.53	1.36	1.42	1.27	1.62	1.53	1.62	1.84	1.80
25	0.93	0.31	0.48	0.76	0.86	0.83	1.01	1.12	1.11
50	0.47	-0.53	-0.25	0.29	0.07	0.16	0.48	0.46	0.50
75	0.33	-0.42	-0.25	0.15	0.17	0.18			
100	0.29	-0.34	-0.20						
T=1000	$\beta_0 = 0.95$			$\beta_0 = 0.90$			$\beta_0 = 0.85$		
5	-0.67	-0.71	-0.66	-1.03	-1.20	-1.10	-1.09	-1.22	-1.16
15	-0.59	-0.56	-0.56	-0.76	-0.94	-0.88	-0.66	-0.83	-0.78
25	-0.56	-0.44	-0.46	-0.72	-0.87	-0.81	-0.58	-0.75	-0.70
50	-0.55	-0.42	-0.45	-0.68	-0.79	-0.76	-0.51	-0.65	-0.62
75	-0.55	-0.42	-0.45	-0.66	-0.82	-0.77	-0.48	-0.65	-0.62
100	-0.55	-0.42	-0.44	-0.66	-0.79	-0.73	-0.48	-0.63	-0.59
T=2500	$\beta_0 = 0.95$			$\beta_0 = 0.90$			$\beta_0 = 0.85$		
5	-0.35	-0.33	-0.33	-0.52	-0.53	-0.52	-0.74	-0.75	-0.75
15	-0.25	-0.24	-0.24	-0.34	-0.41	-0.38	-0.36	-0.43	-0.41
25	-0.23	-0.18	-0.19	-0.31	-0.36	-0.34	-0.30	-0.35	-0.34
50	-0.21	-0.13	-0.15	-0.27	-0.28	-0.28	-0.23	-0.26	-0.26
75	-0.21	-0.15	-0.16	-0.26	-0.30	-0.29			
100	-0.21	-0.15	-0.16						

Table 4: Ratios of RMSE of estimators of α and β, scalar RDCC model

n	$\mathbf{T = 1000}$	α_0=0.03	β_0=0.95	α_0=0.05	β_0=0.90	α_0=0.10	β_0=0.85
5	CL2/L	1.17	1.22	1.23	1.33	1.15	1.29
5	CL3/L	1.08	1.09	1.08	1.12	1.05	1.11
5	CL3/CL2	0.93	0.89	0.87	0.84	0.91	0.86
15	CL2/L	2.02	1.34	2.04	1.73	1.65	1.80
15	CL3/L	1.54	1.13	1.59	1.28	1.44	1.37
15	CL3/CL2	0.76	0.85	0.78	0.74	0.87	0.76
25	CL2/L	3.15	1.20	3.05	1.75	2.30	1.89
25	CL3/L	2.32	1.04	2.24	1.42	1.77	1.53
25	CL3/CL2	0.73	0.87	0.74	0.81	0.77	0.81
50	CL2/L	6.34	1.14	6.08	1.69	3.97	1.90
50	CL3/L	4.35	0.99	4.31	1.42	2.94	1.59
50	CL3/CL2	0.69	0.87	0.71	0.84	0.74	0.84
75	CL2/L	8.80	1.13	9.18	1.75	5.70	1.97
75	CL3/L	6.15	0.96	6.26	1.40	3.97	1.56
75	CL3/CL2	0.70	0.85	0.68	0.80	0.70	0.80
100	CL2/L	10.77	1.08	11.36	1.67	6.71	1.88
100	CL3/L	7.88	0.95	7.95	1.36	4.77	1.53
100	CL3/CL2	0.73	0.89	0.70	0.81	0.71	0.82
n	$\mathbf{T = 2500}$	α_0=0.03	β_0=0.95	α_0=0.05	β_0=0.90	α_0=0.10	β_0=0.85
5	CL2/L	1.19	1.27	1.12	1.27	1.05	1.26
5	CL3/L	1.04	1.08	1.03	1.10	1.01	1.09
5	CL3/CL2	0.88	0.85	0.92	0.86	0.96	0.87
15	CL2/L	1.93	1.63	1.85	1.90	1.42	1.97
15	CL3/L	1.50	1.31	1.48	1.51	1.28	1.51
15	CL3/CL2	0.78	0.80	0.80	0.79	0.90	0.77
25	CL2/L	2.66	1.57	2.58	2.13	1.88	2.28
25	CL3/L	1.99	1.27	1.96	1.66	1.50	1.77
25	CL3/CL2	0.75	0.81	0.76	0.78	0.80	0.78
50	CL2/L	4.55	1.48	4.82	2.30	3.08	2.59
50	CL3/L	3.22	1.19	3.37	1.74	2.25	1.96
50	CL3/CL2	0.71	0.80	0.70	0.76	0.73	0.76
75	CL2/L	6.87	1.53	7.18	2.36	4.58	2.74
75	CL3/L	4.62	1.17	5.06	1.78	3.16	2.02
75	CL3/CL2	0.67	0.77	0.70	0.75	0.69	0.74
100	CL2/L	7.78	1.45	8.86	2.27	5.32	2.64
100	CL3/L	5.46	1.12	6.05	1.68	3.66	1.93
100	CL3/CL2	0.70	0.77	0.68	0.74	0.69	0.73

A value larger (smaller) than 1 indicates that the estimator in the numerator is less (more) efficient than the estimator in the denominator.

Table 5: Ratios of RMSE of estimators of α and β, scalar cRDCC model

n	**T = 1000**	α_0=0.03	β_0=0.95	α_0=0.05	β_0=0.90	α_0=0.10	β_0=0.85
5	CL2/L	1.16	1.18	1.29	1.29	1.20	1.22
5	CL3/L	1.08	1.07	1.11	1.11	1.08	1.09
5	CL3/CL2	0.93	0.91	0.86	0.86	0.90	0.89
15	CL2/L	2.33	1.28	2.35	1.60	1.92	1.56
15	CL3/L	1.76	1.13	1.81	1.34	1.56	1.34
15	CL3/CL2	0.76	0.88	0.77	0.84	0.81	0.86
25	CL2/L	3.35	1.16	3.33	1.70	2.62	1.72
25	CL3/L	2.46	1.02	2.44	1.39	1.99	1.44
25	CL3/CL2	0.74	0.88	0.73	0.82	0.76	0.84
50	CL2/L	6.23	1.10	6.08	1.65	4.59	1.79
50	CL3/L	4.42	0.98	4.34	1.38	3.34	1.50
50	CL3/CL2	0.71	0.89	0.71	0.84	0.73	0.84
75	CL2/L	6.23	1.10	9.36	1.75	7.26	1.92
75	CL3/L	4.42	0.98	6.48	1.43	5.12	1.58
75	CL3/CL2	0.71	0.89	0.69	0.82	0.71	0.82
100	CL2/L	11.26	1.07	10.75	1.67	8.15	1.85
100	CL3/L	7.79	0.95	7.42	1.37	5.69	1.52
100	CL3/CL2	0.69	0.89	0.69	0.82	0.70	0.82
n	**T = 2500**	α_0=0.03	β_0=0.95	α_0=0.05	β_0=0.90	α_0=0.10	β_0=0.85
5	CL2/L	1.20	1.17	1.20	1.23	1.09	1.15
5	CL3/L	1.06	1.06	1.07	1.09	1.03	1.06
5	CL3/CL2	0.89	0.90	0.89	0.88	0.95	0.92
15	CL2/L	2.11	1.46	2.16	1.76	1.64	1.60
15	CL3/L	1.63	1.23	1.68	1.43	1.37	1.35
15	CL3/CL2	0.77	0.85	0.78	0.81	0.84	0.84
25	CL2/L	3.05	1.45	3.12	1.96	2.18	1.82
25	CL3/L	2.27	1.20	2.33	1.56	1.72	1.49
25	CL3/CL2	0.74	0.83	0.75	0.80	0.79	0.82
50	CL2/L	5.50	1.38	5.80	2.15	3.73	2.19
50	CL3/L	3.89	1.12	4.13	1.67	2.73	1.72
50	CL3/CL2	0.71	0.81	0.71	0.77	0.73	0.79
75	CL2/L	7.77	1.42	8.49	2.19		
75	CL3/L	5.45	1.14	5.94	1.68		
75	CL3/CL2	0.70	0.80	0.70	0.77		
100	CL2/L	9.39	1.42				
100	CL3/L	6.59	1.14				
100	CL3/CL2	0.70	0.80				

A value larger (smaller) than 1 indicates that the estimator in the numerator is less (more) efficient than the estimator in the denominator.

estimator. The efficiency loss is very large especially for α_0 and large n. Remember that for $n = 50$ or more, not all triplets are used, implying that the inefficiency ratios are not really comparable to those for $n \leq 25$. Actually for $n = 100$, 5000 pairs only correspond to 3 per cent of all pairs.

2) The efficiency of CL2 is smaller than that of CL3, but the gap is not so high as with respect to L (at most 33 per cent, and in most cases much less).

3) The efficiency gap in favor of the L estimator increases with n.

4) Inefficiency ratios tend to be lower for $T = 2500$ than for $T = 1000$, especially when they are high.

5) The efficiency ratios for the two types of models are generally very close to each other.

7.4 Computing time, models and estimators

The computing time for estimating the RDCC model is smaller than for the cRDCC model. For $n = 5$, the ratio of the former to the latter is FILL, for $n = 50$, FILL, and for $n = 100$, FILL. This pleads in favor of using the scalar RDCC model in empirical work, given the comparable bias and efficiency properties of the two models.

The computing time for estimating the scalar RDCC model (with 2500 observations) is about six times smaller by CL2 than by L for $n = 50$ (2.5 for the cRDCC). As a function of n, this ratio has an inverse U-shape.

8 Empirical illustration

We consider stock returns from 50 assets traded in the NYSE and NASDAQ, their tickers being shown in Table 6. The sample period spans January 05, 1999 to May 22, 2007, which amounts to 2084 trading days. The dataset has been cleaned from weekends, holidays and early closing days. Days with many consecutive missing values or constant prices have also been removed. Rare missing values have been linearly interpolated. The realized conditional covariances are based on intraday returns computed from 6-minute intervals last mid-quotes. Since the daily trading period of the NYSE and NASDAQ is 6.5 hours,

this amounts to 65 intraday observations per day. Relying on the same arguments of Andersen, Bollerslev, Frederiksen, and Nielsen (2010), we estimate the scalar DCC model using daily open-to-close realized covariances.

Table 6: Tickers

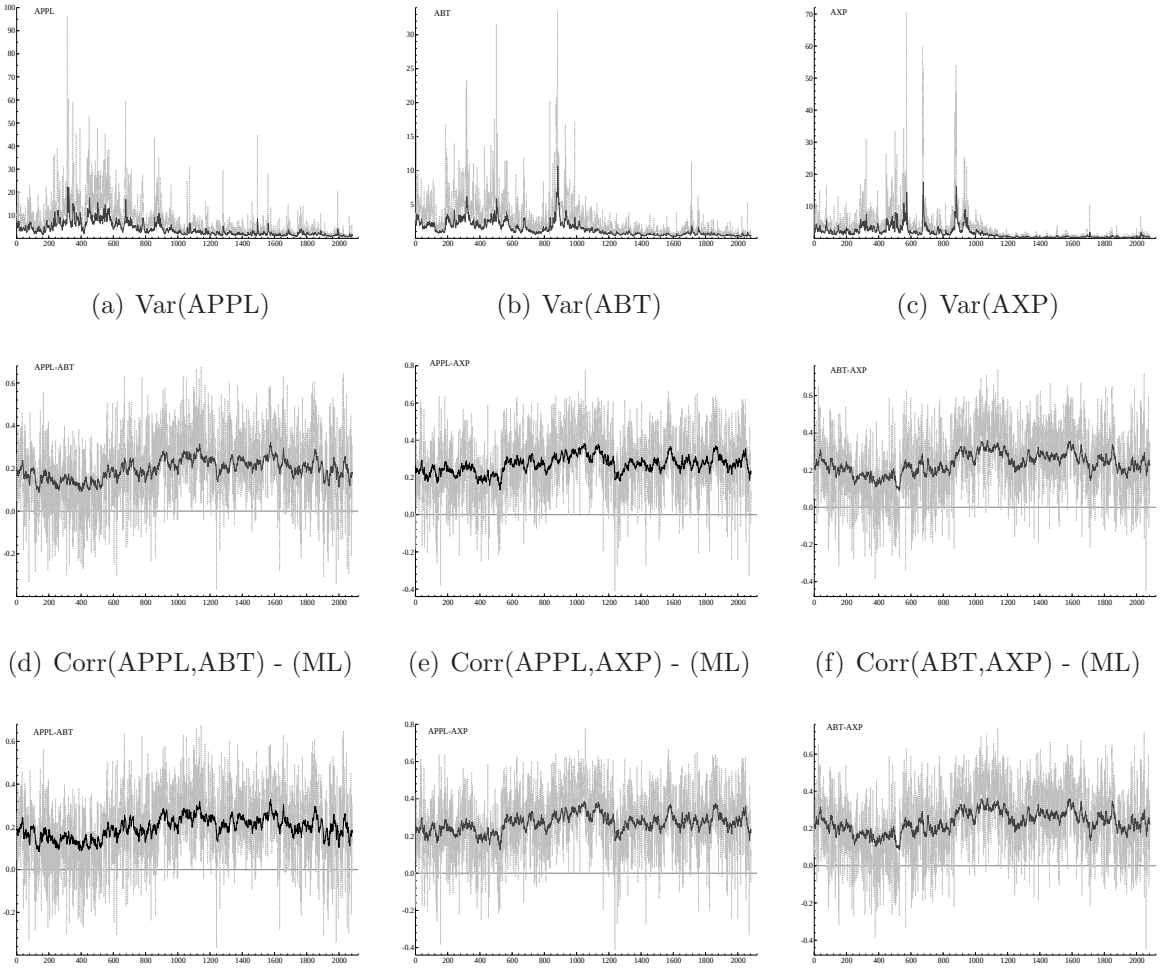
AAPL	BMY	CSCO	EXC	HD	JNJ	MMM	SLB	ORCL	WFC
ABT	BP	CVX	F	HNZ	JPM	MOT	T	PEP	WMT
AXP	C	DELL	FDX	HON	KO	MRK	TWX	PFE	WYE
BA	CAT	DIS	GE	IBM	LLY	MS	UN	PG	XOM
BAC	CL	EK	GM	INTC	MCD	MSFT	VZ	QCOM	XRX

The model is defined by (1), (4) and (5) where the elements on the diagonal of S_t in (4) are specified as $S_{ii,t} = \omega_i + \gamma_i C_{ii,t-1} + \delta_i S_{ii,t-1}$, $i = 1, ..., n$. Table 7 reports parameter estimates and standard deviations for a portfolio made of AAPL, ABT, AXP. The parameters driving the correlation dynamics have been estimated using the QML and the CL approach respectively (we call the former QML because we do not assume that the Wishart assumption is correct, but actually it is the same estimator as what we call L in Section 7). All estimates are significant at standard confidence levels and we do not observe sensible differences between the QMLE and the CLE. Figure 1 shows that the model can fit well the data and suggests that the RDCC can adequately smooth the noisiness and track fairly well the dynamics of realized variances and correlations.

Table 7: AAPL, ABT and APX

	Variances				Correlation	
	γ_i	δ_i	ω_i		α	β
AAPL	0.1453	0.6891	0.4901	QML	0.0564	0.8988
	(0.0274)	(0.0588)	(0.1333)		(0.0238)	(0.0475)
ABT	0.1095	0.7720	0.1076	CL	0.0591	0.8831
	(0.0208)	(0.0427)	(0.0472)		(0.0169)	(0.0333)
AXP	0.1625	0.6764	0.0585			
	(0.0265)	(0.0509)	(0.0277)			

Note: robust standard errors in parentheses.



(a) Var(APPL) (b) Var(ABT) (c) Var(AXP)

(d) Corr(APPL,ABT) - (ML) (e) Corr(APPL,AXP) - (ML) (f) Corr(ABT,AXP) - (ML)

(g) Corr(APPL,ABT) - (CL) (h) Corr(APPL,AXP) - (CL) (i) Corr(ABT-AXP) - (CL)

Figure 1: Fitted variances and correlations (solid) vs. realizations (dashed)

Engle and Sheppard (2001) suggest that for GARCH-DCC models estimated on daily returns, QML estimates of the dynamic parameters seem biased when the cross-sectional dimension n becomes large. The QML and CL_2 (based on all possible pairs) estimates of the correlation dynamics for the 50 assets are reported in Table 8. The QML estimator shows clear signs of bias as the cross-sectional dimension increases. When modeling the covariance of the portfolio of 50 assets, the evidence of bias is striking. In this case, the QML estimator produces correlations that are close to be constant. A random sample of correlation paths generated under the QML and the CL estimators are shown in Figure 2.

Table 8: Scalar RDCC: QML vs. CL estimates

	QML		CL	
n	α	β	α	β
5	0.0412	0.9308	0.0625	0.8952
	(0.0123)	(0.0231)	(0.0089)	(0.0162)
15	0.0201	0.9664	0.0574	0.9052
	(0.0027)	(0.0051)	(0.0027)	(0.0048)
30	0.0139	0.9743	0.0591	0.9024
	(0.0012)	(0.0024)	(0.0013)	(0.0023)
50	0.0105	0.9794	0.0575	0.9058
	(0.0006)	(0.0015)	(0.0008)	(0.0014)

Note: robust standard errors in parentheses. The CL estimator is based on CL_2, defined in (37).

Unlike the QML estimator, the CL estimator does not seem to suffer from the usual downward bias in the innovation parameter α and upward bias in the smoothing parameter β, even when the cross-sectional dimension becomes large. Both parameters seem to be insensitive to the cross-sectional dimension. This result is in line with Engle, Shephard, and Sheppard (2008).

To better illustrate this point we generalize the results reported in Table 8 by contrasting the behaviour of the QML and Cl estimators obtained for a large number of random portfolios of various dimensions (n =2, 3, 5, 10 and 20 respectively[1]) selected from the pool of 50 assets considered in this Section. The aim is to validate the evidence reported in Table 8, i.e., whether the discrepancies reported between the QML and the Cl estimates appear systematically as the cross sectional dimension increases and independently of the composition of the portfolio, and to assess to what extent these estimators are affected by parameter heterogeneity under (possible) model misspecification, i.e., when imposing common dynamics for the conditional correlations[2]. To this end, in Figures 3, 4 and 5 we

[1]The number of random portfolios considered is 1,225 for $n = 2$, 19,600 for $n = 3$, 150,212 for $n = 5$, 60,505 for $n = 10$ and 30,653 for $n = 20$.

[2]Notice that when $n = 2$, parameter heterogeneity is captured to the highest extent since we model individually each correlation. Also, in this case QML and CL coincide.

report distributions, box-plot and rank order representations of the parameters estimates for each set of portfolios. Results are compared with the parameter estimates obtained using the entire set of available assets ($n = 50$).

Consistently with the results reported in Table 8, Figure 4 shows that, as the cross sectional dimension increases, the CL estimator clearly tends to average correlation dynamics. Indeed, average dynamics, measured by means (medians) of the correlation parameters estimated for different portfolio compositions of fixed size, are consistent across cross sectional dimensions and in line with the estimates obtained for a portfolio of $n = 50$ assets (Figure 5(c)-(d)).

Contrary, when n is sufficiently large, the QML estimator seems to be unable to capture correlation dynamics. The discrepancy between the QML and CL estimators becomes striking. The QML estimator appears unable to absorb the heterogeneity in the correlation dynamics. In particular, the innovation parameter shows clear signs of downward bias and, as n increases, it drifts well below the smallest value spanned by α when estimated over portfolios of two assets (see Figure 5(a)). The smoothing parameter, β, shows a spectacular behaviour Figure 5(b)).

9 Conclusions

We have proposed a new dynamic model for realized covariance matrices. The model can be specified and estimated in two steps, the first one for the variances, and the second for the correlation matrix. The first step can also be split into individual steps. This enables to apply the model to matrices of large dimension, where large in this context means of the order of fifty. This is a significant progress relative to existing models. The possibility to split the estimation in steps comes from the use of a scalar DCC model, and from the use of the Wishart distribution. The latter assumption also allows us to use a composite likelihood approach which might be especially relevant for very large dimensions (since the usual ML may then be infeasible computationally). The Wishart assumption should not be viewed as a big drawback given that the estimation has a quasi-likelihood

interpretation. Simulations show that there is, as expected, an efficiency loss to using a composite likelihood estimation rather than a full likelihood. Several extensions are on our research agenda.

References

Aielli, G. (2008): "Consistent estimation of large scale dynamic conditional correlations.," University of Messina, Department of Economics, Statistics, Mathematics and Sociology, Working paper n. 47.

Andersen, T., T. Bollerslev, F. Diebold, and H. Ebens (2001): "The Distribution of Realized Stock Return Volatility," *Journal of Financial Economics*, 61, 43–76.

Andersen, T., T. Bollerslev, F. Diebold, and P. Labys (2001): "The Distribution of Exchange Rate Volatility," *Journal of the American Statistical Association*, 96, 42–55.

——— (2003): "Modeling and Forecasting Realized Volatility," *Econometrica*, 71, 579–625.

Andersen, T., T. Bollerslev, P. Frederiksen, and M. Nielsen (2010): "Continuous-time Models, Realized Volatilities and Testable Distributional Implications for Daily Stock Returns," *Journal of Applied Econometrics*, 25, 233261.

Anderson, T. (1984): *An Introduction to Multivariate Statistical Analysis*. Wiley, second edn.

Bauer, G. H., and K. Vorkink (2006): "Multivariate Realized Stock Market Volatility," Working Paper 2007-20, Bank of Canada.

Bollerslev, T., and J. Wooldridge (1992): "Quasi-maximum Likelihood Estimation and Inference in Dynamic Models with Time-varying Covariances," *Econometric Reviews*, 11, 143–172.

Bonato, M., M. Caporin, and A. Ranaldo (2009): "Forecasting realized (co)variances with a block structure Wishart autoregressive model," Working Papers 2009-03, Swiss National Bank.

Chiriac, R., and V. Voev (2010): "Modelling and Forecasting Multivariate Realized Volatility," *Journal of Applied Econometrics*, Published online in Wiley InterScience, DOI: 10.1002/jae.1152.

Corsi, F. (2009): "A simple approximate long-memory model of realized volatility," *Journal of Financial Econometrics*, 7, 174–196.

Engle, R. (2002a): "Dynamic Conditional Correlation - a Simple Class of Multivariate GARCH Models," *Journal of Business and Economic Statistics*, 20, 339–350.

Engle, R., and F. Kroner (1995): "Multivariate Simultaneous Generalized ARCH," *Econometric Theory*, 11, 122–150.

Engle, R., and J. Rangel (2008): "The spline-GARCH model for low-frequency volatility and its global macroeconomic causes," *Review of Financial Studies*, 21, 1187–1222.

Engle, R., N. Shephard, and K. Sheppard (2008): "Fitting vast dimensional time-varying covariance models," NYU Working Paper No. FIN-08-009.

Engle, R., and K. Sheppard (2001): "Theorical 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., and G. M. Gallo (2006): "A multiple indicators model for volatility using intra-daily data," *Journal of Econometrics*, 131, 3–27.

Engle, R. F., and J. Russell (1998): "Autoregressive Conditional Duration; A New Model for Irregularly Spaced Transaction Data," *Econometrica*, 66, 1127–1162.

Golosnoy, V., B. Gribisch, and R. Liesenfeld (2010): "The Conditional Autoregressive Wishart Model for Multivariate Stock Market Volatility," Economics Working paper 2010-07, Christian-Albrechts-University of Kiel, Department of Economics.

Gouriéroux, C., J. Jasiak, and R. Sufana (2009): "The Wishart Autoregressive process of multivariate stochastic volatility," *Journal of Econometrics*, 150, 167–181.

Jin, X., and J. M. Maheu (2010): "Modelling Realized Covariances and Returns," WP 11-08, The Rimini Center for Economic Analysis.

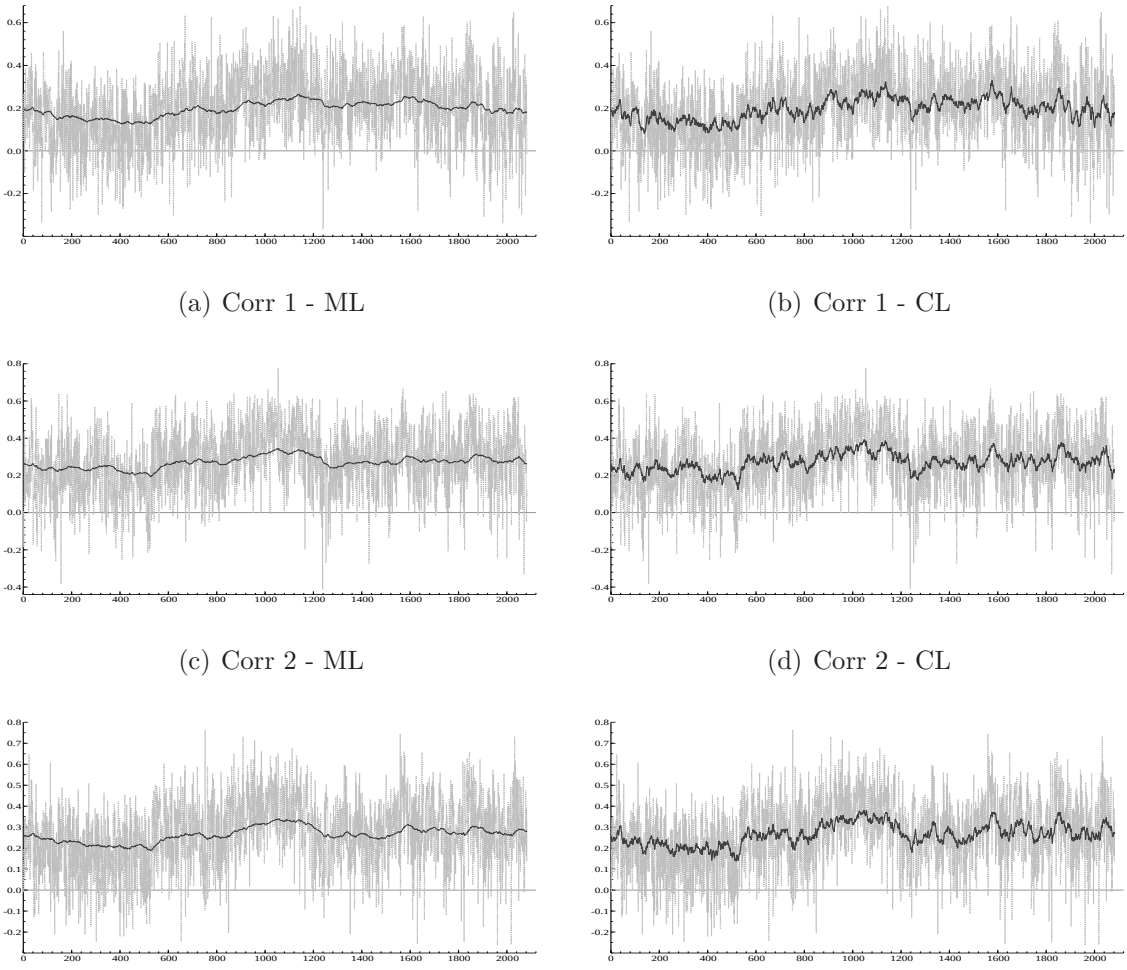
Newey, W., and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, ed. by R. Engle, and D. McFadden, vol. IV, chap. 36, pp. 2113–2245. Eslevier.

Noureldin, D., N. Shephard, and K. Sheppard (2011): "Multivariate High-Frequency-Based Volatility (HEAVY) Models," Working Paper, Department of Economics, University of Oxford.

Tse, Y., and A. Tsui (2002): "A Multivariate GARCH Model with Time-Varying Correlations," *Journal of Business and Economic Statistics*, 20, 351–362.

White, H. (1994): *Estimation, Inference and Specification Analysis*, Econometric Society Monographs No. 22. Cambridge University Press.

Wooldridge, J. (1994): "Estimation and Inference for Dependent Processes," in *Handbook of Econometrics*, ed. by R. Engle, and D. McFadden, vol. IV, pp. 2639–2738. Elsevier.



(a) Corr 1 - ML

(b) Corr 1 - CL

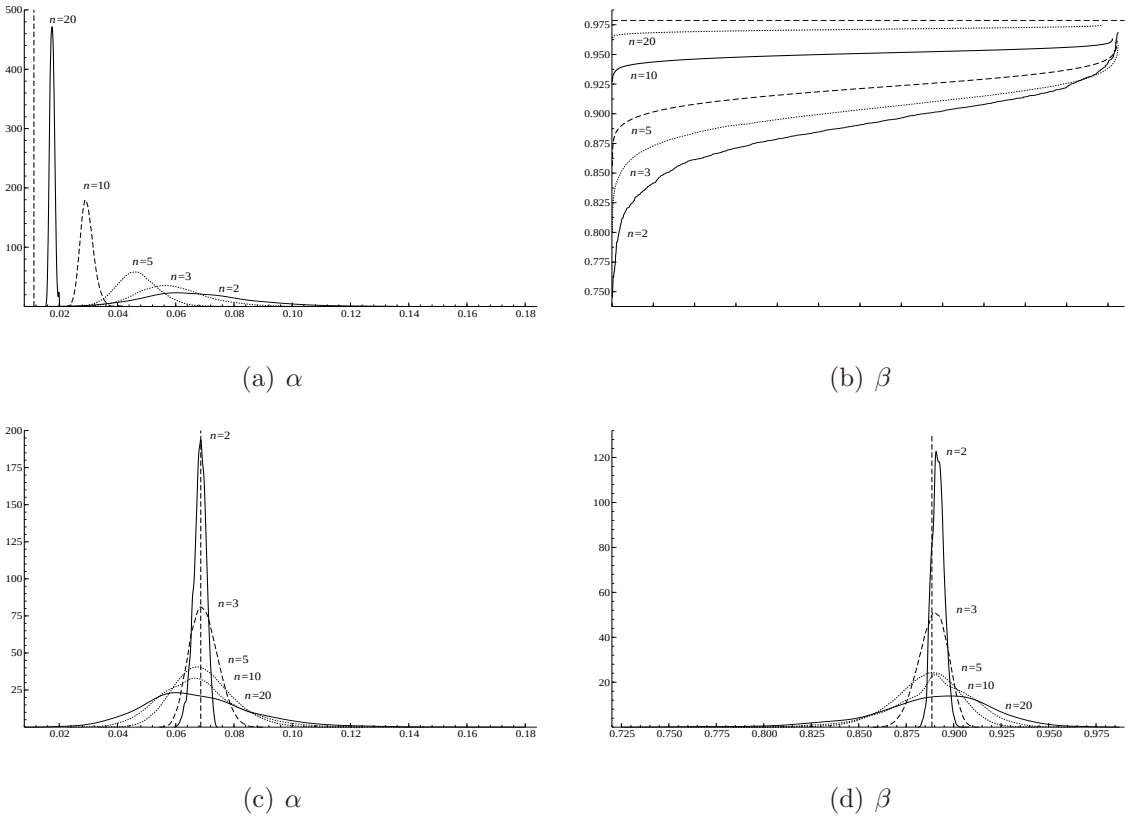
(c) Corr 2 - ML

(d) Corr 2 - CL

(e) Corr 3 - ML

(f) Corr 3 - CL

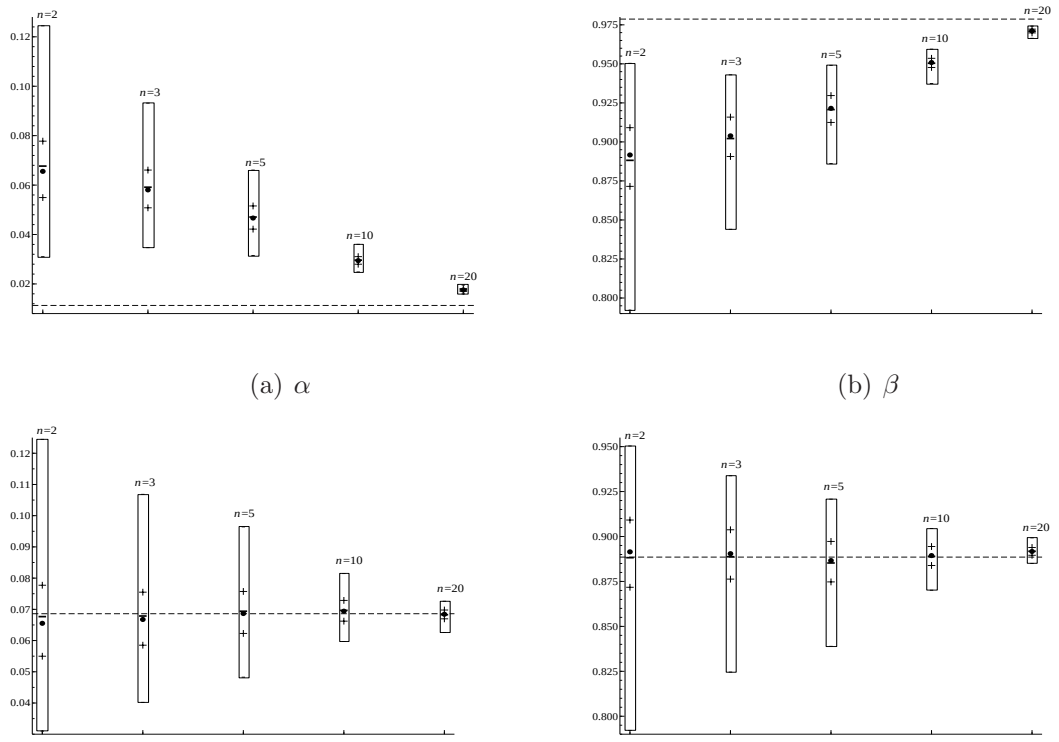
Figure 2: Fitted variances and correlations (solid) vs. realizations (dashed). Parameters estimated using QML (left) and CL (right).



(a) α (b) β

(c) α (d) β

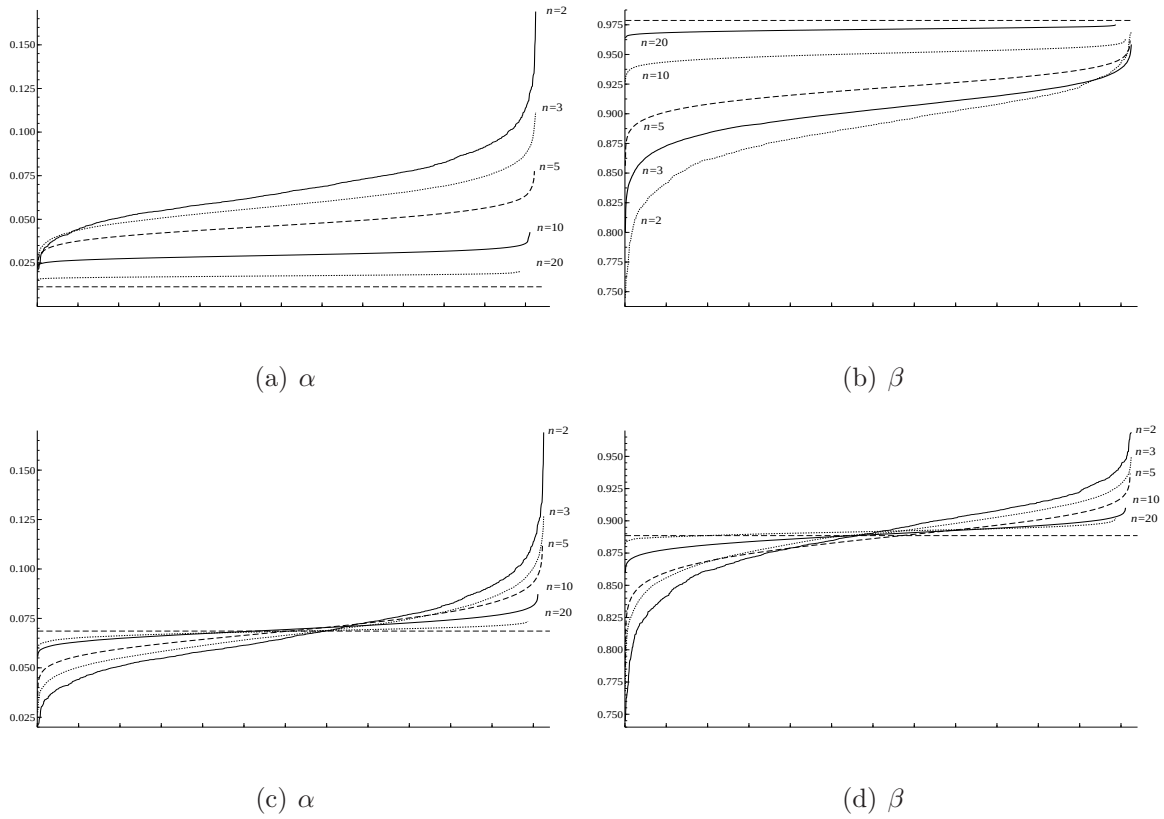
Figure 3: Distribution of the QML (top) and CL (bottom) correlation parameters estimates for a large number of random portfolios of different sizes n. The vertical dashed line represents the parameter estimates for the portfolio made of all the available assets ($n = 50$).



(a) α (b) β

(c) α (d) β

Figure 4: Box-plot representation of the QML (top) and CL (bottom) correlation parameters estimates for a large number of random portfolios of different sizes n. The bounds of the box represent the 1% and 99% quantiles respectively. The +s represent the 25% and 75% quantiles, while the dot and the dash represent the median and mean respectively. The horizontal dashed line represents the parameter estimates for the portfolio made of all the available assets ($n = 50$).



(a) α (b) β

(c) α (d) β

Figure 5: Rank representation of the QML (top) and CL (bottom) correlation parameters estimates for a large number of random portfolios of different sizes n. The horizontal dashed line represents the parameter estimates for the portfolio made of all the available assets ($n = 50$).