Non-scalar GARCH models: Composite likelihood estimation and empirical comparisons

Luc Bauwens[1], Lyudmila Grigoryeva[2], and Juan-Pablo Ortega[3]

April 19, 2015

Abstract

Multivariate volatility models are widely used in the description of the dynamics of time-varying asset correlations and covariances. Among the well-known drawbacks of many of these parametric families one can name the so called curse of dimensionality and the non-linear parameter constraints that need to be imposed at the time of estimation and that are difficult to handle. In this paper we use a Bregman divergences based optimization technique to tackle the quasi-maximum likelihood (QML) estimation of the DVEC (Diagonal VEC) family for various non-scalar specifications. Additionally, we implement a composite likelihood (CL) method to estimate several non-scalar DCC and DVEC model specifications. The use of the CL approach motivates the in-depth study of different model reduction questions and the analysis of the closedness of the considered families under the reduction operation. The availability of both the QML and CL estimation tools makes possible the empirical out-of-sample performance comparison of the non-scalar DCC and DVEC models under study. We discuss an important estimation bias issue related to the use of covariance targeting and its impact on the empirical performance of the considered multivariate volatility models.

Keywords: multivariate GARCH, diagonal VEC (DVEC), dynamic conditional correlations (DCC), non-scalar models, composite likelihood estimation, Bregman divergences, constrained optimization.

JEL Classification: C13, C32, G17.

[1]Université catholique de Louvain, CORE, Voie du Roman Pays 34 L1.03.01, B-1348 Louvain-La-Neuve. Belgium; SKEMA Business School - Université de Lille. France luc.bauwens@uclouvain.be

[2]Laboratoire de Mathématiques de Besançon, Université de Franche-Comté, UFR des Sciences et Techniques. 16, route de Gray. F-25030 Besançon cedex. France. Lyudmyla.Grygoryeva@univ-fcomte.fr

[4]Centre National de la Recherche Scientifique, Laboratoire de Mathématiques de Besançon, Université de Franche-Comté, UFR des Sciences et Techniques. 16, route de Gray. F-25030 Besançon cedex. France. Juan-Pablo.Ortega@univ-fcomte.fr

Acknowledgments: Lyudmila Grigoryeva and Juan-Pablo Ortega acknowledge partial financial support of the Région de Franche-Comté (Convention 2013C-5493). Lyudmila Grigoryeva acknowledges financial support from the Faculty for the Future Program of the Schlumberger Foundation. Luc Bauwens acknowledges the support of "Projet d'Actions de Recherche Concertées" 12/17-045 of the "Communauté française de Belgique", granted by the "Académie universitaire Louvain".

1 Introduction

Many practical problems that arise in the areas of portfolio and risk management motivate the study and development of multivariate volatility models that, apart from being manageable and empirically effective, allow for a better understanding of the dynamics behind the time-varying conditional covariances and correlations of asset returns. This circumstance has motivated extensive research in this direction in recent years and has resulted in the emergence of a great variety of dynamical models in the literature. Despite their diversity, the choice of a particular model for a given specific task faced by practitioners remains an obstacle. The need of handling a large number of assets and, at the same time, the absence of adequate estimation techniques for high dimensional models with non-scalar specifications results in practice in the systematic use of scalar simplified models which describe poorly the dynamics of the processes in question.

In this work we aim at four major goals: (i) put forward optimization techniques that would allow for the constrained quasi-maximum likelihood (QML) estimation of richly parametrized multivariate volatility models, (ii) improve the estimation of high dimensional models using the composite likelihood (CL) estimator implemented via what we will call reduction procedures, (iii) provide details as for the reduction of each of the non-scalar models of interest, and (iv) address the question of the empirical performance of the estimated models and the possibility to solve various problematic estimation issues documented in the literature by using the proposed estimation techniques.

Among many available multivariate volatility models, the dynamic conditional correlation (DCC) model developed in the works of Engle [Eng02] and Tse and Tsui [TT02] as a generalization of the constant conditional correlation (CCC) model of Bollerslev [Bol90], has become a process of choice in the econometrics literature. This can be explained by several well-known advantages of the DCC family like the availability of a two-step estimation procedure and the possibility to use an approximate covariance targeting which makes them easy to use in the presence of a high number of assets. We emphasize here that it is only the scalar specification of this model (of one or both parameters) that up until now has been mostly discussed in the literature as it is the only one for which the QML estimator has been implemented. The main assumption underlying the use of this model is that the dynamical behavior of the conditional correlation is the same for all the asset pairs, which is well-known to be empirically violated.

There exists a substantial literature (see for instance [HF09], [CES06], [BCG03, BCG06], [FdR05]), [Pel06], [BC05], [CEG11], and [BO13]) where different modifications of the DCC family are proposed in order to overcome the built-in rigidities of the scalar description. However, none of the listed references suggests findings that would allow to handle the DCC models with the general matrix Hadamard-type model parameterizations originally proposed by Engle in [Eng02] or other richer specifications.

In [BGO15] we carefully addressed this issue by providing adequate estimation tools for several non-scalar DCC models and by empirically studying how these richer parameterizations perform with respect to each other and with respect to the scalar model. Our main conclusions were that the estimation of these models is practically feasible in high dimensions and that the subsequent comparative assessment of the empirical performance of the estimated non-scalar models for a particular dataset can be used as a template for the model choice depending on the practical task of interest. At the same time we evidenced that the richer parametrized DCC models suffer from various issues already reported for the scalar specification like, for example, an estimation bias in the model parameters that becomes more visible with the increase of the assets dimension, irrespective of the type of the model prescription considered. This problem has been fixed by Aielli in [Aie13] by modifying the model prescription in such a way that the approximate and statistically inconsistent targeting procedure associated to DCC (see the papers by Caporin and McAleer [CM12]) becomes exact.

In relation with this problem, in this work we address the following questions: (i) whether other non-scalar multivariate volatility families of models that have a one-stage estimation procedure and direct covariance targeting can propose a better performance; (ii) if the bias presence issue for the non-scalar DCC models can be eliminated or at least reduced by using the composite likelihood estimation method which has already proved to improve the quality of the parameter estimation in the context of high dimensional scalar multivariate volatility models; (iii) which of the model families under study, in general, and what parameter specifications, in particular, combined with one of the considered estimation techniques, namely QML or CL, shows the best empirical performance.

In order to tackle the first question, we consider, apart from the DCC model recalled in Section 2.1, the Diagonal VEC (DVEC) dynamical model for the conditional covariance process first introduced by Bollerslev in [BEW88]. This model (see Section 2.2 for its detailed description) directly prescribes the dynamics of the conditional covariance matrix process by using iteratively its lagged values and the values of the lagged asset returns. Unlike in the DCC situation, DVEC estimation is handled in one stage and the covariance targeting implemented beforehand is not approximate. In Section B we make explicit the QML estimation for non-scalar DVEC models, deliver explicit formulas for the gradient of the associated log-likelihood function and refer the reader to the paper [BGO15] when the results in the DVEC context are identical to those already provided for the DCC family.

In order to tackle (ii) we consider the so called composite likelihood (CL) estimation method. This approach consists, roughly speaking, in using an approximation of the joint marginal density based on lower dimensional marginal densities. This idea was introduced by Lindsay [Lin88] and produces an estimator which is shown to be potentially more robust to misspecification issues which are of great importance when dealing with high dimensional models. CL estimation has been extensively discussed in literature over the years (see for instance the papers [VHsS05, ZJ05, VV05, Var08, XR11, PSS11, LYS11, VRF11, WYZ13] and references therein). We use the theoretical framework introduced in [PESS14] for in the context of multivariate volatility models. This reference contains all the relevant proofs for the consistency and asymptotic normality of the concentrated (or profile) likelihood estimator in the presence of incidental parameters and we hence do not need to provide here any details concerning CL estimation theory.

We implement the composite likelihood estimation method (see Section 3) for both the DCC and DVEC models via the construction of the reduced models for the each composite assets subset. We show in detail in Section 4 that various parameterizations of both the (semi-strong) DCC and the DVEC families are closed under the reduction procedure. This approach provides us with an efficient way to use previous knowledge about the scores associated to each of the reduced models in order to formulate the optimization problem associated to their optimization and hence to solve it using the same Bregman divergences based technique presented in [CO14, BGO15].

Regarding (iii), we use a real dataset in order to empirically assess the out-of-sample performance of all the models under consideration estimated either with the QML or with the CL methods. In Section 5 we provide evidence of the fact that the composite likelihood method indeed allows us to improve the quality of the estimation for non-scalar models of both types by solving in part estimation bias issue. At the same time we conclude that the question of giving preference to one or another class of models has to be treated with caution since the results can be different depending on the dataset and the applications considered.

Mention the supplementary online appendix (called SoA) that contains technical details and additional empirical results.

2 Multivariate GARCH: DCC and DVEC models

Consider the n-dimensional conditionally heteroscedastic discrete-time process $\{\mathbf{r}_t\}$ defined by:

$$\mathbf{r}_t|\mathcal{F}_{t-1} \sim \mathrm{N}(\mathbf{0}_n, H_t), \quad t = 1, 2, \ldots, T, \tag{2.1}$$

where for each $t \in \mathbb{N}$, H_t is a positive semidefinite symmetric (PSDS) matrix that is $\mathcal{F}_{t-1}$-measurable, $\mathcal{F}_{t-1} := \sigma(\mathbf{r}_1, \ldots, \mathbf{r}_{t-1})$ being the information set generated by $\{\mathbf{r}_1, \ldots, \mathbf{r}_{t-1}\}$. Thus, H_t is the conditional covariance matrix of $\mathbf{r}_t$ (a vector of returns), that is, $\mathrm{Cov}(\mathbf{r}_t|\mathcal{F}_{t-1}) = H_t$.

Different families of multivariate GARCH (MGARCH) models differ in the way in which the dynamics of the conditional covariance matrix process $\{H_t\}$ is prescribed. For an extensive overview of most of the existing MGARCH models see, for example, [BLR06] and ???Silvennoinen and Terasvirta(200?). Among MGARCH models, one may distinguish two main classes. The first type prescribes the behavior of the conditional covariance matrix process by specifying GARCH processes for the conditional variances and a dynamic process for the conditional correlations of the "deGARCHed" returns [Eng09]. The second group of models directly specifies the dynamics of the conditional variances and covariances as functions of their lagged values and the lagged values of the asset returns. In this work, we focus on a wide subset of the first class, namely the Dynamic Conditional Correlation (DCC) models, and on a particular sub-class of the second, namely the Diagonal VEC (DVEC) models. In the following subsections, we define the general DCC and DVEC setups and the specific parameterizations that we use, together with sufficient parametric restrictions that guarantee the stationarity of the joint process $\{\mathbf{r}_t, H_t\}$ and ensure that for all t the matrix H_t is PSDS.

2.1 The Dynamical Conditional Correlation model

The DCC model class was introduced by [Eng02]. It has become very popular in the multivariate GARCH applied literature, especially in its scalar version, due to the possibility of carrying out estimation using a two-stage procedure. This procedure, together with the availability of an approximate correlation targeting one, makes DCC models applicable and performant in various empirical applications in the context of risk and portfolio management for more than a handful of assets; see, for example, EXTEND LIST: [BCG03, BO13]. The main idea in constructing a DCC model consists in specifying the evolution of the conditional covariance matrix process $\{H_t\}$ through the conditional variances and subsequently a conditional correlation matrix process. The first step produces the so-called "deGARCHed" return vector $\boldsymbol{\varepsilon}_t \in \mathbb{R}^n$ for each $t \in \{1, \ldots, T\}$ via the component-wise assignment $\varepsilon_{i,t} := r_{i,t}/\sigma_{i,t}$, where the conditional variances $\sigma_{i,t}^2$ of each component $r_{i,t}$ of $\mathbf{r}_t$ are obtained by fitting a stationary univariate GARCH model, for example the GARCH(1,1) model of [Bol86]. Let $D_t := \operatorname{diag}(\sigma_{1,t}, \ldots, \sigma_{n,t})$ denote the corresponding diagonal matrix of conditional standard deviations.

The second step of the DCC model construction specifies the dynamics of the conditional correlation matrix R_t of the degarched returns $\boldsymbol{\varepsilon}_t$, embodied in the relation $H_t = D_t R_t D_t$ as:

$$
\begin{aligned}
R_t &= Q_t^{*-1/2} Q_t Q_t^{*-1/2}, \quad Q_t^* := \operatorname{Diag}(Q_t) = I_n \odot Q_t, && (2.2)\\
Q_t &= (\mathbf{i}_n \mathbf{i}_n^\top - A - B) \odot Q + A \odot (\boldsymbol{\varepsilon}_{t-1} \boldsymbol{\varepsilon}_{t-1}^\top) + B \odot Q_{t-1}, && (2.3)
\end{aligned}
$$

where the parameter matrices A and B are symmetric of order n, Q is a PSDS parameter matrix of order n, and $\mathbf{i}_n$ is the column vector of n elements all equal to unity. Equation (2.3) is the most general prescription proposed by [Eng02]. It is reduced to a scalar model when all the elements of A are identical and likewise those of B.

The correlation targeting procedure mentioned above consists in assuming that $Q = \mathrm{E}\left[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top\right]$ and can thus be estimated by $\hat{Q} := \sum_{t=1}^T \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top / T$ prior to estimating A and B. Despite the fact that Q is not equal to the second moment matrix of $\boldsymbol{\varepsilon}_t$, and as a consequence, $\hat{Q}$ is not a consistent estimator of Q (see [Aie13]), the described targeting procedure is used in almost all applications of the DCC model. For this reason, we proceed under this targeting procedure, which, according to simulation results in [Aie13], does not lead to strong biases in practice.

In order to guarantee the stationarity of the joint process $\{\mathbf{r}_t, H_t\}$ and to ensure that the matrix $H_t = D_t R_t D_t$ is PSDS, the following sufficient conditions on the parameters A and B are imposed (see [BGO15] for the details):

(PSD) Positivity constraints:

$$
A \succeq 0, \quad B \succeq 0, \quad \left(\mathbf{i}_n \mathbf{i}_n^\top - A - B\right) \odot S \succ 0, \quad \text{and} \quad Q_0 \succeq 0, \tag{2.4}
$$

where Q_0 is the initial value in the iterative equation (2.3), and $A \succeq (\succ) 0$ means that A is PSD (strictly PSD). As it is shown in Proposition 2.1 of [BGO15], these positive definiteness conditions do not only imply the positivity of $\{H_t\}$, but also that

$$
|A_{ij} + B_{ij}| < 1, \quad j \leq i \in \{1, \ldots, n\}. \tag{2.5}
$$

The latter set of conditions is usually associated to the existence of stationary solutions for the DCC process, even though there is no proof that it actually guarantees that property. Such proof is only available for the cDCC model of [Aie13]. Notice that the conditions (2.5) need to be imposed in addition to the **(PSD)** constraints if Q is not targeted but estimated jointly with the other parameters.

2.2 The Diagonal VEC model

The DVEC family of MGARCH models was introduced by Bollerslev in [BEW88]. DVEC is a multivariate generalization of the GARCH process [Bol86], which is obtained as a particular case of the VEC model in [BEW88]. DVEC is much more parsimonious in parameters than VEC, and in this respect it is comparable to DCC even if the latter is a little less parsimonious that DVEC. The DVEC dynamic equation for H_t is specified by the recursive relation:

$$H_t = (\mathbf{i}_n \mathbf{i}_n^\top - A - B) \odot S + A \odot (\mathbf{r}_{t-1} \mathbf{r}_{t-1}^\top) + B \odot H_{t-1}, \tag{2.6}$$

where the parameter matrices A and B are symmetric of order n, and S is a PSDS parameter matrix of order n. In this model, S is equal to $\mathrm{E}\left[\mathbf{r}_t \mathbf{r}_t^\top\right]$ provided that the expectation exists, so that a consistent targeting estimator of S is given by $\hat{S} = \sum_{t=1}^{T} \mathbf{r}_t \mathbf{r}_t^\top / T$. This consistent targeting procedure will be used in the sequel of this article.

The stationarity of the joint process $\{\mathbf{r}_t, H_t\}$, hence the existence of $\mathrm{E}\left[\mathbf{r}_t \mathbf{r}_t^\top\right]$), and the positive definiteness of the elements of $\{H_t\}$ are guaranteed if the following sufficient conditions on the parameters A and B are satisfied:

(PSD) Positivity constraints:

$$A \succeq 0, \quad B \succeq 0, \quad \left(\mathbf{i}_n \mathbf{i}_n^\top - A - B\right) \odot S \succ 0, \quad \text{and} \quad H_0 \succeq 0, \tag{2.7}$$

where H_0 is the initializing value in the iterative equation (2.6). These conditions imply by Proposition 2.1 in [BGO15] that $|A_{ij} + B_{ij}| < 1$, $j \leq i \in \{1, \ldots, n\}$ which, in turn, guarantee the stationarity of the process $\{\mathbf{r}_t\}$.

2.3 Particular DVEC and DCC subfamilies and associated constraints

We make use of four specific parametrization of symmetric matrices in the DCC and DVEC setups, namely, the Hadamard, the rank one, the Almon, and the scalar versions. For a detailed discussion of the considered models in the DCC case we refer the reader to [BGO15]. We adopt the same conventions for the DVEC family. Since the dynamic equations of Q_t in the case of the DCC model and of H_t in the case of the DVEC model are formally similar, we use in this subsection the symbols K_t, K, and $\boldsymbol{\rho}_t$ to denote either Q_t, Q, and $\boldsymbol{\varepsilon}_t$ for DCC, or H_t, S, and $\mathbf{r}_t$ for DVEC. Each paragraph of this subsection contains the dynamic equation for K_t in terms of an intrinsic set of parameters, and the associated constraints on these parameters which ensure that the **(PSD)** constraints stated in the previous subsections are satisfied. In two cases, identification constraints are also needed.

2.3.1 Hadamard model

The symmetric matrices A and B in (2.3) can be parametrized with $(\boldsymbol{a}, \boldsymbol{b}) \in \mathbb{R}^N \times \mathbb{R}^N$, where $N = \frac{1}{2} n\,(n+1)$, by setting $A := \mathrm{math}\,(\boldsymbol{a})$ and $B := \mathrm{math}\,(\boldsymbol{b})$. The operator $\mathrm{math} : \mathbb{R}^N \longrightarrow \mathbb{S}_n$ is the inverse of the vech operator that stacks the elements on and below the main diagonal of a symmetric matrix of order n into a vector of length N. Using $\boldsymbol{a}$ and $\boldsymbol{b}$ as intrinsic parameters, the Hadamard prescription of the DCC model is

$$K_t = (\mathbf{i}_n\mathbf{i}_n^\top - \mathrm{math}(\boldsymbol{a}) - \mathrm{math}(\boldsymbol{b})) \odot K + \mathrm{math}(\boldsymbol{a}) \odot (\boldsymbol{\rho}_{t-1}\boldsymbol{\rho}_{t-1}^\top) + \mathrm{math}(\boldsymbol{b}) \odot K_{t-1} \tag{2.8}$$

In terms of the intrinsic parameters, the parameter constraints are:

(PSD) Positivity constraints:

$$\mathrm{math}\,(\boldsymbol{a}) \succeq 0, \quad \mathrm{math}\,(\boldsymbol{b}) \succeq 0, \quad \Big(\mathbf{i}_n\mathbf{i}_n^\top - \mathrm{math}\,(\boldsymbol{a}) - \mathrm{math}\,(\boldsymbol{b})\Big) \odot S \succ 0, \quad \text{and} \quad K_0 \succeq 0. \tag{2.9}$$

2.3.2 Rank one model

In this version of the DCC model, the matrices A and B in (2.3) are assumed to have their rank equal to unity. They are parametrized in terms of intrinsic parameter vectors $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n$, such that $A := \widetilde{\boldsymbol{a}}\widetilde{\boldsymbol{a}}^\top$, $B := \widetilde{\boldsymbol{b}}\widetilde{\boldsymbol{b}}^\top$. In this notation the rank one model specification can be written as

$$K_t = (\mathbf{i}_n\mathbf{i}_n^\top - \widetilde{\boldsymbol{a}}\widetilde{\boldsymbol{a}}^\top - \widetilde{\boldsymbol{b}}\widetilde{\boldsymbol{b}}^\top) \odot K + (\widetilde{\boldsymbol{a}}\widetilde{\boldsymbol{a}}^\top) \odot (\boldsymbol{\rho}_{t-1}\boldsymbol{\rho}_{t-1}^\top) + (\widetilde{\boldsymbol{b}}\widetilde{\boldsymbol{b}}^\top) \odot K_{t-1}. \tag{2.10}$$

For the model to be well identified, it suffices to impose that one element of $\boldsymbol{a}$ and one of $\boldsymbol{b}$ are positive. The constraints are:

(IC) Identification constraints:

$$a_1 > 0, \quad b_1 > 0. \tag{2.11}$$

(PSD) Positivity constraints:

$$(\mathbf{i}_n\mathbf{i}_n^\top - \widetilde{A}\widetilde{A}^\top - \widetilde{B}\widetilde{B}^\top) \odot K \succ 0, \quad K_0 \succeq 0. \tag{2.12}$$

2.3.3 Almon model

Given $n \in \mathbb{N}$ and $\mathbf{v} \in \mathbb{R}^3$, the Almon lag operator $\mathrm{alm}_n : \mathbb{R}^3 \longrightarrow \mathbb{R}^n$ generates a vector $\mathrm{alm}_n(\mathbf{v})$ whose entries are $(\mathrm{alm}_n(\mathbf{v}))_i = v_1 + \exp(v_2 i + v_3 i^2), i \in \{1, \ldots, n\}$. Next, let $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^3$ be the intrinsic parameters and define $\widetilde{\boldsymbol{a}} := \mathrm{alm}_n(\boldsymbol{a})$, $\widetilde{\boldsymbol{b}} := \mathrm{alm}_n(\boldsymbol{b}) \in \mathbb{R}^n$. The parameter matrices A, B in (2.3) can be written as $A := \widetilde{\boldsymbol{a}}\widetilde{\boldsymbol{a}}^\top$, $B := \widetilde{\boldsymbol{b}}\widetilde{\boldsymbol{b}}^\top$, and hence the Almon model specification is given by:

$$K_t = (\mathbf{i}_n\mathbf{i}_n^\top - \widetilde{\boldsymbol{a}}\widetilde{\boldsymbol{a}}^\top - \widetilde{\boldsymbol{b}}\widetilde{\boldsymbol{b}}^\top) \odot K + (\widetilde{\boldsymbol{a}}\widetilde{\boldsymbol{a}}^\top) \odot (\boldsymbol{\rho}_{t-1}\boldsymbol{\rho}_{t-1}^\top) + (\widetilde{\boldsymbol{b}}\widetilde{\boldsymbol{b}}^\top) \odot K_{t-1}. \tag{2.13}$$

The corresponding constraints are:

(IC) Identification constraints:

$$\widetilde{a}_1 > 0, \ \widetilde{b}_1 > 0, \quad \text{i.e. } a_1 + \exp(a_2 + a_3) > 0, \ b_1 + \exp(b_2 + b_3) > 0. \tag{2.14}$$

(PSD) Positivity constraints:

$$(\mathbf{i}_n\mathbf{i}_n^\top - \widetilde{\boldsymbol{a}}\widetilde{\boldsymbol{a}}^\top - \widetilde{\boldsymbol{b}}\widetilde{\boldsymbol{b}}^\top) \odot K \succ 0, \quad K_0 \succeq 0. \tag{2.15}$$

2.3.4 Scalar model

The parameter matrices A, B in (2.3) are of the form $A = a\mathbf{i}_n\mathbf{i}_n^\top$, $B = b\mathbf{i}_n\mathbf{i}_n^\top$, with $a, b \in \mathbb{R}$ the intrinsic parameters. The scalar model prescription and the associated constraints can be written as:

$$K_t = (1 - a - b)\, K + a\ \boldsymbol{\rho}_{t-1}\boldsymbol{\rho}_{t-1}^\top + b\ K_{t-1}. \tag{2.16}$$

(PSD) Positivity constraints:

$$a \geq 0, \quad b \geq 0, \quad a + b < 1, \quad K \succeq 0, \quad K_0 \succeq 0. \tag{2.17}$$

3 Constrained composite likelihood estimation

The goal of this section is to provide an overview of the "composite likelihood (CL)" estimation technique that we use for the DCC and DVEC models presented in the previous section, keeping in mind that estimation is performed under the constraints exposed in that section. The CL method is used and studied in [PESS14] for the **scalar** version of the DCC and BEKK models (scalar BEKK is equivalent to scalar DCC) but it has not been used for the estimation of *non-scalar* versions of these models. In the next section, we provide evidence of the theoretical validity of applying CL estimation to **non-scalar** models and in Section 5 we show its empirical relevance.[5]

To deal with the positivity (PSD) parameter constraints that substantially complicate the estimation problem, especially in heavily parametrised model versions and when the number of assets is more than a handful, we use the Bregman divergence based optimization technique first introduced in [CO14] for the QML estimation of full VEC models, and extended to DCC models in [BGO15]. This approach is working well to impose the constraints in QML estimation and therefore we apply it also for CL estimation.

The construction of the composite likelihood function for a given sample of returns and model requires first introducing the definition of a data subset or, equivalently, a reduced vector of returns. Let $\mathbf{r}_t \in \mathbb{R}^n$ be a vector of n returns and let $\Sigma^{\mathcal{L}}$ be a selection matrix of dimension $d \times n$ for the d returns with labels $\mathcal{L} := \{l_1, \ldots, l_d\}$, defined as having entries

$$\Sigma^{\mathcal{L}}_{ij} := \delta_{jl_i},\ \ i \in \{1, \ldots, d\}\,,\ \ j \in \{1, \ldots, n\}\,, \tag{3.1}$$

where δ_{jl_i} is Kronecker's delta. Then the corresponding reduced vector is obtained as

$$\mathbf{r}_t^{\mathcal{L}} := \Sigma^{\mathcal{L}}\mathbf{r}_t = ((r_t)_{l_1}, \ldots, (r_t)_{l_d})^\top \in \mathbb{R}^d. \tag{3.2}$$

[5]The CL method is based on the pseudo-likelihood in [Bes74] and the partial likelihood in [Cox75]. The notion of composite likelihood was introduced by Lindsay in [Lin88]. The use of the composite likelihood has received increasing attention in recent years due to its simplicity at the time of defining the objective function and to its computational advantages when dealing with data with complex structure: see for instance [KN00, ZJ05, FV06, DL09]). Reviews can be found in [Var08, VRF11, PSS11] and [Pak14]. We refer the reader to [PESS14] for a detailed proof of the consistency and the asymptotic normality of the CL estimator in the scalar models. Extending these results to non-scalar models is not our objective.

Even though it is mainly the case $d = 2$ that is usually considered in practice, we formulate in what follows the general case. The log-likelihood function associated to the reduced process $\left\{\mathbf{r}_t^{\mathcal{L}}, H_t^{\mathcal{L}}\right\}$ corresponding to the reduced vector $\mathbf{r}_t^{\mathcal{L}}$ of d returns with labels $\mathcal{L}$ can be written as

$$\log L_{\mathcal{L}}(\mathbf{\Theta}; \mathbf{r}^{\mathcal{L}}) = \sum_{t=1}^{T} l_{\mathcal{L},t}(\mathbf{\Theta}; \mathbf{r}_t^{\mathcal{L}}), \tag{3.3}$$

where

$$l_{\mathcal{L},t}(\mathbf{\Theta}; \mathbf{r}_t^{\mathcal{L}}) = -\frac{1}{2}\left(\log(2\pi) + \log\det(H_t^{\mathcal{L}}) + \mathbf{r}_t^{\mathcal{L}\top} H_t^{\mathcal{L}-1}\mathbf{r}_t^{\mathcal{L}}\right). \tag{3.4}$$

Given N different sets of labels $\{\mathcal{L}_1, \ldots, \mathcal{L}_N\}$ of returns, the corresponding composite likelihood $CL\left(\mathbf{\Theta}; \mathbf{r}\right)$, where $\mathbf{r}$ denotes the relevant set of T observations is defined as

$$CL\left(\mathbf{\Theta}; \mathbf{r}\right) := \sum_{i=1}^{N} \log L_{\mathcal{L}_i}(\mathbf{\Theta}; \mathbf{r}^{\mathcal{L}_i}) = \sum_{t=1}^{T}\sum_{i=1}^{N} l_{\mathcal{L}_i,t}(\mathbf{\Theta}; \mathbf{r}_t^{\mathcal{L}_i}). \tag{3.5}$$

For example, when only the distinct pairs ($d = 2$) of asset returns are chosen to construct the log-likelihood $CL\left(\mathbf{\Theta}; \mathbf{r}\right)$, this function is made of $N := n(n-1)/2$ composite components of the type $\log L_{\mathcal{L}_i}(\mathbf{\Theta}; \mathbf{r}^{\mathcal{L}_i})$.

In order to implement the Bregman based CL estimation, we use the analytical expression of the gradient of the function $CL\left(\mathbf{\Theta}; \mathbf{r}\right)$, keeping in mind the dependence of $\mathbf{\Theta}$ on the intrinsic parameters (denoted generically by $\boldsymbol{\theta}$) for each model specification. We proceed by noticing that, by construction, each of the functions $\log L_{\mathcal{L}_i}(\mathbf{\Theta}; \mathbf{r}^{\mathcal{L}_i})$, $i = \{1, \ldots, N\}$ is associated to the corresponding reduced sample $\mathbf{r}^{\mathcal{L}_i}$, as well as to the model that prescribes the dynamics of the corresponding reduced matrix process $\{H_t^{\mathcal{L}_i}\}$ obtained from the initial one via the use of the particular selection matrix $\Sigma^{\mathcal{L}_i}$ at each moment of time t. We refer to the operation that transforms the process $\mathbf{r}$ to $\mathbf{r}^{\mathcal{L}_i}$ as **model reduction**.

In the following section we study in detail the model reduction for the DCC and DVEC families and we address the question of the closure of each of the specifications considered under this procedure, as well as the question of how to optimally impose the constraints that ensure that the resulting reduced model satisfies the positivity and stationarity requirements. The findings in that section allow us to use the explicit expressions of the gradients of the log-likelihood function provided in [BGO15] for the DCC models and in Proposition B.1 of the SoA for the DVEC, in order to determine the gradients $\nabla_{\boldsymbol{\theta}^{\Sigma^{\mathcal{L}_i}}} \log L_{\mathcal{L}_i}(\mathbf{\Theta}; \mathbf{r}^{\Sigma^{\mathcal{L}_i}})$ of the log-likelihood functions $\log L_{\mathcal{L}_i}(\mathbf{\Theta}; \mathbf{r}^{\Sigma^{\mathcal{L}_i}})$, $i = \{1, \ldots, N\}$ with respect to the intrinsic parameters $\boldsymbol{\theta}^{\Sigma^{\mathcal{L}_i}}$ corresponding to each reduced model. Having these results at hand, we can explicitly formulate the functional connection between $\nabla_{\boldsymbol{\theta}^{\Sigma^{\mathcal{L}_i}}} \log L_{\mathcal{L}_i}(\mathbf{\Theta}; \mathbf{r}^{\Sigma^{\mathcal{L}_i}})$ and $\nabla_{\boldsymbol{\theta}} \log L_{\mathcal{L}_i}(\mathbf{\Theta}; \mathbf{r}^{\Sigma^{\mathcal{L}_i}})$ for each $i = \{1, \ldots, N\}$, and hence $\nabla_{\boldsymbol{\theta}} CL\left(\mathbf{\Theta}(\boldsymbol{\theta}); \mathbf{r}\right)$ can be immediately obtained as the sum of all the latter for each specification that remains closed under reduction. The implementation of the CL estimation procedure is then identical to the QML one applicable to the initial non-reduced models and briefly explained in Section B.2 of the SoA. For more details concerning the Bregman divergences based constrained estimation, we refer the reader to [BGO15] and to its Technical Appendix.

4 Reduction of non-scalar DCC and DVEC models

Since the CL approach involves subsets of the data, one of the questions that needs to be addressed in the case of non-scalar models is whether each of the models defined in Section 2 remains closed under such a reduction procedure. In this section we show that this is indeed the case through two main propositions, one for the DCC family, and one for the DVEC family. Additional results are provided in Section C of the SoA for each of the particular models (Hadamard, rank one, and Almon) of each class. All these results justify the CL estimation method presented in the previous section in the context of the non-scalar DCC and DVEC families.

4.1 Reduction of DCC models

In this subsection we provide the main result concerning the closure of the DCC family under the reduction procedure.

Proposition 4.1 (Reduction of non-scalar DCC models) *Consider the DCC model defined by (2.1)-(2.3) together with $H_t = D_t R_t D_t$. Let $\Sigma^{\mathcal{L}}$ be the selection matrix for the set $\mathcal{L} := \{l_1, \ldots, l_d\}$ of $d \in \mathbb{N}$ labels associated to the indices of the asset returns $\mathbf{r}_t \in \mathbb{R}^n$ having the entries defined in (3.1). Then,*

(i) *The reduced process $\{\mathbf{r}_t^{\mathcal{L}}\}$ defined by $\mathbf{r}_t^{\mathcal{L}} := \Sigma^{\mathcal{L}} \mathbf{r}_t \in \mathbb{R}^d$ for all $t \in \{1, \ldots, T\}$ is a DCC process whose conditional covariance matrix process $\{H_t^{\mathcal{L}}\}$ can be written as*

$$H_t^{\mathcal{L}} = \Sigma^{\mathcal{L}} H_t \Sigma^{\mathcal{L}\top}, \text{ with } H_t^{\mathcal{L}} \in \mathbb{S}_d^+. \tag{4.1}$$

The dynamics of the matrix process $\{H_t^{\mathcal{L}}\}$ is determined by the reduced equations:

$$H_t^{\mathcal{L}} = D_t^{\mathcal{L}} R_t^{\mathcal{L}} D_t^{\mathcal{L}}, \tag{4.2}$$

$$R_t^{\mathcal{L}} = (Q_t^{\mathcal{L}})^{*-1/2} Q_t^{\mathcal{L}} (Q_t^{\mathcal{L}})^{*-1/2}, \tag{4.3}$$

$$Q_t^{\mathcal{L}} = \left(\mathbf{i}_d \mathbf{i}_d^\top - A^{\mathcal{L}} - B^{\mathcal{L}}\right) \odot Q^{\mathcal{L}} + A^{\mathcal{L}} \odot \left(\boldsymbol{\varepsilon}_{t-1}^{\mathcal{L}} \boldsymbol{\varepsilon}_{t-1}^{\mathcal{L}}\right) + B^{\mathcal{L}} \odot Q_{t-1}^{\Sigma\mathcal{L}}, \tag{4.4}$$

where $D_t^{\mathcal{L}} = \Sigma^{\mathcal{L}} D_t \Sigma^{\mathcal{L}\top}$, $R_t^{\mathcal{L}} = \Sigma^{\mathcal{L}} R_t \Sigma^{\mathcal{L}\top}$, $Q_t^{\mathcal{L}} = \Sigma^{\mathcal{L}} Q_t \Sigma^{\mathcal{L}\top}$, $(Q_t^{\mathcal{L}})^* = \Sigma^{\mathcal{L}} Q_t^* \Sigma^{\mathcal{L}\top}$, $A^{\mathcal{L}} = \Sigma^{\mathcal{L}} A \Sigma^{\mathcal{L}\top}$, $B^{\mathcal{L}} = \Sigma^{\mathcal{L}} B \Sigma^{\mathcal{L}\top}$, $Q^{\mathcal{L}} = \Sigma^{\mathcal{L}} Q \Sigma^{\mathcal{L}\top}$, and $\boldsymbol{\varepsilon}_t^{\Sigma} = \Sigma^{\mathcal{L}} \boldsymbol{\varepsilon}_t$. Notice that $D_t^{\mathcal{L}}$, $(Q_t^{\mathcal{L}})^* \in \mathbb{D}_d$, $R_t^{\mathcal{L}}$, $Q_t^{\mathcal{L}}$, $A^{\mathcal{L}}$, $B^{\mathcal{L}}$, $Q^{\mathcal{L}} \in \mathbb{S}_d$, and $\boldsymbol{\varepsilon}_t^{\mathcal{L}} \in \mathbb{R}^d$.

(ii) *If the parameters $(A, B) \in \mathbb{S}_n \times \mathbb{S}_n$ satisfy the positive semidefiniteness* **(PSD)** *constraints (2.4), then so do the parameters $(A^{\mathcal{L}}, B^{\mathcal{L}}) \in \mathbb{S}_d \times \mathbb{S}_d$ of the reduced model associated to $\Sigma^{\mathcal{L}}$.*

Proof. **(i)** In order to prove the relation (4.2), we first write down the conditional covariance of the asset returns $\left\{\mathbf{r}_t^{\Sigma^{\mathcal{L}}}\right\}$ using its definition in terms of the conditional expectation $E_{t-1}[\cdot] := \mathrm{E}[\cdot|\mathcal{F}_{t-1}]$ with respect to the information set $\mathcal{F}_{t-1}$ generated by the returns set $\{\mathbf{r}_1, \ldots, \mathbf{r}_{t-1}\}$. Indeed,

$$H_t^{\mathcal{L}} = \mathrm{E}_{t-1}\left[\mathbf{r}_t^{\mathcal{L}} \mathbf{r}_t^{\mathcal{L}\top}\right] = \mathrm{E}_{t-1}\left[\Sigma^{\mathcal{L}} \mathbf{r}_t \mathbf{r}_t^\top \Sigma^{\mathcal{L}\top}\right] = \Sigma^{\mathcal{L}} \mathrm{E}_{t-1}\left[\mathbf{r}_t \mathbf{r}_t^\top\right] \Sigma^{\mathcal{L}\top} = \Sigma^{\mathcal{L}} H_t \Sigma^{\mathcal{L}\top}. \tag{4.5}$$

We next use (**??**) and point **(iii)** of Lemma 6.1 in the Appendix. Since $D_t \in \mathbb{D}_n$, we hence have

$$H_t^{\mathcal{L}} = \Sigma^{\mathcal{L}} H_t \Sigma^{\mathcal{L}\top} = \Sigma^{\mathcal{L}} D_t R_t D_t \Sigma^{\mathcal{L}\top} = \Sigma^{\mathcal{L}} D_t \Sigma^{\mathcal{L}\top} \Sigma^{\mathcal{L}} R_t \Sigma^{\mathcal{L}\top} \Sigma^{\mathcal{L}} D_t \Sigma^{\mathcal{L}\top} = D_t^{\mathcal{L}} R_t^{\mathcal{L}} D_t^{\mathcal{L}}.$$

We use the same reasoning to prove the relation (4.3). We first write

$$R_t^{\mathcal{L}} = \Sigma^{\mathcal{L}} R_t \Sigma^{\mathcal{L}\top} = \Sigma^{\mathcal{L}} Q_t^{*-1/2} Q_t Q_t^{*-1/2} \Sigma^{\mathcal{L}\top} = \Sigma^{\mathcal{L}} Q_t^{*-1/2} \Sigma^{\mathcal{L}\top} \Sigma^{\mathcal{L}} Q_t \Sigma^{\mathcal{L}\top} \Sigma^{\mathcal{L}} Q_t^{*-1/2} \Sigma^{\mathcal{L}\top}. \quad (4.6)$$

Notice that by **(iv)** of Lemma 6.1 we can define

$$(Q_t^{\mathcal{L}})^* := \mathrm{Diag}(Q_t^{\Sigma^{\mathcal{L}}}) = \mathrm{Diag}(\Sigma^{\mathcal{L}} Q_t \Sigma^{\mathcal{L}\top}) = \Sigma^{\mathcal{L}} Q_t^* \Sigma^{\mathcal{L}\top},$$

and hence

$$(Q_t^{\mathcal{L}})^{*-1/2} = \Sigma^{\mathcal{L}} Q_t^{*-1/2} \Sigma^{\mathcal{L}\top}. \quad (4.7)$$

Consequently, using (4.7) in (4.6) we immediately obtain

$$R_t^{\mathcal{L}} = (Q_t^{\mathcal{L}})^{*-1/2} Q_t^{\mathcal{L}} (Q_t^{\mathcal{L}})^{*-1/2}, \quad (4.8)$$

which proves (4.3). Finally, in order to prove (4.4), we use (**??**) and part **(ii)** in Lemma 6.1, which yield (4.4).

(ii) First, if the parameter matrix A is positive semidefinite, that is $A \succeq 0$, then the reduced parameter matrix $A^{\mathcal{L}} \in \mathbb{S}_d$ of the reduced model associated to $\Sigma^{\mathcal{L}}$ satisfies that $A^{\mathcal{L}} = \Sigma^{\mathcal{L}} A \Sigma^{\mathcal{L}\top} \succeq 0$. Indeed, for any $\mathbf{v} \in \mathbb{R}^d$

$$\langle \mathbf{v}, \Sigma^{\mathcal{L}} A \Sigma^{\mathcal{L}\top} \mathbf{v} \rangle = \langle \Sigma^{\mathcal{L}\top} \mathbf{v}, A(\Sigma^{\mathcal{L}\top} \mathbf{v}) \rangle \succeq 0. \quad (4.9)$$

The proof that $B^{\mathcal{L}} \succeq 0$ is identical. It is also straightforward to show that $\left(\mathbf{i}_d \mathbf{i}_d^\top - A^{\mathcal{L}} - B^{\mathcal{L}}\right) \odot Q^{\mathcal{L}} \succeq 0$ by mimicking (4.9) for the relation $(\mathbf{i}_n \mathbf{i}_n^\top - A - B) \odot S \succeq 0$. ■

4.2 Reduction of non-scalar DVEC models

In this subsection we formulate the closure of the DVEC family under the reduction procedure. The proof is not added since it is similar to the corresponding proof for the DCC model.

Proposition 4.2 (Reduction of DVEC models) *Consider the strong DVEC model defined by (2.1)-(2.6). Let $\Sigma^{\mathcal{L}}$ be the selection matrix for the set $\mathcal{L} := \{l_1, \ldots, l_d\}$ of $d \in \mathbb{N}$ labels associated to the indices of the asset returns $\mathbf{r}_t \in \mathbb{R}^n$ having the entries defined in (3.1). Then,*

(i) *The reduced process $\{\mathbf{r}_t^{\mathcal{L}}\}$ defined by $\mathbf{r}_t^{\mathcal{L}} := \Sigma^{\mathcal{L}} \mathbf{r}_t \in \mathbb{R}^d$, for all $t \in \{1, \ldots, T\}$, is a DVEC model whose conditional covariance matrix process $\{H_t^{\mathcal{L}}\}$ can be written as*

$$H_t^{\mathcal{L}} = \Sigma^{\mathcal{L}} H_t \Sigma^{\mathcal{L}\top}, \ \text{with} \ H_t^{\Sigma^{\mathcal{L}}} \in \mathbb{S}_d^+, \quad (4.10)$$

and whose dynamics is determined by the equation

$$H_t^{\mathcal{L}} = \left(\mathbf{i}_d \mathbf{i}_d^\top - A^{\mathcal{L}} - B^{\mathcal{L}}\right) \odot S^{\mathcal{L}} + A^{\mathcal{L}} \odot \left(\mathbf{r}_{t-1}^{\mathcal{L}} \mathbf{r}_{t-1}^{\mathcal{L}}\right) + B^{\mathcal{L}} \odot H_{t-1}^{\mathcal{L}}, \quad (4.11)$$

where $A^{\mathcal{L}} = \Sigma^{\mathcal{L}} A \Sigma^{\mathcal{L}\top}$, $B^{\mathcal{L}} = \Sigma^{\mathcal{L}} B \Sigma^{\mathcal{L}\top}$, and $S^{\mathcal{L}} = \Sigma^{\mathcal{L}} S \Sigma^{\mathcal{L}\top} \in \mathbb{S}_d$.

(ii) *If the parameters* $(A, B) \in \mathbb{S}_n \times \mathbb{S}_n$ *satisfy the positive semidefiniteness* **(PSD)** *constraints* (2.7)*, then so do the parameters* $(A^{\mathcal{L}}, B^{\mathcal{L}}) \in \mathbb{S}_d \times \mathbb{S}_d$ *of the reduced model associated to* $\Sigma^{\mathcal{L}}$.

To complement Propositions 4.1 and 4.2, Section C of the SoA provides in detail the explicit functional connection between the parameters of the original DCC or DVEC models and the reduced counterparts for each of the model parameter specifications defined in Section 2.

As final remark, we point out that the models defined in Sections 2.1 and 2.2 together with (2.1) are semi-strong according to Definition 2 in [DN93]. The reduced models obtained in Proposition 4.1 and Proposition 4.2 are also semi-strong. If the initial models are strong according to Definition 1 in [DN93] (see also Definition 2.2 in [FZ10]), it cannot be established that the reduced models are also strong.

5 Empirical study

In this section we report the results of various experiments that allow us to the compare the empirical performances of the different DCC and DVEC models described in Section 2.3. We use the same dataset as in [BGO15] and estimate the models for various dimensions (from 5 to 30) using the QML and CL methods discussed in Section 3. We report the point estimates of the models in order to study the bias problem and we assess the out-of-sample performances of the different models using model confidence sets [HLN03, HLN11] with loss functions that measure the ability of the models to forecast covariance and correlation matrices.

5.1 Dataset and competing models

We use two datasets that consist of daily price quotes of the thirty components of the Dow Jones Industrial Average Index (DJIA) as of October 2013, downloaded from the Yahoo Finance database.[6] We consider two different periods:

- **Period I**: the price quotes are taken from January 19th, 1996 to December 21st, 2010. This amounts to 3750 observations in the sample. The first 3000 observations (January 19th, 1996 - December 31st, 2007) are reserved for model estimation and the remaining 750 are used for an out-of-sample study.

- **Period II**: the price quotes are taken from August 25th, 1998 to August 1st, 2013. The resulting sample contains 3750 observations. The first 3000 quotes (August 25th, 1998 - August 9th, 2010) are kept for model estimation and the last 750 for out-of-sample testing.

Period I contains the 2008-09 high volatility events in the out-of-sample interval, while Period II includes them in the interval used for estimation.

[6]The Yahoo tickers of the stocks used in the study are AA, AXP, BA, BAC, CAT, CSCO, CVX, DD, DIS, GE, HD, HPQ, IBM, INTC, JNJ, JPM, KO, MCD, MMM, MRK, MSFT, PFE, PG, T, TRV, UNH, UTX, VZ, WMT, XOM. The dataset has been prepared adjusting the quotes with respect to stock splits and dividend payments and the dates at which at least one of the constituents was not quoted were removed.

The Capital Asset Pricing Model (CAPM) [Sha64] based data preprocessing. In order to account in the modeling for the common dynamical factors that influence all the assets under consideration, we use for each asset i (with $i = 1, 2, \ldots, n$) a static unconditional CAPM one-factor model of the form $Y_{i,t} = \alpha_i + \beta_i Z_t + R_{i,t}$, where for time index t, $Y_{i,t}$ is the log-return of asset i, Z_t is the value of the chosen common factor, $R_{i,t}$ is the regression error term, and α_i, β_i are the intercept and slope coefficients, respectively.

In the DCC/DVEC empirical experiments presented in this section, we use the CAPM regression in the following way: let T_{est} and T_{out} be the sample lengths taken for in-sample estimation and out-of-sample testing, respectively, and let $T := T_{est} + T_{out}$ be the total number time series observations. We take the returns of the S&P500 index as the common factor Z_t, estimate by ordinary least-squares the CAPM regression for each asset i, using the observations $t \in \{1, \ldots, T_{est}\}$, and store the OLS residuals $r_{i,t}$. Following the same approach as in Chapter 8 of [Eng09], we then estimate the DCC/DVEC models via QML or CL on these residual returns $(r_{1,t}\, r_{2,t} \ldots r_{n,t})^\top$. In order to perform the out-of-sample analysis, we compute the out-of-sample residual returns $\mathbf{r} := \{\mathbf{r}_{T_{est}+1}, \ldots, \mathbf{r}_T\}$ according to the relations

$$r_{i,t} = Y_{i,t} - \widehat{\alpha}_i - \widehat{\beta}_i Z_t, \quad i \in \{1, \ldots, n\}, \quad t \in \{T_{est} + 1, \ldots, T\}, \tag{5.1}$$

where $\widehat{\alpha}_i$ and $\widehat{\beta}_i$ are obtained using the T_{est} in-sample observations, and use them for the out-of-sample assessment of the empirical performances of the DCC/DVEC models. For the generation of the out-of-sample forecasts of the conditional covariance matrices H_t associated to each DCC/DVEC model, we also keep the values of the parameter estimates obtained using only the in-sample observations, that is, we do not re-estimate each model by adding one observation at a time in the out-of-sample period.

The CAPM based data preprocessing step is not absolutely necessary for the models to show good performance; its relevance depends on the nature of the data and it is up to the practitioner to carry out this step. The empirical study reported in this section has also been implemented without CAPM preprocessing and the conclusions drawn from these results are broadly similar to those drawn from the results presented in the sequel.

The competing models. In the case of the QML estimation we report results for six different DCC/DVEC model parameterizations, namely: **(i)** the Hadamard model, **(ii)** the rank deficient model with rank $r = 2$, **(iii)** the rank deficient model with rank $r = 1$, **(iv)** the Almon model, **(v)** the Almon shuffle model (briefly described in the following remark), and **(vi)** the scalar model. Some of these models are particular cases of others according to the inclusion relations represented in Figure 1. In the case of the CL estimation, the number of the studied models is five since for the rank deficient model with rank $r = 2$ the composite likelihood estimation cannot be implemented. In Figure 1 we mark in red the rank two deficient model that is excluded from the set when the composite likelihood estimator is applied.

Remark 5.1 The Almon shuffle DCC/DVEC specification is a variant of the Almon model in which the different components of the process $\{\mathbf{r}_t\}$ are ordered in order to enhance the performance of the Almon parameterization. Indeed, experience shows that the modeling performance of the Almon prescription is much influenced by the ability of the Almon function to fit the entry values of the parameter vectors $\widetilde{\boldsymbol{a}}$ and $\widetilde{\boldsymbol{b}}$ that one would obtain by using an unrestricted rank deficient model with $r = 1$. This fit can be improved by first carrying out a reordering of the process components so that the vector entries are as monotonous as possible, hence fostering a good match between the typical profiles of Almon curves and the entry values of $\widetilde{\boldsymbol{a}}$ and $\widetilde{\boldsymbol{b}}$. The proposed reordering (shuffle) consists in arranging the components in descending order according to the magnitude of their projection onto the first principal component (the one that corresponds to the largest eigenvalue) computed using the unconditional covariance matrix of the available sample. Once the reordering is implemented, the Almon model of the Subsection 2.3 is used, and the results regarding the reduction procedure in Section C.3 remain valid. The results in Section 5 show that the Almon shuffle DCC/DVEC models often exhibit a better performance than the corresponding Almon models.

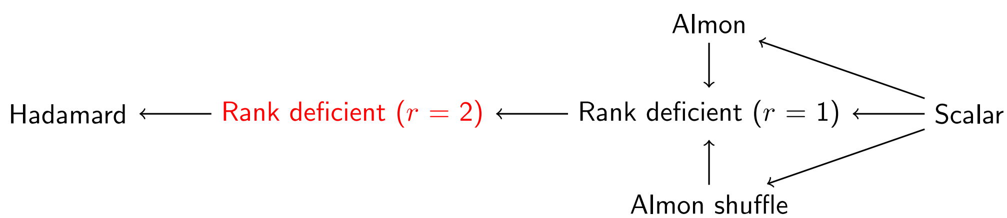

Figure 1: Inclusion hierarchy of models. The symbol $B \longleftarrow A$ should be read as "A is a particular case of B".

5.2 In-sample results

The estimations are performed in several dimensions ranging from 5 up to 30 in such a way that for the n-dimensional case, the first n assets are picked in the DJIA dataset arranged in alphabetical order. For the case of the DCC models the estimations are done conditionally on the estimated standardized ("degarched") residuals and the approximate targeting estimator of S, which are the same for all models. In the case of the DVEC models, the estimations are implemented directly on the CAPM residuals using an exact variance targeting based on them.

Features of the point estimates. Tables **??**-**??** provide the in-sample estimates of the matrices A and B for the two periods for both the DCC and the DVEC families, as well as for the QML and the CL estimators. These tables also contain the estimates obtained using the McGyver method in[Eng08]. For dimension 30, this method consists in estimating all (i.e. 435) bivariate scalar DCC or DVEC models, thus getting as many point estimates of a and b. For each dimension smaller than 30 (except 5), the estimates for the corresponding group of assets are used (e.g. the first ten assets for dimension 10, which provide 45 estimates that are a subset of the 435 ones for dimension 30). For dimension 5, six groups of five assets are formed, (1-5, 6-10, until 26-30) which yield 60 estimates altogether for the six groups. For each dimension, the McGyver estimates exhibit much more heterogeneity than the different DCC/DVEC models; this is a consequence of the positive semidefiniteness constraints imposed on the DCC/DVEC models, to which the McGyver estimator is not exposed. Additionally, unusual values arise sometimes in the McGyver case in comparison with the DCC/DVEC estimates; more specifically, there are values of b close to zero (see the minimum McGyver values) and of a farther away from zero (see the McGyver maximum values). The mean values of the McGyver estimates of a are therefore larger than the corresponding medians (reported in the tables), and the reverse is true for the b estimates. The median values of the a and b estimates are hardly influenced by the dimension.

5.3 Out-of-sample specification tests

The specification tests used in order to assess the out-of-sample one-step ahead forecasting performance of the competing DCC/DVEC models estimated via QML and CL methods are presented in this subsection. The first test is based on the use of multivariate variance standardized returns, and the next three on the use of portfolio returns.

5.3.1 Model confidence sets based on correlation and covariance loss functions

The different models are compared by computing the model confidence set (MCS) of [HLN03, HLN11] using the following loss functions:

$$d_t^{\mathrm{corr}} := \frac{2}{n(n-1)} \sum_{i<j=2,\ldots,n} \left(\varepsilon_{i,t}\varepsilon_{j,t} - \rho_{ij,t}\right)^2, \tag{5.2}$$

and

$$d_t^{\mathrm{cov}} := \frac{2}{n(n-1)} \sum_{i<j=2,\ldots,n} \left(r_{i,t}r_{j,t} - h_{ij,t}\right)^2, \tag{5.3}$$

where $\boldsymbol{\varepsilon}_t$ are the GARCH standardized returns, $\rho_{ij,t}$ and $h_{ij,t}$ are the (i,j)-entries of the model dependent conditional correlation R_t and conditional covariance H_t matrices, respectively.

The results of these tests are provided in Tables **??**-**??**, **??**-**??** for both periods, two families of models, and both estimation methods considered.

5.3.2 Tests based on portfolio returns

The performances of the competing DCC models can be compared indirectly by running tests on portfolios constructed using the assets whose returns are modeled. Let $\mathbf{w}_t \in \mathbb{R}^n$ denote a vector of portfolio weights at date t, $p_t = \mathbf{w}_t^\top \mathbf{r}_t$ the portfolio return, and $\sigma_{p,t}^2 = \mathbf{w}_t^\top H_t \mathbf{w}_t$ the corresponding portfolio variance, where H_t is the relevant conditional covariance matrix of $\mathbf{r}_t$ (see (2.1)). Two kinds of portfolios are constructed:

- The minimum variance portfolio (MVP), defined by choosing a weight vector $\mathbf{w}_t$ that minimizes $\mathbf{w}^\top H_t \mathbf{w}$ subjected to the constraint $\mathbf{i}_n^\top \mathbf{w}_t = 1$. The solution of this problem is given by $\mathbf{w}_t = H_t^{-1}\mathbf{i}_n / \mathbf{i}_n^\top H_t^{-1}\mathbf{i}_n$. This expression is used to construct the sequence of variance minimizing portfolios associated to each model.

- The equally weighted portfolio (EWP), implied by the weight vector $\mathbf{w}_t := \mathbf{i}_n/n$ for each date t.

Three tests are considered, based on the observation that under a correct specification of the type (2.1), the standardized portfolio return $y_t = \mathbf{w}_t^\top \mathbf{r}_t / \sqrt{\mathbf{w}_t^\top H_t \mathbf{w}_t}$, has unconditional variance equal to one. The tests assess the validity of different hypotheses for the series $\widehat{y}_t = \mathbf{w}_t^\top \mathbf{r}_t / \sqrt{\mathbf{w}_t^\top \widehat{H}_t \mathbf{w}_t}$ constructed using the one-step ahead forecast of the conditional covariance matrices $\widehat{H}_t$ implied by each of the estimated models under consideration.

Model confidence set (MCS) based on the predictive ability for squared portfolio returns: The one-step ahead predictive ability of the models is evaluated by computing model confidence sets using the loss function

$$d_t := \left((\mathbf{w}_t^\top \mathbf{r}_t)^2 - \mathbf{w}_t^\top \widehat{H}_t \mathbf{w}_t \right)^2 . \tag{5.4}$$

The results of this procedure are provided in Tables **??**-**??** for Period I and in Tables **??**-**??** for Period II. The information that they contain is organized in the same way that was already described at the end of Section 5.3.1.

6 Conclusions

In this work we achieve four major goals: (i) show that the optimization techniques proposed in [CO14, BGO15] allow for the further extension to the constrained quasi-maximum likelihood (QML) estimation of different richly parametrized multivariate volatility models; in particular, we consider DVEC family (ii) improve the estimation of high dimensional models using the composite likelihood (CL) estimator implemented via what we call model reduction procedures, (iii) provide details as to the reduction of each of the non-scalar models of interest, and (iv) address the question of the empirical performance of the estimated models and the possibility to solve various estimation issues documented in the literature by using the proposed techniques.

We also showed that (i) the bias presence issue for the non-scalar DCC models can be eliminated or at least reduced by selecting other non-scalar multivariate volatility families of models that use a one-stage estimation procedure and for which exact covariance targeting is available, and also by using the composite likelihood estimation method, which has already been shown to improve the quality of the parameter estimation in the context of high dimensional scalar multivariate volatility models; (ii) we provided empirical results that can help practitioners in choosing which of the model families under study and the estimation techniques considered, namely QML or CL, are the most appropriate to be used in connection to a particular dataset and application needed.

Appendix

Lemma 6.1 *Let $A, B \in \mathbb{S}_n$, $C \in \mathbb{S}_d$, $D \in \mathbb{D}_n$, and let Σ be the selection matrix for the $d \leq n \in \mathbb{N}$ labels $\{l_1, \ldots, l_d\}$. Then*

(i) $\left(\Sigma A \Sigma^\top\right)_{ij} = A_{l_i l_j}$, $i, j \in \{1, \ldots, d\}$.

(ii) $\Sigma\left(A\odot B\right)\Sigma^{\top}=\Sigma A\Sigma^{\top}\odot\Sigma B\Sigma^{\top}$.

(iii) $\Sigma\Sigma^{\top}=\mathbb{I}_d$.

(iv) $\left(\Sigma^{\top}\Sigma\right)_{ij}=\begin{cases}\delta_{ij}, & j\in\{l_1,\ldots,l_d\}\\ 0, & \text{otherwise}\end{cases}, \quad for\ \ i,j\in\{1,\ldots,n\}$.

(v) $\Sigma D\Sigma^{\top}\Sigma A\Sigma^{\top}\Sigma D\Sigma^{\top}=\Sigma DAD\Sigma^{\top}$.

(vi) $\mathrm{Diag}(\Sigma A\Sigma^{\top})=\Sigma\mathrm{Diag}(A)\Sigma^{\top}$.

(vii) $\mathrm{Diag}(\Sigma^{\top}C\Sigma)=\mathrm{Diag}(\Sigma^{\top}\mathrm{Diag}(C)\Sigma)=\Sigma^{\top}\mathrm{Diag}(C)\Sigma$.

Proof 6.2 **(i)** Is straightforward.

(ii) On the one hand, for any $i,j\in\{1,\ldots,d\}$

$$\left(\Sigma\left(A\odot B\right)\Sigma^{\top}\right)_{ij}=\sum_{k=1}^{n}\sum_{m=1}^{n}\Sigma_{ik}(A\odot B)_{km}\Sigma_{jm}=\sum_{k=1}^{n}\sum_{m=1}^{n}\delta_{kl_i}(A\odot B)_{km}\delta_{ml_j}=\left(A\odot B\right)_{l_il_j}=A_{l_il_j}B_{l_il_j} \tag{6.1}$$

On the other hand, it is clear that

$$\left(\Sigma A\Sigma^{\top}\odot\Sigma B\Sigma^{\top}\right)_{ij}=A_{l_il_j}B_{l_il_j}, \tag{6.2}$$

as required.

(iii) Let $i,j\in\{1,\ldots,d\}$, then

$$(\Sigma\Sigma^{\top})_{ij}=\sum_{k=1}^{n}\Sigma_{ik}\Sigma_{jk}=\sum_{k=1}^{n}\delta_{kl_i}\delta_{kl_j}=\delta_{ij}, \tag{6.3}$$

where we use the fact that

$$\delta_{ij}\delta_{ij}=\begin{cases}1, & k=l_i,k=l_j,\iff\ l_i=l_j\iff\ i=j\\ 0, & \text{otherwise.}\end{cases} \tag{6.4}$$

(iv) Let $i,j\in\{1,\ldots,n\}$, then

$$(\Sigma^{\top}\Sigma)_{ij}=\sum_{k=1}^{d}\Sigma_{ki}\Sigma_{kj}=\sum_{k=1}^{d}\delta_{il_k}\delta_{jl_k}=\delta_{ij}, \tag{6.5}$$

where we use the fact that

$$\delta_{il_k}\delta_{jl_k}=\begin{cases}1, & i=l_k,j=l_k,\iff\ i=j,j\in\{l_1,\ldots,l_d\}\\ 0, & \text{otherwise.}\end{cases} \tag{6.6}$$

(v) On the one hand, for $i, j \in \{1, \ldots, d\}$

$$\begin{aligned}\left(\Sigma D \Sigma^\top \Sigma A \Sigma^\top \Sigma D \Sigma^\top\right)_{ij} &= \sum_{q,t,s,k=1}^{n} (\Sigma D)_{ik} \left(\Sigma^\top \Sigma\right)_{ks} A_{st} \left(\Sigma^\top \Sigma\right)_{tq} \left(D \Sigma^\top\right)_{qj} \\ &= \sum_{q,t,s,k=1}^{n} \sum_{u,v=1}^{d} (\Sigma D)_{ik} \delta_{kl_u} \delta_{sl_u} A_{st} \delta_{tl_v} \delta_{ql_v} \left(D \Sigma^\top\right)_{qj} \\ &= \sum_{u,v=1}^{d} (\Sigma D)_{il_u} A_{l_u l_v} \left(D \Sigma^\top\right)_{l_v j} = \sum_{u,v=1}^{d} \sum_{s,k=1}^{n} \Sigma_{ik} D_{kl_u} A_{l_u l_v} D_{l_v s} \Sigma_{js} \\ &= \sum_{u,v=1}^{d} \sum_{s,k=1}^{n} \delta_{kl_i} D_{kl_u} A_{l_u l_v} D_{l_v s} \delta_{sl_j} = \sum_{u,v=1}^{d} D_{l_i l_u} A_{l_u l_v} D_{l_v l_j} = d_{l_i} A_{l_i l_j} d_{l_j}.\end{aligned} \tag{6.7}$$

On the other hand, for $i, j \in \{1, \ldots, d\}$

$$\left(\Sigma D A D \Sigma^\top\right)_{ij} = \sum_{q,v,s,k=1}^{n} \Sigma_{ik} D_{ks} A_{sq} D_{dv} \Sigma_{jv} = \sum_{q,v,s,k=1}^{n} \delta_{kl_i} D_{ks} A_{sq} D_{qv} \delta_{vl_j} = \sum_{q,s=1}^{n} D_{l_i s} A_{sq} D_{ql_j} = d_{l_i} A_{l_i l_j} d_{l_j},$$

which is equivalent to (6.7), as required.

(vi) By **(i)** we have that for $i, j \in \{1, \ldots, d\}$

$$\left(\text{Diag}\left(\Sigma A \Sigma^\top\right)\right)_{ij} = \delta_{ij} A_{l_i l_j}. \tag{6.8}$$

Analogously,

$$\left(\Sigma \text{Diag}\left(A\right) \Sigma^\top\right)_{ij} = \left(\text{Diag}(A)\right)_{l_i l_j} = \delta_{l_i l_j} A_{l_i l_j} = \delta_{ij} A_{l_i l_j}, \tag{6.9}$$

which is equivalent to (6.9), as required.

(vii) By part **(iv)** we automatically have that for any $C \in \mathbb{S}_d$, $d \leq n$

$$\left(\Sigma^\top C \Sigma\right)_{ij} = \begin{cases} \delta_{ij} C_{l_i l_j}, & i, j \in \{l_1, \ldots, l_d\} \\ 0, & \text{otherwise} \end{cases}, \quad \text{for } i, j \in \{1, \ldots, n\},$$

which provides the following relations for $i, j \in \{1, \ldots, n\}$:

$$(\text{Diag}(\Sigma^\top C \Sigma))_{ij} = \delta_{ij} C_{l_i l_j}, \tag{6.10}$$

$$(\text{Diag}(\Sigma^\top \text{Diag}(C) \Sigma))_{ij} = \delta_{ij} \delta_{l_i l_j} C_{l_i l_j}, \tag{6.11}$$

$$(\Sigma^\top \text{Diag}(C) \Sigma)_{ij} = \delta_{l_i l_j} C_{l_i l_j}. \tag{6.12}$$

and form the fact that $i = j \iff l_i = l_j$ we conclude that (6.10)=(6.11)=(6.12), as required.

References

[Aie13] Gian Piero Aielli. Dynamic Conditional Correlation: on Properties and Estimation. *Journal of Business & Economic Statistics*, 31(3):282–299, July 2013.

[BC05] Monica Billio and Massimiliano Caporin. Multivariate Markov switching dynamic conditional correlation GARCH representations for contagion analysis. *Statistical Methods & Applications*, 14:145–161, 2005.

[BCG03] M. Billio, Massimiliano Caporin, and M. Gobbo. Block dynamic conditional correlation multivariate GARCH models. 2003.

[BCG06] Monica Billio, Massimiliano Caporin, and Michele Gobbo. Flexible Dynamic Conditional Correlation multivariate GARCH models for asset allocation. *Applied Financial Economics Letters*, 2(2):123–130, 2006.

[Bes74] Julian Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 36(2):192–236, 1974.

[BEW88] Tim Bollerslev, Robert F. Engle, and J. M. Wooldridge. A capital asset pricing model with time varying covariances. *Journal of Political Economy*, 96:116–131, 1988.

[BGO15] Luc Bauwens, Lyudmila Grigoryeva, and Juan-Pablo Ortega. Estimation and empirical performance of non-scalar dynamic conditional correlation models. *To appear in Computational Statistics and Data Analysis*, 2015.

[BLR06] Luc Bauwens, Sébastien Laurent, and Jeroen V. K. Rombouts. Multivariate GARCH models: a survey. *Journal of Applied Econometrics*, 21(1):79–109, January 2006.

[BO13] Luc Bauwens and Edoardo Otranto. Modeling the dependence of conditional correlations on volatility. 2013.

[Bol86] Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, 1986.

[Bol90] Tim Bollerslev. Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model. *Review of Economics and Statistics*, 72(3):498–505, 1990.

[CEG11] Riccardo Colacito, RF Robert F. Engle, and Eric Ghysels. A component model for dynamic correlations. *Journal of Econometrics*, 164:45–59, 2011.

[CES06] Lorenzo Cappiello, Robert F. Engle, and Kevin K. Sheppard. Asymmetric dynamics in the correlations of global equity and bond returns. *Journal of Financial Economics*, 4(4):537–572, 2006.

[CM12] Massimiliano Caporin and Michael McAleer. Do we really need both BEKK and DCC? A tale of two multivariate GARCH models. *Journal of Economic Surveys*, 26(4):736–751, September 2012.

[CO14] Stéphane Chrétien and Juan-Pablo Ortega. Multivariate GARCH estimation via a Bregman-proximal trust-region method. *Computational Statistics and Data Analysis*, 76:210–236, 2014.

[Cox75] D. R. Cox. Partial likelihood. *Biometrika*, 62:269–276, 1975.

[DL09] J.V. Dillon and G. Lebanon. Statistical and computational tradeoffs in stochastic composite likelihood. *Proceedings of the 12th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 5:129–136, 2009.

[DN93] Feike C Drost and Theo E Nijman. Temporal aggregation of GARCH processes. *Econometrica*, 61(4):909–927, 1993.

[Eng02] Robert F Engle. Dynamic conditional correlation -a simple class of multivariate GARCH models. *Journal of Business and Economic Statistics*, 20:339–350, 2002.

[Eng08] Robert F. Engle. High dimensional dynamic correlations. In J. L. Castle and N. Shephard, editors, *The Methodology and Practice of Econometrics: Papers in Honour of David F Hendry*. Oxford University Press, Oxford, 2008.

[Eng09] Robert Engle. *Anticipating Correlations*. Princeton University Press, Princeton, NJ, 2009.

[FdR05] M. Fernandes, B. de Sa Mota, and G. Rocha. A multivariate conditional autoregressive range model. *Economics Letters*, 86:435–440, 2005.

[FV06] S. Fieuws and G. Verbeke. Pairwise fitting of mixed models for the joint modeling of multivariate longitudinal profiles. *Biometrics*, 62:424–431, 2006.

[FZ10] Christian Francq and Jean-Michel Zakoian. *GARCH Models: Structure, Statistical Inference and Financial Applications*. Wiley, 2010.

[HF09] Christian M. Hafner and P. H. Franses. A generalized Dynamic Conditional Correlation model: simulation and application to many assets. *Econometric Reviews*, 28(6):612–631, 2009.

[HLN03] Peter Reinhard Hansen, Asger Lunde, and James M. Nason. Choosing the best volatility models: the model confidence set approach. *Oxford Bulletin of Economics and Statistics*, 65(s1):839–861, December 2003.

[HLN11] Peter Reinhard Hansen, Asger Lunde, and James M. Nason. The model confidence set. *Econometrica*, 79(2):453–497, 2011.

[KN00] Anthony Y.C. Kuk and David J. Nott. A pairwise likelihood approach to analyzing correlated binary data. *Statistics & Probability Letters*, 47:329–335, 2000.

[Lin88] Bruce G. Lindsay. Composite likelihood methods. *Contemporary Mathematics*, 80:221–239, 1988.

[LYS11] Bruce G. Lindsay, Grace Y. Yi, and Jianping Sun. Issues and strategies in the selection of composite likelihoods. *Statistica Sinica*, 21:71–105, 2011.

[Pak14] Cavit Pakel. *Bias Reduction in Nonlinear and Dynamic Panels in the Presence of Cross-Section Dependence*. PhD thesis, 2014.

[Pel06] Denis Pelletier. Regime switching for dynamic correlations. *Journal of Econometrics*, 131(1-2):445–473, 2006.

[PESS14] Cavit Pakel, Robert F. Engle, Neil Shephard, and Kevin K. Sheppard. Fitting Vast Dimensional Time-Varying Covariance Models, 2014.

[PSS11] Cavin Pakel, Neil Shephard, and Kevin Sheppard. Nuisance parameters, composite likelihoods and a panel of GARCH models. *Statistica Sinica*, 21:307–329, 2011.

[Sha64] William F. Sharpe. Capital asset prices: a theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3):425–442, 1964.

[TT02] Y. K. Tse and A. K. C. Tsui. A multivariate GARCH with time-varying correlations. *Journal of Business and Economic Statistics*, 20:351–362, 2002.

[Var08] Cristiano Varin. On composite marginal likelihoods. *AStA Advances in Statistical Analysis*, 92(1):1–28, February 2008.

[VHsS05] Cristiano Varin, Gudmund Hø st, and Ø ivind Skare. Pairwise likelihood inference in spatial generalized linear mixed models. *Computational Statistics & Data Analysis*, 49(4):1173–1191, June 2005.

[VRF11] Cristiano Varin, Nancy Reid, and David Firth. An overview of composite likelihood methods. *Statistica Sinica*, 21:5–42, 2011.

[VV05] Cristiano Varin and Paolo Vidoni. A note on composite likelihood inference and model selection. *Biometrika*, 92(3):519–528, 2005.

[WYZ13] Billy Wu, Qiwei Yao, and Shiwu Zhu. Estimation in the presence of many nuisance parameters: composite likelihood and plug-in likelihood. 2013.

[XR11] Ximing Xu and Nancy Reid. On the robustness of maximum composite likelihood estimate. *Journal of Statistical Planning and Inference*, 141(9):3047–3054, September 2011.

[ZJ05] Yinshan Zhao and Harry Joe. Composite likelihood estimation in multivariate data analysis. *Canadian Journal of Statistics*, 33(3):335–356, September 2005.

Technical Appendix to “Non-scalar GARCH models: Composite likelihood estimation and empirical comparisons”

Luc Bauwens[1], Lyudmila Grigoryeva[2], and Juan-Pablo Ortega[3]

April 23, 2015

Abstract

This document contains supplementary material to the paper [BGO15]. Sections in this appendix are indexed by letters and propositions and formulas by a letter followed by a number (e.g. A.1). Sections, propositions, and formulas in the paper are referenced by numbers.

[1]Université catholique de Louvain, CORE, Voie du Roman Pays 34 L1.03.01, B-1348 Louvain-La-Neuve. Belgium; SKEMA Business School - Université de Lille. France `luc.bauwens@uclouvain.be`

[2]Laboratoire de Mathématiques de Besançon, Université de Franche-Comté, UFR des Sciences et Techniques. 16, route de Gray. F-25030 Besançon cedex. France. `Lyudmyla.Grygoryeva@univ-fcomte.fr`

[3]Centre National de la Recherche Scientifique, Laboratoire de Mathématiques de Besançon, Université de Franche-Comté, UFR des Sciences et Techniques. 16, route de Gray. F-25030 Besançon cedex. France. `Juan-Pablo.Ortega@univ-fcomte.fr`

Acknowledgments: Lyudmila Grigoryeva and Juan-Pablo Ortega acknowledge partial financial support of the Région de Franche-Comté (Convention 2013C-5493). Lyudmila Grigoryeva acknowledges financial support from the Faculty for the Future Program of the Schlumberger Foundation. Luc Bauwens acknowledges the support of “Projet d’Actions de Recherche Concertées” 12/17-045 of the “Communauté francaise de Belgique”, granted by the “Académie universitaire Louvain”.

A Notation and preliminaries

In this section we set the notation that we use in this appendix and provide some general results that are used in the sequel.

A.1 Vectors and matrices

Vector notation: a column vector is denoted by a bold lower case symbol like $\mathbf{v}$ and $\mathbf{v}^\top$ indicates its transpose. Given a vector $\mathbf{v} \in \mathbb{R}^n$, we denote its entries by v_i, with $i \in \{1, \ldots, n\}$; we also write $\mathbf{v} = (v_i)_{i \in \{1,\ldots,n\}}$. The symbols $\mathbf{i}_n, \mathbf{0}_n \in \mathbb{R}^n$ stand for the vectors of length n consisting of ones and zeros, respectively. Additionally, given $n \in \mathbb{N}$, we define the vectors $\mathbf{k}_n^1 := (1, 2, \ldots, n)^\top$, $\mathbf{k}_n^2 := \left(1, 2^2, \ldots, n^2\right)^\top \in \mathbb{R}^n$.

Matrix notation: we denote by $\mathbb{M}_{n,m}$ the space of real $n \times m$ matrices with $m, n \in \mathbb{N}$. When $n = m$, we use the symbols $\mathbb{M}_n$ and $\mathbb{D}_n$ to refer to the space of square and diagonal matrices of order n, respectively. Given a matrix $A \in \mathbb{M}_{n,m}$, we denote its components by A_{ij} and we write $A = (A_{ij})$, with $i \in \{1, \ldots, n\}$, $j \in \{1, \ldots m\}$. The symbol $\mathbb{L}_{n,m}$ denotes the subspace of lower triangular matrices, that is, matrices that have zeros above the main diagonal:

$$\mathbb{L}_{n,m} = \{A \in \mathbb{M}_{n,m} \mid A_{ij} = 0, j > i\} \subset \mathbb{M}_{n,m}.$$

We denote by $\mathbb{L}_{n,m}^+ \subset \mathbb{L}_{n,m}$ (respectively $\mathbb{L}_{n,m}^- \subset \mathbb{L}_{n,m}$) the cone of matrices in $\mathbb{L}_{n,m}$ whose elements in the main diagonal are all positive (respectively negative). We use $\mathbb{S}_n$ to denote the subspace $\mathbb{S}_n \subset \mathbb{M}_n$ of symmetric matrices:

$$\mathbb{S}_n = \left\{A \in \mathbb{M}_n \mid A^\top = A\right\},$$

and we use $\mathbb{S}_n^+$ (respectively $\mathbb{S}_n^-$) to refer to the cone $\mathbb{S}_n^+ \subset \mathbb{S}_n$ (respectively $\mathbb{S}_n^- \subset \mathbb{S}_n$) of positive (respectively negative) semidefinite matrices. We write $A \succeq 0$ (respectively $A \preceq 0$) when $A \in \mathbb{S}_n^+$ (respectively $A \in \mathbb{S}_n^-$). The symbol $\mathbb{I}_n \in \mathbb{D}_n$ denotes the identity matrix.

The Hadamard product of matrices: given two matrices $A, B \in \mathbb{M}_{n,m}$, we denote by $A \odot B \in \mathbb{M}_{n,m}$ their elementwise multiplication matrix or Hadamard product, that is:

$$(A \odot B)_{ij} := A_{ij} B_{ij} \text{ for all } i \in \{1, \ldots, n\}, \ j \in \{1, \ldots, m\}. \tag{A.1}$$

The main properties of the Hadamard product that are used in the sequel are the following:

(i) ***The Hadamard product of two vectors:*** given two arbitrary vectors $\mathbf{u}, \mathbf{w} \in \mathbb{R}^n$, the following relation holds true

$$\mathbf{u} \odot \mathbf{w} = U\mathbf{w}, \tag{A.2}$$

where $U \in \mathbb{D}_n$ is defined by $U_{ii} := u_i$, for all $i \in \{1, \ldots, n\}$, that is, $U := \mathrm{diag}(\mathbf{u})$ and where the operator diag is defined in the following subsection.

(ii) ***The Hadamard product trace property:*** consider the matrices $A, B, C \in \mathbb{M}_{n,m}$. Then the following relation holds (see for instance [HJ94, page 304])

$$((A \odot B)C^\top)_{ii} = ((A \odot C)B^\top)_{ii} \text{ for all } i \in \{1, \ldots, n\}.$$

This leads to the equality

$$\operatorname{tr}\left(\left(A \odot B\right) C^{\top}\right) = \operatorname{tr}\left(\left(A \odot C\right) B^{\top}\right), \tag{A.3}$$

which we refer to as *the Hadamard product trace property*.

(iii) ***Schur Product Theorem***: let $A, B \in \mathbb{M}_n$ be positive semidefinite matrices. Then $A \odot B$ is also positive semidefinite. See [BR97] for a proof.

The selection matrices, reduced vectors, and reduced matrices: let $\mathcal{N} := \{1, \ldots, n\}$ and let $\mathcal{L} := \{l_1, \ldots, l_d\}$ be a subset of $\mathcal{N}$ of cardinality $d \leq n$. We define the ***selection matrix*** $\Sigma^{\mathcal{L}} \in \mathbb{M}_{d,n}$ for the labels $\mathcal{L}$ out of $\mathcal{N}$ as

$$\Sigma^{\mathcal{L}}_{ij} := \delta_{jl_i}, \ \ i \in \{1, \ldots, d\}, \ \ j \in \{1, \ldots, n\}, \tag{A.4}$$

where δ_{jl_i} is Kronecker's delta. Given a vector $\mathbf{v} \in \mathbb{R}^n$ and a selection matrix $\Sigma^{\mathcal{L}} \in \mathbb{M}_{d,n}$ for the labels $\mathcal{L}$, we define the ***reduced vector*** $\mathbf{v}^{\Sigma^{\mathcal{L}}} \in \mathbb{R}^d$ as

$$\mathbf{v}^{\mathcal{L}} := \Sigma^{\mathcal{L}} \mathbf{v} = \left(v_{l_1}, \ldots, v_{l_d}\right)^{\top}. \tag{A.5}$$

Given a square matrix $A \in \mathbb{M}_n$ and a selection matrix $\Sigma^{\mathcal{L}} \in \mathbb{M}_{d,n}$ for the labels $\mathcal{L}$, we define the ***reduced matrix*** $A^{\mathcal{L}} \in \mathbb{M}_d$ as

$$A^{\mathcal{L}} := \Sigma^{\mathcal{L}} A \Sigma^{\mathcal{L}\top}. \tag{A.6}$$

A.2 Operators and their adjoints

We recall some standard matrix operators and introduce several new ones that we use in the following sections.

The diag and Diag operators: we denote as Diag the operator $\operatorname{Diag} : \mathbb{M}_n \longrightarrow \mathbb{D}_n$ that sets equal to zero all the components of a square matrix except for those that are on the main diagonal. The operator $\operatorname{diag} : \mathbb{R}^n \longrightarrow \mathbb{D}_n$ takes a given vector and constructs a diagonal matrix with its entries in the main diagonal.

The vech and math operators: we denote by vech the operator that stacks the elements on and below the main diagonal of a symmetric matrix into a vector of length $N := \frac{1}{2} n \left(n+1\right)$, that is,

$$\operatorname{vech} : \mathbb{S}_n \longrightarrow \mathbb{R}^N, \quad \operatorname{vech}\left(A\right) = \left(A_{11}, \ldots, A_{n1}, A_{22}, \ldots, A_{n2}, \ldots, A_{nn}\right)^{\top}, A \in \mathbb{S}_n,$$

and we denote the inverse of this operator by $\operatorname{math} : \mathbb{R}^N \longrightarrow \mathbb{S}_n$.
The adjoint map of vech (respectively math) with respect to the Frobenius inner product is denoted by $\operatorname{vech}^* : \mathbb{R}^N \longrightarrow \mathbb{S}_n$ (respectively $\operatorname{math}^* : \mathbb{S}_n \longrightarrow \mathbb{R}^N$). In [CO14] it is shown that given $A \in \mathbb{S}_n$ and $\mathbf{v} \in \mathbb{R}^N$, the following relations hold true:

$$\operatorname{vech}^*\left(\mathbf{v}\right) = \frac{1}{2}\left(\operatorname{math}\left(\mathbf{v}\right) + \operatorname{Diag}\left(\operatorname{math}\left(\mathbf{v}\right)\right)\right), \tag{A.7}$$

$$\operatorname{math}^*\left(A\right) = 2 \operatorname{vech}(A - \frac{1}{2} \operatorname{Diag}\left(A\right)). \tag{A.8}$$

The Almon lag operator and its tangent map: in [BGO15] we introduced the Almon lag operator $\mathrm{alm}_n : \mathbb{R}^3 \longrightarrow \mathbb{R}^n$, $n \in \mathbb{N}$, $\mathbf{v} \in \mathbb{R}^3$, by using the using the Almon lag function [Alm65] as

$$(\mathrm{alm}_n(\mathbf{v}))_i := v_1 + \exp(v_2 i + v_3 i^2), \text{ for } i \in \{1, \ldots, n\}. \tag{A.9}$$

The tangent map $T_{\mathbf{v}}\mathrm{alm}_n : \mathbb{R}^3 \longrightarrow \mathbb{R}^n$ is determined by the equality

$$T_{\mathbf{v}}\mathrm{alm}_n \cdot \boldsymbol{\delta}\mathbf{v} = K_{\mathbf{v}} \cdot \boldsymbol{\delta}\mathbf{v}, \text{ with } K_{\mathbf{v}} := \left(\mathbf{i}_n \mid \mathbf{k}_n^1 \odot \mathrm{alm}_n(\bar{\mathbf{v}}) \mid \mathbf{k}_n^2 \odot \mathrm{alm}_n(\bar{\mathbf{v}})\right) \in \mathbb{M}_{n,3}, \boldsymbol{\delta}\mathbf{v} \in \mathbb{R}^3, \tag{A.10}$$

where the symbol $|$ denotes vertical concatenation of matrices (or vectors), the vectors $\mathbf{i}_n, \mathbf{k}_n^1, \mathbf{k}_n^2 \in \mathbb{R}^n$ are introduced in Subsection A.1, and $\bar{\mathbf{v}} \in \mathbb{R}^3$ is obtained out of the vector $\mathbf{v} \in \mathbb{R}^3$ by setting $\bar{\mathbf{v}} := (0, v_2, v_3)^\top$.
The adjoint $T_{\mathbf{v}}^*\mathrm{alm}_n : \mathbb{R}^n \longrightarrow \mathbb{R}^3$ of the tangent map $T_{\mathbf{v}}\mathrm{alm}_n$ is determined by the relation

$$T_{\mathbf{v}}^*\mathrm{alm}_n(\mathbf{u}) = K_{\mathbf{v}}^\top \cdot \mathbf{u}, \quad \text{for any } \mathbf{u} \in \mathbb{R}^n. \tag{A.11}$$

A.3 Intrinsic parametrization of symmetric matrices

The multivariate volatility models which are at the core of this paper, typically prescribe the dynamics of the conditional correlation or covariance process by using various types of symmetric matrices as parameters. Since for any matrix in $\mathbb{S}_n$, its $\frac{1}{2}n(n-1)$ supradiagonal elements are redundant in the sense that they can be deduced from the symmetry, we may exploit this feature in order to reduce the parameter spaces of the models by using what we call **intrinsic parametrizations**.

We denote by $\mathcal{P}$ an intrinsic parameter subspace, s is its dimension, $\boldsymbol{\theta} \in \mathcal{P}$ a generic element, and by $\boldsymbol{\Theta}$ the operator $\boldsymbol{\Theta} : \mathcal{P} \longrightarrow \mathbb{S}_n$ that assigns to any $\boldsymbol{\theta} \in \mathcal{P}$ its corresponding element in $\mathbb{S}_n$. We list below four instances of such intrinsic parametrizations $\boldsymbol{\Theta}(\boldsymbol{\theta})$ that are of much use for the multivariate volatility models that we consider in the following sections. We also provide their tangent maps and corresponding adjoints for each of the cases under consideration. Notice that in all the cases below we provide the intrinsic parameterization for only one parameter case even though two parameter matrices or more are typically used for most multivariate volatility families. In those cases, the relevant maps are obtained in a straightforward way by extending both their domain and range spaces.

(i) Hadamard case: given a full rank matrix $A \in \mathbb{S}_n$, it can be naturally parametrized with the vector $\boldsymbol{a} \in \mathbb{R}^N$ with $N = \frac{1}{2}n\,(n+1)$, by setting $A := \mathrm{math}\,(\boldsymbol{a})$ (see Subsection A.2 for the definition of the operator math). In this case $\boldsymbol{\Theta}(\boldsymbol{\theta})$ is defined as

$$\begin{array}{rccc} \boldsymbol{\Theta}: & \mathbb{R}^N & \longrightarrow & \mathbb{S}_n \\ & \boldsymbol{\theta} & \longmapsto & \mathrm{math}\,(\boldsymbol{\theta}), \end{array} \tag{A.12}$$

the tangent map $T_{\boldsymbol{\theta}}\boldsymbol{\Theta} : \mathbb{R}^N \longrightarrow \mathbb{S}_n$ is determined by

$$T_{\boldsymbol{\theta}}\boldsymbol{\Theta} \cdot \boldsymbol{\delta\theta} = \mathrm{math}\,(\boldsymbol{\delta\theta}), \quad \boldsymbol{\delta\theta} \in \mathbb{R}^N, \tag{A.13}$$

and its adjoint is the differential operator $T^*_{\boldsymbol{\theta}}\boldsymbol{\Theta} : \mathbb{S}_n \longrightarrow \mathbb{R}^N$ which for any given $\Delta \in \mathbb{S}_n$ is determined by the expression

$$T^*_{\boldsymbol{\theta}}\boldsymbol{\Theta} \cdot \Delta = \mathrm{math}^*(\Delta), \tag{A.14}$$

with the operator math^* provided in Subsection A.2.

(ii) Rank deficient case: let $A \in \mathbb{S}_n$ be the matrix of rank $r < n \in \mathbb{N}$ then, as such, it can be expressed as $A := \widetilde{A}\widetilde{A}^\top$, where $\widetilde{A} := \mathrm{mat}_r(\boldsymbol{a}) \in \mathbb{L}_{n,r}$ and $\boldsymbol{a} \in \mathbb{R}^{N^*}$, $N^* = nr - \frac{1}{2}r(r-1)$ (see Subsection A.2 for the definition of the operator $\mathrm{mat}_r : \mathbb{R}^{N^*} \longrightarrow \mathbb{L}_{n,r}$). In this case $\boldsymbol{a} \in \mathbb{R}^{N^*}$ is chosen as the intrinsic parameter vector. We showed in Proposition 3.1 of [BGO15] that in order to ensure that $A \in \mathbb{S}_n$ is well identified, it suffices to require that $\widetilde{A} \in \mathbb{L}^+_{n,r}$. In this case $\boldsymbol{\Theta}(\boldsymbol{\theta})$ is defined as

$$\begin{array}{rccc} \boldsymbol{\Theta}: & \mathbb{R}^{N^*} & \longrightarrow & \mathbb{S}_n \\ & \boldsymbol{\theta} & \longmapsto & \mathrm{mat}_r(\boldsymbol{\theta})(\mathrm{mat}_r(\boldsymbol{\theta}))^\top, \end{array} \tag{A.15}$$

the tangent map $T_{\boldsymbol{\theta}}\boldsymbol{\Theta} : \mathbb{R}^{N^*} \longrightarrow \mathbb{S}_n$ is given by

$$T_{\boldsymbol{\theta}}\boldsymbol{\Theta} \cdot \boldsymbol{\delta\theta} = \mathrm{mat}_r(\boldsymbol{\delta\theta})(\mathrm{mat}_r(\boldsymbol{\theta}))^\top, \;\; \boldsymbol{\delta\theta} \in \mathbb{R}^N, \tag{A.16}$$

and its adjoint is the differential operator $T^*_{\boldsymbol{\theta}}\boldsymbol{\Theta} : \mathbb{S}_n \longrightarrow \mathbb{R}^{N^*}$ which for any $\Delta \in \mathbb{S}_n$ is determined by the expression

$$T^*_{\boldsymbol{\theta}}\boldsymbol{\Theta} \cdot \Delta = 2\ \mathrm{vec}_r(\Delta\ \mathrm{mat}_r(\boldsymbol{\theta})), \tag{A.17}$$

with the operator vec_r provided in Subsection A.2.

(iii) Almon case: this is a particular case of **(ii)** where the vector in $\mathbb{R}^n$ that generates the matrix $A \in \mathbb{S}_n$ of rank $r = 1$ is parametrized using the Almon lag [Alm65] operator alm_n introduced in [BGO15] and recalled in (A.9). More explicitly, take $\boldsymbol{a} \in \mathbb{R}^3$ as intrinsic parameter vector and define $\widetilde{\boldsymbol{a}} := \mathrm{alm}_n(\boldsymbol{a}) \in \mathbb{R}^n$, then the matrix $A \in \mathbb{S}_n$ can be written as $A := \widetilde{\boldsymbol{a}}\widetilde{\boldsymbol{a}}^\top$. In this case $\boldsymbol{\Theta}(\boldsymbol{\theta})$ is defined as

$$\begin{array}{rccc} \boldsymbol{\Theta}: & \mathbb{R}^3 & \longrightarrow & \mathbb{S}_n \\ & \boldsymbol{\theta} & \longmapsto & \mathrm{alm}_n(\boldsymbol{\theta})\,(\mathrm{alm}_n(\boldsymbol{\theta}))^\top, \end{array} \tag{A.18}$$

the tangent map $T_{\boldsymbol{\theta}}\boldsymbol{\Theta} : \mathbb{R}^3 \longrightarrow \mathbb{S}_n$ is given by

$$T_{\boldsymbol{\theta}}\boldsymbol{\Theta} \cdot \boldsymbol{\delta\theta} = (K_{\boldsymbol{\theta}} \cdot \boldsymbol{\delta\theta})(\mathrm{alm}_n(\boldsymbol{\theta}))^\top + \mathrm{alm}_n(\boldsymbol{\theta})\,(K_{\boldsymbol{\theta}} \cdot \boldsymbol{\delta\theta})^\top, \;\; \boldsymbol{\delta\theta} \in \mathbb{R}^3, \tag{A.19}$$

where $K_{\boldsymbol{\theta}} = \left(\mathbf{i}_n \mid \mathbf{k}^1_n \odot \mathrm{alm}_n(\bar{\boldsymbol{\theta}}) \mid \mathbf{k}^2_n \odot \mathrm{alm}_n(\bar{\boldsymbol{\theta}})\right) \in \mathbb{M}_{n,3}$, $\bar{\boldsymbol{\theta}} := (0, \theta_2, \theta_3)^\top$, the symbol $\mid$ denotes vertical concatenation, and $\mathbf{k}^1_n := (1, 2, \ldots, n)^\top$, $\mathbf{k}^2_n := \left(1, 2^2, \ldots, n^2\right)^\top \in \mathbb{R}^n$. The adjoint of $T_{\boldsymbol{\theta}}\boldsymbol{\Theta} : \mathbb{R}^3 \longrightarrow \mathbb{S}_n$ is the differential operator $T^*_{\boldsymbol{\theta}}\boldsymbol{\Theta} : \mathbb{S}_n \longrightarrow \mathbb{R}^3$ which for any $\Delta \in \mathbb{S}_n$ is determined by the expression

$$T^*_{\boldsymbol{\theta}}\boldsymbol{\Theta} \cdot \Delta = 2K^\top_{\boldsymbol{\theta}} \cdot \Delta \cdot \mathrm{alm}_n(\boldsymbol{\theta}). \tag{A.20}$$

(iv) Scalar case: in this instance the intrinsic parameter is a scalar $a \in \mathbb{R}$ and we set $A := a\mathbf{i}_n\mathbf{i}_n^\top \in \mathbb{S}_n$. The map $\boldsymbol{\Theta}(\boldsymbol{\theta})$ is given in this case by

$$\begin{array}{rccc} \boldsymbol{\Theta}: & \mathbb{R} & \longrightarrow & \mathbb{S}_n \\ & \theta & \longmapsto & \theta\ \mathbf{i}_n\mathbf{i}_n^\top, \end{array} \tag{A.21}$$

its tangent map $T_{\boldsymbol{\theta}}\boldsymbol{\Theta} : \mathbb{R} \longrightarrow \mathbb{S}_n$ is

$$T_{\boldsymbol{\theta}}\boldsymbol{\Theta} \cdot \delta\theta = \delta\theta \; \mathbf{i}_n \mathbf{i}_n^\top, \;\; \delta\theta \in \mathbb{R}, \tag{A.22}$$

And for $\Delta \in \mathbb{S}_n$ arbitrary we have

$$T^*_{\boldsymbol{\theta}}\boldsymbol{\Theta} \cdot \Delta = \langle \Delta, \mathbf{i}_n \mathbf{i}_n^\top \rangle. \tag{A.23}$$

B Constrained QML estimation of DCC and DVEC models

In this section we discuss the constrained quasi-maximum likelihood (QML) method for the estimation of the parameters of the two families of MGARCH models considered in the paper. We direct the reader to the work [BGO15] that contains all the details concerning the QML method applied to the constrained estimation of the DCC model (2.1) with correlation dynamics determined by (2.3). All the necessary details concerning the computation of the score function associated to different prescriptions of the DCC model are provided there, as well as the explicit formulas for the Bregman divergences constructed in order to handle the positive semidefiniteness (PSD) and the identification (IC) constraints imposed on the parameters in each relevant case. We adopt the same constrained estimation strategy for the DVEC models. In this section we provide the log-likelihood function associated to the DVEC model and the explicit expressions for the components of the gradient of the log-likelihood function related to each of the DVEC model subfamilies presented in Section 2.3.

B.1 Log-likelihood function for the DVEC model

Let $\mathbf{r} = \{\mathbf{r}_1, \ldots, \mathbf{r}_T\}$ be a sample of size T of n-dimensional observations of the process $\{\mathbf{r}_t\}$ and let $\boldsymbol{\Theta} := (A, B) \in \mathbb{S}_n \times \mathbb{S}_n$ denote the parameters to be estimated. We implement variance targeting by pre-estimating S in (2.6) by $\sum_{t=1}^{T} \mathbf{r}_t \mathbf{r}_t^\top / T$. The log-likelihood function associated to the process (2.1) is

$$\log L\left(\boldsymbol{\Theta}; \mathbf{r}\right) = \sum_{t=1}^{T} l_t\left(\boldsymbol{\Theta}; \mathbf{r}_t\right), \tag{B.1}$$

with

$$l_t\left(\boldsymbol{\Theta}; \mathbf{r}_t\right) = -\frac{1}{2}\left(n \log(2\pi) + \log\det\left(H_t\right) + \mathbf{r}_t^\top H_t^{-1} \mathbf{r}_t\right), \tag{B.2}$$

where the conditional covariance matrices H_t follow the dynamics specified by (2.6).

The QML estimation problem for the DVEC model is to obtain the parameter values $\widehat{\boldsymbol{\Theta}}$ that maximize the log-likelihood function (B.1)-(B.2) and that satisfy the necessary constraints imposed on the model parameters. In Section 2.3 we provid the list of the DVEC model subfamilies that we are interested in together with the **(IC)** and **(PSD)** constraints associated to each of them. We emphasize that in all the cases we consider the relevant intrinsic parametrizations $\boldsymbol{\Theta}(\boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \mathcal{P} \times \mathcal{P}$ is defined in the corresponding intrinsic parameter space of dimension s specific to each subfamily considered (see also Section A.3).

B.2 Constrained optimization of the log-likelihood function using Bregman divergences

This technique was proposed in [CO14] for the case of QML estimation of full VEC models and extended to DCC models in [BGO15]. This estimation method is based on the optimization of a sequence of penalized local functions that incorporate the Bregman divergences that guarantee that the corresponding constraints associated the model are satisfied at each iteration. More specifically, the solution $\boldsymbol{\theta}^{(k+1)}$ of the local optimization problem after k iterations is defined by

$$\boldsymbol{\theta}^{(k+1)} = \underset{\boldsymbol{\theta}\in\mathcal{P}\times\mathcal{P}}{\arg\min}\ \tilde{f}^{(k)}(\boldsymbol{\theta}), \tag{B.3}$$

where the local objective function $\tilde{f}^{(k)}$ at $\boldsymbol{\theta}^{(k)}$ is constructed at each iteration k as

$$\begin{aligned} \tilde{f}^{(k)}(\boldsymbol{\theta}) \quad := \quad & f(\boldsymbol{\theta}^{(k)}) + \nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta}^{(k)})(\boldsymbol{\theta}-\boldsymbol{\theta}^{(k)}) + \frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}^{(k)})^\top H^{(k)}(\boldsymbol{\theta}-\boldsymbol{\theta}^{(k)}) \\ & + \sum_{j=1}^{s_1} l_1^j D_{M_j}(\boldsymbol{\theta},\boldsymbol{\theta}^{(k)}) + \sum_{j=1}^{s_2} l_2^j \mathbf{i}_{q_j}^\top D_{N_j}(\boldsymbol{\theta},\boldsymbol{\theta}^{(k)}) + \sum_{j=1}^{s_3} l_3^j \mathbf{i}_{m_j}^\top D_{L_j}(\boldsymbol{\theta},\boldsymbol{\theta}^{(k)}). \end{aligned} \tag{B.4}$$

In this expression $\nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta}^{(k)}) = -\nabla_{\boldsymbol{\theta}} \log L(\boldsymbol{\theta}^{(k)};\mathbf{r})$ is the gradient of minus the log-likelihood function given in (B.1) and $H^{(k)}$ is its Hessian computed at the point $\boldsymbol{\theta}^{(k)}$. The integers s_1, s_2, s_3 are the numbers of positive semidefiniteness, nonlinear, and linear constraints, respectively; the symbols $D_{M_j}(\boldsymbol{\theta},\boldsymbol{\theta}^{(k)}) \in \mathbb{R}$, $j \in \{1,\ldots,s_1\}$, $D_{N_j}(\boldsymbol{\theta},\boldsymbol{\theta}^{(k)}) \in \mathbb{R}^{q_j}$, $j \in \{1,\ldots,s_2\}$, and $D_{L_j}(\boldsymbol{\theta},\boldsymbol{\theta}^{(k)}) \in \mathbb{R}^{m_j}$, $j \in \{1,\ldots,s_3\}$ denote the Bregman divergences whose explicit expressions are provided in [BGO15]; $\boldsymbol{l}_i \in \mathbb{R}^{s_i}, i = \{1,2,3\}$ are vectors whose components control the strength of the corresponding Bregman penalizations, and $\mathbf{i}_{q_j} \in \mathbb{R}^{q_j}$, $j \in \{1,\ldots,s_2\}$, $\mathbf{i}_{m_j} \in \mathbb{R}^{m_j}$, $j \in \{1,\ldots,s_3\}$ are vectors of ones.

The local optimization problem in (B.3) is solved by finding the value $\boldsymbol{\theta}_0$ for which

$$\nabla_{\boldsymbol{\theta}} \tilde{f}^{(k)}(\boldsymbol{\theta}_0) = 0, \tag{B.5}$$

where the gradient $\nabla_{\boldsymbol{\theta}} \tilde{f}^{(k)}(\boldsymbol{\theta})$ of the local model (B.4) can be explicitly computed. The equation (B.5) for each iteration k can be solved involving for instance the Newton-Raphson method. We refer the reader to the paper [BGO15] for an exhaustive explanation and detailed computations of all the necessary ingredients.

B.3 Gradient of the log-likelihood function for the DVEC model

The computation of the estimator $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ via the constrained maximum likelihood optimization procedure that we use, requires the gradient $\nabla_{\boldsymbol{\theta}} \log L(\boldsymbol{\theta};\mathbf{r})$ for each specific intrinsic dependence $\boldsymbol{\Theta}(\boldsymbol{\theta})$ that we work with. In the following proposition we provide explicit expressions of the score, where we use several operators and vectors that we defined in Section A; the proof is analogous to the one provided in the Technical Appendix of [BGO15] for the DCC model and hence it is not repeated here.

Proposition B.1 *Let $\mathbf{r} = \{\mathbf{r}_1, \ldots, \mathbf{r}_T\}$ be a sample with $\mathbf{r}_t \in \mathbb{R}^n$, $t \in \{1, \ldots, T\}$. Let $\boldsymbol{\theta} := (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \in \mathcal{P} \times \mathcal{P}$, $\boldsymbol{\Theta}(\boldsymbol{\theta}) := (A(\boldsymbol{\theta}_1), B(\boldsymbol{\theta}_2)) \in \mathbb{S}_n \times \mathbb{S}_n$, and let $\log L(\boldsymbol{\theta}; \mathbf{r})$ be the log-likelihood in (3.5)-(B.2). Then,*

$$\nabla_{\boldsymbol{\theta}} \log L(\boldsymbol{\theta}; \mathbf{r}) = \sum_{t=1}^{T} \nabla_{\boldsymbol{\theta}} l_t(\boldsymbol{\theta}; \mathbf{r}_t) = \sum_{t=1}^{T} T_{\boldsymbol{\theta}}^* \boldsymbol{\Theta} \cdot T_{\boldsymbol{\Theta}}^* H_t \cdot \nabla_{H_t} l_t(\boldsymbol{\theta}; \mathbf{r}_t), \tag{B.6}$$

with

$$\nabla_{H_t} l_t(\boldsymbol{\theta}; \mathbf{r}_t) = -\frac{1}{2}\left[H_t^{-1} - H_t^{-1}\mathbf{r}_t\mathbf{r}_t^\top H_t^{-1}\right]. \tag{B.7}$$

In the relation (B.6), the differential operator $T_{\boldsymbol{\Theta}}^* H_t : \mathbb{S}_n \times \mathbb{S}_n \longrightarrow \mathbb{S}_n \times \mathbb{S}_n$ is the adjoint of the map $T_{\boldsymbol{\Theta}} H_t : \mathbb{S}_n \times \mathbb{S}_n \longrightarrow \mathbb{S}_n \times \mathbb{S}_n$. For each component $A(\boldsymbol{\theta}_1)$ and $B(\boldsymbol{\theta}_2)$ of $\boldsymbol{\Theta}$ and for any $\Delta \in \mathbb{S}_n$, $T_{\boldsymbol{\Theta}}^* H_t$ is determined by the recursions:

$$T_A^* H_t \cdot \Delta = \Delta \odot \left(\mathbf{r}_{t-1}\mathbf{r}_{t-1}^\top - S\right) + T_A^* H_{t-1}[\Delta \odot B], \tag{B.8}$$

$$T_B^* H_t \cdot \Delta = \Delta \odot (H_{t-1} - S) + T_B^* H_{t-1}[\Delta \odot B], \tag{B.9}$$

that are initialized by setting $T_A^* H_0 = 0$ and $T_B^* H_0 = 0$. Finally, the differential operator $T_{\boldsymbol{\theta}}^* \boldsymbol{\Theta} : \mathbb{S}_n \times \mathbb{S}_n \longrightarrow \mathcal{P} \times \mathcal{P}$ is the adjoint of the map $T_{\boldsymbol{\theta}} \boldsymbol{\Theta} : \mathcal{P} \times \mathcal{P} \longrightarrow \mathbb{S}_n \times \mathbb{S}_n$, with $\mathcal{P} \times \mathcal{P}$ the intrinsic $\boldsymbol{\theta}$ parameter space and parameterization $\boldsymbol{\Theta}(\boldsymbol{\theta})$ associated to each of the model subfamilies considered in Section 2.3. For a given pair $\Delta_1, \Delta_2 \in \mathbb{S}_n$, these maps are determined by the following expressions:

(i) ***The Hadamard DVEC family:*** *let $n \in \mathbb{N}$, $N := \frac{1}{2}n(n+1)$. In this case, the intrinsic parameter subspace $\mathcal{P}$ is $\mathbb{R}^N$, $\boldsymbol{\theta} := (\boldsymbol{a}, \boldsymbol{b})$, and $\boldsymbol{\Theta}(\boldsymbol{\theta}) := (\mathrm{math}(\boldsymbol{a}), \mathrm{math}(\boldsymbol{b}))$, for any $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^N$. Moreover,*

$$\begin{array}{rccc} T_{\boldsymbol{\theta}}^* \boldsymbol{\Theta} : & \mathbb{S}_n \times \mathbb{S}_n & \longrightarrow & \mathbb{R}^N \times \mathbb{R}^N \\ & (\Delta_1, \Delta_2) & \longmapsto & (\mathrm{math}^*(\Delta_1), \mathrm{math}^*(\Delta_2)). \end{array} \tag{B.10}$$

(ii) ***The rank deficient DVEC family with rank r:*** *let $r < n \in \mathbb{N}$, $N^* := nr - \frac{1}{2}r(r-1)$. In this case the intrinsic parameter subspace $\mathcal{P}$ is $\mathbb{R}^{N^*}$, $\boldsymbol{\theta} := (\boldsymbol{a}, \boldsymbol{b})$, and*

$$\boldsymbol{\Theta}(\boldsymbol{\theta}) := (\mathrm{mat}_r(\boldsymbol{a})(\mathrm{mat}_r(\boldsymbol{a}))^\top, \mathrm{mat}_r(\boldsymbol{b})(\mathrm{mat}_r(\boldsymbol{b}))^\top), \quad \textit{for any} \quad \boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^{N^*}.$$

Moreover,

$$\begin{array}{rccc} T_{\boldsymbol{\theta}}^* \boldsymbol{\Theta} : & \mathbb{S}_n \times \mathbb{S}_n & \longrightarrow & \mathbb{R}^{N^*} \times \mathbb{R}^{N^*} \\ & (\Delta_1, \Delta_2) & \longmapsto & 2\ (\mathrm{vec}_r(\Delta_1 \mathrm{mat}_r(\boldsymbol{a})), \mathrm{vec}_r(\Delta_2 \mathrm{mat}_r(\boldsymbol{b}))). \end{array} \tag{B.11}$$

(iii) ***The Almon DVEC family:*** *in this case the intrinsic parameter subspace $\mathcal{P}$ is $\mathbb{R}^3$ and $\boldsymbol{\Theta}(\boldsymbol{\theta}) := (\mathrm{alm}_n(\boldsymbol{\theta}_1)(\mathrm{alm}_n(\boldsymbol{\theta}_1))^\top, \mathrm{alm}_n(\boldsymbol{\theta}_2)(\mathrm{alm}_n(\boldsymbol{\theta}_2))^\top)$, with $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \mathbb{R}^3$, $\boldsymbol{\theta} := (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$. Moreover,*

$$\begin{array}{rccc} T_{\boldsymbol{\theta}}^* \boldsymbol{\Theta} : & \mathbb{S}_n \times \mathbb{S}_n & \longrightarrow & \mathbb{R}^3 \times \mathbb{R}^3 \\ & (\Delta_1, \Delta_2) & \longmapsto & 2\ (K_{\boldsymbol{\theta}_1}^\top \Delta_1 \mathrm{alm}_n(\boldsymbol{\theta}_1), K_{\boldsymbol{\theta}_2}^\top \Delta_2 \mathrm{alm}_n(\boldsymbol{\theta}_2)), \end{array} \tag{B.12}$$

where $K_{\boldsymbol{\theta}_i} = (\ \mathbf{i}_n \mid \mathbf{k}_n^1 \odot \mathrm{alm}_n(\bar{\boldsymbol{\theta}}_i) \mid \mathbf{k}_n^2 \odot \mathrm{alm}_n(\bar{\boldsymbol{\theta}}_i)) \in \mathbb{M}_{n,3}$*,* $\bar{\boldsymbol{\theta}}_i := (0, (\boldsymbol{\theta}_i)_2, (\boldsymbol{\theta}_i)_3)^\top$*,* $i \in \{1, 2\}$*, the symbol* $\mid$ *denotes vertical concatenation, and* $\mathbf{k}_n^1 := (1, 2, \ldots, n)^\top$*,* $\mathbf{k}_n^2 := \left(1, 2^2, \ldots, n^2\right)^\top \in \mathbb{R}^n$*.*

(iv) ***The scalar DVEC family:*** *the intrinsic parameter subspace is* $\mathbb{R}$ *and* $\boldsymbol{\Theta}(\boldsymbol{\theta}) := (a\mathbf{i}_n\mathbf{i}_n^\top, b\mathbf{i}_n\mathbf{i}_n^\top)$*, with* $a, b \in \mathbb{R}$*,* $\boldsymbol{\theta} = (a, b)$*. Moreover,*

$$T^*_{\boldsymbol{\theta}}\boldsymbol{\Theta} : \begin{array}{ccc} \mathbb{S}_n \times \mathbb{S}_n & \longrightarrow & \mathbb{R} \times \mathbb{R} \\ (\Delta_1, \Delta_2) & \longmapsto & \left(\langle \Delta_1, \mathbf{i}_n\mathbf{i}_n^\top \rangle, \langle \Delta_2, \mathbf{i}_n\mathbf{i}_n^\top \rangle\right). \end{array} \tag{B.13}$$

In order to algorithmically implement Proposition B.1, the operator recursions (B.8)-(B.9) together with (B.10)-(B.13) have to be reformulated in terms of matrix recursions. We refer the reader to [BGO15] where that work has been carried out for the DCC case; the obtained results remain valid in the context of Proposition B.1.

C Reduction procedure and CL implementation for different DCC and DVEC subfamilies

The goal of this section is to provide the explicit functional connection between the parameters of the original DCC or DVEC models and the reduced counterparts for each of the model parameter specifications defined in Section 2. We show that those remain closed under the reduction procedure and provide all the necessary ingredients to implement the composite likelihood estimation using the expressions of the corresponding components of the score discussed in Proposition 4.1 in [BGO15] for the DCC model and in Proposition B.1 for the DVEC model, respectively.

We recall that since the dynamic equations for the matrix processes $\{Q_t\}$ for DCC and $\{H_t\}$ for DVEC, respectively, have the same iterative pattern, we proceed in the same way as in Section 2.3, that is, we denote by the letter K_t the matrices H_t or Q_t and use the symbol $\boldsymbol{\rho}_t$ to indicate the vectors of degarched returns $\boldsymbol{\varepsilon}_t$ for DCC or the v vectors of the original returns $\mathbf{r}_t$ for DVEC, respectively. Likewise, we use the letter K for either Q or S. We emphasize that each of the specifications discussed in this section corresponds to either the DCC model defined by (2.1)-(2.3), or the DVEC model given in (2.1) and (2.6) with a certain intrinsic subparametrization in either (2.3) or (2.6), respectively. Additionally, the common notion of reduced model is used to designate either the reduced DCC model (4.2)-(4.4) in Proposition 4.1 or the reduced DVEC model (4.10)-(4.11) in Proposition 4.2 depending on the context. The results in this subsection hold true for both model families unless the opposite is stated.

C.1 Reduction of the Hadamard model

We start with the Hadamard model prescription (2.8) provided in Subsection 2.3 with the corresponding positivity constraints (2.9). In the sequel we call initial the non-reduced model of the highest available dimension and reduced the one which is obtained via the reduction procedure described in Section 4.1.

Proposition C.1 *Let $\mathbf{r} = \{\mathbf{r}_1, \ldots, \mathbf{r}_T\}$ be a sample of n-dimensional asset returns $\mathbf{r}_t \in \mathbb{R}^n$, $t \in \{1, \ldots, T\}$. Let $\mathcal{L} = \{l_1, \ldots, l_d\}$ be a set of d asset labels with $d \leq n$ and $\Sigma^{\mathcal{L}} \in \mathbb{M}_{d,n}$ be the selection matrix associated to $\mathcal{L}$ with the entries defined as in (A.4), and $\mathbf{r}^{\mathcal{L}}$ be a corresponding sample of reduced vectors of the asset returns $\mathbf{r}_t^{\mathcal{L}} \in \mathbb{R}^d$, $t \in \{1, \ldots, T\}$ defined as in (A.5). Consider a n-dimensional initial Hadamard model (2.8) with parameters $A, B \in \mathbb{S}_n$, then the following statements hold true:*

(i) *Let $\boldsymbol{\theta} = (\boldsymbol{a}, \boldsymbol{b}) \in \mathbb{R}^N \times \mathbb{R}^N$, $N = n(n+1)/2$, $\boldsymbol{\theta}^{\mathcal{L}} = (\boldsymbol{a}^{\mathcal{L}}, \boldsymbol{b}^{\mathcal{L}}) \in \mathbb{R}^{N_d} \times \mathbb{R}^{N_d}$, $N_d = d(d+1)/2$ be the intrinsic parameters associated to the initial model parameters $\boldsymbol{\Theta} = (A, B) \in \mathbb{S}_n \times \mathbb{S}_n$ and to the reduced model parameters $\boldsymbol{\Theta}^{\mathcal{L}} = (A^{\mathcal{L}}, B^{\mathcal{L}}) \in \mathbb{S}_d \times \mathbb{S}_d$, respectively. Let $L_{\mathcal{L}} = \log L_{\mathcal{L}}(\boldsymbol{\Theta}^{\mathcal{L}}(\boldsymbol{\theta}^{\mathcal{L}}); \mathbf{r}^{\mathcal{L}})$ be the log-likelihood function associated to the reduced process $\{\mathbf{r}_t^{\mathcal{L}}, H_t^{\mathcal{L}}\}$ with $\{H_t^{\mathcal{L}}\}$ the reduced conditional covariance matrix process. Then, the intrinsic parameters $\boldsymbol{\theta}^{\mathcal{L}}$ and $\boldsymbol{\theta}$ of the reduced model and the initial models, respectively, are related via*

$$\boldsymbol{a}^{\mathcal{L}} = \boldsymbol{a}^{\mathcal{L}}(\boldsymbol{a}) = \mathrm{vech}(\Sigma^{\mathcal{L}} \mathrm{math}(\boldsymbol{a}) \Sigma^{\mathcal{L}\top}), \tag{C.1}$$

$$\boldsymbol{b}^{\mathcal{L}} = \boldsymbol{b}^{\mathcal{L}}(\boldsymbol{b}) = \mathrm{vech}(\Sigma^{\mathcal{L}} \mathrm{math}(\boldsymbol{b}) \Sigma^{\mathcal{L}\top}) \tag{C.2}$$

and the components of the gradient of $L_{\mathcal{L}}$ with respect to the corresponding intrinsic parameters $\boldsymbol{\theta}$ of the initial model are given by

$$\nabla_{\boldsymbol{a}} L_{\mathcal{L}} = \mathrm{vech}\left(\Sigma^{\mathcal{L}\top} \mathrm{math}\left(\nabla_{\boldsymbol{a}^{\mathcal{L}}} L_{\mathcal{L}}\right) \Sigma^{\mathcal{L}}\right), \tag{C.3}$$

$$\nabla_{\boldsymbol{b}} L_{\mathcal{L}} = \mathrm{vech}\left(\Sigma^{\mathcal{L}\top} \mathrm{math}\left(\nabla_{\boldsymbol{b}^{\mathcal{L}}} L_{\mathcal{L}}\right) \Sigma^{\mathcal{L}}\right), \tag{C.4}$$

where $\nabla_{\boldsymbol{a}^{\mathcal{L}}} L_{\mathcal{L}}$ and $\nabla_{\boldsymbol{b}^{\mathcal{L}}} L_{\mathcal{L}}$ are the components of the gradient of the log-likelihood function $L_{\mathcal{L}}$ of the reduced model with respect to its intrinsic parameters, that is, to $\boldsymbol{a}^{\mathcal{L}}$ and $\boldsymbol{b}^{\mathcal{L}}$, respectively.

(ii) *If the intrinsic parameters $\boldsymbol{\theta} = (\boldsymbol{a}, \boldsymbol{b}) \in \mathbb{R}^N \times \mathbb{R}^N$ of the initial model satisfy the positive semidefiniteness* **(PSD)** *constraint (2.9), then so do the intrinsic parameters $\boldsymbol{\Theta}^{\mathcal{L}} = (A^{\mathcal{L}}, B^{\mathcal{L}}) \in \mathbb{S}_d \times \mathbb{S}_d$ of the reduced Hadamard model associated to $\Sigma^{\mathcal{L}}$.*

Proof. **(i)** The relations (C.1) and (C.2) can be easily verified. For example,

$$\boldsymbol{a}^{\mathcal{L}} = \mathrm{vech}(A^{\mathcal{L}}) = \mathrm{vech}(\Sigma^{\mathcal{L}} A \Sigma^{\mathcal{L}\top}) = \mathrm{vech}(\Sigma^{\mathcal{L}} \mathrm{math}(\boldsymbol{a}) \Sigma^{\mathcal{L}\top}).$$

(ii) In order to prove (C.3), we use the relation (C.1) and the the chain rule. Indeed, for any $\mathbf{v} \in \mathbb{R}^N$ we write:

$$d_{\boldsymbol{a}} L_{\mathcal{L}} \cdot \mathbf{v} = d_{\boldsymbol{a}^{\mathcal{L}}} L_{\mathcal{L}} \cdot T_{\boldsymbol{a}} \boldsymbol{a}^{\mathcal{L}} \cdot \mathbf{v} = d_{\boldsymbol{a}^{\mathcal{L}}} L_{\mathcal{L}} \cdot \left[\mathrm{vech}\left(\Sigma^{\mathcal{L}} \mathrm{math}(\mathbf{v}) \Sigma^{\mathcal{L}\top}\right)\right], \tag{C.5}$$

or, equivalently,

$$\begin{aligned}\langle \nabla_{\boldsymbol{a}} L_{\mathcal{L}}, \mathbf{v} \rangle &= \langle \nabla_{\boldsymbol{a}^{\mathcal{L}}} L_{\mathcal{L}}, \mathrm{vech}(\Sigma^{\mathcal{L}} \mathrm{math}(\mathbf{v}) \Sigma^{\mathcal{L}\top}) \rangle = \langle \mathrm{vech}^*(\nabla_{\boldsymbol{a}^{\mathcal{L}}} L_{\mathcal{L}}), \Sigma^{\mathcal{L}} \mathrm{math}(\mathbf{v}) \Sigma^{\mathcal{L}\top} \rangle = \\ &= \mathrm{tr}\left(\mathrm{vech}^*(\nabla_{\boldsymbol{a}^{\mathcal{L}}} L_{\mathcal{L}}) \cdot \Sigma^{\mathcal{L}} \mathrm{math}(\mathbf{v}) \Sigma^{\mathcal{L}\top}\right) = \mathrm{tr}\left(\Sigma^{\mathcal{L}\top} \mathrm{vech}^*(\nabla_{\boldsymbol{a}^{\mathcal{L}}} L_{\mathcal{L}}) \cdot \Sigma^{\mathcal{L}} \mathrm{math}(\mathbf{v})\right) = \\ &= \langle \Sigma^{\mathcal{L}\top} \mathrm{vech}^*(\nabla_{\boldsymbol{a}^{\mathcal{L}}} L_{\mathcal{L}}) \Sigma^{\mathcal{L}}, \mathrm{math}(\mathbf{v}) \rangle = \langle \mathrm{math}^*(\Sigma^{\mathcal{L}\top} \mathrm{vech}^*(\nabla_{\boldsymbol{a}^{\mathcal{L}}} L_{\mathcal{L}}) \Sigma^{\mathcal{L}})), \mathbf{v} \rangle,\end{aligned} \tag{C.6}$$

where $\mathrm{vech}^* : \mathbb{R}^{N_d} \longrightarrow \mathbb{S}_d$ and $\mathrm{math}^* : \mathbb{S}_d \longrightarrow \mathbb{R}^{N_d}$ are the adjoint maps of the operators vech and math, respectively. Using the properties (A.7) and (A.8) we obtain that for any $\mathbf{w} \in \mathbb{R}^{N_d}$

$$\begin{aligned}
\mathrm{math}^*(\Sigma^{\mathcal{L}\top}\mathrm{vech}^*(\mathbf{w})\Sigma^{\mathcal{L}}) &= \frac{1}{2}\mathrm{math}^*\left[\Sigma^{\mathcal{L}\top}\cdot(\mathrm{math}(\mathbf{w})+\mathrm{Diag}(\mathrm{math}(\mathbf{w})))\cdot\Sigma^{\mathcal{L}}\right]\\
&= \frac{1}{2}\mathrm{math}^*(\Sigma^{\mathcal{L}\top}\mathrm{math}(\mathbf{w})\Sigma^{\mathcal{L}}) + \frac{1}{2}\mathrm{math}^*(\Sigma^{\mathcal{L}\top}\mathrm{Diag}(\mathrm{math}(\mathbf{w}))\Sigma^{\mathcal{L}})\\
&= \mathrm{vech}(\Sigma^{\mathcal{L}\top}\mathrm{math}(\mathbf{w})\Sigma^{\mathcal{L}} - \frac{1}{2}\mathrm{Diag}(\Sigma^{\mathcal{L}\top}\mathrm{math}(\mathbf{w})\Sigma^{\mathcal{L}}))\\
&+ \mathrm{vech}(\Sigma^{\mathcal{L}\top}\mathrm{Diag}(\mathrm{math}(\mathbf{w}))\Sigma^{\mathcal{L}})\\
&- \frac{1}{2}\mathrm{vech}\left(\mathrm{Diag}\left(\Sigma^{\mathcal{L}\top}\mathrm{Diag}(\mathrm{math}(\mathbf{w}))\Sigma^{\mathcal{L}}\right)\right). \qquad \text{(C.7)}
\end{aligned}$$

By part **(vii)** of Lemma 6.1, the relation (C.7) immediately yields that

$$\mathrm{math}^*(\Sigma^{\mathcal{L}\top}\mathrm{vech}^*(\mathbf{w})\Sigma^{\mathcal{L}}) = \mathrm{vech}(\Sigma^{\mathcal{L}\top}\mathrm{math}(\mathbf{w})\Sigma^{\mathcal{L}}), \qquad \text{(C.8)}$$

which substituted in (C.6) proves (C.3). The relation (C.4) is proved analogously. ■

C.2 Reduction of the rank one model

We consider the rank one model prescription (2.10) described in Section 2.3 with the corresponding identification **(IC)** and positivity **(PSD)** constraints (2.11)-(2.12). The following proposition states the closure of the rank one specification under reduction and provides the functional link between the expressions for the gradient of the log-likelihood function of the reduced rank one model with respect to its intrinsic parameters and those of the initial model.

Proposition C.2 *In the same conditions as in Proposition C.1, consider an initial rank one model with parameter matrices* $\mathbf{\Theta} = (A, B) \in \mathbb{S}_n \times \mathbb{S}_n$ *intrinsically paramtrized with* $\boldsymbol{\theta} = (\boldsymbol{a}, \boldsymbol{b}) \in \mathbb{R}^n \times \mathbb{R}^n$*. Then the following statements hold true:*

(i) *The reduced model associated to* $\Sigma^{\mathcal{L}}$ *is a rank one model with parameter matrices* $\mathbf{\Theta}^{\mathcal{L}} = (A^{\mathcal{L}}, B^{\mathcal{L}}) \in \mathbb{S}_d \times \mathbb{S}_d$ *intrinsically parametrized by* $\boldsymbol{\theta}^{\mathcal{L}} = (\boldsymbol{a}^{\mathcal{L}}, \boldsymbol{b}^{\mathcal{L}}) \in \mathbb{R}^d \times \mathbb{R}^d$ *which are related to the parameters* $\boldsymbol{\theta}$ *of the initial rank one model via the relations*

$$\boldsymbol{a}^{\mathcal{L}} = \Sigma^{\mathcal{L}}\boldsymbol{a} \ \text{ and } \ \boldsymbol{b}^{\mathcal{L}} = \Sigma^{\mathcal{L}}\boldsymbol{b}. \qquad \text{(C.9)}$$

(ii) *If the initial rank one model parameters* $\boldsymbol{\theta}$ *satisfy the positivity constraints* **(PSD)** *(2.12), then so do the parameters* $\boldsymbol{\theta}^{\mathcal{L}}$ *of the reduced rank one model. This is not necessarily true for the identification constraints* **(IC)** *(2.11).*

(iii) *The statement in part* **(ii)** *of Proposition C.1 applies to the rank one case with (C.3)-(C.4) replaced by*

$$\nabla_{\boldsymbol{a}} L_{\mathcal{L}} = \Sigma^{\mathcal{L}\top}\nabla_{\boldsymbol{a}^{\mathcal{L}}} L_{\mathcal{L}}, \qquad \text{(C.10)}$$

$$\nabla_{\boldsymbol{b}} L_{\mathcal{L}} = \Sigma^{\mathcal{L}\top}\nabla_{\boldsymbol{b}^{\mathcal{L}}} L_{\mathcal{L}}. \qquad \text{(C.11)}$$

Proof. **(i)** It is obtained by a straightforward verification. Indeed, for the parameter matrix $A(\mathbf{PSD}) \in \mathbb{S}_d$ of the reduced model associated to $\Sigma^{\mathcal{L}}$ we write

$$A^{\mathcal{L}} = \Sigma^{\mathcal{L}} A \Sigma^{\mathcal{L}\top} = \Sigma^{\mathcal{L}} \boldsymbol{a}\boldsymbol{a}^{\top} \Sigma^{\mathcal{L}\top} = \boldsymbol{a}^{\mathcal{L}} \boldsymbol{a}^{\mathcal{L}\top}. \tag{C.12}$$

The same argument applies to the parameter $\boldsymbol{b}^{\mathcal{L}}$.
(ii) It is straightforward.
(iii) In order to prove (C.10) and (C.11), we proceed in the same way as in the proof of the part **(ii)** in Proposition C.1. We write for any $\mathbf{v} \in \mathbb{R}^n$

$$d_{\boldsymbol{a}} L_{\mathcal{L}} \cdot \mathbf{v} = d_{\boldsymbol{a}^{\mathcal{L}}} L_{\mathcal{L}} \cdot T_{\boldsymbol{a}} \boldsymbol{a}^{\mathcal{L}} \cdot \mathbf{v} = d_{\boldsymbol{a}^{\mathcal{L}}} L_{\mathcal{L}} \cdot \Sigma^{\mathcal{L}} \mathbf{v}, \tag{C.13}$$

which immediately yields

$$\nabla_{\boldsymbol{a}} L_{\mathcal{L}} = \Sigma^{\mathcal{L}\top} \cdot \nabla_{\boldsymbol{a}^{\mathcal{L}}} L_{\mathcal{L}}, \tag{C.14}$$

as required. The relation (C.11) is proved analogously. ■

C.3 Reduction of the Almon model

Consider the Almon model prescription (2.13) provided in Subsection 2.3 subjected to the corresponding identification **(IC)** (2.14) and positivity **(PSD)** constraints (2.15). In the following proposition we show that the Almon model is, in general, not closed under reduction but it is nevertheless transformed into a specifically "subparametrized" rank one model for which the composite likelihood estimation can be implemented. Additionally, we provide the relevant details on how to derive the expressions of the gradient of the log-likelihood function of the reduced rank one model with respect to the intrinsic parameters of the initial Almon model.

Proposition C.3 *In the same conditions as in Proposition C.1, consider an initial Almon model with parameter matrices* $\boldsymbol{\Theta} = (A, B) \in \mathbb{S}_n \times \mathbb{S}_n$ *intrinsically paramtrized with* $\boldsymbol{\theta} = (\boldsymbol{a}, \boldsymbol{b}) \in \mathbb{R}^3 \times \mathbb{R}^3$ *via the assignments* $\widetilde{\boldsymbol{a}} = \mathrm{alm}_n(\boldsymbol{a})$, $\widetilde{\boldsymbol{b}} = \mathrm{alm}_n(\boldsymbol{b})$ *or, analogously,* $\widetilde{a}_i = \mathrm{alm}_n(\boldsymbol{a}, i)$, $\widetilde{b}_i = \mathrm{alm}_n(\boldsymbol{b}, i)$, $i \in \{1, \ldots, n\}$ *and* $A = \widetilde{\boldsymbol{a}}\widetilde{\boldsymbol{a}}^{\top}$, $B = \widetilde{\boldsymbol{b}}\widetilde{\boldsymbol{b}}^{\top}$ *as in* (2.13). *Then the following statements hold true:*

(i) *The reduced model associated to* $\Sigma^{\mathcal{L}}$ *is a rank one model with matrices* $\boldsymbol{\Theta}^{\mathcal{L}} = (A^{\mathcal{L}}, B^{\mathcal{L}}) \in \mathbb{S}_d \times \mathbb{S}_d$ *that are intrinsically parametrized by the same parameters* $\boldsymbol{\theta} = (\boldsymbol{a}, \boldsymbol{b}) \in \mathbb{R}^3 \times \mathbb{R}^3$ *as the original Almon model via the relations* $A^{\mathcal{L}} = \widetilde{\boldsymbol{a}}^{\mathcal{L}} \widetilde{\boldsymbol{a}}^{\mathcal{L}\top}$, $B^{\mathcal{L}} = \widetilde{\boldsymbol{b}}^{\mathcal{L}} \widetilde{\boldsymbol{b}}^{\mathcal{L}\top}$, *where*

$$\widetilde{\boldsymbol{a}}^{\mathcal{L}} = \Sigma^{\mathcal{L}} \widetilde{\boldsymbol{a}} \quad \text{with} \quad (\widetilde{\boldsymbol{a}}^{\mathcal{L}})_j = \mathrm{alm}_n(\boldsymbol{a}, l_j), \quad j \in \{1, \ldots, d\}, \tag{C.15}$$

$$\widetilde{\boldsymbol{b}}^{\mathcal{L}} = \Sigma^{\mathcal{L}} \widetilde{\boldsymbol{b}} \quad \text{with} \quad (\widetilde{\boldsymbol{b}}^{\mathcal{L}})_j = \mathrm{alm}_n(\boldsymbol{b}, l_j), \quad j \in \{1, \ldots, d\}. \tag{C.16}$$

(ii) *If the initial Almon model satisfies the positive semidefiniteness constraints* **(PSD)** (2.15), *then so does the reduced rank one model. This is not necessarily true for the identification constraints* **(IC)** *in* (2.14).

(iii) *Let $L_{\mathcal{L}} = \log L_{\mathcal{L}}(\boldsymbol{\Theta}^{\mathcal{L}}(\boldsymbol{\theta}); \mathbf{r}^{\Sigma^{\mathcal{L}}})$ be the log-likelihood function associated to the reduced rank one model corresponding to $\Sigma^{\mathcal{L}}$. The components of the gradient of $L_{\mathcal{L}}$ with respect to the intrinsic parameters $\boldsymbol{\theta}$ of the initial Almon model is computed using the expression for the gradient of the log-likelihood function for the Almon case replacing the map $\boldsymbol{\Theta}(\boldsymbol{\theta})$ given in (A.18) by*

$$\begin{array}{rccc} \boldsymbol{\Theta}^{\mathcal{L}}: & \mathbb{R}^3 \times \mathbb{R}^3 & \longrightarrow & \mathbb{S}_d \times \mathbb{S}_d \\ & \boldsymbol{\theta} = (\boldsymbol{a}, \boldsymbol{b}) & \longmapsto & (\Sigma^{\mathcal{L}} \mathrm{alm}_n(\boldsymbol{a}) (\mathrm{alm}_n(\boldsymbol{a}))^\top \Sigma^{\mathcal{L}\top}, \Sigma^{\mathcal{L}} \mathrm{alm}_n(\boldsymbol{b}) (\mathrm{alm}_n(\boldsymbol{b}))^\top \Sigma^{\mathcal{L}\top}), \end{array} \tag{C.17}$$

and replacing the adjoint of the tangent map $T^*_{\boldsymbol{\theta}} \boldsymbol{\Theta} : \mathbb{S}_n \times \mathbb{S}_n \longrightarrow \mathbb{R}^3 \times \mathbb{R}^3$ in (B.12) by

$$\begin{array}{rccc} T^*_{\boldsymbol{\theta}} \boldsymbol{\Theta}^{\Sigma^{\mathcal{L}}}: & \mathbb{S}_d \times \mathbb{S}_d & \longrightarrow & \mathbb{R}^3 \times \mathbb{R}^3 \\ & (\Delta_1, \Delta_2) & \longmapsto & 2\left(K_{\boldsymbol{a}}^\top \Sigma^{\mathcal{L}\top} \Delta_1 \Sigma^{\mathcal{L}} \mathrm{alm}_n(\boldsymbol{a}), K_{\boldsymbol{b}}^\top \Sigma^{\mathcal{L}\top} \Delta_2 \Sigma^{\mathcal{L}} \mathrm{alm}_n(\boldsymbol{b})\right), \end{array} \tag{C.18}$$

where $K_{\boldsymbol{a}} = \left(\mathbf{i}_n \mid \mathbf{k}_n^1 \odot \mathrm{alm}_n(\bar{\boldsymbol{a}}) \mid \mathbf{k}_n^2 \odot \mathrm{alm}_n(\bar{\boldsymbol{a}})\right)$, $K_{\boldsymbol{b}} = \left(\mathbf{i}_n \mid \mathbf{k}_n^1 \odot \mathrm{alm}_n(\bar{\boldsymbol{b}}) \mid \mathbf{k}_n^2 \odot \mathrm{alm}_n(\bar{\boldsymbol{b}})\right)$ $\in \mathbb{M}_{n,3}$, $\bar{\boldsymbol{a}} = (0, a_2, a_3)^\top$, $\bar{\boldsymbol{b}} = (0, b_2, b_3)^\top$, the symbol $\mid$ denotes vertical concatenation, and $\mathbf{k}_n^1 = (1, 2, \ldots, n)^\top$, $\mathbf{k}_n^2 = \left(1, 2^2, \ldots, n^2\right)^\top \in \mathbb{R}^n$.

Proof. **(i)** It is obtained by a straightforward verification. Indeed, for the parameter matrix $A^{\mathcal{L}} \in \mathbb{S}_d$ of the reduced model associated to $\Sigma^{\mathcal{L}}$ we write

$$A^{\mathcal{L}} = \Sigma^{\mathcal{L}} A \Sigma^{\mathcal{L}\top} = \Sigma^{\mathcal{L}} \widetilde{\boldsymbol{a}} \widetilde{\boldsymbol{a}}^\top \Sigma^{\mathcal{L}\top} = \widetilde{\boldsymbol{a}}^{\mathcal{L}} \widetilde{\boldsymbol{a}}^{\mathcal{L}\top}, \tag{C.19}$$

where $\widetilde{\boldsymbol{a}}^{\mathcal{L}} = \Sigma^{\mathcal{L}} \widetilde{\boldsymbol{a}} = \Sigma^{\mathcal{L}} \mathrm{alm}_n(\boldsymbol{a})$ with components $\widetilde{a}^{\mathcal{L}}_j = \delta_{l_j, j} \mathrm{alm}_n(\boldsymbol{a})_j$, $j = \{1, \ldots, d\}$ which can be hence written as $(\widetilde{\boldsymbol{a}}^{\mathcal{L}})_j = \mathrm{alm}_n(\boldsymbol{a}, l_j)$, $j \in \{1, \ldots, d\}$. The same argument applies to $B^{\mathcal{L}}$.
(ii) The expression (C.17) immediately follows from part **(i)** of the proposition. In order to prove (C.18), we proceed in the same way as in the proof of part **(ii)** in Proposition C.1. We write for any $\mathbf{v} \in \mathbb{R}^3$

$$d_{\boldsymbol{a}} L_{\mathcal{L}} \cdot \mathbf{v} = d_{A^{\mathcal{L}}} L_{\mathcal{L}} \cdot T_{\boldsymbol{a}} A^{\mathcal{L}} \cdot \mathbf{v}, \tag{C.20}$$

where $T_{\boldsymbol{a}} A^{\mathcal{L}} : \mathbb{R}^3 \longrightarrow \mathbb{S}_d$ is the tangent of the map $A^{\mathcal{L}}(\boldsymbol{\theta}) = \widetilde{\boldsymbol{a}}^{\mathcal{L}}(\boldsymbol{\theta}) \widetilde{\boldsymbol{a}}^{\mathcal{L}\top}(\boldsymbol{\theta})$. This expression yields:

$$\nabla_{\boldsymbol{a}} L_{\mathcal{L}} = T^*_{\boldsymbol{a}} A^{\mathcal{L}}(\nabla_{A^{\mathcal{L}}} L_{\mathcal{L}}). \tag{C.21}$$

We now determine the map $T_{\boldsymbol{a}} A^{\mathcal{L}} : \mathbb{R}^3 \longrightarrow \mathbb{S}_d$ and obtain

$$T_{\boldsymbol{a}} A^{\mathcal{L}} \cdot \boldsymbol{\delta a} = \Sigma^{\mathcal{L}} (K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a}) \mathrm{alm}_n(\boldsymbol{a})^\top \Sigma^{\mathcal{L}\top} + \Sigma^{\mathcal{L}} \mathrm{alm}_n(\boldsymbol{a}) (K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a})^\top \Sigma^{\mathcal{L}\top}, \tag{C.22}$$

where we used the expressions (C.19), (C.17), and the definition (A.10) of the tangent map $T_{\boldsymbol{\theta}} \mathrm{alm}_n : \mathbb{R}^3 \longrightarrow \mathbb{R}^n$. We now dualize the relation (C.22) in order to determine the adjoint map $T^*_{\boldsymbol{a}} A^{\mathcal{L}} : \mathbb{S}_d \longrightarrow \mathbb{R}^3$. For arbitrary $\Delta \in \mathbb{S}_d$ we write

$$\begin{aligned} \langle T^*_{\boldsymbol{a}} A^{\mathcal{L}}(\Delta), \boldsymbol{\delta a} \rangle =& \langle \Delta, T_{\boldsymbol{a}} A^{\mathcal{L}} \cdot \boldsymbol{\delta a} \rangle = \langle \Delta, \Sigma^{\mathcal{L}} (K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a}) \mathrm{alm}_n(\boldsymbol{a})^\top \Sigma^{\mathcal{L}\top} \rangle \\ & + \langle \Delta, \Sigma^{\mathcal{L}} \mathrm{alm}_n(\boldsymbol{a}) (K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a})^\top \Sigma^{\mathcal{L}\top} \rangle \\ =& 2 \langle \Delta \Sigma^{\mathcal{L}} \mathrm{alm}_n(\boldsymbol{a}), \Sigma^{\mathcal{L}} (K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a}) \rangle = 2 \langle K_{\boldsymbol{a}}^\top \Sigma^{\mathcal{L}\top} \Delta \Sigma^{\mathcal{L}} \mathrm{alm}_n(\boldsymbol{a}), \boldsymbol{\delta a} \rangle \end{aligned} \tag{C.23}$$

which yields

$$\begin{array}{rccc} T_{\boldsymbol{a}}^{*}A^{\mathcal{L}}: & \mathbb{S}_d & \longrightarrow & \mathbb{R}^3 \\ & \Delta & \longmapsto & 2\left(K_{\boldsymbol{a}}^{\top}\Sigma^{\mathcal{L}\top}\Delta\Sigma^{\mathcal{L}}\mathrm{alm}_n(\boldsymbol{a})\right). \end{array} \tag{C.24}$$

An analog expression can be proved for the tangent map $T_{\boldsymbol{b}}B^{\mathcal{L}}: \mathbb{R}^3 \longrightarrow \mathbb{S}_d$ and its adjoint $T_{\boldsymbol{b}}^{*}B^{\Sigma;\mathcal{L}}: \mathbb{S}_d \longrightarrow \mathbb{R}^3$. By combining the resulting two maps we obtain (C.18). ■

References

[Alm65] S. Almon. The distributed lag between capital appropriations and expenditures. *Econometrica*, 33(1):178–196, 1965.

[BGO15] Luc Bauwens, Lyudmila Grigoryeva, and Juan-Pablo Ortega. Estimation and empirical performance of non-scalar dynamic conditional correlation models. *To appear in Computational Statistics and Data Analysis*, 2015.

[BR97] R. B. Bapat and T. E. S. Raghavan. *Nonnegative Matrices and Applications*. Cambridge University Press, 1997.

[CO14] Stéphane Chrétien and Juan-Pablo Ortega. Multivariate GARCH estimation via a Bregman-proximal trust-region method. *Computational Statistics and Data Analysis*, 76:210–236, 2014.

[HJ94] Roger A. Horn and Charles R. Johnson. *Topics in matrix analysis*. Cambridge University Press, Cambridge, 1994.