

Regularized semiparametric estimation of high dimensional dynamic conditional covariance matrices*

Claudio Morana[†]

University of Milano-Bicocca,

Center for European Studies (CefES),

Center for Research on Pensions and Welfare Policies (CeRP)

and Rimini Centre for Economic Analysis (RCEA).

October 2018

Abstract

This paper proposes a three-step estimation strategy for dynamic conditional correlation (DCC) models. In the first step, conditional variances for individual and aggregate series are estimated by means of QML equation by equation. In the second step, conditional covariances are estimated by means of the polarization identity, and conditional correlations are estimated by their usual normalization. In the third step, the two-step conditional covariance and correlation matrices are regularized by means of a new non-linear shrinkage procedure and optimally smoothed. Due to its scant computational burden, the proposed regularized semiparametric DCC model (RSP-DCC) allows to estimate high dimensional conditional covariance and correlation matrices. An application to global minimum variance portfolio is also provided, confirming that RSP-DCC is a simple and viable alternative to existing DCC models.

Keywords: Conditional covariance, Dynamic conditional correlation model, Semiparametric dynamic conditional correlation model, Multivariate GARCH.

JEL classification: C32, C58

*A previous version of this paper has been presented at the 2017 World Banking and Finance Symposium, the 11th International Conference on Computational Financial Econometrics, the XIX Workshop on Quantitative Finance, the 1st Italian Workshop of Econometrics and Empirical Economics: Panel Data Models and Applications, the 2nd International Conference on Econometrics and Statistics, and at seminars at the University of Maastricht and the University of Paris-Dauphine. The author is grateful to conference and seminar participants for constructive comments.

[†]Address for correspondence: Claudio Morana, Università di Milano-Bicocca, Dipartimento di Economia, Metodi Quantitativi e Strategie di Impresa, Piazza dell'Ateneo Nuovo 1, 20126, Milano, Italy. E-mail: claudio.morana@unimib.it.

1 Introduction

Since the seminal contribution of Engle (2002), the literature on Dynamic Conditional Correlation (DCC) models has rapidly developed, providing alternative models and improvements upon the original formulation.¹ The DCC model is a generalization of the Constant Conditional Correlation model of Bollerslev (1990, CCC), which assumes time invariant, but pairwise specific correlations, and posits the correlation matrix to be a transformation of a latent matrix, which is a function of past devolatilized innovations. It is grounded on a two-step estimation procedure, that, in principle, allows implementation in high dimensional frameworks, particularly in its Dynamic Conditional Equi-Correlation (DECO) form (Engle and Kelly, 2008); in the latter framework, however, the time varying correlations are constrained to be equal across series. Similar to CCC, DECO therefore relies on assumptions on conditional correlation dynamics which are unlikely to be supported by the data. An asymmetric DCC specification has also been introduced by Cappiello et al. (2006). While a rigorous asymptotic theory for DCC, as well as for multivariate GARCH models in general, is still missing, the two-step estimation strategy is conjectured to yield consistent and asymptotically normal estimation (Engle and Sheppard, 2001; Engle and Kelly, 2008)².

However, estimating the DCC model in high dimensional frameworks is challenging. Firstly, as recognized in Engle and Sheppard (2001), the parameters of the latent process driving correlation dynamics tend to be severely downward biased, resulting in correlation trajectories which are increasingly smooth and eventually flat. Secondly, as shown by Aielli (2009), the sample covariance matrix estimator of the long-run covariance target is asymptotically distorted. A similar problem also affects the alternative Varying Conditional Correlation model of Tse and Tsui (2002), where the correlation matrix is modelled directly as a function of past correlations of devolatilized innovations. Aielli (2009) fixes the consistency problem in the DCC model by correcting the correlation latent process in such a way that it has martingale difference innovations (cDCC). Thirdly, the inversion of the correlation matrix might become problematic, due to ill-conditioning. In this respect, Patel et al. (2014) use the Aielli (2009) cDCC specification and propose a composite likelihood estimator, based on the averaging of the quasi-likelihood functions constructed from bivariate marginal densities of observations. The approach shows some similarities with the MacGyver strategy of Engle (2009) and the Flexible Multivariate GARCH model of Ledoit et al. (2003), where, however, the bivariate log-likelihoods are separately maximized. The common advantage of the latter approaches is that they avoid the inversion of high dimensional conditional correlation matrices. Patel et al. (2014) also find that the composite likelihood cDCC model is not any longer affected by downward biased parameters estimation.

However, as shown by Hafner and and Reznikova (2012), long-run targeting of the covariance matrix using the sample covariance matrix estimator might still be problematic and exposed to distortions also in the composite likelihood cDCC framework, particularly as the concentration ratio N/T , i.e. the ratio of the cross-sectional (N) to the temporal (T) dimensions, gets close or larger than one. In this context, ill-conditioning manifests the well-known impossibility of estimating $O(N^2)$ parameters from $O(N^2)$ noisy obser-

¹See for instance the survey accounts of Bauwens et al. (2006), Silvennoinen and Terasvirta (2009), Engle (2009). See also Caporin and McAleer (2013).

²See McAleer et al. (2008) for advances in this respect. See also Paoletta and Polak (2015) for alternative formulations of the CCC model in the non-Gaussian setting.

vations. As a solution, they suggest to shrink the sample covariance estimator using the linear shrinkage approach of Ledoit and Wolf (2004). Relative to Patel et al. (2014), the shrunk DCC (DCC-L) would also show superior efficiency properties, since it does not ignore the joint likelihood. The same dimensional limitations of standard DCC estimation however apply for the implementation of DCC-L. Engle et al. (2017) build on Hafner and Reznikova (2012) and Patel et al. (2014), and combine the composite likelihood cDCC model with the nonlinear shrinkage approach of Ledoit and Wolf (2012). Their DCC-NL model is in principle free from the various shortcomings affecting DCC estimation, granting consistent yet not efficient estimation.

In this paper we address the issue of consistent estimation of time-varying conditional correlations in high dimensional settings, within the context of the Semiparametric DCC model proposed by Morana (2015). While sharing a similar sequential approach to DCC, SP-DCC has the advantage of not requiring the direct parameterization of the conditional covariance or correlation processes, therefore also avoiding any assumption on their long-run target. It can be implemented in a high dimensional context, without being exposed to the downward bias estimation issue, since it only requires univariate QML GARCH estimation for individual and pairwise aggregated series in the first step; conditional covariances are then estimated in the second step by means of the polarization identity, and conditional correlations are estimated by their usual scaling. SP-DCC also allows for straightforward estimation of cross-covariance and correlations of any order, opening up interesting new scenarios for time-varying parameter modeling (see Morana, 2017). Monte Carlo results, reported in Morana and Sbrana (2017), yield support to SP-DCC estimation, also in relation to exact ML and DCC, in various parametric settings of empirical interests for financial applications.

Relative to the two-step SP-DCC model, this paper contributes an additional, third step, which entails ex-post regularization of the conditional covariance and correlation matrices and an efficiency gain. In this respect, a new regression-based non-linear shrinkage approach is proposed, which ensures accurate estimation and safe inversion of the conditional covariance and correlation matrices also in high dimensional cases. Moreover, optimal smoothing of the conditional correlation and covariance matrices, grounded on the maximization of the joint likelihood of the model, is performed. Therefore, relative to SP-DCC, the new regularized SP-DCC (RSP-DCC) yields also an efficiency improvement, since the joint likelihood of the model is explicitly considered in estimation.

Relative to available DCC approaches, including DCC-NL, RSP-DCC is likely to yield improvements not only in terms of efficiency, but also in terms of estimation accuracy of second conditional moments. In fact, RSP-DCC allows for more flexible modelling than competing DCC approaches, which, by assuming a diagonal VECH structure, neglect spillover effects without any testing. In this respect, the implied conditional covariance parameterization of RSP-DCC is more general than the one assumed under the usual DVECH-GARCH(1,1) model -including the latter as special case-, accounting for spillover effects of past conditional variances and innovations on the conditional covariances.

We apply the proposed approach to a set of 50 assets, i.e. the Fama-French market return and the 49 industry sector portfolios and compare their estimated dynamic conditional correlations with the market return and their dynamic market betas, using various SP-DCC specifications and Engle (2002) DCC. We also estimate a global minimum variance portfolio (GMV) using the fifty assets in the sample. Overall, the empirical results are strongly encouraging, and confirm that SP-DCC is a simple and viable alternative to existing DCC models for any application of interest.

2 The econometric model

The semiparametric dynamic conditional correlation model (SP-DCC; Morana, 2015) is defined by the following equations

$$\mathbf{y}_t = \boldsymbol{\mu}_t(\boldsymbol{\delta}) + \boldsymbol{\varepsilon}_t \quad (1)$$

$$\boldsymbol{\varepsilon}_t = \mathbf{H}_t^{1/2}(\boldsymbol{\delta})\mathbf{z}_t \quad (2)$$

where \mathbf{y}_t is the $N \times 1$ column vector of the variables of interest, $\boldsymbol{\mu}_t(\boldsymbol{\delta})$ is the $N \times 1$ conditional mean vector $E(\mathbf{y}_t|I_{t-1})$, $\boldsymbol{\delta}$ is a vector of parameters, I_{t-1} is the sigma field, \mathbf{z}_t is an *i.i.d.* random vector of dimension $N \times 1$, with first two moments $E(\mathbf{z}_t) = \mathbf{0}$ and $Var(\mathbf{z}_t) = \mathbf{I}_N$, $\mathbf{H}_t(\boldsymbol{\delta})$ is the $N \times N$ conditional variance-covariance matrix $Var(\mathbf{y}_t|I_{t-1})$.

The conditional mean process We assume that the mean process is linear in the parameters and covariance stationary. For instance, it could be specified in terms of N univariate AR(p) processes or as a N -variate VAR(p) process

$$\boldsymbol{\mu}_t(\boldsymbol{\delta}) = \sum_{j=1}^p \Phi_j \mathbf{y}_{t-j}, \quad (3)$$

where all the roots of the polynomial matrix in the lag operator $\mathbf{I}_N - \sum_{j=1}^p \Phi_j L^j$ are outside the unit circle.

The conditional variance-covariance process We assume that the elements along the main diagonal of $\mathbf{H}_t(\boldsymbol{\delta})$, i.e. the conditional variances $Var(y_{i,t}|I_{t-1}) \equiv h_{i,t}$, $i = 1, \dots, N$, follow a univariate GARCH(1,1) process; hence

$$\mathbf{D}_t^2 = \text{diag}(\omega_i) + \text{diag}(\alpha_i) \circ \boldsymbol{\varepsilon}_{t-1} \boldsymbol{\varepsilon}_{t-1}' + \text{diag}(\beta_i) \circ \mathbf{D}_{t-1}^2, \quad (4)$$

where $\text{diag}(\cdot)$ denotes an $N \times N$ diagonal matrix and \circ denotes element by element (Adamart) multiplication. The generic i th equation is then

$$h_{i,t} = \omega_i + \alpha_i \varepsilon_{i,t-1}^2 + \beta_i h_{i,t-1}, \quad (5)$$

subject to the usual restrictions to ensure that the conditional variances are positive almost surely at any point in time.

The off-diagonal elements of \mathbf{H}_t , i.e. the conditional covariances $Cov(y_{i,t}, y_{j,t}|I_{t-1}) \equiv h_{ij,t}$, $i, j = 1, \dots, N$ $i \neq j$, are defined according to the polarization identity of the covariance operator³

$$\begin{aligned} h_{ij,t} &= \frac{1}{4} [Var(y_{ij,t}^+|I_{t-1}) - Var(y_{ij,t}^-|I_{t-1})] \\ &= \frac{1}{4} [h_{ij,t}^+ - h_{ij,t}^-] \quad i, j = 1, \dots, N \quad i \neq j, \end{aligned} \quad (6)$$

where also the conditional variances for the aggregate variables $y_{ij,t}^+ \equiv y_{i,t} + y_{j,t}$ and $y_{ij,t}^- \equiv y_{i,t} - y_{j,t}$, i.e $Var(y_{ij,t}^+|I_{t-1}) \equiv h_{ij,t}^+$ and $Var(y_{ij,t}^-|I_{t-1}) \equiv h_{ij,t}^-$, are assumed to follow a univariate GARCH(1,1) specification. The generic ij th elements are therefore

$$h_{ij,t}^+ = \omega_{ij}^+ + \alpha_{ij}^+ \varepsilon_{ij,t-1}^{+2} + \beta_{ij}^+ h_{ij,t-1}^+ \quad (7)$$

³For any two random variables A and B , one has $Cov(A, B) \equiv \frac{1}{4} [Var(A+B) - Var(A-B)]$, since $Var(A \pm B) = Var(A) + Var(B) \pm 2Cov(A, B)$.

$$h_{ij,t}^- = \omega_{ij}^- + \alpha_{ij}^- \varepsilon_{ij,t-1}^{-2} + \beta_{ij}^- h_{ij,t-1}^- \quad (8)$$

with $\varepsilon_{ij,t}^+ = \varepsilon_{i,t} + \varepsilon_{j,t}$ and $\varepsilon_{ij,t}^- = \varepsilon_{i,t} - \varepsilon_{j,t}$.⁴

The conditional correlation matrix \mathbf{R}_t is then defined as

$$\mathbf{R}_t = \mathbf{D}_t^{-1} \mathbf{H}_t \mathbf{D}_t^{-1} \quad (9)$$

where $\mathbf{D}_t = \text{diag} \left(h_{1,t}^{1/2}, \dots, h_{N,t}^{1/2} \right)$, and the generic i, j element in \mathbf{R}_t is $\rho_{ij,t} = \frac{h_{ij,t}}{\sqrt{h_{i,t} h_{j,t}}}$.

2.1 The implied conditional covariance process

As shown in the Appendix, the implied parametric structure for the generic conditional covariance process can be worked out using the ARMA(1,1) representation of the GARCH(1,1) process for the aggregated series and the polarization identity.

This yields

$$h_{ij,t} = \omega_{ij} + \theta_{1,ij} (\varepsilon_{i,t-1}^2 + \varepsilon_{j,t-1}^2) + \theta_{2,ij} \varepsilon_{i,t-1} \varepsilon_{j,t-1} + \delta_{1,ij} (h_{i,t-1} + h_{j,t-1}) + \delta_{2,ij} h_{ij,t-1} \quad (10)$$

for the generic i, j processes, where $\omega_{ij} = \frac{1}{4} (\omega_{ij}^+ - \omega_{ij}^-)$, $\theta_{1,ij} = \frac{1}{4} (\alpha_{ij}^+ - \alpha_{ij}^-)$, $\theta_{2,ij} = \frac{1}{2} (\alpha_{ij}^+ + \alpha_{ij}^-)$, $\delta_{1,ij} = \frac{1}{4} (\beta_{ij}^+ - \beta_{ij}^-)$, $\delta_{2,ij} = \frac{1}{2} (\beta_{ij}^+ + \beta_{ij}^-)$.

Equation (10) then shows that the implied conditional covariance parameterization is more general than the one assumed under the usual DVECH-GARCH(1,1) model, implicitly accounting for some spillover effects in conditional covariance.

SP-DCC for the generic i, j processes can then be seen as a halfway model between the bivariate VECH-GARCH(1,1) and DVECH-GARCH(1,1) models. Similar to the DVECH-GARCH(1,1), it assumes a univariate GARCH(1,1) structure for the conditional variance of the individual series; similar to the non-diagonal VECH-GARCH(1,1) model, it allows for some spillovers effects of past conditional variances and innovations on the conditional covariance for the involved individual series. Also in the light of the well-known linkage between asset correlations and volatilities (Solnik et al., 1996), conditional to the validity of the GARCH(1,1) specification for the individual and aggregated series (which can be assessed empirically), SP-DCC should then grant more flexible and accurate modelling of second conditional moments than competing DCC approaches, which, by assuming a diagonal VECH structure, neglect spillover effects without any testing.

The standard DVECH-GARCH(1,1) parameterization for the conditional covariance can then be obtained from (10) by imposing the restrictions

$$\begin{aligned} \alpha_{ij}^+ &= \alpha_{ij}^- = \alpha_{ij} \\ \beta_{ij}^+ &= \beta_{ij}^- = \beta_{ij}, \end{aligned} \quad (11)$$

yielding, after simplification and rearranging,

$$h_{ij,t} = \omega_{ij} + \alpha_{ij} \varepsilon_{i,t-1} \varepsilon_{j,t-1} + \beta_{ij} h_{ij,t-1}, \quad (12)$$

where $\omega_{ij} = \frac{1}{4} (\omega_{ij}^+ - \omega_{ij}^-) = \frac{1}{2} (\omega_i + \omega_j)$.

As shown in the Appendix, the restrictions in (11) have implications also for the parameters of the GARCH(1,1) process for the individual series, implying

$$\begin{aligned} \alpha_{ij} &= \alpha_i = \alpha_j \quad i, j = 1, \dots, N \\ \beta_{ij} &= \beta_i = \beta_j \quad i, j = 1, \dots, N, \end{aligned} \quad (13)$$

⁴Notice that $\varepsilon_{ij,t}^{-2}$ stands for $(\varepsilon_{ij,t}^-)^2 \equiv (\varepsilon_{i,t} - \varepsilon_{j,t})^2$ and $\varepsilon_{ij,t}^{+2}$ stands for $(\varepsilon_{ij,t}^+)^2 \equiv (\varepsilon_{i,t} + \varepsilon_{j,t})^2$.

i.e. common scalar innovation and persistence parameters across all N individual and $N(N - 1)$ aggregate GARCH(1,1) process. Under the latter restrictions, (4) and (12) then yield a constrained DVECH-GARCH(1,1) process.

2.1.1 Relation with the DVECH-GARCH(1,1) process

The above results are consistent with available evidence on the aggregation properties of the bivariate DVECH-GARCH(1,1) process (Nijman and Sentana, 1996; Sbrana, 2012; Morana and Sbrana, 2017). For instance, Morana and Sbrana (2017) show that, for the bivariate DVECH-GARCH(1,1) model, the aggregates $y_{ij,t}^+ = y_{i,t} + y_{j,t}$ and $y_{ij,t}^- = y_{i,t} - y_{j,t}$ also have an ARMA(1,1) representation under the root cancellation condition

$$(\alpha_i + \beta_i) = (\alpha_{ij} + \beta_{ij}) = (\alpha_j + \beta_j) = \gamma, \quad (14)$$

i.e.,

$$(1 - \gamma L) \varepsilon_{ij,t}^{+2} = \omega_{ij}^+ + (1 - \beta_i L) \eta_{i,t} + (1 - \beta_j L) 2\eta_{ij,t} + (1 - \beta_j L) \eta_{j,t}, \quad (15)$$

where $\omega_{ij}^+ = \omega_i + 2\omega_{ij} + \omega_j$, and

$$(1 - \gamma L) \varepsilon_{ij,t}^{-2} = \omega_{ij}^- + (1 - \beta_i L) \eta_{i,t} - (1 - \beta_j L) 2\eta_{ij,t} + (1 - \beta_j L) \eta_{j,t}, \quad (16)$$

where $\omega_{ij}^- = \omega_i - 2\omega_{ij} + \omega_j$.

The root cancellation condition is therefore the condition under which the SP-DCC model is consistent with a DVECH-GARCH(1,1) model, concerning the *order* of the ARMA process describing the aggregates (ARMA(1,1)).

However, the ARMA(1,1) representation for the aggregates assumed by SP-DCC show, in general, a different parameterization from DVECH-GARCH(1,1) model in (15) and (16), i.e.

$$(1 - \gamma_{ij}^+ L) \varepsilon_{ij,t}^{+2} = \omega_{ij}^+ + (1 - \beta_{ij}^+ L) (\eta_{i,t} + 2\eta_{ij,t} + \eta_{j,t}) \quad (17)$$

and

$$(1 - \gamma_{ij}^- L) \varepsilon_{ij,t}^{-2} = \omega_{ij}^- + (1 - \beta_{ij}^- L) (\eta_{i,t} - 2\eta_{ij,t} + \eta_{j,t}), \quad (18)$$

allowing the aggregates to show different persistence and innovation parameters, i.e. $\beta_{ij}^+ \neq \beta_{ij}^-$ and $\alpha_{ij}^+ \neq \alpha_{ij}^-$, in the unconstrained case. Hence, the SP-DCC model is equivalent to the DVECH-GARCH(1,1) model only in the constrained case, where common scalar innovation and persistence parameters hold for the conditional variances of individual and aggregated series, as well as for the conditional covariances.

3 Two-step estimation of the Gaussian SP-DCC model

Consistent and asymptotically Normal estimation is obtained by *QML*, following a two-step procedure. Consider the Gaussian log-likelihood for the model in (1) and (2),

$$L(\boldsymbol{\delta}) = -\frac{1}{2} \sum_{t=1}^T (N \log(2\pi) + \log |\mathbf{H}_t| + \boldsymbol{\varepsilon}_t' \mathbf{H}_t^{-1} \boldsymbol{\varepsilon}_t). \quad (19)$$

Following Engle (2002), the latter can be written as

$$L(\boldsymbol{\delta}) = -\frac{1}{2} \sum_{t=1}^T \left(N \log(2\pi) + 2 \log |\mathbf{D}_t| + \boldsymbol{\varepsilon}'_t \mathbf{D}_t^{-1} \mathbf{D}_t^{-1} \boldsymbol{\varepsilon}_t \right) + \left(-\mathbf{z}'_t \mathbf{z}_t + \log |\mathbf{R}_t| + \mathbf{z}'_t \mathbf{R}_t^{-1} \mathbf{z}_t \right) \quad (20)$$

where \mathbf{D}_t^2 is defined as in (4) and $\mathbf{z}_t = \mathbf{D}_t^{-1} \boldsymbol{\varepsilon}_t$ is the devolatilized disturbance vector.

The log-likelihood function in (20) is then the sum of a *volatility part*

$$L_v(\boldsymbol{\theta}) = -\frac{1}{2} \sum_{t=1}^T \left(N \log(2\pi) + 2 \log |\mathbf{D}_t| + \boldsymbol{\varepsilon}'_t \mathbf{D}_t^{-1} \mathbf{D}_t^{-1} \boldsymbol{\varepsilon}_t \right) \quad (21)$$

and a *correlation part*

$$L_c(\boldsymbol{\theta}, \boldsymbol{\phi}) = -\frac{1}{2} \sum_{t=1}^T \left(-\mathbf{z}'_t \mathbf{z}_t + \log |\mathbf{R}_t| + \mathbf{z}'_t \mathbf{R}_t^{-1} \mathbf{z}_t \right). \quad (22)$$

Estimation is performed in the following steps. Firstly, the mean equation model in (1) is estimated equation by equation by *QML*, i.e. the misspecified likelihood

$$L_m(\boldsymbol{\vartheta}) = -\frac{1}{2} \sum_{t=1}^T \sum_{i=1}^N \log(2\pi) + \log \sigma_i^2 + \frac{\varepsilon_{i,t}^2}{\sigma_i^2} \quad (23)$$

is maximized by separately maximizing each term.

Then, using the estimated conditional mean residuals $\hat{\boldsymbol{\varepsilon}}_t$, the volatility part of the likelihood (21) is maximized with respect to the conditional variance parameters; since (21) is the sum of individual GARCH likelihoods, i.e.

$$L_v(\hat{\boldsymbol{\vartheta}}, \boldsymbol{\varsigma}) = -\frac{1}{2} \sum_{t=1}^T \sum_{i=1}^N \log(2\pi) + \log h_{i,t} + \frac{\hat{\varepsilon}_{i,t}^2}{h_{i,t}}, \quad (24)$$

the volatility part is maximized by separately maximizing each term.

Finally, rather than maximizing the correlation part in (22), conditional to the estimated mean residuals and conditional variances delivered by the former two steps, *SP-DCC* maximizes the sum of individual GARCH likelihoods for the aggregate series $y_{ij,t}^+$ and $y_{ij,t}^-$, i.e.

$$\begin{aligned} L_{SP}(\hat{\boldsymbol{\vartheta}}, \boldsymbol{\phi}) &= -\sum_{t=1}^T \sum_{i=1}^N \sum_{j>i}^N \left(\log(2\pi) + \log h_{ij,t}^+ + \frac{\hat{\varepsilon}_{ij,t}^{+2}}{h_{ij,t}^+} \right) \\ &\quad - \sum_{t=1}^T \sum_{i=1}^N \sum_{j>i}^N \left(\log(2\pi) + \log h_{ij,t}^- + \frac{\hat{\varepsilon}_{ij,t}^{-2}}{h_{ij,t}^-} \right), \end{aligned} \quad (25)$$

which is jointly maximized by separately maximizing each term. Hence, the conditional variances for the aggregates $h_{ij,t}^+$, $h_{ij,t}^-$, $i, j = 1, \dots, N$, $i \neq j$, are estimated equation by equation by means of *QML*, using the aggregates of the conditional mean residuals $\hat{\varepsilon}_{ij,t}^+ = \hat{\varepsilon}_{i,t} + \hat{\varepsilon}_{j,t}$ and $\hat{\varepsilon}_{ij,t}^- = \hat{\varepsilon}_{i,t} - \hat{\varepsilon}_{j,t}$.

The conditional covariances are then estimated by means of the polarization identity, i.e. the off-diagonal elements of \mathbf{H}_t , $h_{ij,t}$, $i, j = 1, \dots, N$, $i \neq j$, are computed as

$$\hat{h}_{ij,t} = \frac{1}{4} \left[\hat{h}_{ij,t}^+ - \hat{h}_{ij,t}^- \right] \quad i, j = 1, \dots, N \quad i \neq j. \quad (26)$$

The conditional correlation matrix \mathbf{R}_t is finally estimated as

$$\hat{\mathbf{R}}_t = \hat{\mathbf{D}}_t^{-1} \hat{\mathbf{H}}_t \hat{\mathbf{D}}_t^{-1} \quad (27)$$

where

$$\hat{\mathbf{D}}_t = \text{diag} \left(\hat{h}_{1,t}^{1/2}, \dots, \hat{h}_{N,t}^{1/2} \right). \quad (28)$$

Hence, the proposed approach to maximize the log-likelihood function is to find

$$\hat{\boldsymbol{\vartheta}} = \arg \max \{L_m(\boldsymbol{\vartheta})\} \quad (29)$$

$$\hat{\boldsymbol{\theta}} = \arg \max \left\{ L_v(\hat{\boldsymbol{\vartheta}}, \boldsymbol{\varsigma}) \right\} \quad (30)$$

$$\hat{\boldsymbol{\phi}} = \arg \max \left\{ L_{SP}(\hat{\boldsymbol{\vartheta}}, \boldsymbol{\phi}) \right\} \quad (31)$$

and then use these values to evaluate $L(\boldsymbol{\delta})$ in (20), provided $\hat{\mathbf{R}}_t$ is positive definite at each point in time. As the latter property is not granted by construction, in this paper we propose a new ex-post shrinkage procedure, which allows for effective regularization of the conditional covariance and correlation matrices.

3.1 The non-Gaussian SP-DCC model

By assuming a conditional multivariate Student's t -distribution, with v degrees of freedom ($v > 2$), in the place of the multivariate Gaussian distribution, the log-likelihood function for the model is then written as

$$L = \sum_{t=1}^T \left(\log \left[\frac{\Gamma(\frac{v+N}{2})}{[\pi(v-2)]^{N/2} \Gamma(\frac{v}{2})} \right] - \frac{1}{2} \log |\mathbf{H}_t| - \frac{v+N}{2} \log \left[1 + \frac{\boldsymbol{\varepsilon}'_t \mathbf{H}_t^{-1} \boldsymbol{\varepsilon}_t}{v-2} \right] \right) \quad (32)$$

or, equivalently,

$$L = \sum_{t=1}^T \left(\log \left[\frac{\Gamma(\frac{v+N}{2})}{[\pi(v-2)]^{N/2} \Gamma(\frac{v}{2})} \right] - \frac{1}{2} \log |\mathbf{R}_t| - \log |\mathbf{D}_t| - \frac{v+N}{2} \log \left[1 + \frac{\boldsymbol{\varepsilon}'_t \mathbf{D}_t^{-1} \mathbf{R}_t^{-1} \mathbf{D}_t^{-1} \boldsymbol{\varepsilon}_t}{v-2} \right] \right), \quad (33)$$

since $\mathbf{R}_t = \mathbf{D}_t^{-1} \mathbf{H}_t \mathbf{D}_t^{-1}$.

Maximization of the log-likelihood in (33) can be performed as for the Gaussian case, following a multi-step approach. In particular, steps 1 and 2 can be carried out using the Student- t version of the volatility likelihoods

$$L_v(\hat{\boldsymbol{\vartheta}}, \boldsymbol{\varsigma}) = \sum_{i=1}^N \sum_{t=1}^T \log \left[\frac{\Gamma(\frac{v+1}{2})}{[\pi(v-2)]^{1/2} \Gamma(\frac{v}{2})} \right] - \frac{1}{2} \log h_{i,t} - \frac{v+1}{2} \log \left[1 + \frac{\hat{\varepsilon}_{i,t}^2}{h_{i,t}(v-2)} \right] \quad (34)$$

and

$$L_{SP}(\hat{\boldsymbol{\vartheta}}, \boldsymbol{\phi}) = \sum_{t=1}^T \sum_{i=1}^N \sum_{j>i}^N \left(\log \left[\frac{\Gamma(\frac{v+1}{2})}{[\pi(v-2)]^{1/2} \Gamma(\frac{v}{2})} \right] - \frac{1}{2} \log h_{ij,t}^+ - \frac{v+1}{2} \log \left[1 + \frac{\hat{\varepsilon}_{ij,t}^{+2}}{h_{ij,t}^+(v-2)} \right] \right) \\ + \sum_{t=1}^T \sum_{i=1}^N \sum_{j>i}^N \left(\log \left[\frac{\Gamma(\frac{v+1}{2})}{[\pi(v-2)]^{1/2} \Gamma(\frac{v}{2})} \right] - \frac{1}{2} \log h_{ij,t}^- - \frac{v+1}{2} \log \left[1 + \frac{\hat{\varepsilon}_{ij,t}^{-2}}{h_{ij,t}^-(v-2)} \right] \right) \quad (35)$$

where $\hat{\varepsilon}_{i,t}$, $i = 1, \dots, N$, are the estimated conditional mean residuals for the individual series, $\hat{\varepsilon}_{ij,t}^+ = \hat{\varepsilon}_{i,t} + \hat{\varepsilon}_{j,t}$ and $\hat{\varepsilon}_{ij,t}^- = \hat{\varepsilon}_{i,t} - \hat{\varepsilon}_{j,t}$, $i, j = 1, \dots, N$, $i \neq j$ are their aggregates. Conditional covariances can then be computed by means of the polarization identity in (26).

Alternatively, by relying on the *QML* principle and following the current practice, steps 1 and 2 can be performed using the conditional Gaussian log-likelihoods in (24) and (25); this avoids the potential conflict concerning the estimated degrees of freedom for the marginal and the joint likelihood.

3.2 Constrained SP-DCC estimation

As already discussed above, constrained SP-DCC is equivalent to the constrained DVECH-GARCH(1,1) model

$$vech(\mathbf{H}_t) = vech(\mathbf{\Omega}) + \alpha vech(\boldsymbol{\varepsilon}_{t-1} \boldsymbol{\varepsilon}'_{t-1}) + \beta vech(\mathbf{H}_{t-1}), \quad (36)$$

where conditional variance-covariance matrices are uniformly positive definite if $\mathbf{\Omega}$ is positive definite and α and β are non-negative (Engle and Shephard, 2007). Under covariance stationarity, $\mathbf{\Omega}$ can be concentrated out of the evolution of \mathbf{H}_t and replaced with a consistent estimate, $(\boldsymbol{\iota} \boldsymbol{\iota}' - \bar{\alpha} - \bar{\beta}) \odot \bar{\mathbf{H}}$, where $\boldsymbol{\iota}$ is a $N \times 1$ unitary vector and $\bar{\mathbf{H}}$ is the unconditional sample variance-covariance matrix. Conditional covariances will be positive definite if $\bar{\mathbf{H}}$ is positive definite, $\bar{\alpha} \geq 0$, $\bar{\beta} \geq 0$ and $\bar{\alpha} + \bar{\beta} < 1$. Similarly positive definite are the Risk-Metrics covariances, obtained from (36) by omitting the intercept matrix $\mathbf{\Omega}$ and setting $\alpha = 1 - \beta$ and $0 < \beta < 1$. See Engle and Shephard (2007) for additional details.

Constrained SP-DCC estimation can then be implemented as follows.

- Maximize the log-likelihood function in (21), in order to obtain estimates of the scalar innovation (α_i) and smoothing (β_i) parameters for the individual N series in the sample, imposing stationarity and non-negativity constraints. Then, compute average (or median) values across series, i.e. $\bar{\beta} = \frac{1}{N} \sum_{i=1}^N \hat{\beta}_i$, $\bar{\alpha} = \frac{1}{N} \sum_{i=1}^N \hat{\alpha}_i$.
- Filter the individual and aggregate series using a GARCH(1,1) model, using the estimated mean (or median) parameters and a consistent estimate of the intercept parameter, i.e. compute

$$\begin{aligned} \hat{h}_{i,t} &= \hat{\sigma}_i^2 (1 - \bar{\alpha} - \bar{\beta}) + \bar{\alpha} \hat{\varepsilon}_{i,t-1}^2 + \bar{\beta} \hat{h}_{i,t-1} \\ \hat{h}_{ij,t}^+ &= \hat{\sigma}_{ij}^{+2} (1 - \bar{\alpha} - \bar{\beta}) + \bar{\alpha} \hat{\varepsilon}_{ij,t-1}^{+2} + \bar{\beta} \hat{h}_{ij,t-1}^+ \\ \hat{h}_{ij,t}^- &= \hat{\sigma}_{ij}^{-2} (1 - \bar{\alpha} - \bar{\beta}) + \bar{\alpha} \hat{\varepsilon}_{ij,t-1}^{-2} + \bar{\beta} \hat{h}_{ij,t-1}^- \end{aligned} \quad (37)$$

where $\hat{\sigma}_i^2 = \frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_{i,t}^2$, $i = 1, \dots, N$; $\hat{\sigma}_{ij}^{+2} = \frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_{ij,t}^{+2}$; $\hat{\sigma}_{ij}^{-2} = \frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_{ij,t}^{-2}$, $i, j = 1, \dots, N$.

- Estimate the conditional variance and correlation matrices as in (26) and (27). The above implementation relies on the positive definiteness of $\bar{\mathbf{H}}$, which can be checked ex-post; if required, as for instance when the concentration ratio N/T gets close or larger than one, positive definiteness of the conditional variance and covariance matrices at any point in time can be imposed ex-post, as for the unconstrained estimation case.

3.3 Asymptotic properties of QML estimation of the SP-DCC model

While a rigorous asymptotic theory for multivariate GARCH processes is still missing, a conjecture on the consistency and asymptotic Normality of the *QML* two-step estimator is provided in Engle (2002), Engle and Kelly (2012) and Patel et al. (2014) for the standard DCC and its composite likelihood formulations. In this respect, despite the proposed SP-DCC two-step procedure does not ensure the maximization of the joint log-likelihood in (20), consistent and asymptotically Normal *QML* estimation can also be conjectured, under the usual standard assumptions, and grounded on the consistency and asymptotic Normality of *QML* univariate estimation of the GARCH(1,1) conditional variance model for the individual and aggregated series. For the latter model the optimal *QML* properties have been shown to hold not only for the stationary case, but also for the integrated and the (mildly) explosive ones, also when the devolatilized innovation is non Gaussian and non i.i.d., provided its fourth moment is bounded. See Lee and Hansen (1994) and Francq and Zakořian (2009).

Hence, for the individual series $i = 1, \dots, N$, under standard regularity conditions, it follows

$$T^{1/2} \left(\hat{\boldsymbol{\delta}}_i - \boldsymbol{\delta}_{i,0} \right) \overset{asy}{\approx} N\{\mathbf{0}, \mathbf{A}(\boldsymbol{\delta}_{i,0})^{-1} \mathbf{B}(\boldsymbol{\delta}_{i,0}) \mathbf{A}(\boldsymbol{\delta}_{i,0})^{-1}\},$$

where $\boldsymbol{\delta}_{i,0}$ is the true value of the vector of parameters in the GARCH equations for the aggregates; $\mathbf{A}(\boldsymbol{\delta}_{i,0})$ and $\mathbf{B}(\boldsymbol{\delta}_{i,0})$ are the Hessian and the outer product gradient matrices, evaluated at the true parameter values.

Moreover, as shown in equation (10), the (implied) parameters of the conditional covariance equation for the generic i, j processes $\boldsymbol{\zeta}_{ij}$ are linear combinations of the scalar innovation and smoothing parameters of the GARCH(1,1) models for the aggregated series, i.e.

$$\boldsymbol{\zeta}_{ij} = \mathbf{R} \begin{bmatrix} \boldsymbol{\delta}_{ij}^+ \\ \boldsymbol{\delta}_{ij}^- \end{bmatrix}, \quad (38)$$

where $\boldsymbol{\zeta}_{ij} = (\omega_{ij} \ \theta_1 \ \theta_2 \ \delta_1 \ \delta_2)'$, $\boldsymbol{\delta}_{ij}^+ = (\omega_{ij}^+ \ \alpha_{ij}^+ \ \beta_{ij}^+)'$ and $\boldsymbol{\delta}_{ij}^- = (\omega_{ij}^- \ \alpha_{ij}^- \ \beta_{ij}^-)'$
and $\mathbf{R} = \begin{bmatrix} \frac{1}{4} & 0 & 0 & -\frac{1}{4} & 0 & 0 \\ 0 & \frac{1}{4} & 0 & 0 & -\frac{1}{4} & 0 \\ 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{4} & 0 & 0 & -\frac{1}{4} \\ 0 & 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} \end{bmatrix}$.

Hence, under standard regularity conditions, it also follows

$$\begin{aligned} T^{1/2} \left(\hat{\boldsymbol{\delta}}_{ij}^+ - \boldsymbol{\delta}_{ij,0}^+ \right) &\overset{asy}{\approx} N\{\mathbf{0}, \mathbf{A}(\boldsymbol{\delta}_{ij,0}^+)^{-1} \mathbf{B}(\boldsymbol{\delta}_{ij,0}^+) \mathbf{A}(\boldsymbol{\delta}_{ij,0}^+)^{-1}\} \\ T^{1/2} \left(\hat{\boldsymbol{\delta}}_{ij}^- - \boldsymbol{\delta}_{ij,0}^- \right) &\overset{asy}{\approx} N\{\mathbf{0}, \mathbf{A}(\boldsymbol{\delta}_{ij,0}^-)^{-1} \mathbf{B}(\boldsymbol{\delta}_{ij,0}^-) \mathbf{A}(\boldsymbol{\delta}_{ij,0}^-)^{-1}\}, \end{aligned} \quad (39)$$

where $\boldsymbol{\delta}_{ij,0}^+$ and $\boldsymbol{\delta}_{ij,0}^-$ are true value of the vector of parameters in the GARCH equations for the aggregates; $\mathbf{A}(\boldsymbol{\delta}_{ij,0}^+)$ and $\mathbf{A}(\boldsymbol{\delta}_{ij,0}^-)$ the Hessian matrices, $\mathbf{B}(\boldsymbol{\delta}_{ij,0}^+)$ and $\mathbf{B}(\boldsymbol{\delta}_{ij,0}^-)$ the outer product gradient matrices, evaluated at the true parameter values.

From (38) and (39) it then also follows

$$T^{1/2} \left(\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_0 \right) \overset{asy}{\approx} N\{\mathbf{0}, \mathbf{A}(\boldsymbol{\zeta}_0)^{-1} \mathbf{B}(\boldsymbol{\zeta}_0) \mathbf{A}(\boldsymbol{\zeta}_0)^{-1}\}, \quad (40)$$

where $\hat{\zeta} = \mathbf{R} \begin{bmatrix} \hat{\delta}_{ij}^+ \\ \hat{\delta}_{ij}^- \end{bmatrix}$, $\zeta_0 = \mathbf{R} \begin{bmatrix} \delta_{ij,0}^+ \\ \delta_{ij,0}^- \end{bmatrix}$ is the true value of the vector of parameters, $\mathbf{A}(\zeta_0) = \mathbf{R} \begin{bmatrix} \mathbf{A}(\delta_{ij,0}^+) & \mathbf{0} \\ \mathbf{0} & \mathbf{A}(\delta_{ij,0}^-) \end{bmatrix}$ \mathbf{R}' is the Hessian and $\mathbf{B}(\zeta_0) = \mathbf{R} \begin{bmatrix} \mathbf{B}(\delta_{ij,0}^+) & \mathbf{0} \\ \mathbf{0} & \mathbf{B}(\delta_{ij,0}^-) \end{bmatrix}$ \mathbf{R}' is the outer product gradient evaluated at the true parameter values.

Support for the above conjecture is provided by the Monte Carlo results in Morana and Sbrana (2017), which show that the two-step SP-DCC estimator fares very well, also in relation to the exact *ML* estimator and the two-step Engle (2002) DCC estimator, in various parametric cases of interest for real data applications. See Morana and Sbrana (2017) for details.

4 Third step QML estimation: regularization and smoothing

The two-step SP-DCC approach does not grant the maximization of the correlation part and therefore of the Gaussian log-likelihood for the model in (20). Unless constrained (and for N sufficiently smaller than T), it does not also grants, in general, that the 2-step covariance and correlation matrices are positive definite at each point in time. Both drawbacks can however be fixed through the implementation of a third step, which consists of a *regularization* part, imposing positive definiteness of the conditional correlation and covariance matrix ex post at each point in time; a *maximization/smoothing* part, determining the optimal smoothing intensity to be applied to the conditional covariance and correlation matrix, using the sequence of consistent two-step estimates to initialize the maximization of the joint likelihood. The latter step might be expected to improve the properties of the estimates, by removing noisiness and increasing efficiency.

4.1 Regularization step

In the case of violation of the condition of positive definiteness, some of the eigenvalues of the conditional covariance matrix are non-positive. The proposed shrinkage procedure then replaces the non-positive eigenvalues with predicted strictly positive values, obtained by OLS estimation of the decay pattern of the positive eigenvalue sequence, using a quadratic specification.⁵ The latter procedure is then implemented at any point in time. Hence, the shrinkage procedure works as follows:

- Firstly, the spectral decomposition of the correlation matrix $\hat{\mathbf{R}}_t$ is performed at each point in time; for the generic time period t , this yields

$$\hat{\mathbf{R}}_t = \hat{\mathbf{E}}_t \hat{\mathbf{V}}_t \hat{\mathbf{E}}_t', \quad (41)$$

where $\hat{\mathbf{V}}_t$ is the $N \times N$ diagonal matrix containing the ordered (descending) eigenvalues along the main diagonal and $\hat{\mathbf{E}}_t$ is the $N \times N$ matrix containing the associated orthogonal eigenvectors.

⁵Assuming a quadratic law of decay for the eigenvalues is loosely related to the seminal contribution of Marcenko and Pastur (1967), concerning the asymptotic linkage between population and sample eigenvalues of random positive definite matrices. See Ledoit and Wolf (2012) for regularization methods based on Marcenko and Pastur (1967) results.

- Assuming s non-positive eigenvalues, the following OLS regression of the $N - s$ positive eigenvalues is then run

$$\hat{\boldsymbol{\lambda}}_t^{1/2} = \mathbf{X}_t \boldsymbol{\tau}_t + \mathbf{v}_t \quad t = 1, \dots, T, \quad (42)$$

where $\hat{\boldsymbol{\lambda}}_t^{1/2}$ is the $N - s \times 1$ vector of (square root-transformed) positive eigenvalues, $\mathbf{X}_t = [\boldsymbol{\iota} \quad \mathbf{o} \quad \mathbf{d}_{1,t} \quad \dots \quad \mathbf{d}_{k-2,t}]$ is the $N - s \times k$ design matrix ($k < N - s$), $\boldsymbol{\iota}$ is a $N - s \times 1$ identity vector, $\mathbf{o} = [1 \quad 2 \quad \dots \quad N - s]'$ is a $N - s \times 1$ index vector, and $\mathbf{d}_{i,t}$ $i = 1, \dots, k - 2$ are impulse dummies included to account for (outlying) positive eigenvalues observations.⁶ If required, also a quadratic component \mathbf{o}^2 could be included in the specification, i.e. $\mathbf{X}_t = [\boldsymbol{\iota} \quad \mathbf{o} \quad \mathbf{o}^2 \quad \mathbf{d}_{1,t} \quad \dots \quad \mathbf{d}_{k-2,t}]$.

Predicted values for the missing s eigenvalues are then computed as

$$\hat{\boldsymbol{\lambda}}_{*t}^{1/2} = \mathbf{X}_* \hat{\boldsymbol{\tau}}_t, \quad (43)$$

where $\mathbf{X}_* = [\boldsymbol{\iota}_* \quad \mathbf{o}_*]$ (or $\mathbf{X}_* = [\boldsymbol{\iota}_* \quad \mathbf{o}_* \quad \mathbf{o}_*^2]$), $\mathbf{o}_* = [N - s + 1 \quad N - s + 2 \quad \dots \quad N]'$, and $\boldsymbol{\iota}_*$ is a $s \times 1$ unity vector.⁷

- Then, by then denoting $\hat{\mathbf{V}}_t^*$ the $N \times N$ diagonal matrix containing the ordered original $N - s$ positive eigenvalues ($\hat{\boldsymbol{\lambda}}_t$) and the s shrunk/predicted eigenvalues ($\hat{\boldsymbol{\lambda}}_{*t}$)

$$\hat{\mathbf{V}}_t^* = \text{diag}(\hat{\boldsymbol{\lambda}}_t, \hat{\boldsymbol{\lambda}}_{*t}), \quad (44)$$

the new estimate of the conditional correlation matrix can be computed as

$$\hat{\mathbf{R}}_t^* = \hat{\mathbf{E}}_t \hat{\mathbf{V}}_t^* \hat{\mathbf{E}}_t' \quad (45)$$

which, by construction, is positive definite at each point in time. Since the elements along the main diagonal of $\hat{\mathbf{R}}_t^*$ might slightly deviate from unity, a normalization is required, i.e., every column and every row of the $\hat{\mathbf{R}}_t^*$ matrix has to be divided by the square root of the corresponding diagonal entry, i.e.

$$\hat{\mathbf{R}}_t^{**} = \hat{\mathbf{D}}_{R^*,t}^{-1/2} \hat{\mathbf{R}}_t^* \hat{\mathbf{D}}_{R^*,t}^{-1/2} \quad (46)$$

where $\hat{\mathbf{D}}_{R,t}$ is the $N \times N$ diagonal matrix containing the elements on the main diagonal of the matrix $\hat{\mathbf{R}}_t^*$.

⁶For instance, beyond $3 \times \hat{\sigma}_t$, where $\hat{\sigma}_t = \left(\sum_{j=1}^{N-s} \frac{(\hat{\lambda}_{j,t} - \bar{\lambda}_t)^2}{N-s-1} \right)^{1/2}$ and $\bar{\lambda}_t = \sum_{j=1}^{N-s} \frac{\hat{\lambda}_{j,t}}{N-s}$, i.e. the sample standard deviation and the sample mean of the positive eigenvalues for time period t , respectively.

⁷Notice that the predicted positive eigenvalues are still required to be in descending order ($\hat{\lambda}_{*N-s+1,t} > \hat{\lambda}_{*N-s+2,t} > \dots > \hat{\lambda}_{*N,t}$) and of a sufficient magnitude to avoid ill-conditioning and related numerical issues with the inversion of the conditional correlation and covariance matrices. To this purpose, a lower bound for the smallest predicted eigenvalue can be set such that one might loose up to κ digits of accuracy (on top of what would be lost to the numerical method due to loss of precision from arithmetic methods). For instance, one can set $\hat{\lambda}_{SR,N,t} = 10^{-\kappa} \times \hat{\lambda}_{SR,1,t}$, i.e. $\rho(\kappa) = \frac{\hat{\lambda}_{SR,1,t}}{\hat{\lambda}_{SR,N,t}} = 10^\kappa$. In the case the lower bound were hit, any positive eigenvalue below the threshold value would then be set equal to the $10^{-\kappa} \times \hat{\lambda}_{SR,1,t}$ cut-off value. Setting a lower bound threshold value is similar to the regularized estimator of Won et al. (2013), which has the extreme eigenvalues winsorized at given constants and the other eigenvalues left unchanged at their sample values.

The implied, well-behaved conditional covariance process at time period t is then obtained as

$$\hat{\mathbf{H}}_t^{**} = \hat{\mathbf{D}}_t \hat{\mathbf{R}}_t^{**} \hat{\mathbf{D}}_t, \quad (47)$$

which is also positive definite by construction.⁸

The proposed regression procedure also allows to shrink the largest eigenvalues if required. In the latter case, predicted eigenvalues are computed for all the entries of the $\boldsymbol{\lambda}_t^{1/2}$ vector, i.e.

$$\hat{\boldsymbol{\lambda}}_{*t}^{1/2} = \mathbf{X}_* \hat{\boldsymbol{\tau}}_t, \quad (48)$$

where $\mathbf{X}_* = [\boldsymbol{\iota} \quad \mathbf{o}_*]$ (or $\mathbf{X}_* = [\boldsymbol{\iota} \quad \mathbf{o}_* \quad \mathbf{o}_*^2]$), $\mathbf{o}_* = [1 \quad 2 \quad \dots \quad N]'$ is a $N \times 1$ index vector, and $\boldsymbol{\iota}$ is a $N \times 1$ identity vector. This might become particularly desirable when the dimension N gets close to the sample size T , or when $N > T$, i.e. when the concentration ratio N/T gets close or larger than unity.⁹

In the latter cases, the eigenvalue regression in (42) might also be run using only a subset of the positive eigenvalues, obtained by trimming out the smallest and largest figures, i.e. only considering positive eigenvalues within the α and $(1 - \alpha)$ percentiles. All this various possibilities are explored in the Monte Carlo analysis. As show by the Monte Carlo results, the suggested shrinkage strategy is not only straightforward to implement, but also fully effective in a variety of scenarios of empirical interest.

4.1.1 Monte Carlo evidence

We consider two profiles. In the first profile the generated data are independent and multivariate Gaussian, i.e. $\mathbf{y}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_N)$, where $\boldsymbol{\Sigma}_N$ is generated each time. In the second profile the data are independent and multivariate Student- t with 5 degrees of freedom, i.e. $\mathbf{y}_t = \boldsymbol{\Sigma}_N \mathbf{z}_t$, where the \mathbf{z}_t vector contains independent standardized t_5 random variables. Concerning the generation of $\boldsymbol{\Sigma}_N$, we consider the case of a non diagonal covariance matrix; in order to achieve generality, we use the average of a Toeplitz and a Hub structure (see Hardin, 2013).

For each profile, we use various sample sizes ($T = 200, 400, 800, 1600, 3200$), dimensions ($N = 10, 25, 50, 100, 200, 400, 800$), condition numbers $\rho(\kappa) = 10^\kappa$ with $\kappa = 1, \dots, 5$, and 500 replications. The investigated concentration ratio N/T then ranges between 0.003 ($T = 3200; N = 10$) and 4 ($T = 200; N = 800$). For the trimmed cases, trimming depends on the dimension N , i.e. the threshold percentiles α and $(1 - \alpha)$ are 10% and 90% for $N < 20$; 5% and 95% for $20 < N < 100$; 1%-99% for $N > 100$.

Moreover, covariance estimation risk is assessed using two loss functions, i.e. the *entropy* loss function

$$L_{ent}(\hat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) = tr(\boldsymbol{\Sigma}^{-1} \hat{\boldsymbol{\Sigma}}) - \ln(\boldsymbol{\Sigma}^{-1} \hat{\boldsymbol{\Sigma}}) - N, \quad (49)$$

and the *quadratic* loss function

$$L_q(\hat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) = \left\| \boldsymbol{\Sigma}^{-1} \hat{\boldsymbol{\Sigma}} - \mathbf{I}_N \right\|_F^2, \quad (50)$$

⁸Notice that the shrinkage procedure can also be directly applied to the conditional variance-covariance matrix, following a similar procedure to the one detailed above.

⁹For instance, a single very large eigenvalue might be indicative of a common factor driving the temporal evolution of the various processes. Concerning stock returns, this would be consistent with asset pricing theories such as CAPM. In the latter case, correcting the largest eigenvalues would not appear to be desirable.

where $\hat{\Sigma}$ is the shrunk sample covariance matrix.

Finally, the fit of the eigenvalues regression function is assessed by means of its coefficient of determination.

Selected results are reported in Figures 1-2. In particular, consistent with the results discussed below, we report the results for the linear and quadratic regressions for the Gaussian model (Figure 1) and for the trimmed linear and quadratic regressions for the Student- t case (Figure 2). In both cases the conditioning number is $\rho(\kappa) = 10^\kappa$ with $\kappa = 1$. A full set of results is reported in the Online Appendix to the paper.

As shown in the plots, *ERS* performance is highly satisfactory in both the Gaussian and Student- t frameworks, depending, as it can be expected, on the concentration ratio N/T , as well as on N and T themselves. In particular, *ERS* yields uniformly better results for the Gaussian than the Student- t case according to both loss functions; moreover, for both cases, the performance of *ERS* improves as the concentration ratio decreases, independently of the value of N , i.e. as the sample size T increases for any given dimension N . The very satisfactory *ERS* performance is also documented by the coefficient of determination of the eigenvalues regressions, in general taking values larger than 0.9 (0.8) for the quadratic (linear) specifications.

ERS performance also appears to be relatively more stable when evaluated in terms of L_q . In fact, according to the latter metric, holding T constant, *ERS* performance appears to be independent of N starting from a concentration ratio value N/T as low as 0.125 ($N = 25$; $T = 200$). On the other hand, the L_{ent} metric points to $N/T = 1$ as threshold value for the performance of the shrinkage regression; this is due to the term $\ln(\Sigma^{-1}\hat{\Sigma})$ in the loss function, which inflates the value of the statistic in the most challenging scenarios, i.e. when $N/T > 1$.

Moreover, by comparing the performance across specifications (not reported), it appears that non trimmed specifications performs best for the Gaussian case, while trimmed specifications for the Student- t case. In this respect, according to the quadratic loss function, for the Gaussian case a quadratic (non trimmed) specification appears to work best for $N/T < 0.5$, while a linear (non trimmed) specification for $N/T \geq 0.5$. Similar results hold for the the Student- t case, albeit the threshold value for selecting the quadratic trimmed specification is lower, i.e. $N/T < 0.125$.

On the other hand, for the entropy loss functions the findings are less uniform. For instance, for the Gaussian model the quadratic (non trimmed specification) is preferred to the linear (non trimmed) specification only for $N/T < 1$ and the case $\kappa = 1$; moreover, for the Student- t case the threshold value for selecting the quadratic trimmed specification is $N/T < 0.25$ for $\kappa = 1$. For both models, for $\kappa \geq 2$ the (non trimmed) quadratic specification always yields the best performance.

4.2 Maximization/smoothing step

The maximization part of the third-step is carried out by suitably parameterizing the joint likelihood in (19) as

$$L(\boldsymbol{\delta}; \varphi) = -\frac{1}{2} \sum_{t=1}^T (N \log(2\pi) + \log |\mathbf{H}(\varphi)_t| + \hat{\boldsymbol{\varepsilon}}_t' \mathbf{H}(\varphi)_t^{-1} \hat{\boldsymbol{\varepsilon}}_t) \quad (51)$$

where $\boldsymbol{\delta} = (\boldsymbol{\vartheta}', \boldsymbol{\varsigma}', \boldsymbol{\phi}')'$ and φ is a smoothing parameter.

The maximization problem is then

$$\hat{\varphi} = \arg \max \left\{ L \left(\varphi | \hat{\vartheta}, \hat{\varsigma}, \hat{\phi} \right) \right\}, \quad (52)$$

where, by assuming exponential smoothing, one has

$$\hat{\mathbf{H}}_t^{sm} = \left\{ \begin{array}{ll} \hat{\mathbf{H}}_t^{**} & t = 1 \\ (1 - \varphi) \hat{\mathbf{H}}_t^{**} + \varphi \hat{\mathbf{H}}_{t-1}^{sm} & t > 1 \end{array} \right\}, \quad (53)$$

with generic elements

$$\begin{aligned} \hat{h}_{i,t}^{sm} &= \left\{ \begin{array}{ll} \hat{h}_{i,t}^{**} & t = 1 \\ (1 - \varphi) \hat{h}_{i,t}^{**} + \varphi \hat{h}_{i,t-1}^{sm} & t > 1 \end{array} \right\} & \text{for } i = 1, \dots, N \\ \hat{h}_{ij,t}^{sm} &= \left\{ \begin{array}{ll} \hat{h}_{ij,t}^{**} & t = 1 \\ (1 - \varphi) \hat{h}_{ij,t}^{**} + \varphi \hat{h}_{ij,t-1}^{sm} & t > 1 \end{array} \right\} & \text{for } i, j = 1, \dots, N \text{ and } i \neq j. \end{aligned} \quad (54)$$

The joint likelihood in (51) can then be maximized with respect to φ , conditional on $\hat{\vartheta}$, $\hat{\varsigma}$ and $\hat{\phi}$, using the sequence of shrunk conditional covariance matrices $\left\{ \hat{\mathbf{H}}_t^{**} \right\}_{t=1}^T$ to initialize the grid search procedure, i.e. setting $\varphi = 0$ as starting value in the grid search procedure, where $0 \leq \varphi < 1$.

Once $\hat{\varphi}_{ML}$, i.e. $\hat{\mathbf{H}}_{ML,t}$, is determined, the smoothed conditional correlation process is then computed as $\hat{\mathbf{R}}_{ML,t} = \hat{\mathbf{D}}_{ML,t}^{-1} \hat{\mathbf{H}}_{ML,t} \hat{\mathbf{D}}_{ML,t}^{-1}$, where $\hat{\mathbf{D}}_{ML,t} = \text{diag} \left(\hat{h}_{1,t}^{sm1/2}, \dots, \hat{h}_{N,t}^{sm1/2} \right)$.¹⁰

4.2.1 Non Gaussian extension

The procedure can be similarly implemented also for the non Gaussian case, by parameterizing the likelihood in (32) as

$$L(\hat{\boldsymbol{\delta}}; \varphi, v) = \sum_{t=1}^T \left(\log \left[\frac{\Gamma \left(\frac{v+N}{2} \right)}{[\pi(v-2)]^{N/2} \Gamma \left(\frac{v}{2} \right)} \right] - \frac{1}{2} \log |\mathbf{H}(\varphi)_t| - \frac{v+N}{2} \log \left[1 + \frac{\boldsymbol{\varepsilon}_t' \mathbf{H}(\varphi)_t^{-1} \boldsymbol{\varepsilon}_t}{v-2} \right] \right), \quad (55)$$

where $\hat{\boldsymbol{\delta}} = \left(\hat{\vartheta}', \hat{\varsigma}', \hat{\phi}' \right)'$ and φ is the exponential smoothing parameter. The joint likelihood in (55) is then maximized with respect to φ and v , conditional to $\hat{\vartheta}$, $\hat{\varsigma}$ and $\hat{\phi}$, using the sequence of shrunk conditional covariance matrices $\left\{ \hat{\mathbf{H}}_t^{**} \right\}_{t=1}^T$, i.e. $\varphi = 0$, and a suitable value for the degrees of freedom parameter v , to initialize the grid search procedure.

5 Empirical application

SP-DCC estimation is implemented using the Fama-French 49 industry portfolios plus the market factor excess returns.¹¹ The sample frequency is daily and the time span is from 1980:01:02 through 2017:10:31, for a total of 9543 observations.

¹⁰Alternatively, smoothing can be applied to the conditional correlations, yielding $\hat{\mathbf{R}}_{ML,t}$, and then $\hat{\mathbf{H}}_{ML,t} = \hat{\mathbf{D}}_t \hat{\mathbf{R}}_{ML,t} \hat{\mathbf{D}}_t$.

¹¹The data are available from the French's data library, at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

Univariate analysis is based on AR(1)-GARCH(1,1) models, where excess returns, previous to conditional variance analysis, are linearly filtered.

In Figure 3 we report box-plots for selected estimated coefficients and the coefficient of determination of the shrinkage regressions. As reported in the top plots, the estimated GARCH parameters for the individual and aggregate series show very similar mean values and dispersion. For instance, the mean (standard deviation) value for the coefficient on the lagged conditional variance is 0.907 (0.020) for the 50 individual series and 0.925 (0.030) for the 2450 distinct aggregated series; the interquartile range (*IQR*) is [0.894 – 0.920] and [0.899 – 0.952], for the individual and aggregated series, respectively. Figures for the coefficient on the lagged squared innovation are 0.080 (0.016) and 0.064 (0.024), [0.070 – 0.088] and [0.042 – 0.085], for the individual and aggregated series, respectively. The latter results then naturally invite also the implementation of the constrained version of the SP-DCC model, imposing the same scalar (α_i) and smoothing (β_i) parameters across all the individual and aggregated series.

As shown in the bottom plots, the fit of the shrinkage regressions, implemented using a quadratic specification and condition number $\rho(\kappa) = 10^3$, is also very satisfactory. For instance, for the unconstrained estimation case, the mean (standard deviation) value of the coefficient of determination of the shrinkage regressions (using raw covariances) is 0.949 (0.029; *IQR* = [0.932 – 0.972]); similar figures hold for all the other cases.

Coherently, by comparing raw and shrunk estimated conditional correlations, it appears that the adjustment required to impose positive definiteness at any point in time is almost negligible. In fact, the mean (standard deviation) Theil’s U coefficient (computed for the $N(N - 1)/2 = 1225$ distinct raw and regularized estimates) is only 0.016 (0.005; *IQR* = [0.012, 0.018]), with maximum (minimum) value of 0.043 (0.005), yielding therefore strong support to SP-DCC estimation (not reported).

The various models can be easily ranked according to the value attained by the mean log-likelihood function. We find that the log-likelihood function raises from a value of $L = -206.661$, as yield by 2-step SP-DCC (raw), to $L = -90.213$, as delivered by regularized 3-step SP-DCC (only regularization implemented), and $L = -64.014$, as yield by regularized-smoothed 3-step SP-DCC (the estimated smoothing parameter φ is 0.98), therefore providing strong empirical support to the 3-step SP-DCC approach.

On the other hand, less sizable is the improvement in the log-likelihood function value yield by the implementation of the third step for the restricted estimation case. In fact, for the latter case, the log-likelihood function raises from a value of $L = -75.670$ for 2-step SP-DCC, to $L = -73.116$ for regularized 3-step SP-DCC, and to $L = -71.686$ for regularized-smoothed 3-step SP-DCC (the estimated smoothing parameter φ is 0.96). The latter finding is not surprising, given that regularization is likely to be redundant for the constrained estimation case, as in the current application the temporal dimension T is much larger than the N dimension ($N/T = 0.005$).

Due to computational issues, Engle (2002) DCC estimation has been performed in the simplified scalar framework, where the correlation driving process is driven by common scalar innovation (a) and smoothing (b) parameters. This yields point estimates of 0.003 for the innovation parameter a and 0.995 for the smoothing parameter b , and therefore fairly smooth conditional correlations. Not surprisingly, the latter model yields the highest value for the mean log-likelihood function, i.e. $L = -54.691$, given that the parameters of the latent correlation process are estimated by directly maximizing the joint log-likelihood function.

Due to the well-known bias problem in high dimensional frameworks, we repeated the

estimation of the DCC model in a restricted framework, i.e. we have estimated the distinct $N(N-1)/2$ (1225) bivariate DCC model first and then used the median estimates for the innovation and smoothing parameters to filter the latent correlation process for the 50 assets model. The estimated median innovation parameter a is 0.022, with interquartile range $IQR = [0.017 - 0.027]$, while the estimated median smoothing parameter β is 0.972, with interquartile range $IQR = [0.965 - 0.979]$. Finally, the value of the log-likelihood function is slightly higher than for the unconstrained case, i.e. $L = -58.306$.

A detailed comparison of estimates across models is reported in the Online Appendix; for expository purposes, in Figures 4 and 5 we report selected results for the banking sector. Plots refer to the last ten years of data for graphical convenience. As shown in Figure 4, the estimated conditional correlations show very similar patterns and levels across models. The most noticeable feature is the lower dispersion shown by the regularized-smoothed 3-step SP-DCC estimates relative to all the other models, apart from the (multivariate) DCC model. Moreover, as shown in Figure 5, for the case at hand, constrained estimation appears to hardly affect 2-step and regularized 3-step SP-DCC conditional correlations; more noticeable is the impact on the regularized-smoothed 3-step SP-DCC estimates, which, however, is likely to depend on the different smoothing coefficient selected for the constrained ($\hat{\varphi} = 0.96$) and unconstrained estimates ($\hat{\varphi} = 0.98$). Interestingly, constrained 3-step regularized-smoothed SP-DCC (*c-sm*) and constrained DCC (*c-DCC*) estimates are closest among all the series considered in the comparison. As shown in the Appendix, this finding is not peculiar to the banking sector, but holds on average for the 49 sectors assessed. In fact, the average correlation coefficient for the *c-sm* and *c-DCC* series is 0.91; moreover, the average level and dispersion discrepancy is negligible (see Table A7 in the online Appendix). Similar findings hold for the estimated market- β coefficients ($\beta_{i,mk,t} = \frac{h_{i,mk;t}}{h_{mk,t}}$, $i = 1, \dots, 49$). While an in-depth assessment of the properties of the dynamic conditional correlations and market- β coefficients for the Fama-French industry portfolios is beyond the scope of the current paper, the findings are surely interesting and invite further explorations using the SP-DCC model within this framework. This is also in the light of the evidence that conditional correlations and betas appears to show time varying properties for all the industry sectors, also when assessed using regularized-smoothed 3-step SP-DCC estimates, which are the least dispersed across all models (Figure 6).

5.1 Optimal portfolio allocation

We finally consider the issue of estimating the global minimum variance portfolio (GMV). In the absence of short-sale constraints, the problem is formulated as

$$\min_{\mathbf{w}} \mathbf{w}' \mathbf{H}_t \mathbf{w} \quad (56)$$

$$s.t. \mathbf{w}' \mathbf{1} = 1 \quad (57)$$

where \mathbf{H}_t is the time period t conditional variance-covariance matrix, \mathbf{w} is the vector of optimal weights and $\mathbf{1}$ denotes a unitary vector. Its solution yields

$$\mathbf{w} = \frac{\mathbf{H}_t^{-1} \mathbf{1}}{\mathbf{1}' \mathbf{H}_t \mathbf{1}}. \quad (58)$$

The vector of optimal weights can then be estimated as

$$\hat{\mathbf{w}} = \frac{\hat{\mathbf{H}}_t^{-1} \mathbf{1}}{\mathbf{1}' \hat{\mathbf{H}}_t \mathbf{1}}, \quad (59)$$

where in our application $\hat{\mathbf{H}}_t$ is the estimated conditional variance-covariance matrix delivered by regularized and regularized-smoothed 3-step SP-DCC, for the case of both constrained and unconstrained estimation, and by the Engle (2002) DCC model. The out of sample period ranges from 25/10/1995 through 02/10/2017, yielding a total of 263 months (each defined by 21 consecutive working days) or a total of 5523 observations. Optimal weights are updated monthly, holding the weights fixed from one day to the next during each month. At each investment date $s = 1, \dots, 263$, a conditional covariance matrix is estimated using the most recent 4,000 daily observations.

We then consider the following five portfolios.

- EW: the equally weighted portfolio ($1/N$; $N = 50$), which is a standard benchmark.
- SP-sk: portfolio (59), where $\hat{\mathbf{H}}_t$ is obtained from unconstrained regularized SP-DCC.
- SP-sm: portfolio (59), where $\hat{\mathbf{H}}_t$ is obtained from unconstrained regularized-smoothed SP-DCC.
- c-SP-sk: portfolio (59), where $\hat{\mathbf{H}}_t$ is obtained from constrained regularized SP-DCC.
- c-SP-sm: portfolio (59), where $\hat{\mathbf{H}}_t$ is obtained from constrained regularized-smoothed SP-DCC.
- DCC: portfolio (59), where $\hat{\mathbf{H}}_t$ is obtained from Engle (2002) DCC.

Moreover, we consider optimal allocations also under short-sale constraints; for the latter case the generic element in the optimal portfolio weights vector is \hat{w}_i^{**} , where $\hat{w}_i^{**} = \hat{w}_i^* / \mathbf{1}' \hat{\mathbf{w}}^*$ and

$$\hat{w}_i^{**} = \begin{cases} \hat{w}_i & \text{if } \hat{w}_i > 0 \\ 0 & \text{if } \hat{w}_i \leq 0 \end{cases}. \quad (60)$$

The above short-cut sets at the admissible lower-bound zero value the estimated optimal negative weights, avoiding the numerical solution of the constrained optimization problem in (56), where (57) is modified in order to allow only for non-negative portfolio weights, i.e. $\mathbf{w} \geq \mathbf{0}$ and $\mathbf{w}' \mathbf{1} = 1$. We then consider the five additional below portfolios.

- SP-sk*: SP-sk with the short-sale constraints imposed.
- SP-sm*: SP-sm with the short-sale constraints imposed.
- c-SP-sk*: c-SP-sk with the short-sale constraints imposed.
- c-SP-sm*: c-SP-sm with the short-sale constraints imposed.
- DCC*: DCC with the short-sale constraints imposed.

Portfolios are evaluated by means of the following three out of sample statistics (annualized and in percentage terms).

- MR: the average of the 5523 out-of-sample daily log returns, which is then multiplied by 252 to annualize.
- SD: the standard deviation of the 5523 out-of-sample daily log returns, which is then multiplied by $\sqrt{252}$ to annualize.
- IR: the annualized information ratio, computed as the MR/SD ratio.

As noticed by Engle et al. (2017), out of the three above performance measures, the out-of-sample standard deviation SD is the most important one, as the global minimum variance portfolio is designed to minimize the variance, rather than to maximize the expected return or the information ratio. High out-of-sample MR and IR are also desirable, but should be held as subordinate objectives to a low out-of-sample SD when assessing a variance-covariance matrix estimator. The results are reported in Table 1. The following findings are noteworthy.

- Firstly, in terms of our preferred performance measure SD, all the DCC and SP-DCC based portfolios outperform EW by a wide margin. The gain is particularly sizable when the short-selling constraints are not imposed. In this respect, DCC and c-SP-sk/c-SP-sm are the best performer, yielding a 37% and 34% reduction in portfolio risk, relative to the equally weighted portfolio, respectively. The latter models still yield a sizable improvement in risk reduction (18% and 16%, respectively) also when the short-selling constraints are imposed (DCC* and c-SP-sk*/c-SP-sm*). Despite the difference in the point figures, a one-sided F-test for the null hypothesis of equal portfolio return variances, does not allow to reject the null hypothesis that the DCC and Sp-DCC portfolios yield the same risk. The value of the test and p-values are reported in Table 1, Panel C. In order to control for non-normality, p-values have been computed by means of the bootstrap (see Ledoit and Wolf (2011) for a similar approach). As shown in Panel C, the risk improvement yield by both DCC and SP-DCC relative to EW are statistically significant, while the ranking across constrained DCC and SP-DCC models is not statistically significant at the 5% level, independently of the imposition of short-selling constraints.
- Secondly, DCC and c-SP-sk*/c-SP-sm* outperform EW also according to the IR statistics; yet, the improvement in the information ratio relative to EW is more sizable for DCC (12%) than for the other models (3% to 5%).
- Finally, while EW is the best performing portfolio in terms of average return, constrained SP-DCC models, under short-selling constraints, yield figures just 12% to 14% below the EW return; similarly for DCC. Among the SP-DCC models, c-SP-sk* and c-SP-sm* are again the best performer.

Overall, the results are strongly encouraging, showing that SP-DCC is a valuable tool also for optimal portfolio allocation. Moreover, the superior performance shown by constrained over unconstrained models also suggests that very large portfolios are in principle easily tractable in the proposed framework, due to the scant computational burden required.

In this respect, the implementation of c-SP-sk just requires the estimation of common scalar innovation (α) and smoothing (β) GARCH parameters across the N series, while c-SP-sm requires the estimation of an additional smoothing parameter (φ). For both cases,

positive definiteness of the conditional variance-covariance matrix is likely to hold even without regularization, in so far as the temporal dimension T is sufficiently larger than the N dimension. The computational burden is surely not larger than for the scalar DCC model, which requires, in addition to the estimation of the parameters of the individual GARCH equations, the estimation of the scalar innovation and smoothing parameters for the latent process driving conditional correlation dynamics.

6 Conclusions

This paper deals with consistent estimation of time varying conditional covariance and correlation matrices in high dimensional settings, within the context of the Semiparametric DCC model (Morana, 2015). Over available DCC methods, SP-DCC has the advantage of not requiring the parameterization of the conditional covariances or correlations, therefore also avoiding any assumption on their long-run target. It can be used in a high dimensional context, without being exposed to the downward bias estimation issue, since it only requires univariate *QML* GARCH estimation for individual and pairwise aggregated series in the first step; conditional covariances are then estimated in the second step by means of the polarization identity, and conditional correlations obtained by their usual scaling. Monte Carlo results, reported in Morana and Sbrana (2017), yield support to SP-DCC estimation, also in relation to exact ML and DCC, in various parametric setting of empirical interests for financial applications.

Relative to the two-step SP-DCC model, this paper contributes an additional, third step, which entails ex-post regularization of conditional covariance and correlation matrices, by means of a new regression-based shrinkage approach, and their optimal smoothing, through the maximization of the joint likelihood of the model. Relative to SP-DCC, the new regularized SP-DCC (RSP-DCC) yields, therefore, also an efficiency improvement, since the joint likelihood of the model is explicitly considered in estimation. Moreover, the non-linear shrinkage procedure ensures accurate estimation and safe inversion of the conditional covariance and correlation matrices also in high dimensional cases, which are easily tractable due to the scant computational burden.

Relative to available DCC approaches, including the recent DCC-NL of Engle et al. (2017), RSP-DCC is likely to yields improvements not only in terms of efficiency, but also in terms of estimation accuracy of second conditional moments. RSP-DCC, in fact, allows for more flexible modelling than competing DCC approaches, which, by assuming a diagonal VEC structure, neglect spillover effects without any testing. The implied conditional covariance parameterization of RSP-DCC is, in fact, more general than the one assumed under the usual DVECH-GARCH(1,1) model -including the latter as special case-, accounting for spillover effects of past conditional variances and innovations on the conditional covariance for the involved individual series.

We apply the proposed approach to a set of 50 assets, i.e. the Fama-French market return and the 49 industrial sector portfolios and compare the estimated dynamic conditional correlations of the industry portfolios with the market return, and their market- β , using various SP-DCC specifications and with Engle (2002) DCC. We also provide an application to global minimum variance portfolio (GMV) estimation. The empirical results are strongly encouraging, and confirm that SP-DCC is a simple and viable alternative to existing DCC models for any application of interest.

7 Appendix 1: The implied conditional covariance process

The implied parametric structure for the generic conditional covariance process can be easily worked out through the ARMA representation of the GARCH(1,1) processes for the aggregated series and the polarization identity. Detailed derivations of the below results is contained in the Online Appendix.

Then, following Bollerslev (1986), the ARMA(1,1) representation for the GARCH(1,1) models assumed for the aggregates $y_{ij,t}^+ \equiv y_{i,t} + y_{j,t}$ and $y_{ij,t}^- \equiv y_{i,t} - y_{j,t}$ in (7) and (8), i.e.

$$h_{ij,t}^+ = \omega_{ij}^+ + \alpha_{ij}^+ \varepsilon_{ij,t-1}^{+2} + \beta_{ij}^+ h_{ij,t-1}^+ \quad (61)$$

$$h_{ij,t}^- = \omega_{ij}^- + \alpha_{ij}^- \varepsilon_{ij,t-1}^{-2} + \beta_{ij}^- h_{ij,t-1}^-, \quad (62)$$

is

$$\begin{bmatrix} (1 - (\alpha_{ij}^+ + \beta_{ij}^+)L) & 0 \\ 0 & (1 - (\alpha_{ij}^- + \beta_{ij}^-)L) \end{bmatrix} \begin{bmatrix} \varepsilon_{ij,t}^{+2} \\ \varepsilon_{ij,t}^{-2} \end{bmatrix} = \begin{bmatrix} \omega_{ij}^+ \\ \omega_{ij}^- \end{bmatrix} + \begin{bmatrix} (1 - \beta_{ij}^+)L & 0 \\ 0 & (1 - \beta_{ij}^-)L \end{bmatrix} \begin{bmatrix} \eta_{ij,t}^+ \\ \eta_{ij,t}^- \end{bmatrix}, \quad (63)$$

where

$$h_{ij,t}^+ \equiv (h_{i,t} + 2h_{ij,t} + h_{j,t}) = \varepsilon_{ij,t}^{+2} - \eta_{ij,t}^+, \quad (64)$$

$$h_{ij,t}^- \equiv (h_{i,t} - 2h_{ij,t} + h_{j,t}) = \varepsilon_{ij,t}^{-2} - \eta_{ij,t}^-, \quad (65)$$

and $\eta_{ij,t}^+$, $\eta_{ij,t}^-$ are zero mean and serially uncorrelated disturbances by definition, with $E[\eta_{ij,t}^{+2}]$, $E[\eta_{ij,t}^{-2}]$ positive, finite constants. Moreover,

$$\varepsilon_{ij,t}^{+2} = (\varepsilon_{i,t}^2 + 2\varepsilon_{i,t}\varepsilon_{j,t} + \varepsilon_{j,t}^2) \quad (66)$$

$$\varepsilon_{ij,t}^{-2} = (\varepsilon_{i,t}^2 - 2\varepsilon_{i,t}\varepsilon_{j,t} + \varepsilon_{j,t}^2) \quad (67)$$

$$\eta_{ij,t}^+ = \eta_{i,t} + \eta_{j,t} + 2\eta_{ij,t} \quad (68)$$

$$\eta_{ij,t}^- = \eta_{i,t} + \eta_{j,t} - 2\eta_{ij,t}. \quad (69)$$

By denoting $\gamma_{ij}^+ = (\alpha_{ij}^+ + \beta_{ij}^+)$ and $\gamma_{ij}^- = (\alpha_{ij}^- + \beta_{ij}^-)$, and subtracting the second equation from the first one in (63), i.e. using the polarization identity, one then has

$$(1 - \gamma_{ij}^+ L)\varepsilon_{ij,t}^{+2} - (1 - \gamma_{ij}^- L)\varepsilon_{ij,t}^{-2} = (\omega_{ij}^+ - \omega_{ij}^-) + (1 - \beta_{ij}^+ L)\eta_{ij,t}^+ - (1 - \beta_{ij}^- L)\eta_{ij,t}^-. \quad (70)$$

Moreover, by substituting (66), (67), (68) and (69) into (70), and rearranging, one then has the ARMA representation for the implied conditional covariance process

$$\begin{aligned} & \left[1 - \frac{1}{2}(\gamma_{ij}^+ + \gamma_{ij}^-)L \right] (\varepsilon_{i,t}\varepsilon_{j,t}) = \\ & \frac{1}{4}(\omega_{ij}^+ - \omega_{ij}^-) + \frac{1}{4}(\gamma_{ij}^+ - \gamma_{ij}^-)L(\varepsilon_{i,t}^2 + \varepsilon_{j,t}^2) \\ & - \frac{1}{4}(\beta_{ij}^+ - \beta_{ij}^-)L(\eta_{i,t} + \eta_{j,t}) + \left[1 - \frac{1}{2}(\beta_{ij}^+ + \beta_{ij}^-)L \right] \eta_{ij,t}. \end{aligned} \quad (71)$$

By further rearranging, one then has

$$\begin{aligned} h_{ij,t} &= \frac{1}{4}(\omega_{ij}^+ - \omega_{ij}^-) + \frac{1}{2}(\beta_{ij}^+ + \beta_{ij}^-)h_{ij,t-1} + \frac{1}{2}(\alpha_{ij}^+ + \alpha_{ij}^-)\varepsilon_{i,t-1}\varepsilon_{j,t-1} \\ & + \frac{1}{4}(\beta_{ij}^+ - \beta_{ij}^-)(h_{i,t-1} + h_{j,t-1}) + \frac{1}{4}(\alpha_{ij}^+ - \alpha_{ij}^-)(\varepsilon_{i,t-1}^2 + \varepsilon_{j,t-1}^2). \end{aligned} \quad (72)$$

Equation (72) then shows that the implied conditional covariance parameterization is more general than the one assumed under the usual DVECH-GARCH(1,1) model, implicitly accounting for some spillover effects in conditional covariance. Yet, under the restrictions

$$\begin{aligned}\alpha_{ij}^+ &= \alpha_{ij}^- = \alpha_{ij} \\ \beta_{ij}^+ &= \beta_{ij}^- = \beta_{ij}\end{aligned}\quad (73)$$

equation (72) yields

$$h_{ij,t} = \omega_{ij} + \alpha_{ij}\varepsilon_{i,t-1}\varepsilon_{j,t-1} + \beta_{ij}h_{ij,t-1}, \quad (74)$$

where $\omega_{ij} = \frac{1}{4}(\omega_{ij}^+ - \omega_{ij}^-)$. Hence, the DVECH-GARCH(1,1) model is nested in the SP-DCC model and obtained under the restrictions $\alpha_{ij}^+ = \alpha_{ij}^- = \alpha_{ij}$ and $\beta_{ij}^+ = \beta_{ij}^- = \beta_{ij}$. As shown below, the latter also implies that the innovation (α_i) and smoothing (β_i) parameters of the GARCH(1,1) equations for the individual series are common across both individual and aggregated series, i.e.

$$\begin{aligned}\beta_i &= \beta_j = \beta_{ij}^- = \beta_{ij}^+ = \beta_{ij} \\ \alpha_i &= \alpha_j = \alpha_{ij}^- = \alpha_{ij}^+ = \alpha_{ij}.\end{aligned}\quad (75)$$

7.1 Additional implications

In order to gauge the implications for the GARCH parameters of the individual series, the derivation of the ARMA(1,1) representation for the conditional covariance $h_{ij,t}$ can be obtained also exploiting the following identity

$$h_{ij,t}^+ = 2h_{i,t} + 2h_{j,t} - h_{ij,t}^- \quad (76)$$

Then, the polarization identity can be written as

$$h_{ij,t}^+ - h_{ij,t}^- = 2h_{i,t} + 2h_{j,t} - 2h_{ij,t}^- \quad (77)$$

Given the ARMA(1,1) representations for the squared disturbances $\varepsilon_{i,t}^2$ and $\varepsilon_{j,t}^2$

$$\begin{aligned}\begin{bmatrix} (1 - (\alpha_i + \beta_i)L) & 0 \\ 0 & (1 - (\alpha_j + \beta_j)L) \end{bmatrix} \begin{bmatrix} \varepsilon_{i,t}^2 \\ \varepsilon_{j,t}^2 \end{bmatrix} = \\ \begin{bmatrix} \omega_i \\ \omega_j \end{bmatrix} + \begin{bmatrix} (1 - \beta_i L) & 0 \\ 0 & (1 - \beta_j L) \end{bmatrix} \begin{bmatrix} \eta_{i,t} \\ \eta_{j,t} \end{bmatrix},\end{aligned}\quad (78)$$

and for $\varepsilon_{ij,t}^{-2}$, as in (63), i.e.

$$(1 - (\alpha_{ij}^- + \beta_{ij}^-)L)\varepsilon_{ij,t}^{-2} = \omega_{ij}^- + (1 - \beta_{ij}^-L)\eta_{ij,t}^-$$

by substituting into the polarization identity in (77) and simplifying, one then has

¹²This follows from the identity $VAR(A + B) = 2VAR(A) + 2VAR(B) - VAR(A - B)$, for any two random variables A and B .

$$\begin{aligned}
& [-(\alpha_i + \beta_i)L + (\alpha_{ij}^- + \beta_{ij}^-)L] \varepsilon_{i,t}^2 + \\
& [-(\alpha_j + \beta_j)L + (\alpha_{ij}^- + \beta_{ij}^-)L] \varepsilon_{j,t}^2 + \\
& 2(1 - (\alpha_{ij}^- + \beta_{ij}^-)L) (\varepsilon_{i,t}\varepsilon_{j,t}) = \\
& (\omega_i + \omega_j - \omega_{ij}^-) + [-(\beta_i - \beta_{ij}^-)L] \eta_{i,t} \\
& + [-(\beta_j - \beta_{ij}^-)L] \eta_{j,t} + 2(1 - \beta_{ij}^-L) (\eta_{i,t}\eta_{j,t}), \tag{79}
\end{aligned}$$

which, by rearranging yields

$$\begin{aligned}
h_{ij,t} &= \frac{1}{2} (\omega_i + \omega_j - \omega_{ij}^-) + \beta_{ij}^- h_{ij,t-1} + \alpha_{ij}^- (\varepsilon_{i,t-1}\varepsilon_{j,t-1}) \\
&+ \frac{1}{2} (\beta_i - \beta_{ij}^-) h_{i,t-1} + \frac{1}{2} (\alpha_i - \alpha_{ij}^-) \varepsilon_{i,t-1}^2 \\
&+ \frac{1}{2} (\beta_j - \beta_{ij}^-) h_{j,t-1} + \frac{1}{2} (\alpha_j - \alpha_{ij}^-) \varepsilon_{j,t-1}^2, \tag{80}
\end{aligned}$$

requiring the restrictions

$$\begin{aligned}
\beta_i &= \beta_j = \beta_{ij}^- = \beta_{ij} \\
\alpha_i &= \alpha_j = \alpha_{ij}^- = \alpha_{ij}, \tag{81}
\end{aligned}$$

in order to yield the DVECH-GARCH(1,1) parameterization for the conditional covariance, i.e.

$$h_{ij,t} = \frac{1}{2} (\omega_i + \omega_j - \omega_{ij}^-) + \beta_{ij}^- h_{ij,t-1} + \alpha_{ij}^- (\varepsilon_{i,t-1}\varepsilon_{j,t-1}). \tag{82}$$

Notice that the same result follows by writing (76) as

$$h_{ij,t}^- = 2h_{i,t} + 2h_{j,t} - h_{ij,t}^+, \tag{83}$$

yielding

$$\begin{aligned}
h_{ij,t} &= \frac{1}{2} (\omega_{ij}^+ - \omega_i - \omega_j) + \beta_{ij}^+ h_{ij,t-1} + \alpha_{ij}^+ (\varepsilon_{i,t-1}\varepsilon_{j,t-1}) \\
&+ \frac{1}{2} (\beta_i - \beta_{ij}^+) h_{i,t-1} + \frac{1}{2} (\alpha_i - \alpha_{ij}^+) \varepsilon_{i,t-1}^2 \\
&+ \frac{1}{2} (\beta_j - \beta_{ij}^+) h_{j,t-1} + \frac{1}{2} (\alpha_j - \alpha_{ij}^+) \varepsilon_{j,t-1}^2, \tag{84}
\end{aligned}$$

as well as

$$h_{ij,t} = \frac{1}{2} (\omega_{ij}^+ - \omega_i - \omega_j) + \beta_{ij}^+ h_{ij,t-1} + \alpha_{ij}^+ (\varepsilon_{i,t-1}\varepsilon_{j,t-1}), \tag{85}$$

under the restrictions

$$\begin{aligned}
\beta_i &= \beta_j = \beta_{ij}^+ = \beta_{ij} \\
\alpha_i &= \alpha_j = \alpha_{ij}^+ = \alpha_{ij}. \tag{86}
\end{aligned}$$

By comparing the two processes in (82), (85), and the two sets of restriction in (81), (86), it then also follows

$$\begin{aligned}
\beta_{ij}^+ &= \beta_{ij}^- \\
\alpha_{ij}^+ &= \alpha_{ij}^-. \tag{87}
\end{aligned}$$

Finally, it is easy to check that the processes in (74) and (82) or, alternatively, in (74) and (85), are fully coherent, since

$$\begin{aligned}\omega_{ij} &= \frac{1}{4}(\omega_{ij}^+ - \omega_{ij}^-) \\ &= \frac{1}{2}(\omega_i + \omega_j)\end{aligned}\tag{88}$$

and

$$\begin{aligned}\beta_{ij}^- &= \beta_{ij}^+ = \beta_{ij} \\ \alpha_{ij}^- &= \alpha_{ij}^+ = \alpha_{ij}.\end{aligned}\tag{89}$$

To sum up, given a GARCH(1,1) specification for both the aggregates and individual series, the usual DVECH-GARCH(1,1) parameterization for the conditional covariance is obtained only for the constrained model case, where

$$\begin{aligned}\beta_i &= \beta_j = \beta_{ij}^- = \beta_{ij}^+ = \beta_{ij} \\ \alpha_i &= \alpha_j = \alpha_{ij}^- = \alpha_{ij}^+ = \alpha_{ij},\end{aligned}$$

i.e. the scalar innovation and smoothing parameters are common across the GARCH(1,1) conditional variances for all the N individual and $N(N - 1)$ aggregated series, as well as for the $N(N - 1)/2$ implied conditional covariances. As shown, in equation (72) above, in the unconstrained case, the implied conditional covariance parameterization is more general than the one assumed under the usual DVECH-GARCH(1,1) model, allowing for some spillover effects in conditional covariance.

References

- [1] Aielli, G.P., 2013. Dynamic conditional correlations: On properties and estimation. *Journal of Business and Economic Statistics* 31, 282-299.
- [2] Alexander, C.O., 2002. Principal component models for generating large covariance matrices, *Economic Notes*, 31, 337-359.
- [3] Alexander, C., 2007. Moving Average Models for Volatility and Correlation, and Covariance Matrices. In Fabozzi, F.J. (ed.), *Handbook of Finance*, Volume 1., Wiley.
- [4] Bauwens, L., Laurent, S., Rombouts, L., 2006. Multivariate GARCH models: A survey, *Journal of Applied Econometrics* 21, 79-109.
- [5] Bollerslev, T., Engle, R., Wooldridge, J., 1988. A capital asset pricing model with time varying covariances, *Journal of Political Economy* 96, 116-131.
- [6] Bollerslev, T., 1990. Modeling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model. *Review of Economics and Statistics* 72, 498-505.
- [7] Bollerslev, T., Wooldridge, J., 1992. Quasi-Maximum Likelihood estimation and inference in dynamic models with time varying covariances. *Econometrics Reviews* 11, 143-172.
- [8] Caporin, M., McAleer, M., 2013. Ten things you should know about the Dynamic Conditional Correlation representation. *Econometrics* 1, 115-126.

- [9] Cappiello, L., Engle, R., Sheppard K., 2006. Asymmetric dynamics in the correlations of global equity and bond returns. *Journal of Financial Econometrics* 4, 537-572.
- [10] Engle, R., Gonzalez-Rivera, G., 1991. Semiparametric ARCH models. *Journal of Business and Economic Statistics* 9, 345-59.
- [11] Engle, R., Kroner, F., 1995. Multivariate simultaneous generalized ARCH. *Econometric Theory* 11, 122-150.
- [12] Engle, R.F. and Sheppard, K., 2001. Theoretical and empirical properties of Dynamic Conditional Correlation Multivariate GARCH. NBER Working Paper no. 8554.
- [13] Engle, R.F., 2002. Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business and Economic Statistics* 20, 339-350.
- [14] Engle, R.F., Shephard, K., 2007. Evaluating the specification of covariance models for large portfolios. mimeo, University of Oxford.
- [15] Engle, R., Kelly, B. 2008. Dynamic Equicorrelation, *Journal of Business & Economic Statistics* 30, 212-228.
- [16] Engle, R.F., 2009. High dimension dynamic correlations. In *The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry, J. Castle and N. Shephard* (eds.). Oxford University Press.
- [17] Engle, R.F., 2016. Dynamic conditional beta. *Journal of Financial Econometrics* 14, 643-667.
- [18] Engle, R.F., Ledoit, O., Wolf, M., 2017. Large dynamic covariance matrices. *Journal of Business and Economic Statistics*, forthcoming.
- [19] Fan, J., Wang, M., Yao, Q., 2008. Modelling multivariate volatilities via conditionally uncorrelated components. *Journal of Royal Statistical Society B* 70, 679-702.
- [20] Francq, C., Zakořan, J.M., 2009. A tour in the asymptotic theory of GARCH estimation. In: Mikosch, T., Kreiß, J.P., Davis, R., Andersen, T. (eds). *Handbook of Financial Time Series*. Springer, Berlin, Heidelberg.
- [21] Haffner, C.M. and Reznikova, O., 2012. On the estimation of dynamic conditional correlation models. *Computational Statistics and Data Analysis* 56, 3533-3545.
- [22] Hardin, J., Garcia, S.R., Golan, D., 2013. A method for generating realistic correlation matrices. *The Annals of Applied Statistics* 7, 1733-1762.
- [23] J.P.Morgan, 1996. Riskmetrics Technical Document, 4th ed. J.P.Morgan, New York.
- [24] Ledoit, O., Santa-Clara, P., Wolf, M., 2003. Flexible multivariate GARCH modeling with an application to international stock markets. *Review of Economics and Statistics* 85, 735-747.
- [25] Ledoit, O., Wolf, M., 2004. Honey, I shrunk the sample covariance matrix. *Journal of Portfolio Management* 30, 10-119.
- [26] Ledoit, O., Wolf, M., 2011. Robust performance hypothesis testing with the variance. *Willmott Magazine* 9, 86-89.
- [27] Ledoit, O., Wolf, M., 2012. Nonlinear shrinkage estimation of large-dimensional covariance matrices. *The Annals of Statistics* 40, 1024-1060.
- [28] Lee, S.W., Hansen, B.E., 1994. Asymptotic theory for the GARCH(1,1) quasi maximum likelihood estimator. *Econometric Theory* 10, 29-52.

- [29] Marcenko, V., Pastur, L., 1967. Distribution of eigenvalues for some sets of random matrices. *Sbornik: Mathematics* 1, 457-483.
- [30] McAleer, M., Chan, F., Hoti, S., Lieberman, O., 2008. Generalized autoregressive conditional correlation. *Econometric Theory* 24, 1554-1583.
- [31] Morana, C. 2015. Semiparametric estimation of multivariate GARCH models. *Open Journal of Statistics* 5, 852-858.
- [32] Morana, C., 2017. Macroeconomic and financial effects of oil price shocks: Evidence for the euro area. *Economic Modelling* 64, 82-96.
- [33] Morana, C., Sbrana, G., 2017. Some financial implications of global warming. An empirical assessment. University of Milan-Bicocca, DEMS Working Paper Series No. 377.
- [34] Nijman, T., Sentana, E., 1996. Marginalization and contemporaneous aggregation in multivariate GARCH processes. *Journal of Econometrics* 71, 71-87.
- [35] Pakel, C., Shephard, D., Sheppard, K., Engle, R.F., 2014. Fitting vast diensional time-varying covariance models. Mimeo. New York University.
- [36] Paoletta, M.S., Polak, P., 2015. COMFORT: A common market factor non-Gaussian returns model. *Journal of Econometrics*, 187, 593-605.
- [37] Sbrana, G., 2012. Aggregation and marginalization of GARCH processes: Some further results. *Metron* LXX, 1-8.
- [38] Silvennoinen, A., Terasvirta, T., 2009. Multivariate GARCH Models. In *Handbook of Financial Time Series*, Andersen, T., Davis, R., Kreiss, J.F., Mikosch, T. (eds). Springer.
- [39] Solnik, B., Boucrelle, C., Le Fur, Y., 1996. International market correlation and volatility, *Financial Analysts Journal* 52, 17-34.
- [40] Tse, Y., Tsui, A., 2002. A multivariate GARCH model with time-varying correlations, *Journal of Business and Economic Statistics* 20, 351-362.
- [41] van der Weide, R., 2002, GO-GARCH: A multivariate generalized orthogonal GARCH model, *Journal of Applied Econometrics* 17, 549-564.
- [42] Vrontos, I., Dellaportas, P., Politis, V., 2003. A full-factor multivariate GARCH model, *Econometrics Journal* 6, 311-333.
- [43] Won, J.H., Lin, J., Kim, S.J., Rajaratnam, B., 2013. Condition-number regularized covariance estimation. *Journal of the Royal Statistical Association, Series B, Statistical Methodology* 75, 427-450.