2014/12

■

Estimation and empirical performance of non-scalar dynamic conditional correlation models

Luc Bauwens, Lyudmila Grigoryeva
and Juan-Pablo Ortega

DISCUSSION PAPER

Center for Operations Research
and Econometrics

Voie du Roman Pays, 34
B-1348 Louvain-la-Neuve
Belgium
http://www.uclouvain.be/core

CORE DISCUSSION PAPER
2014/12

Estimation and empirical performance of non-scalar dynamic conditional correlation models

Luc BAUWENS [1], Lyudmila GRIGORYEVA [2]
and Juan-Pablo ORTEGA [3]

March 2014

Abstract

This paper presents a method capable of estimating richly parametrized versions of the dynamic conditional correlation (DCC) model that go beyond the standard scalar case. The algorithm is based on the maximization of a Gaussian quasi-likelihood using a Bregman-proximal trust-region method to handle the various non-linear stationarity and positivity constraints that arise in this context. We consider the general matrix Hadamard DCC model with full rank, rank equal to two and, additionally, two different rank one matrix specifications. In the last mentioned case, the elements of the vectors that determine the rank one parameter matrices are either arbitrary or parsimoniously defined using the Almon lag function. We use actual stock returns data in dimensions up to thirty in order to carry out performance comparisons according to several in- and out-of-sample criteria. Our empirical results show that the use of richly parametrized models adds value with respect to the conventional scalar case.

Keywords: multivariate volatility modeling, dynamic conditional correlations (DCC), non-scalar DCC models, constrained optimization, Bregman divergences, Bregman-proximal trust-region method.

JEL classification: C13, C32, G17.

[1] Université catholique de Louvain, CORE, B-1348 Louvain-la-Neuve, Belgium; University of Johannesburg, Department of Economics, Johannesburg, South Africa. E-mail: luc.bauwens@uclouvain.be

[2] Laboratoire de Mathématiques de Besançon, Université de Franche-Comté, UFR des Sciences et Techniques, F-25030 Besançon, France. E-mail: 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, F-25030 Besançon, France. E-mail: juan-pablo.ortega@univ-fcomte.fr

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". Lyudmila Grigoryeva acknowledges financial support from the Faculty for the Future Program of the Schlumberger Foundation. Luc Bauwens and Juan-Pablo Ortega acknowledge partial financial support of the Région de Franche-Comté (Convention 2013C-5493).

1 Introduction

The choice of the dynamic conditional correlation (DCC) model has become very common in the multivariate GARCH applied literature, where the goal is often to fit the dynamics of time-varying conditional variances and correlations of asset returns and to forecast their future values. The DCC family was developed in the works of Engle [Engl 02] and Tse and Tsui [Tse 02] as a generalization of the constant conditional correlation (CCC) model of Bollerslev [Boll 90]. The main advantage of DCC models is the availability of a two-step estimation procedure which, combined with correlation targeting, makes their use feasible even when the number of assets is high. It is worth noting that even though in the original paper [Engl 02] a general matrix Hadamard-type model parameterization is proposed, it is almost exclusively the scalar prescription that is used in applications. This simplified version of the model imposes the same correlation dynamics to all the pairs of assets that are considered which, for sizable dimensions, may constitute a too restrictive homogeneity assumption.

The main contribution of this paper is to provide adequate estimation tools for several non-scalar richly parameterized DCC models and to empirically evaluate how they perform with respect to each other and to the scalar model. The non-scalar models that we consider are the general matrix Hadamard-type model and four more parsimonious particular cases thereof. Our main conclusions are that the estimation of these models is practically feasible in high dimensions, and that two of the non-scalar models that we consider are worth using in practice.

We emphasize that we restrict ourselves to DCC models in which only an approximate correlation targeting as suggested by Engle [Engl 02] is possible. This is a widespread approach that aims at reducing the number of parameters in the likelihood maximization by replacing the constant term matrix of the quasi-correlation process by a moment estimate. This approach is controversial in the DCC context since the approximate targeting procedure is statistically inconsistent (see the papers by Caporin and McAleer [Capo 12] for an extensive review). This has motivated the introduction by Aielli [Aiel 13] of a corrected DCC (cDCC) model that carefully addresses this issue, but the price to pay is that estimation is more convoluted. Despite the theoretical interest of the cDCC model, we do not consider it in this work since existing empirical findings and simulation results in [Aiel 13] reveal that the cDCC and DCC models do not differ much in the one-step-ahead forecasting context that we focus on.

The need to go beyond the scalar DCC prescription has been partly addressed in the works of Franses and Hafner [Hafn 09] and Cappiello *et al.* [Capp 06].[4] In these papers, a so-called diagonal DCC model that allows for asset-specific heterogeneity in the correlation structure is used. This extension specifies the parameter matrix associated to the lagged innovations as a rank one matrix, and the same applies to the parameter matrix of the lagged quasi-correlation term although that one is also kept as a scalar parameter in [Hafn 09]. The results in those papers provide empirical evidence supporting that these richer DCC models exhibit improved performance with respect to the scalar model. Another relevant work in the same direction is the paper by Noureldin *et al.* [Nour 14] who introduce (among others) a model, called Rotated DCC (RDCC), which uses similar specifications as in the two previously cited papers, after applying an estimated orthogonal transformation to the devolatized returns, and illustrate that it yields increased performance when compared to existing models like, for instance, the OGARCH model [Alex 98]. Our contribution with respect to these papers is that we consider non-scalar models other than the diagonal DCC cases mentioned above.

The results in [Hafn 09] and [Capp 06] lead us to believe that non-scalar DCC models with asset specific dynamics in the correlation structure, like the one associated to the original matrix Hadamard-type parameterization [Engl 02], can yield a superior performance in practical applications when compared

[4] Other extensions or variations of the scalar DCC model have been proposed in the literature: we can mention the papers of Billio *et al.* [Bill 03, Bill 06], Fernandes *et al.* [Fern 05]), Pelletier [Pell 06], Billio and Caporin [Bill 05], Colacito *et al.* [Cola 11], and Bauwens and Otranto [Bauw 13]. We do not discuss these models in detail because they are not directly related to the non-scalar DCC models that we consider in this paper.

with the widespread scalar DCC model. This conjecture can only be verified when effective estimation procedures are available for the richer models, the absence of which explains in part statements in the literature (see for example [Bill 06] or Chapter 7 of [Engl 09]) about the lack of empirical interest of these more general models. Two difficulties in this respect arise: the first one is the quadratic dependence of the number of parameters on the model dimension and the second is the need to impose in the estimation the nonlinear constraints that ensure the positivity and the stationarity of the dynamic correlation process.

We therefore turn first our attention to the optimization techniques that can be used for the estimation of non-scalar DCC models. More specifically, we develop the approach proposed by Chrétien and Ortega in [Chre 13] based on the use of Bregman divergences. These authors have successfully applied this approach to the estimation of the heavily parametrized VEC-GARCH model subjected to stationarity and positivity constraints. In our work we extend this optimization method to the DCC models and develop explicit estimation tools for a variety of non-scalar DCC specifications originating from the general Hadamard-type DCC prescription in [Engl 02]. Even though the DCC family has a much smaller parameter space than VEC, the use of the Bregman divergences approach is extremely advantageous in the treatment of the DCC highly nonlinear optimization constraints. The paper is therefore organized around the following three topics.

DCC model specifications: We consider the Hadamard DCC family for which the two parameter matrices of the lagged innovations and lagged quasi-correlation terms are symmetric with full rank, as well as four other subfamilies where these matrices have smaller ranks. More specifically, in one case the rank is set equal to two, and in three other cases it is equal to one. The first of these three "rank one" cases is equivalent to the diagonal DCC model considered also in [Hafn 09] and Cappiello *et al.* [Capp 06] where the parameter matrices are built as outer products of vectors of parameters of dimension n. The last two "rank one" models are new and called Almon DCC and Almon shuffle DCC. In these models, the elements of the vectors that generate the rank deficient parameter matrices are defined using an Almon function [Almo 65]. Thus, like in the scalar model, the number of parameters that need to be estimated does not depend on the dimension n but, like in the diagonal model, the correlation dynamics differs for all pairs of assets. The Almon DCC models are therefore more flexible than the scalar DCC model, while determined by a comparable number of parameters.

Section 3 is devoted to presenting the general setup for DCC models, to describing in detail the different parameterizations under study, as well as the constraints that are imposed on each of them in order to ensure the stationarity of the process and the positive definiteness of the resulting conditional correlation matrices. As a prerequisite, Section 2 spells out the notation and contains several mathematical results used in the paper.

DCC model estimation: this is the subject of Section 4 where we provide explicit expressions for the log-likelihood function and its gradient that are used for estimation. Following the scheme proposed in [Chre 13] we use Bregman divergences in order to handle the model constraints. This approach is much used in the context of machine learning (see for instance [Dhil 07, Kuli 09a] and references therein). In our context, it is particularly advantageous because it allows to treat the nonlinear optimization constraints that our problem is exposed to without resorting to Lagrange duality or other techniques that demand the solution of supplementary optimization problems. Section 4 contains a comprehensive description of the ingredients necessary to implement this optimization algorithm for each of the DCC models that we consider. In a forthcoming work we show that the non-scalar DCC models considered in this paper can be estimated by the composite quasi-maximum likelihood (CQML) method, as done by Engle *et al.* [Engl 08b] for the scalar model, and that this method is advantageous for processes in high dimensions.

DCC specifications performance assessment: the possibility of estimating non-scalar DCC models with the tools that we just mentioned allows us to empirically study their performance and, ultimately, to assess the need for those models in the processing of financial data. In Section 5 we carry out an in- and and out-of-sample study using two datasets that contain the historical evolution of the returns of the thirty assets listed in the Dow Jones Industrial Average Index during two different time periods. The results that we obtain reveal that: **(i)** no model (whether scalar or not) dominates systematically the others in terms of in-sample and out-of-sample performance; **(ii)** the full rank Hadamard model is in most cases dominated by the less richly parametrized versions; **(iii)** among the less rich models, the rank one (or diagonal) model and the Almon (shuffle) models are in many cases performing better than the scalar model. We nevertheless underline that for a given dataset, the choice between the models has to be made by taking into account the specifics of time series in question, the available sample length, dimension, and other considerations.

A technical appendix is available online or by request to one of the authors. It contains the proofs of the results in the paper and other technical details.

2 Notation and preliminaries

In this section we specify the notation that is used throughout the paper and we provide various general results that are quoted in the sequel.

2.1 Vectors and matrices

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

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 $i \in \{1, \dots, n\}$, $j \in \{1, \dots m\}$. 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 $\mathbf{I}_n \in \mathbb{D}_n$ denotes the identity matrix and $\mathbb{O}_n \subset \mathbb{M}_n$ is the subspace of orthogonal matrices, that is, $\mathbb{O}_n := \{A \in \mathbb{M}_n \mid AA^\top = \mathbf{I}_n\}$.

The Frobenius inner product is defined on the space $\mathbb{M}_{n,m}$ as:

$$\langle A, B\rangle := \operatorname{tr}\left(AB^\top\right) = \operatorname{tr}\left(A^\top B\right), \quad A, B \in \mathbb{M}_{n,m}. \tag{2.1}$$

The symbol tr denotes the trace of the matrix. This inner product induces the Frobenius norm that we denote as $\|A\| := \langle A, A\rangle^{1/2}$. Given a linear operator $\mathcal{A} : \mathbb{M}_{n,m} \longrightarrow \mathbb{M}_{p,q}$, we denote by $\mathcal{A}^* : \mathbb{M}_{p,q} \longrightarrow \mathbb{M}_{n,m}$ its adjoint operator with respect to (2.1) by the relation

$$\langle B, \mathcal{A}(C)\rangle = \langle \mathcal{A}^*(B), C\rangle, \quad \text{for any} \quad B \in \mathbb{M}_{n,m}, \text{ and } C \in \mathbb{M}_{p,q}.$$

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:

$$(A \odot B)_{ij} := A_{ij}B_{ij} \text{ for all } i \in \{1, \ldots, n\}, \ j \in \{1, \ldots, m\}. \tag{2.2}$$

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

$$\mathbf{u} \odot \mathbf{w} = U\mathbf{w}, \tag{2.3}$$

where $U \in \mathbb{D}_n$ is defined by $U_{ii} := u_i$, for all $i \in \{1, \ldots, n\}$, that is, $U := \text{diag}(\mathbf{u})$, 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 [Horn 94, 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

$$\text{tr}\left((A \odot B)\, C^\top\right) = \text{tr}\left((A \odot C)\, B^\top\right), \tag{2.4}$$

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 [Bapa 97] for a proof.

2.2 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 $\text{Diag} : \mathbb{M}_n \longrightarrow \mathbb{D}_n$ that sets equal to zero all the components of a square matrix except for those that are on the main diagonal. The operator $\text{diag} : \mathbb{R}^n \longrightarrow \mathbb{D}_n$ takes a given vector and constructs a diagonal matrix with its entries in the main diagonal. We denote by $\text{diag}^{-1} : \mathbb{D}_n \longrightarrow \mathbb{R}^n$ the inverse of the diag operator. The adjoint operator of Diag (respectively diag) is denoted by $\text{Diag}^* : \mathbb{D}_n \longrightarrow \mathbb{M}_n$ (respectively $\text{diag}^* : \mathbb{D}_n \longrightarrow \mathbb{R}^n$); it is easy to see that Diag^* is just the injection $\mathbb{D}_n \hookrightarrow \mathbb{M}_n$ and that $\text{diag}^* = \text{diag}^{-1}$.

The vec and mat operators: Given a matrix $A \in \mathbb{M}_{n,m}$, we denote by vec the operator that transforms A into a vector of length nm by stacking all its columns, namely,

$$\text{vec} : \mathbb{M}_{n,m} \longrightarrow \mathbb{R}^{nm}, \quad \text{vec}(A) = (A_{11}, \ldots, A_{n1}, \ldots, A_{1m}, \ldots, A_{nm})^\top.$$

The inverse of this operator is denoted as $\text{mat} : \mathbb{R}^{nm} \longrightarrow \mathbb{M}_{n,m}$.

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 $N := \frac{1}{2}n(n+1)$, that is,

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

and we denote the inverse of this operator by $\text{math} : \mathbb{R}^N \longrightarrow \mathbb{S}_n$.
The adjoint map of vech (respectively math) is denoted by $\text{vech}^* : \mathbb{R}^N \longrightarrow \mathbb{S}_n$ (respectively $\text{math}^* : \mathbb{S}_n \longrightarrow \mathbb{R}^N$). In [Chre 13] it is shown that given $A \in \mathbb{S}_n$ and $\mathbf{v} \in \mathbb{R}^N$, the following relations hold true:

$$\text{vech}^*(\mathbf{v}) = \frac{1}{2}\left(\text{math}(\mathbf{v}) + \text{Diag}\left(\text{math}(\mathbf{v})\right)\right), \tag{2.5}$$

$$\text{math}^*(A) = 2\ \text{vech}(A - \frac{1}{2}\ \text{Diag}(A)). \tag{2.6}$$

The relation between the vech, math, vec, and Diag operators: given a matrix $A \in \mathbb{S}_n$ and $N := \frac{1}{2}n(n+1)$, we denote by $L_n \in \mathbb{M}_{N,n^2}$ and by $D_n \in \mathbb{M}_{n^2,N}$ the elimination and the duplication matrices [Lutk 05], respectively. These matrices satisfy:

$$\text{vech}(A) = L_n \text{vec}(A), \tag{2.7}$$

$$\text{vec}(A) = D_n \text{vech}(A). \tag{2.8}$$

Given $A \in \mathbb{M}_n$, we define the diagonalization matrix $P_n^d \in \mathbb{M}_n$ via the relation:

$$\text{math}\left(P_n^d \text{vech}(A)\right) = \text{Diag}(A). \tag{2.9}$$

A well-known property of the vec operator that is exploited in the following sections is

$$\text{vec}(ABC) = \left(C^\top \otimes A\right)\text{vec}(B). \tag{2.10}$$

The mat_r and vec_r operators: let $r \leq n \in \mathbb{N}$, $N^* := nr - \frac{1}{2}r(r-1)$ and define the operator $\text{mat}_r : \mathbb{R}^{N^*} \longrightarrow \mathbb{L}_{n,r}$ that transforms a vector of length N^* into the lower triangular $n \times r$ matrix defined by,

$$\text{mat}_r(\mathbf{v}) = \begin{pmatrix} v_1 & 0 & \cdots & 0 \\ v_2 & v_{n+1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ v_r & v_{n+r-1} & \cdots & v_{N^*-n+r} \\ \vdots & \vdots & \ddots & \vdots \\ v_n & v_{2n-1} & \cdots & v_{N^*} \end{pmatrix}, \quad \text{for any } \mathbf{v} \in \mathbb{R}^{N^*}. \tag{2.11}$$

We denote the inverse of this operator as $\text{vec}_r : \mathbb{L}_{n,r} \longrightarrow \mathbb{R}^{N^*}$. We note for future reference that the elements in the main diagonal of (2.11) are given by

$$\{v_{i_1}, \ldots, v_{i_r}\} \quad \text{with} \quad i_j = n(j-1) + \frac{1}{2}j(3-j), \quad j \in \{1, \ldots, r\}. \tag{2.12}$$

The following proposition characterizes the adjoint maps of mat_r and vec_r, respectively. Its proof is provided in the Technical Appendix A.1.

Proposition 2.1 *Given $r \leq n \in \mathbb{N}$ and $N^* = nr - \frac{1}{2}r(r-1)$, let $A \in \mathbb{L}_{n,r}$ and $\mathbf{v} \in \mathbb{R}^{N^*}$ arbitrary. Let $\text{mat}_r^* : \mathbb{L}_{n,r} \longrightarrow \mathbb{R}^{N^*}$ and $\text{vec}_r^* : \mathbb{R}^{N^*} \longrightarrow \mathbb{L}_{n,r}$ be the adjoint maps of mat_r and vec_r, respectively. Then, the following relations hold true:*

$$\text{mat}_r^*(A) = \text{vec}_r(A), \tag{2.13}$$

$$\text{vec}_r^*(\mathbf{v}) = \text{mat}_r(\mathbf{v}). \tag{2.14}$$

The Almon lag operator and its tangent map: using the Almon lag function [Almo 65], we define the Almon lag operator $\mathrm{alm}_n : \mathbb{R}^3 \longrightarrow \mathbb{R}^n$, with $n \in \mathbb{N}$ and $\mathbf{v} \in \mathbb{R}^3$, as

$$(\mathrm{alm}_n(\mathbf{v}))_i := v_1 + \exp(v_2 i + v_3 i^2), \quad \text{for } i \in \{1, \dots, n\}\,. \tag{2.15}$$

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

$$T_{\mathbf{v}}\mathrm{alm}_n \cdot \boldsymbol{\delta}\mathbf{v} = K_{\mathbf{v}} \cdot \boldsymbol{\delta}\mathbf{v}, \quad \text{with } K_{\mathbf{v}} := \left(\mathbf{i}_n \mid \mathbf{k}_n^1 \odot \mathrm{alm}_n(\bar{\mathbf{v}}) \mid \mathbf{k}_n^2 \odot \mathrm{alm}_n(\bar{\mathbf{v}})\right) \in \mathbb{M}_{n,3}, \boldsymbol{\delta}\mathbf{v} \in \mathbb{R}^3, \tag{2.16}$$

where the symbol $\mid$ denotes vertical concatenation of matrices (or vectors), the vectors $\mathbf{i}_n, \mathbf{k}_n^1, \mathbf{k}_n^2 \in \mathbb{R}^n$ were introduced in Subsection 2.1, and $\bar{\mathbf{v}} \in \mathbb{R}^3$ is obtained out of the vector $\mathbf{v} \in \mathbb{R}^3$ by setting its first component equal to zero, namely, $\bar{\mathbf{v}} := (0, v_2, v_3)^\top$.
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

$$T_{\mathbf{v}}^*\mathrm{alm}_n(\mathbf{u}) = K_{\mathbf{v}}^\top \cdot \mathbf{u}, \quad \text{for any } \mathbf{u} \in \mathbb{R}^n. \tag{2.17}$$

3 Dynamic conditional correlation models

In this section, we present the various dynamic conditional correlation (DCC) model specifications that we study in the paper. In the first subsection, we define the most general model that we consider in the paper which is referred to as the (general) Hadamard DCC model. The following sections define more particular DCC models obtained by imposing parametric restrictions on the general Hadamard model. For each model, we explain the parametric constraints that need to be imposed so that the correlation process admits a stationary solution and the resulting conditional correlation matrices are positive definite. In some cases identification constraints are also required.

All these DCC models are based on the following assumption for the n-dimensional conditionally heteroscedastic discrete-time process $\{\mathbf{r}_t\}$:

$$\mathbf{r}_t = H_t^{1/2}\boldsymbol{\xi}_t, \quad \{\boldsymbol{\xi}_t\} \sim \mathrm{IN}(\mathbf{0}_n, \mathbf{I}_n), \quad t = 1, 2, \dots, T, \tag{3.1}$$

meaning that $\{\boldsymbol{\xi}_t\}$ is a set of n-dimensional independent normally distributed vectors with mean $\mathbf{0}_n$ and identity covariance matrix $\mathbf{I}_n$. $\{H_t\}$ is a predictable positive semidefinite matrix process, that is, for each $t \in \mathbb{N}$, H_t is a random matrix that takes values in $\mathbb{S}_n^+$; predictable means that the random variable H_t is $\mathcal{F}_{t-1}$-measurable, where $\mathcal{F}_{t-1} := \sigma(\mathbf{r}_1, \dots, \mathbf{r}_{t-1})$ is the information set generated by $\{\mathbf{r}_1, \dots, \mathbf{r}_{t-1}\}$.

In the first stage of the DCC model construction, a dynamic process is chosen for the conditional variances $\sigma_{i,t}^2$ of each component $r_{i,t}$ of $\mathbf{r}_t$, for example the GARCH(1,1) model [Engl 82, Boll 86]

$$\sigma_{i,t}^2 = \alpha_{0,i} + \alpha_{1,i} r_{i,t-1}^2 + \beta_{1,i}\sigma_{i,t-1}^2, \quad i \in \{1, \dots, n\}\,, \tag{3.2}$$

where the parameters $\alpha_{0,i}$, $\alpha_{1,i}$, and $\beta_{1,i}$ for all $i \in \{1, \dots, n\}$ satisfy the inequalities $\alpha_{0,i} > 0$, $\alpha_{1,i}, \beta_{1,i} \geq 0$, and $\alpha_{1,i} + \beta_{1,i} < 1$ in order to ensure the stationarity of the process and the positivity of the conditional variances $\sigma_{i,t}^2$. We define the standardized returns $\varepsilon_{i,t}$ of the i-th asset at time t as $\varepsilon_{i,t} := r_{i,t}/\sigma_{i,t}$ and assemble them in the vector $\boldsymbol{\varepsilon}_t$.

The second stage of the DCC model construction consists in specifying a dynamic equation for the conditional correlation matrices R_t of the standardized returns $\boldsymbol{\varepsilon}_t$. The matrix R_t is related to the covariance matrix H_t by the relation

$$H_t = D_t R_t D_t, \tag{3.3}$$

where $D_t := \mathrm{diag}(\sigma_{1,t}, \dots, \sigma_{n,t})$. The dynamic behavior of the conditional correlation process $\{R_t\}$ is modeled through a dynamic matrix process $\{Q_t\}$ to which it is connected by the relation

$$R_t = Q_t^{*-1/2} Q_t Q_t^{*-1/2}, \tag{3.4}$$

where the matrix Q_t^* is defined as $Q_t^* := \mathrm{Diag}(Q_t)$. The previous relation is needed to transform the matrix Q_t into a correlation matrix since the diagonal elements of Q_t are not necessarily equal to 1. The different DCC models that we consider in the paper differ in the way in which the dynamics for Q_t is parametrized.

3.1 The general Hadamard DCC model and its parameter constraints

The most general dynamic prescription that we use for Q_t is (see [Engl 02]):

$$Q_t = (\mathbf{i}_n\mathbf{i}_n^\top - A - B) \odot S + A \odot (\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}^\top) + B \odot Q_{t-1}. \tag{3.5}$$

In this equation, the parameters A and B are symmetric matrices, that is $A, B \in \mathbb{S}_n$, and $S := E\left[\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^\top\right] \in \mathbb{S}_n^+$. The particular expression chosen for the first summand of (3.5) leads to an approximate variance targeting of the model that is implemented by estimating S in a preliminary step (i.e. prior to the estimation of the parameters in A and B) by

$$S := \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^\top, \tag{3.6}$$

where T is the sample size (see [Capo 12, Aiel 13] for additional details). Since the parameter matrices A and B are generic elements in $\mathbb{S}_n$, they can be naturally 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 $B := \mathrm{math}\,(\boldsymbol{b})$. We refer to $\mathbb{R}^N \times \mathbb{R}^N$ as the **intrinsic parameter space** of the Hadamard DCC model. This space is generically denoted by $\mathcal{P} \times \mathcal{P}$ and the dimension of the **intrinsic parameter subspace** $\mathcal{P}$ by P. Using the intrinsic parameters that we just described, we write (3.5) as

$$Q_t = \left(\mathbf{i}_n\mathbf{i}_n^\top - \mathrm{math}\,(\boldsymbol{a}) - \mathrm{math}\,(\boldsymbol{b})\right) \odot S + \mathrm{math}\,(\boldsymbol{a}) \odot \left(\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}^\top\right) + \mathrm{math}\,(\boldsymbol{b}) \odot Q_{t-1}. \tag{3.7}$$

The dynamic equation that we just described guarantees neither that the resulting joint process $\{\mathbf{r}_t, H_t\}$ is stationary nor that $\{H_t\}$ takes values in the cone of positive semidefinite matrices $\mathbb{S}_n^+$ and hence consists of covariance matrices. Consequently, we state sufficient conditions on the parameters $\boldsymbol{a}$ and $\boldsymbol{b}$ that ensure those features:

(SC) Stationarity constraints: (see [Aiel 13, Proposition 2.3])

$$|a_i + b_i| < 1, \quad i \in \{1, \ldots, N\}\,. \tag{3.8}$$

(PSD) Positivity constraints:

$$\mathrm{math}\,(\boldsymbol{a}) \succeq 0, \quad \mathrm{math}\,(\boldsymbol{b}) \succeq 0, \tag{3.9}$$

$$\left(\mathbf{i}_n\mathbf{i}_n^\top - \mathrm{math}\,(\boldsymbol{a}) - \mathrm{math}\,(\boldsymbol{b})\right) \odot S \succ 0, \quad Q_0 \succeq 0. \tag{3.10}$$

The sufficiency of the positivity constraints follow from an inductive argument using the expression (3.7) that defines $\{Q_t\}$, together with the Schur Product Theorem recalled in Section 2. We stress that the first constraint in (3.10) cannot be replaced by $\mathbf{i}_n\mathbf{i}_n^\top - \mathrm{math}\,(\boldsymbol{a}) - \mathrm{math}\,(\boldsymbol{b}) \succ 0$ even if we know that $S \succ 0$ holds (which is the case due to the targeting). Indeed, if the constraints (3.9) hold, they prevent that $\mathbf{i}_n\mathbf{i}_n^\top - \mathrm{math}\,(\boldsymbol{a}) - \mathrm{math}\,(\boldsymbol{b}) \succ 0$ since any vector $\mathbf{v}$ in the kernel of $\mathbf{i}_n\mathbf{i}_n^\top$ satisfies $\langle \mathbf{v}, (\mathbf{i}_n\mathbf{i}_n^\top - A - B)\mathbf{v}\rangle \leq 0$.

Given that the number of parameters in (3.7) exhibits a quadratic dependence on the dimension n of the process $\{\mathbf{r}_t\}$, we dedicate the following subsections to formulate several more parsimonious parameterizations of this general Hadamard DCC model, and to characterize the associated identification, positivity, and stationarity constraints. In Table 6.1 we report the number of parameters of the matrices A and B in the different models for process dimensions going from five to thirty.

3.2 Rank deficient DCC models

The rank deficient DCC models constitute a subfamily of the Hadamard DCC family where the matrices A and $B \in \mathbb{S}_n$ in (3.5) are constrained to have a common prescribed rank $r \in \{1, \ldots, n-1\}$, that is, $\text{rank}(A) = \text{rank}(B) = r$. A natural parameterization for A and B consists of using $\widetilde{A}, \widetilde{B} \in \mathbb{M}_{n,r}$ such that $A = \widetilde{A}\widetilde{A}^\top$ and $B = \widetilde{B}\widetilde{B}^\top$. This choice poses an identification problem due to the invariance properties of the product $\widetilde{A}\widetilde{A}^\top$ because given $A \in \mathbb{S}_n$ with rank r and $\widetilde{A} \in \mathbb{M}_{n,r}$ such that $A = \widetilde{A}\widetilde{A}^\top$, this equality also holds true for any other matrix $\tilde{\widetilde{A}} := \widetilde{A}O$, with $O \in \mathbb{O}(r)$ an arbitrary element of the orthogonal group. Indeed, $\tilde{\widetilde{A}}\tilde{\widetilde{A}}^\top = \widetilde{A}OO^\top\widetilde{A}^\top = \widetilde{A}\widetilde{A}^\top = A$. This observation indicates that the intrinsic parameter subspace that we are looking for is not $\mathbb{M}_{n,r}$ but the orbit space $\mathbb{M}_{n,r}/\mathbb{O}(r)$ of the Lie group action of $\mathbb{O}(r)$ on the set of rectangular matrices $\mathbb{M}_{n,r}$ via:

$$\begin{array}{rccc} \Phi: & \mathbb{O}(r) \times \mathbb{M}_{n,r} & \longrightarrow & \mathbb{M}_{n,r} \\ & (O, A) & \longmapsto & AO^{-1}. \end{array} \tag{3.11}$$

The next proposition provides a convenient model subspace for the quotient $\mathbb{M}_{n,r}/\mathbb{O}(r)$ and, as a corollary, a characterization of the intrinsic parameter subspace of the rank deficient DCC model. A proof of this result can be found in the Technical Appendix A.2.

Proposition 3.1 *In the setup that we just described, consider the cone $\mathbb{L}^+_{n,m}$ of lower triangular matrices with positive elements in the main diagonal. Then, the map*

$$\begin{array}{rccc} \Psi: & \mathbb{L}^+_{n,m} & \longrightarrow & \mathbb{M}_{n,r}/\mathbb{O}(r) \\ & A & \longmapsto & [A], \end{array} \tag{3.12}$$

is a bijection. The symbol $[A]$ in (3.12) denotes the orbit in $\mathbb{M}_{n,r}/\mathbb{O}(r)$ corresponding to the element $A \in \mathbb{M}_{n,r}$ with respect to the action (3.11).

This proposition implies that the intrinsic parameter subspace of the rank deficient family that we are looking for is $\mathbb{L}^+_{n,m}$. This can be described as $\mathbb{R}^{N^*}$, $N^* = nr - \frac{1}{2}r(r-1)$, via the operator $\text{mat}_r : \mathbb{R}^{N^*} \longrightarrow \mathbb{L}_{n,r}$ introduced in Subsection 2.2 just by adding a positivity constraint on the entries of the vector in $\mathbb{R}^{N^*}$ that constitute the main diagonal of the corresponding matrix via the mat_r operator. More explicitly, let $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^{N^*}$ and define $\widetilde{A} := \text{mat}_r(\boldsymbol{a})$ and $\widetilde{B} := \text{mat}_r(\boldsymbol{b})$. The matrices A and B are hence given by $A := \widetilde{A}\widetilde{A}^\top$, $B := \widetilde{B}\widetilde{B}^\top$. Using the notation that we just adopted, the rank deficient DCC model specification can be written as

$$Q_t = (\mathbf{i}_n\mathbf{i}_n^\top - \widetilde{A}\widetilde{A}^\top - \widetilde{B}\widetilde{B}^\top) \odot S + (\widetilde{A}\widetilde{A}^\top) \odot (\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}^\top) + (\widetilde{B}\widetilde{B}^\top) \odot Q_{t-1}. \tag{3.13}$$

In order to ensure that $\widetilde{A}, \widetilde{B} \in \mathbb{L}^+_{n,r}$ and that, by Proposition 3.1, the model is well identified we need to impose the following constraints:

(IC) Identification constraints: let $\{i_1, \ldots, i_r\}$ with $i_j = n(j-1) + \frac{1}{2}j(3-j)$, $j \in \{1, \ldots, r\}$, be the entries spelled out in (2.12) of any vector in $\mathbb{R}^{N^*}$ that amount to the main diagonal of the corresponding matrix in $\mathbb{L}_{n,r}$ via the mat_r representation. Then

$$a_{i_j} > 0, \quad b_{i_j} > 0, \quad j \in \{1, \ldots, r\}. \tag{3.14}$$

Additionally, the stationarity and positivity constraints are expressed as:

(SC) Stationarity constraints:

$$\sum_{k=1}^{r} \left| \widetilde{A}_{ik}\widetilde{A}_{jk} + \widetilde{B}_{ik}\widetilde{B}_{jk} \right| < 1, \;\; i, j \in \{1 \ldots, n\}, \;\; i \geq j. \tag{3.15}$$

(PSD) Positivity constraints:

$$(\mathbf{i}_n\mathbf{i}_n^\top - \widetilde{A}\widetilde{A}^\top - \widetilde{B}\widetilde{B}^\top) \odot S \succ 0, \quad Q_0 \succeq 0. \tag{3.16}$$

3.3 The Almon DCC model

The Almon DCC model specification is a particular case of the rank deficient DCC model with $r = 1$, where the vectors that generate the matrices $A, B \in \mathbb{S}_n$ in (3.5) are parametrized using the Almon lag function [Almo 65] alm_n introduced in Subsection 2.2. More explicitly, given $n \in \mathbb{N}$ and $\mathbf{v} \in \mathbb{R}^3$, the Almon lag operator $\mathrm{alm}_n : \mathbb{R}^3 \longrightarrow \mathbb{R}^n$ can be used to generate a vector $\mathrm{alm}_n(\mathbf{v})$ whose entries are $(\mathrm{alm}_n(\mathbf{v}))_i = v_1 + \exp(v_2 i + v_3 i^2), i \in \{1, \ldots, n\}$.

Let $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^3$ 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 $A, B \in \mathbb{S}_n$ can be written as $A := \widetilde{\boldsymbol{a}}\widetilde{\boldsymbol{a}}^\top$, $B := \widetilde{\boldsymbol{b}}\widetilde{\boldsymbol{b}}^\top$, and hence the Almon DCC model specification is given by

$$Q_t = (\mathbf{i}_n\mathbf{i}_n^\top - \widetilde{\boldsymbol{a}}\widetilde{\boldsymbol{a}}^\top - \widetilde{\boldsymbol{b}}\widetilde{\boldsymbol{b}}^\top) \odot S + (\widetilde{\boldsymbol{a}}\widetilde{\boldsymbol{a}}^\top) \odot (\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}^\top) + (\widetilde{\boldsymbol{b}}\widetilde{\boldsymbol{b}}^\top) \odot Q_{t-1}. \tag{3.17}$$

and the associated constraints by

(IC) Identification constraints:

$$\widetilde{a}_1 > 0, \ \ \widetilde{b}_1 > 0, \quad \text{i.e. } \ a_1 + \exp(a_2 + a_3) > 0, \ \ b_1 + \exp(b_2 + b_3) > 0. \tag{3.18}$$

(SC) Stationarity constraints:

$$\left|\widetilde{a}_i\widetilde{a}_j + \widetilde{b}_i\widetilde{b}_j\right| < 1, \ \ i, j \in \{1, \ldots, n\}, \ \ i \geq j. \tag{3.19}$$

(PSD) Positivity constraints:

$$(\mathbf{i}_n\mathbf{i}_n^\top - \widetilde{\boldsymbol{a}}\widetilde{\boldsymbol{a}}^\top - \widetilde{\boldsymbol{b}}\widetilde{\boldsymbol{b}}^\top) \odot S \succ 0, \quad Q_0 \succeq 0. \tag{3.20}$$

3.4 The Almon shuffle DCC model

The Almon shuffle DCC specification is a variant of the Almon DCC 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 DCC 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 reordering (shuffle) that we propose 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 DCC model of the previous subsection is used. The results in Section 5 show that the Almon shuffle DCC model often exhibits a better performance than the Almon model.

3.5 The scalar DCC model

The scalar DCC model is by far the most widely used in the literature. In this case, the parameter matrices $A, B \in \mathbb{S}_n$ in (3.5) 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 scalar DCC model specification and the associated constraints can be written as:

$$Q_t = (1 - a - b)\, S + a\ \boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}^\top + b\ Q_{t-1}. \tag{3.21}$$

(SC) Stationarity constraint:

$$a + b < 1. \tag{3.22}$$

(PSD) Positivity constraints:

$$a \geq 0, \quad b \geq 0, \quad S \succeq 0, \quad Q_0 \succeq 0. \tag{3.23}$$

4 Constrained estimation of DCC model parameters

In this section, we present a constrained optimization scheme adapted to the quasi-maximum likelihood (QML) estimation of the DCC models previously described, and we provide all the necessary details to implement it. The QML approach stems from the assumptions that the innovations $\boldsymbol{\xi}_t$ appearing in (3.1) are normally distributed and that the conditional variance and correlation of the data generating process behave according to the Hadamard DCC equation.

4.1 The log-likelihood function and its gradient

In this subsection we provide the log-likelihood function associated to the DCC model (3.1) with correlation dynamics determined by (3.5). We compute the gradient of the log-likelihood function related to each of the DCC model specifications discussed in Section 3. The issues posed by the need to handle the parameter constraints are addressed in detail in Subsection 4.2.

The log-likelihood function. 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, keeping in mind that this is performed after substituting the targeting estimator defined in (3.6) for the parameter S (in the sequel the symbol S stands for the targeting estimator instead of the unknown value). The log-likelihood function associated to the process (3.1) is

$$\log L\left(\boldsymbol{\Theta}; \mathbf{r}\right) = \sum_{t=1}^{T} l_t\left(\boldsymbol{\Theta}; \mathbf{r}_t\right), \tag{4.1}$$

where

$$l_t\left(\boldsymbol{\Theta}; \mathbf{r}_t\right) = -\frac{1}{2}\left(n \log(2\pi) + \log\det\left(H_t\right) + \mathbf{r}_t^\top H_t^{-1} \mathbf{r}_t\right). \tag{4.2}$$

The dependence of l_t on $\boldsymbol{\Theta}$ is materialized through H_t, that explicitly relies on the set of parameters $\boldsymbol{\Theta}$ through Q_t since H_t is equal to $D_t Q_t^{*-1/2} Q_t Q_t^{*-1/2} D_t$. We recall that D_t depends on the parameters of the conditional volatilities (3.2) specified in the first stage of the DCC model construction. These parameters can be estimated consistently in a first stage as explained in [Engl 02] and we hence proceed in this section by assuming that the symbol D_t in (4.2) and all the expressions derived from it stand for the corresponding estimated version. After this substitution, the expression (4.1) is strictly equivalent to the second stage log-likelihood function defined by [Engl 02].

Consequently, the DCC QML (second stage) estimation problem consists of finding the parameter value $\widehat{\boldsymbol{\Theta}}$ that maximizes the log-likelihood (4.1) associated to a particular DCC model subjected to the constraints **(IC)**, **(SC)**, and **(PSD)** associated to it. As we explain in the next subsection, we carry this out using an iterative optimization method proposed in [Chre 13], which we adapt to the different versions of the DCC model that we have defined in the previous section. For each of these versions, the matrices $A, B \in \mathbb{S}_n$ are functions of a parameter vector $\boldsymbol{\theta}$ that belongs to an intrinsic parameter space $\mathcal{P} \times \mathcal{P}$ that is specific to each model, and for which $\boldsymbol{\Theta} = \boldsymbol{\Theta}(\boldsymbol{\theta})$.

Gradient of the log-likelihood function. The computation of the estimator $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ via the constrained maximum likelihood optimization that we use requires the gradient $\nabla_{\boldsymbol{\theta}} \log L(\boldsymbol{\Theta}; \mathbf{r})$. We provide analytical expressions of the gradient in the following proposition, where we use several operators and vectors defined in Section 2; the proof is available in the Technical Appendix A.3.

Proposition 4.1 *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* (4.1)-(4.2). *Then,*

$$\nabla_{\boldsymbol{\theta}} \log L(\boldsymbol{\theta}; \mathbf{r}) = \sum_{t=1}^{T} \nabla_{\boldsymbol{\theta}} l_t(\boldsymbol{\theta}; \mathbf{r}_t) = \sum_{t=1}^{T} T^*_{\boldsymbol{\theta}} \boldsymbol{\Theta} \cdot T^*_{\boldsymbol{\Theta}} H_t \cdot \nabla_{H_t} l_t(\boldsymbol{\theta}; \mathbf{r}_t), \tag{4.3}$$

with

$$\nabla_{H_t} l_t(\boldsymbol{\theta}; \mathbf{r}_t) = -\frac{1}{2}\left[H_t^{-1} - H_t^{-1}\mathbf{r}_t\mathbf{r}_t^\top H_t^{-1}\right]. \tag{4.4}$$

In the relation (4.3), *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* Θ *(that is* $A(\boldsymbol{\theta}_1)$ *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 expression:*

$$T^*_{\Theta} H_t \cdot \Delta = T^*_{\Theta} Q_t \left[Q_t^{*-1} D_t \Delta D_t Q_t^{*-1} - \frac{1}{2}\mathrm{Diag}(Q_t^{*-2}(D_t \Delta D_t Q_t^{*-1} Q_t + Q_t Q_t^{*-1} D_t \Delta D_t) Q_t^{*-1}) \right]. \tag{4.5}$$

Additionally, the differential operator $T^*_{\boldsymbol{\Theta}} Q_t : \mathbb{S}_n \times \mathbb{S}_n \longrightarrow \mathbb{S}_n \times \mathbb{S}_n$ *is the adjoint of the map* $T_{\boldsymbol{\Theta}} Q_t : \mathbb{S}_n \times \mathbb{S}_n \longrightarrow \mathbb{S}_n \times \mathbb{S}_n$; *for each component* $A(\boldsymbol{\theta}_1)$ *and* $B(\boldsymbol{\theta}_2)$ *of* $\boldsymbol{\Theta}$ *and for any* $\Delta \in \mathbb{S}_n$, $T^*_{\boldsymbol{\Theta}} Q_t$ *is determined by the recursions:*

$$T^*_A Q_t \cdot \Delta = \Delta \odot \left(\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}^\top - S\right) + T^*_A Q_{t-1}[\Delta \odot B], \tag{4.6}$$

$$T^*_B Q_t \cdot \Delta = \Delta \odot (Q_{t-1} - S) + T^*_B Q_{t-1}[\Delta \odot B], \tag{4.7}$$

that are initialized by setting $T^*_A Q_0 = 0$ *and* $T^*_B Q_0 = 0$.
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 3. For a given pair* $\Delta_1, \Delta_2 \in \mathbb{S}_n$ *these maps are determined by the following expressions:*

(i) ***The Hadamard DCC 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,*

$$\begin{array}{rccc} T^*_{\boldsymbol{\theta}} \boldsymbol{\Theta} : & \mathbb{S}_n \times \mathbb{S}_n & \longrightarrow & \mathbb{R}^N \times \mathbb{R}^N \\ & (\Delta_1, \Delta_2) & \longmapsto & (\mathrm{math}^*(\Delta_1), \mathrm{math}^*(\Delta_2)). \end{array} \tag{4.8}$$

(ii) ***The rank deficient DCC 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,

$$\begin{array}{rccc} T^*_{\boldsymbol{\theta}} \boldsymbol{\Theta} : & \mathbb{S}_n \times \mathbb{S}_n & \longrightarrow & \mathbb{R}^{N^*} \times \mathbb{R}^{N^*} \\ & (\Delta_1, \Delta_2) & \longmapsto & 2\,(\mathrm{vec}_r(\Delta_1 \mathrm{mat}_r(\boldsymbol{a})), \mathrm{vec}_r(\Delta_2 \mathrm{mat}_r(\boldsymbol{b}))). \end{array} \tag{4.9}$$

(iii) ***The Almon DCC 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,*

$$\begin{array}{rccc} T^*_{\boldsymbol{\theta}}\boldsymbol{\Theta}: & \mathbb{S}_n \times \mathbb{S}_n & \longrightarrow & \mathbb{R}^3 \times \mathbb{R}^3 \\ & (\Delta_1, \Delta_2) & \longmapsto & 2\,(K_{\boldsymbol{\theta}_1}^\top \Delta_1 \mathrm{alm}_n(\boldsymbol{\theta}_1), K_{\boldsymbol{\theta}_2}^\top \Delta_2 \mathrm{alm}_n(\boldsymbol{\theta}_2)), \end{array} \tag{4.10}$$

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 DCC 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,*

$$\begin{array}{rccc} T^*_{\boldsymbol{\theta}}\boldsymbol{\Theta}: & \mathbb{S}_n \times \mathbb{S}_n & \longrightarrow & \mathbb{R} \times \mathbb{R} \\ & (\Delta_1, \Delta_2) & \longmapsto & \left(\langle \Delta_1, \mathbf{i}_n\mathbf{i}_n^\top \rangle, \langle \Delta_2, \mathbf{i}_n\mathbf{i}_n^\top \rangle\right). \end{array} \tag{4.11}$$

In order to algorithmically implement Proposition 4.1, the operator recursions (4.5)-(4.10) have to be materialized in matrix recursions. We provide these results in the Appendix.

4.2 Constrained optimization using Bregman divergences

We present in this subsection the constrained optimization method that we advocate to compute the QML estimator of the DCC model parameters. Handling the linear and nonlinear parameter constraints associated to each of the specifications that we are considering complicates the optimization problem. Standard techniques, like Lagrange duality, often require to solve additional secondary optimization problems. We circumvent this difficulty by implementing a penalized optimization scheme that uses the so-called Bregman matrix divergences [Breg 67] and that provides a solution using the primal space as working setup. This technique has been introduced in the context of machine learning (see for instance [Dhil 07, Kuli 09b] and references therein) and has has shown a good performance in the estimation of the heavily parametrized VEC-GARCH model [Chre 13].

4.2.1 Bregman divergences and constrained optimization problems

Bregman matrix nearness measures and divergences. We start by recalling the definition of Bregman matrix nearness measures.

Definition 4.2 *Let* $X, Y \in \mathbb{S}_n$ *and* $\phi : \mathbb{S}_n \longrightarrow \mathbb{R}$ *a strictly convex and differentiable function. The* **Bregman matrix nearness measure** *associated to* ϕ *is defined by*

$$D_\phi(X, Y) := \phi(X) - \phi(Y) - \mathrm{tr}\left((\nabla\phi(Y))^\top (X - Y)\right).$$

Different choices of ϕ lead to the following examples of this measure:

(i) The squared Frobenius distance: if the function ϕ is the squared Frobenius norm, that is, $\phi(X) := \|X\|^2 = \langle X, X \rangle$ then $D_\phi(X, Y) := \|X - Y\|^2$.

(ii) The von Neumann divergence is the Bregman divergence associated to the entropy of the eigenvalues of a positive definite matrix. more explicitly, if X is a positive definite matrix with eigenvalues $\{\lambda_1, \ldots, \lambda_n\}$, then $\phi(X) := \sum_{i=1}^n (\lambda_i \log \lambda_i - \lambda_i)$ and the resulting divergence is

$$D_{vN}(X, Y) := \mathrm{tr}(X \log X - X \log Y - X + Y),$$

where $\log X$ and $\log Y$ are the matrix logarithms.

(iii) The Burg matrix divergence: it is also referred to in the literature as **LogDet divergence** or **Stein's loss** (mainly in the statistics literature [Jame 61]). This divergence is obtained out of the Burg entropy of the eigenvalues of a positive definite matrix, that is $\phi(X) := -\sum_{i=1}^{n} \log \lambda_i$, or equivalently $\phi(X) := -\log\det(X)$. We define the associated Bregman divergence over the positive definite matrices $X, Y \in \mathbb{S}_n^+$ by

$$D_M(X,Y) := \mathrm{tr}(XY^{-1}) - \log\det(XY^{-1}) - n. \tag{4.12}$$

The last two examples are particular cases of **spectral Bregman matrix nearness measures**, that is, the function ϕ that defines them can be written down as the composition $\phi(X) = (\varphi \circ \lambda)(X)$, where $\varphi : \mathbb{R}^n \longrightarrow \mathbb{R}$ is a differentiable strictly convex function and $\lambda : \mathbb{S}_n \longrightarrow \mathbb{R}^n$ is the map that provides the eigenvalues of the matrix in its argument.

In what follows we use the denomination **Bregman matrix divergence** when we refer to the von Neumann or the Burg divergences. In practice, we use the Burg divergence.

The use of Bregman matrix nearness measures in constrained optimization problems. Bregman divergences are particularly appropriate in optimization at the time of dealing with positive (semi)definiteness constraints. Suppose that we want to minimize a function $f(A)$ defined on a space of square matrices and subjected to $A \succ 0$. We proceed by iteratively solving the optimization problems associated to penalized local models of the form

$$f_A^{(k)}(A) := f(A^{(k)}) + \nabla f(A^{(k)})(A - A^{(k)}) + \frac{1}{2}(A - A^{(k)})^T H(A^{(k)})(A - A^{(k)}) + D_\phi(A, A^{(k)}), \tag{4.13}$$

where $k \in \mathbb{N}$ is the label for the iteration and $H(A^{(k)})$ is the Hessian of the function f computed at the point $A^{(k)}$. If we use as penalization term $D_\phi(A, A^{(k)})$ in (4.13) a Bregman divergence instead of the usual Frobenius distance, it diverges when A approaches the set where the constraints are violated and forces the solution of the penalized local model to automatically lay in the constrained set. Proceeding this way we reduce the constrained optimization problem to a sequence of local unconstrained ones.

4.2.2 Bregman divergences for positivity constraints

We write down the Bregman divergences associated to the constraints that we need to impose on the different DCC models under consideration in terms of the variables $\boldsymbol{\theta}$ of their intrinsic parameter space $\mathcal{P} \times \mathcal{P}$. We recall that $\theta \in \mathcal{P}$ denotes a generic P-dimensional component of $\boldsymbol{\theta} \in \mathcal{P} \times \mathcal{P}$. In view of the different functional character of the constraints considered, we classify them into three groups and provide the corresponding expressions for the divergences.

Positive semidefinite (definite) constraints (PSDC): this group of constraints can be generically written as $M(\boldsymbol{\theta}) \succeq 0$, where $M : \mathcal{P} \times \mathcal{P} \longrightarrow \mathbb{S}_q$, $q \in \mathbb{N}$, is a smooth map. In this case, the Bregman matrix divergence $D_M(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) \in \mathbb{R}$ is given by

$$D_M(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = \mathrm{tr}\left(M(\boldsymbol{\theta}) \cdot M(\boldsymbol{\theta}^{(k)})^{-1}\right) - \log\,\det\left(M(\boldsymbol{\theta}) \cdot M(\boldsymbol{\theta}^{(k)})^{-1}\right) - q. \tag{4.14}$$

Nonlinear positivity constraints (NPC): they are specified by the relation $N(\boldsymbol{\theta}) > \mathbf{0}_q$, where $N : \mathcal{P} \times \mathcal{P} \longrightarrow \mathbb{R}^q$ is a differentiable map and q is the number of components of the constraint. The corresponding entries of the divergence $D_N(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) \in \mathbb{R}^q$ are determined by the relation

$$D_N^i(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = (D_N(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})_i = \frac{(N(\boldsymbol{\theta}))_i}{(N(\boldsymbol{\theta}^{(k)}))_i} - \log\frac{(N(\boldsymbol{\theta}))_i}{(N(\boldsymbol{\theta}^{(k)}))_i} - 1, \quad i = \{1, \ldots, q\}. \tag{4.15}$$

Linear positivity constraints (LPC): they are a particular case of the previous ones, where N is a linear map. More specifically, in the case when $\boldsymbol{\theta} \in \mathcal{P} \times \mathcal{P}$, we consider linear constraints of the form $L(\boldsymbol{\theta}) := \mathbf{f} - \sum_{i=1}^{2} C_{\theta_i}\theta_i > \mathbf{0}_m$, with $C_{\theta_i} \in \mathbb{M}_{m,P}$, $\theta_i \in \mathcal{P}$, $\mathbf{f} \in \mathbb{R}^m$, and $m \leq P \in \mathbb{N}$. The divergence $D_L(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) \in \mathbb{R}^m$ associated to the condition $L(\boldsymbol{\theta}) > \mathbf{0}_m$ has m components given by

$$D_L^i(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = (D_L(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}))_i = \frac{(L(\boldsymbol{\theta}))_i}{(L(\boldsymbol{\theta}^{(k)}))_i} - \log\frac{(L(\boldsymbol{\theta}))_i}{(L(\boldsymbol{\theta}^{(k)}))_i} - 1, \quad i \in \{1, \ldots, m\}. \tag{4.16}$$

4.2.3 The local model, its gradient and Jacobian

As we already explained, the algorithm is based on the optimization of a sequence of penalized local functions that incorporate the Bregman divergences in order to ensure that the constraints are satisfied at each iteration. More specifically, the solution $\boldsymbol{\theta}^{(k+1)}$ of the local optimization problem after k iterations is defined by

$$\boldsymbol{\theta}^{(k+1)} = \underset{\boldsymbol{\theta} \in \mathcal{P}\times\mathcal{P}}{\arg\min}\ \tilde{f}^{(k)}(\boldsymbol{\theta}), \tag{4.17}$$

where the local objective function $\tilde{f}^{(k)}$ at $\boldsymbol{\theta}^{(k)}$ is given by

$$\begin{aligned}\tilde{f}^{(k)}(\boldsymbol{\theta}) \;:=\; & f(\boldsymbol{\theta}^{(k)}) + \nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta}^{(k)})(\boldsymbol{\theta} - \boldsymbol{\theta}^{(k)}) + \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}^{(k)})^\top H^{(k)}(\boldsymbol{\theta} - \boldsymbol{\theta}^{(k)}) \\ & + \sum_{j=1}^{s_1} L_1^j D_{M_j}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) + \sum_{j=1}^{s_2} L_2^j \mathbf{i}_{q_j}^\top D_{N_j}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) + \sum_{j=1}^{s_3} L_3^j \mathbf{i}_{m_j}^\top D_{L_j}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}).\end{aligned} \tag{4.18}$$

In the above 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 which is determined, in the DCC case, by the relation (4.3) in Proposition 4.1. The symbol $H^{(k)}$ denotes the Hessian of the function f 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 defined in (4.14), (4.15), and (4.16), respectively; $\mathbf{L}_1 \in \mathbb{R}^{s_1}, \mathbf{L}_2 \in \mathbb{R}^{s_2}, \mathbf{L}_3 \in \mathbb{R}^{s_3}$ are vectors whose components control the strength of the 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 (4.17) is solved by finding the value $\boldsymbol{\theta}_0$ for which

$$\nabla_{\boldsymbol{\theta}} \tilde{f}^{(k)}(\boldsymbol{\theta}_0) = 0, \tag{4.19}$$

where the gradient $\nabla_{\boldsymbol{\theta}} \tilde{f}^{(k)}(\boldsymbol{\theta})$ of the local model (4.18) is given by

$$\begin{aligned}\nabla_{\boldsymbol{\theta}} \tilde{f}^{(k)}(\boldsymbol{\theta}) \;=\; & \nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta}^{(k)}) + H^{(k)}(\boldsymbol{\theta} - \boldsymbol{\theta}^{(k)}) \\ & + \sum_{j=1}^{s_1} L_1^j \nabla_{\boldsymbol{\theta}} D_{M_j}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) + \sum_{j=1}^{s_2} L_2^j \sum_{i=1}^{q_j} \nabla_{\boldsymbol{\theta}} D_{N_j}^i(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) \\ & + \sum_{j=1}^{s_3} L_3^j \sum_{i=1}^{m_j} \nabla_{\boldsymbol{\theta}} D_{L_j}^i(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}).\end{aligned} \tag{4.20}$$

The gradients of the divergences in (4.20) are given by

$$\nabla_\theta D_{M_j}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = -T_\theta^* M_j (M_j(\boldsymbol{\theta})^{-1} - M_j(\boldsymbol{\theta}^{(k)})^{-1}), \tag{4.21}$$

$$\nabla_\theta D^i_{N_j}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = -\left[\frac{1}{(N_j(\boldsymbol{\theta}))_i} - \frac{1}{(N_j(\boldsymbol{\theta}^{(k)}))_i}\right](T_\theta^* N_j \cdot \mathbf{e}^{(i)}_{q_j}),\ \ i \in \{1, \dots, q_j\}, \tag{4.22}$$

$$\nabla_\theta D^i_{L_j}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = \left[\frac{1}{(L_j(\boldsymbol{\theta}))_i} - \frac{1}{(L_j(\boldsymbol{\theta}^{(k)}))_i}\right](C_\theta^{(j)T} \cdot \mathbf{e}^{(i)}_{m_j}),\ \ i \in \{1, \dots, m_j\}, \tag{4.23}$$

where $\theta \in \mathcal{P}$ is a generic component of the intrinsic parameter $\boldsymbol{\theta} \in \mathcal{P} \times \mathcal{P}$, $C_\theta^{(j)} \in \mathbb{M}_{m_j,S}$, $j \in \{1, \dots, s_3\}$ is the matrix associated to the j-th linear constraint, and $\mathbf{e}^{(i)}_{q_j} \in \mathbb{R}^{q_j}$, $i \in \{1, \dots, q_j\}$ and $\mathbf{e}^{(i)}_{m_j} \in \mathbb{R}^{m_j}$, $i \in \{1, \dots, m_j\}$ are the canonical unit vectors introduced in Subsection 2.1. Additionally, for every $j \in \{1, \dots, s_1\}$, $T^*_{\boldsymbol{\theta}} M_j : \mathbb{S}_q \longrightarrow \mathcal{P} \times \mathcal{P}$ is the adjoint of the tangent map $T_{\boldsymbol{\theta}} M_j : \mathcal{P} \times \mathcal{P} \longrightarrow \mathbb{S}_q$ of $M_j(\boldsymbol{\theta}) : \mathcal{P} \times \mathcal{P} \longrightarrow \mathbb{S}_q$ that determines the j-th positive semidefinite constraint. The symbol $T^*_\theta M_j : \mathbb{S}_q \longrightarrow \mathcal{P}$ in the relation (4.21) denotes the adjoint of the partial derivative $T_\theta M_j : \mathcal{P} \longrightarrow \mathbb{S}_q$ for some component θ of $\boldsymbol{\theta}$. Analogously, for every $j \in \{1, \dots, s_2\}$, the map $T^*_{\boldsymbol{\theta}} N_j : \mathbb{R}^{q_j} \times \mathbb{R}^{q_j} \longrightarrow \mathcal{P} \times \mathcal{P}$ is the adjoint of the tangent map $T_{\boldsymbol{\theta}} N_j : \mathcal{P} \times \mathcal{P} \longrightarrow \mathbb{R}^{q_j} \times \mathbb{R}^{q_j}$ of the function $N_j(\boldsymbol{\theta}) : \mathcal{P} \times \mathcal{P} \longrightarrow \mathbb{R}^{q_j}$ that determines the j-th nonlinear constraint. Again $T^*_\theta N_j$ denotes the adjoint of the corresponding θ partial derivative.

In order to numerically obtain the solution $\boldsymbol{\theta}_0$ of (4.19) using the Newton-Raphson algorithm, we need the Jacobian of the local model and in particular the tangent maps of all the gradients of the divergences in (4.21)-(4.23). Straightforward computations yield for every component $\theta \in \mathcal{P}$ of $\boldsymbol{\theta}$ and any $\delta\theta \in \mathcal{P}$:

$$\begin{aligned}
T_\theta(\nabla_\theta D_{M_j}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})) \cdot \delta\theta &= T^*_\theta M_j (M_j(\boldsymbol{\theta})^{-1} (T_\theta M_j \cdot \delta\theta) M_j(\boldsymbol{\theta})^{-1})) \\
&\quad - (T_\theta(T^*_\theta M_j))\,(M_j(\boldsymbol{\theta})^{-1} - M_j(\boldsymbol{\theta}^{(k)})^{-1}) \cdot \delta\theta,\ \ j \in \{1, \dots, s_1\}, \\
T_\theta(\nabla_\theta D^i_{N_j}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})) \cdot \delta\theta &= \frac{(T_\theta N_j(\boldsymbol{\theta}))_i \cdot \delta\theta}{(N_j(\boldsymbol{\theta}))_i^2}(T^*_\theta N_j \cdot \mathbf{e}^{(i)}_{q_j}) - \left[\frac{1}{(N_j(\boldsymbol{\theta}))_i} - \frac{1}{(N_j(\boldsymbol{\theta}^{(k)}))_i}\right] T_\theta(T^*_\theta N_j \cdot \mathbf{e}^{(i)}_{q_j}) \cdot \delta\theta, \\
&\qquad\qquad i \in \{1, \dots, q_j\},\ \ j \in \{1, \dots, s_2\}, \\
T_\theta(\nabla_\theta D^i_{L_j}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) \cdot \delta\theta &= \frac{(C_\theta^{(j)} \cdot \delta\theta)_i}{(L_j(\boldsymbol{\theta}))_i^2}\,(C_\theta^{(j)T} \cdot \mathbf{e}^{(i)}_{m_j}), \quad i \in \{1, \dots, m_j\},\ \ j \in \{1, \dots, s_3\}.
\end{aligned}$$

Using these identities it is easy to determine the Jacobian of the local model $\tilde{f}^{(k)}(\boldsymbol{\theta})$. Indeed, the expression for the block of the Jacobian that corresponds to the component $\theta \in \mathcal{P}$ is:

$$\begin{aligned}
T_\theta(\nabla_\theta \tilde{f}^{(k)}(\boldsymbol{\theta})) \cdot \delta\theta &= H^{(k)} \cdot \delta\theta + \sum_{j=1}^{s_1} L_1^j \Big[T^*_\theta M_j (M_j(\boldsymbol{\theta})^{-1} (T_\theta M_j \cdot \delta\theta) M_j(\boldsymbol{\theta})^{-1})) \\
&\quad - (T_\theta(T^*_\theta M_j))(M_j(\boldsymbol{\theta})^{-1} - M_j(\boldsymbol{\theta}^{(k)})^{-1}) \cdot \delta\theta\Big] + \sum_{j=1}^{s_2} L_2^j \sum_{i=1}^{q_j} \Big[\frac{(T_\theta N_j(\boldsymbol{\theta}))_i \cdot \delta\theta}{(N_j(\boldsymbol{\theta}))_i^2}(T^*_\theta N_j \cdot \mathbf{e}^{(i)}_{q_j}) \\
&\quad - \left[\frac{1}{(N_j(\boldsymbol{\theta}))_i} - \frac{1}{(N_j(\boldsymbol{\theta}^{(k)}))_i}\right] T_\theta(T^*_\theta N_j \cdot \mathbf{e}^{(i)}_{q_j}) \cdot \delta\theta\Big] + \sum_{j=1}^{s_3} L_3^j \sum_{i=1}^{m_j} \frac{(C_\theta^{(j)} \cdot \delta\theta)_i}{(L_j(\boldsymbol{\theta}))_i^2}\,(C_\theta^{(j)\top} \cdot \mathbf{e}^{(i)}_{m_j}).
\end{aligned} \tag{4.24}$$

4.3 Implementation for the Hadamard DCC model

In this subsection, we provide explicit expressions for the Bregman divergences, the local penalized model, its gradient and Jacobian, for the Hadamard DCC model that we presented in Section 3. Similar

results for the other DCC models are reported in Technical Appendix B. The lengthy computations that lead to these expressions are contained in the Technical Appendices C.1-C.3.

Recall that in this particular case, the matrices $A, B \in \mathbb{S}_n$ are parametrized with two vectors $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^N$, $N = \frac{1}{2}n\,(n+1)$, by setting $A := \mathrm{math}\,(\boldsymbol{a})$, $B := \mathrm{math}\,(\boldsymbol{b})$. Let $\boldsymbol{\theta} := (\boldsymbol{a}, \boldsymbol{b}) \in \mathbb{R}^N \times \mathbb{R}^N$; the intrinsic parameter subspace $\mathcal{P}$ is in this case $\mathbb{R}^N$ and has dimension $P = N$. The dynamics of the conditional correlation matrix for the Hadamard DCC model is given by (3.7) and $\boldsymbol{\theta}$ is subjected to the stationarity (3.8) and positivity constraints (3.9)-(3.10).

Constraints and the local model. The stationarity and positivity constraints imposed on the Hadamard model specification can be written down according to the classification introduced in Subsection 4.2.2 as the following set of two linear positivity and three positive semidefiniteness and definiteness constraints:

$$L_1\,(\boldsymbol{\theta}) := \mathbf{i}_N - C_{\boldsymbol{a}}^{(1)}\boldsymbol{a} - C_{\boldsymbol{b}}^{(1)}\boldsymbol{b} > \mathbf{0}_N, \tag{4.25}$$

$$L_2\,(\boldsymbol{\theta}) := \mathbf{i}_N - C_{\boldsymbol{a}}^{(2)}\boldsymbol{a} - C_{\boldsymbol{b}}^{(2)}\boldsymbol{b} > \mathbf{0}_N, \tag{4.26}$$

$$M_1(\boldsymbol{a}) := \mathrm{math}(\boldsymbol{a}) \succeq 0, \tag{4.27}$$

$$M_2(\boldsymbol{b}) := \mathrm{math}(\boldsymbol{b}) \succeq 0, \tag{4.28}$$

$$M_3\,(\boldsymbol{\theta}) := \left(\mathbf{i}_n\mathbf{i}_n^\top - \mathrm{math}(\boldsymbol{a}) - \mathrm{math}(\boldsymbol{b})\right) \odot S \succ 0, \tag{4.29}$$

where $L_1(\boldsymbol{\theta}), L_2(\boldsymbol{\theta}) \in \mathbb{R}^N$, $M_1(\boldsymbol{a}), M_2(\boldsymbol{b}), M_3\,(\boldsymbol{\theta}) \in \mathbb{S}_n$, $C_{\boldsymbol{a}}^{(1)} = C_{\boldsymbol{b}}^{(1)} = \mathbf{I}_N$, $C_{\boldsymbol{a}}^{(2)} = C_{\boldsymbol{b}}^{(2)} = -\mathbf{I}_N$ with $\mathbf{I}_N$ the identity matrix. The local model corresponding to (4.17) is given in this case by

$$\begin{aligned}\tilde{f}^{(k)}\,(\boldsymbol{\theta}) =& f(\boldsymbol{\theta}^{(k)}) + \nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta}^{(k)})(\boldsymbol{\theta} - \boldsymbol{\theta}^{(k)}) + \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}^{(k)})^\top H^{(k)}(\boldsymbol{\theta} - \boldsymbol{\theta}^{(k)}) \\ &+ \sum_{j=1}^{3} L_1^j D_{M_j}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) + \sum_{j=1}^{2} L_2^j \mathbf{i}_N^\top D_{L_j}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}),\end{aligned} \tag{4.30}$$

where $\mathbf{L}_1 := \left(L_1^1, L_1^2, L_1^3\right)^\top$ and $\mathbf{L}_2 := \left(L_2^1, L_2^2\right)^\top$ are the vectors that contain the penalization strengths, $f(\boldsymbol{\theta}^{(k)})$ is minus the log-likelihood function (4.1), $\nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta}^{(k)})$ is its gradient, that is $\nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta}^{(k)}) = -\nabla_{\boldsymbol{\theta}} \mathrm{log} L(\boldsymbol{\theta}^{(k)}; \mathbf{r})$, which for the Hadamard model is determined by relations (4.3)-(4.8) in Proposition 4.1, and $H^{(k)}$ is its Hessian computed at the point $\boldsymbol{\theta}^{(k)}$. In relation (4.30) the Bregman divergences associated to each of the constraints (4.25)-(4.29) are easily obtained from (4.16) and (4.14), which yields the following expressions:

$$D_{L_1}^i(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = \frac{1 - a_i - b_i}{1 - a_i^{(k)} - b_i^{(k)}} - \log\frac{1 - a_i - b_i}{1 - a_i^{(k)} - b_i^{(k)}} - 1, \;\; i \in \{1, \ldots, N\}\,, \tag{4.31}$$

$$D_{L_2}^i(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = \frac{1 + a_i + b_i}{1 + a_i^{(k)} + b_i^{(k)}} - \log\frac{1 + a_i + b_i}{1 + a_i^{(k)} + b_i^{(k)}} - 1, \;\; i \in \{1, \ldots, N\}\,, \tag{4.32}$$

$$D_{M_1}(\boldsymbol{a}, \boldsymbol{a}^{(k)}) = \mathrm{tr}(M_1(\boldsymbol{a}) \cdot M_1(\boldsymbol{a}^{(k)})^{-1}) - \log\,\det(M_1(\boldsymbol{a}) \cdot M_1(\boldsymbol{a}^{(k)})^{-1}) - n, \tag{4.33}$$

$$D_{M_2}(\boldsymbol{b}, \boldsymbol{b}^{(k)}) = \mathrm{tr}(M_2(\boldsymbol{b}) \cdot M_2(\boldsymbol{b}^{(k)})^{-1}) - \log\,\det(M_2(\boldsymbol{b}) \cdot M_2(\boldsymbol{b}^{(k)})^{-1}) - n, \tag{4.34}$$

$$D_{M_3}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = \mathrm{tr}(M_3(\boldsymbol{a}, \boldsymbol{b}) \cdot M_3(\boldsymbol{a}^{(k)}, \boldsymbol{b}^{(k)})^{-1}) - \log\,\det(M_3(\boldsymbol{a}, \boldsymbol{b}) \cdot M_3(\boldsymbol{a}^{(k)}, \boldsymbol{b}^{(k)})^{-1}) - n. \tag{4.35}$$

Gradient of the local model. A straightforward computation contained in Technical Appendix C.1 provides the following expressions for the components of the gradient of the local model:

$$\begin{aligned}\nabla_{\boldsymbol{a}}\tilde{f}^{(k)}(\boldsymbol{\theta}) &= \nabla_{\boldsymbol{a}} f(\boldsymbol{\theta}^{(k)}) + H_{\boldsymbol{a}}^{(k)}(\boldsymbol{a}-\boldsymbol{a}^{(k)}) - L_1^1\ \mathrm{math}^*(\mathrm{math}(\boldsymbol{a})^{-1} - \mathrm{math}(\boldsymbol{a}^{(k)})^{-1})\\ &+ L_1^3\ \mathrm{math}^*(((\big(\mathbf{i}_n\mathbf{i}_n^\top - \mathrm{math}(\boldsymbol{a}) - \mathrm{math}(\boldsymbol{b})\big)\odot S)^{-1} - ((\mathbf{i}_n\mathbf{i}_n^\top - \mathrm{math}(\boldsymbol{a}^{(k)}) - \mathrm{math}(\boldsymbol{b}^{(k)}))\odot S)^{-1})\odot S)\\ &+ L_2^1\sum_{i=1}^N\left[\frac{1}{1-a_i-b_i} - \frac{1}{1-a_i^{(k)}-b_i^{(k)}}\right]\cdot\mathbf{e}_N^{(i)} - L_2^2\sum_{i=1}^N\left[\frac{1}{1+a_i+b_i} - \frac{1}{1+a_i^{(k)}+b_i^{(k)}}\right]\cdot\mathbf{e}_N^{(i)},\end{aligned} \tag{4.36}$$

$$\begin{aligned}\nabla_{\boldsymbol{b}}\tilde{f}^{(k)}(\boldsymbol{\theta}) &= \nabla_{\boldsymbol{b}} f(\boldsymbol{\theta}^{(k)}) + H_{\boldsymbol{b}}^{(k)}(\boldsymbol{b}-\boldsymbol{b}^{(k)}) - L_1^2\ \mathrm{math}^*(\mathrm{math}(\boldsymbol{b})^{-1} - \mathrm{math}(\boldsymbol{b}^{(k)})^{-1})\\ &+ L_1^3\ \mathrm{math}^*(((\big(\mathbf{i}_n\mathbf{i}_n^\top - \mathrm{math}(\boldsymbol{a}) - \mathrm{math}(\boldsymbol{b})\big)\odot S)^{-1} - ((\mathbf{i}_n\mathbf{i}_n^\top - \mathrm{math}(\boldsymbol{a}^{(k)}) - \mathrm{math}(\boldsymbol{b}^{(k)}))\odot S)^{-1})\odot S)\\ &+ L_2^1\sum_{i=1}^N\left[\frac{1}{1-a_i-b_i} - \frac{1}{1-a_i^{(k)}-b_i^{(k)}}\right]\cdot\mathbf{e}_N^{(i)} - L_2^2\sum_{i=1}^N\left[\frac{1}{1+a_i+b_i} - \frac{1}{1+a_i^{(k)}+b_i^{(k)}}\right]\cdot\mathbf{e}_N^{(i)}.\end{aligned} \tag{4.37}$$

In these relations $\nabla_{\boldsymbol{a}} f(\boldsymbol{\theta}^{(k)})$ and $\nabla_{\boldsymbol{b}} f(\boldsymbol{\theta}^{(k)})$ are the components of the gradient of minus the log-likelihood function computed at the point $\boldsymbol{\theta}^{(k)}$ and $\mathbf{e}_N^{(i)} \in \mathbb{R}^N$, $i \in \{1,\dots,N\}$ are the canonical unit vectors introduced in Subsection 2.1.

Jacobian of the local model. We use the general relation (4.24) to determine the tangent map of $\nabla_{\boldsymbol{\theta}}\tilde{f}^{(k)}(\boldsymbol{\theta})$ for any $\boldsymbol{\delta\theta} := (\boldsymbol{\delta a}, \boldsymbol{\delta b}) \in \mathbb{R}^N \times \mathbb{R}^N$:

$$\begin{aligned}T_{\boldsymbol{\theta}}\nabla_{\boldsymbol{a}}\tilde{f}^{(k)}(\boldsymbol{\delta a},\boldsymbol{\delta b}) =& H_{\boldsymbol{a}}^{(k)}\boldsymbol{\delta a} + L_1^1\ \mathrm{math}^*\left(\mathrm{math}(\boldsymbol{a})^{-1}\ \mathrm{math}(\boldsymbol{\delta a})\ \mathrm{math}(\boldsymbol{a})^{-1}\right)\\ &+ L_1^3\ \mathrm{math}^*(((\big(\mathbf{i}_n\mathbf{i}_n^\top - \mathrm{math}(\boldsymbol{a}) - \mathrm{math}(\boldsymbol{b})\big)\odot S)^{-1}(\mathrm{math}(\boldsymbol{\delta a}+\boldsymbol{\delta b})\odot S)\times\\ &\times((\mathbf{i}_n\mathbf{i}_n^\top - \mathrm{math}(\boldsymbol{a}) - \mathrm{math}(\boldsymbol{b})\big)\odot S)^{-1})\odot S) + L_2^1\sum_{i=1}^N\frac{\delta a_i+\delta b_i}{(1-a_i-b_i)^2}\cdot\mathbf{e}_N^{(i)}\\ &+ L_2^2\sum_{i=1}^N\frac{\delta a_i+\delta b_i}{(1+a_i+b_i)^2}\cdot\mathbf{e}_N^{(i)},\end{aligned} \tag{4.38}$$

$$\begin{aligned}T_{\boldsymbol{\theta}}\nabla_{\boldsymbol{b}}\tilde{f}^{(k)}(\boldsymbol{\delta a},\boldsymbol{\delta b}) =& H_{\boldsymbol{b}}^{(k)}\boldsymbol{\delta b} + L_1^2\ \mathrm{math}^*\left(\mathrm{math}(\boldsymbol{b})^{-1}\ \mathrm{math}(\boldsymbol{\delta b})\ \mathrm{math}(\boldsymbol{b})^{-1}\right)\\ &+ L_1^3\ \mathrm{math}^*(((\big(\mathbf{i}_n\mathbf{i}_n^\top - \mathrm{math}(\boldsymbol{a}) - \mathrm{math}(\boldsymbol{b})\big)\odot S)^{-1}(\mathrm{math}(\boldsymbol{\delta a}+\boldsymbol{\delta b})\odot S)\times\\ &\times((\mathbf{i}_n\mathbf{i}_n^\top - \mathrm{math}(\boldsymbol{a}) - \mathrm{math}(\boldsymbol{b})\big)\odot S)^{-1})\odot S) + L_2^1\sum_{i=1}^N\frac{\delta a_i+\delta b_i}{(1-a_i-b_i)^2}\cdot\mathbf{e}_N^{(i)}\\ &+ L_2^2\sum_{i=1}^N\frac{\delta a_i+\delta b_i}{(1+a_i+b_i)^2}\cdot\mathbf{e}_N^{(i)}.\end{aligned} \tag{4.39}$$

The $(1,1)$ and $(1,2)$ blocks of the Jacobian matrix can be obtained from (4.38) by taking increments $\boldsymbol{\delta\theta}$ of the form $(\boldsymbol{\delta a}, \mathbf{0})$ and $(\mathbf{0}, \boldsymbol{\delta b})$, respectively. Analogously, the $(2,1)$ and $(2,2)$ blocks of the Jacobian follow from (4.39) by using increments $\boldsymbol{\delta\theta}$ of the form $(\boldsymbol{\delta a}, \mathbf{0})$ and $(\mathbf{0}, \boldsymbol{\delta b})$, respectively.

5 Empirical study

The goal of this section is to report the results of experiments that allow us to the compare the empirical performances of the different DCC models that we described in Section 3. For that purpose we select

a dataset and estimate those models with various dimensions (from 5 to 30), using the algorithm presented in Section 4. We then evaluate the in-sample adequacy of the different models with the Akaike information criterion (AIC) in Subsection 5.2, where we also discuss the heterogeneity and other features related to the point estimates of the non-scalar models. We finally assess the out-of-sample performances of the different models using statistical tests described in Subsection 5.3 and discuss their results in Subsection 5.4.

5.1 Datasets and competing models

We use two datasets that consist of daily closing price quotes of the thirty components included in the Dow Jones Industrial Average Index (DJIA) as of October 2013, downloaded from the Yahoo Finance database.[5] 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 crisis events in the out-of-sample interval, while Period II includes these events in the interval used for the estimations.

The Capital Asset Pricing Model (CAPM) [Shar 64] 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 $i \in \{1, 2, \ldots, n\}$ a static unconditional CAPM one-factor model of the form $Y_{i,t} = \alpha_i + \beta_i X_t + Z_{i,t}$, where for each time index t, $Y_{i,t}$ is the log-return of asset i, X_t is the value of the chosen common factor, $Z_{i,t}$ is the regression error term, and α_i, β_i are the intercept and slope coefficients, respectively.

In the DCC 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 log-returns of the S&P500 index as the common factor Z_t, we 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 $Z_{i,t}$. Following the same approach as in Chapter 8 of [Engl 09], we then estimate the DCC models under consideration on these residual returns, that is $(Z_{1,t}, Z_{2,t}, \ldots, Z_{n,t})^\top$ is the observed counterpart of $\mathbf{Z}_t$ appearing in (3.1). In order to perform the out-of-sample analysis, we compute the out-of-sample residual returns $\mathbf{Z} := \{\mathbf{Z}_{T_{est}+1}, \ldots, \mathbf{Z}_T\}$ according to the relations

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

where $\widehat{\alpha}_i$ and $\widehat{\beta}_i$ are obtained using the T_{est} in-sample observations and subsequently kept for the out-of-sample assessment of the empirical performances of the DCC models. When generating the out-of-sample forecasts of the H_t matrices of each DCC 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 adding one observation at a time in the out-of-sample period.

[5] 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 competing models. We report results for six different DCC model parameterizations, namely: **(i)** the Hadamard DCC, **(ii)** the rank deficient DCC with rank $r = 2$, **(iii)** the rank deficient DCC with rank $r = 1$, **(iv)** the Almon DCC, **(v)** the Almon shuffle DCC, and **(vi)** the scalar DCC. Some of these models are particular cases of others according to the inclusion relations represented in Figure 1.

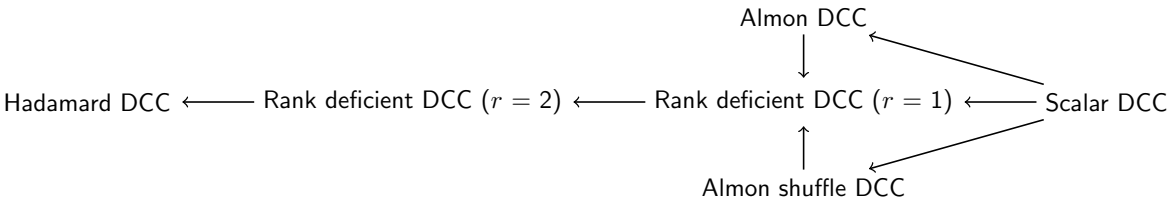

Figure 1: Inclusion hierarchy of models. The symbol $B \longleftarrow A$ should be read as "A is a particular case of B".

5.2 In-sample results

The estimations are performed in various dimensions ranging from 5 up to 30 in such a way that for the n-dimensional case we pick the first n assets in the DJIA datasets arranged in alphabetical order.

Goodness-of-fit comparisons. A consequence of the inclusion relations shown in Figure 1 is that the optimal log-likelihood values obtained in the estimation of these models for a given sample need to be ordered accordingly. More specifically, the optimal value resulting from the estimation procedure for the Hadamard DCC model has to be the smallest[6] among them all, and the others need to respect the hierarchy established by the diagram in Figure 1. In Tables 6.3 and 6.4 we report these values and the associated AIC statistics defined as

$$\text{AIC} := -\frac{2}{T_{est}}\log L + \frac{\log T_{est}}{T_{est}}k,$$

where T_{est} is the sample size reserved for estimation, $\log L$ is the obtained maximal value of the log-likelihood function associated to the model and the sample used for estimation, and k is the number of parameters of the model. A lower AIC value indicates a better trade-off between the model fit and the number of parameters used to achieve it.

The values reported in the tables reveal that the Hadamard DCC model has, as should be the case, the smallest minus log-likelihood values after convergence of the algorithms, and that the hierarchy of models in terms of the maximized values is respected. At the same time, the rank deficient DCC models with rank one (in eight cases) and rank two (in two cases) exhibit the best trade-off between fit and parsimony according to the AIC statistics. In all cases, the Hadamard models have the largest AIC values, which is not a big surprise given that they have many more parameters than the other models (see Table 6.1). It is interesting to notice that the differences in AIC values between the Hadamard model and all the other models are clearly larger than the differences between the other models. If we rank the six models by attributing the score 6 to the worst AIC fitting model (the Hadamard), the score 5 to the next, up to 1 to the best performing model, and we aggregate these scores through the six dimensions, we observe that the best performing model in both periods is the rank one deficient model. In period I, the second best is the Almon shuffle model, followed by the scalar model. In period II, the

[6]Recall that we minimize minus the log-likelihood function. In the tables 6.3 and 6.4 we report the minimal values divided by the number of observations; we call these normalized values.

second best model is the rank two deficient model, the third is the Almon shuffle, and the fourth the scalar. The global ranking (for both periods together, thus the best score can be 6, the worst 72), from best to worst, is rank one deficient (14 points), Almon shuffle (31), rank two deficient (39), scalar (45), Almon (51) and Hadamard (72). Based on these results, we consider that the rank one deficient and Almon shuffle models are worth using in practice, in addition to the scalar model.

Likelihood ratio (LR) tests. For each dimension, the scalar model has the smallest maximal log-likelihood, followed by the Almon model, and next by the Almon shuffle model, with one exception in Period I for the ten assets sample, where the order between the Almon models is reversed. Using the maximal log-likelihood values reported in Tables 6.3 and 6.4, LR-statistics can be computed to test the scalar model as null against all the other more general parameterizations as alternatives. The p-values of the corresponding LR statistics are reported in Tables 6.3 and 6.4. Two main results emerge: firstly, the scalar model is rejected in both periods at the level of 5% (and in many cases 1%) against the Almon shuffle, rank one, and rank two deficient models, with two exceptions that occur in Period I (for $n = 5$ against the rank two deficient model, and for $n = 10$ against the Almon shuffle model). Secondly, the scalar model is not rejected against the full Hadamard model, all p-values being very close to 1 when the dimension is larger than five. The numbers of degrees of freedom of the χ^2 distributions used in the tests are very large (for example, 928 in dimension 30 and 418 in dimension 10) and hence these tests are likely to be not very powerful. On the contrary, those numbers are much smaller for the other tests discussed above (4 for scalar against Almon, for example). In summary, the LR tests show that the more general models, especially the rank deficient ones and the Almon shuffle, are valuable flexible extensions of the scalar model.

Features of the point estimates. Since non-scalar DCC models allow for heterogeneity in the parameters of the conditional correlations, we provide some information on the estimates of these models. In Tables 6.5 and 6.6, we report for each model and dimension, the mean, standard deviation, minimum, and maximum of the estimates of the elements of the matrices A and B of the DCC process written as in (3.5). If these matrices are functions of intrinsic parameters of lower dimension than in the Hadamard case, we compute the estimates of the matrices implied by the intrinsic parameter estimates, and then the statistics. Obviously, for the scalar model, the mean, maximum, and minimum are equal to the point estimate of the corresponding scalar parameters. We observe the following results:

- The A mean values decrease towards zero in all models as the dimension n increases: they are approximately divided by a factor of 2 to 3 in dimension 30 relative to dimension 10. This is a well-known phenomenon for the scalar DCC model, already mentioned in [Engl 02] and discussed in [Engl 08b]. The latter paper shows by a simulation study that the a parameter is subjected to a downward bias and that variance targeting is responsible for that. Our results suggest that the same problem occurs in non-scalar models. We also observe that the mean values are very close to each other across the different models. This is less the case in dimension 5 for both periods and perhaps due to the arbitrary selection of the assets. Discounting these cases, this suggests that the scalar model estimates the average value of the elements of A, and that all models are perhaps subjected to the downward bias problem. Further insight on this issue could be obtained by a a simulation study, which we leave for further research.

- The A standard deviations (which measure the degree of heterogeneity between the elements of A) decrease as the dimension increases (in period I, the decrease is monotone, contrary to period II). The decrease is more pronounced between dimensions 5 and 15 than between dimensions 20 and 30.

- For every dimension, the largest degree of heterogeneity in A among non-scalar models takes place for the Hadamard case (with two minor exceptions where it is the second largest: in Period I for dimension 10, and in Period II for dimension 5). The rank deficient models have a slightly smaller degree of heterogeneity than Hadamard, while the Almon models exhibit much less heterogeneity than the previous three models. This reflects the more parsimonious parameterization of the Almon cases.

- The B mean values increase slightly towards unity until dimension 20, and stay approximately the same for dimensions 20, 25, and 30.

- The B standard deviations decrease globally (comparing dimensions 10 and 30), with some intermediate ups and downs, as the dimension increases. In Period I, this is a bit less clear, and in particular there is a global increase for the rank two and rank one deficient models (discounting for the latter the unusual result in dimension 10).

We also report in the tables the statistics of the McGyver method [Engl 08a]. For dimension 30, this consists in estimating all (in our case 435) the associated bivariate scalar DCC models, thus getting as many point estimates of a and b. For each dimension smaller than 30 (except 5), we use the estimates for the corresponding group of assets (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, we form six groups of five assets (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 models we consider; this is a consequence of the positive semidefiniteness constraints satisfied by DCC to which the McGyver estimator is not exposed. Additionally, unusual values arise sometimes in the McGyver case in comparison with the DCC estimates; more specifically, there are values of b close to zero (see the minimum McGyver values in comparison to other models) 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.

Rank one deficient and Almon shuffle point estimates. Given that the AIC statistics favor these models, we provide more information on the estimates of the corresponding parameters. Figures 2 and 3 represent the values of the entries of the parameter vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ of both models in dimension 30, ordered according to the Almon shuffle estimation. Each figure also shows the corresponding estimate implied by the scalar model, which is the square root of the parameter a or b of (3.21). The figures illustrate the flexibility of the rank one deficient and Almon models with respect to the scalar one. The pattern of estimates of the entries of $\boldsymbol{a}$ of the Almon model (monotone increasing) is different from that of $\boldsymbol{b}$ (concave). In Period II (not represented) the pattern for $\boldsymbol{a}$ is also concave. For the rank one deficient model, the estimates are naturally more variable. The scalar model estimates correspond approximately to the average values of the individual estimates of the other models.

The estimates of the entries of $\boldsymbol{a}$ and $\boldsymbol{b}$ in the Almon shuffle model are computed by using the Almon function given in (2.15). In Table 6.7 we report the estimates of the three parameters of each Almon function for the dimensions considered. As the dimension increases, for the function defining $\boldsymbol{a}$, the estimates of the parameters that intervene inside the exponential function tend to zero, and the estimates of the constant term tend to minus one. This implies that the entries of $\boldsymbol{a}$ tend to zero and are more homogeneous in higher dimensions. As we mentioned above, these trends seem to saturate beyond dimension 15. Regarding the parameters of the Almon function defining $\boldsymbol{b}$, the three parameters tend to zero, so that the elements of $\boldsymbol{b}$ tend to one from below.

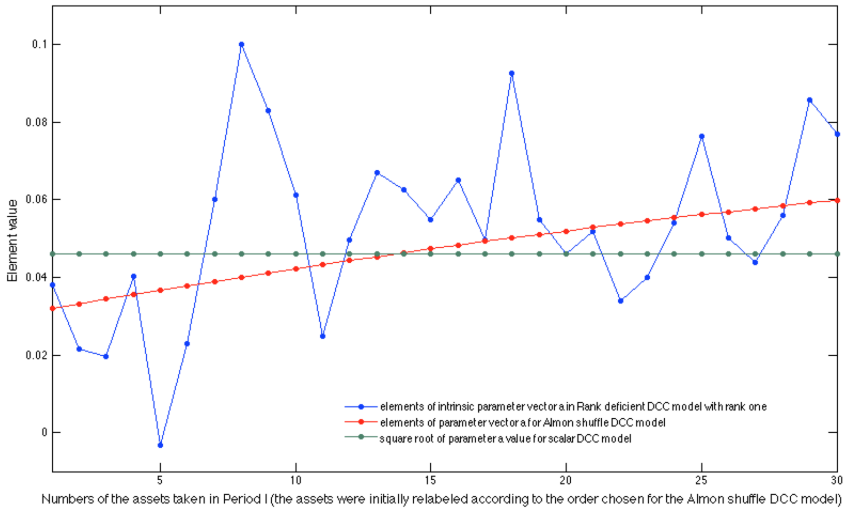

Figure 2: Estimates of the entries of the parameter $\boldsymbol{a}$ of the rank one deficient and Almon shuffle models for the thirty assets case in the order determined by the Almon shuffle estimation (dataset of Period I).

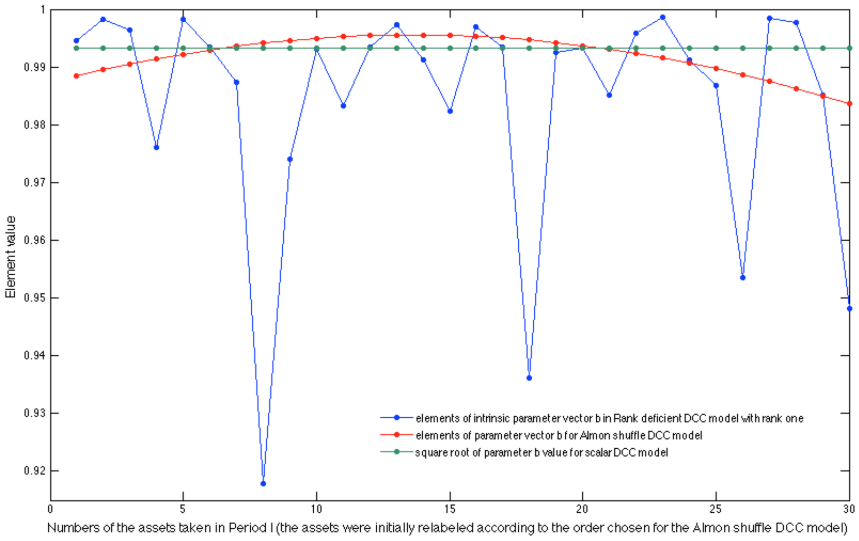

Figure 3: Estimates of the entries of the parameter $\boldsymbol{b}$ of the rank one deficient and Almon shuffle models for the thirty assets case in the order determined by the Almon shuffle estimation (dataset of Period I).

Computational effort. In Table 6.2 we provide the time and the number of gradient calls necessary to carry out the estimation of the various DCC models under consideration in different dimensions when using the Period II sample. The figures provided are relative to the computational effort necessary to estimate the scalar model in the corresponding dimension. There are several considerations that make sometimes those figures not directly comparable and that need to be taken into account at the time of drawing conclusions:

- The initial values used for the estimation of the different models are those obtained after the estimation of the scalar model. More specifically, the scalar model is estimated using a generic

initial condition ($a = 0.2$, $b = 0.7$) and then, the resulting A and B matrices are fed as initial conditions to the optimization algorithm in the estimation of all the other models.

- The stopping tolerance on the function value increments used in the optimization algorithm is the same when handling the different models (10^{-6}). However, those having to do with the change in the variable values had to be tuned for each different parameterization in order to secure reasonable results (it equals 10^{-4} for the Hadamard case and rank deficient parameterizations; for the Almon models it is equal to 10^{-4} in dimensions up to fifteen and 10^{-5} for $n = 20, 25, 30$).

It is interesting to observe that the built-in nonlinearities of some of these parameterizations (exponential for Almon and quadratic for rank deficient models) have sometimes a bigger impact on the computational effort than the dimension of the intrinsic parameter space (compare for example the values exhibited by the Hadamard parameterization with those associated to rank two deficient models).

The main conclusion that we draw from the figures in Table 6.2 is that from a computational effort standpoint, non-scalar models are quite affordable with respect to the scalar model in the dimensions considered.

5.3 Out-of-sample specification tests

We start by presenting the specification tests that we use in order to assess the out-of-sample one-step ahead forecasting performance of the competing DCC models. One test is based on the use of multivariate variance standardized returns, and three others on the use of portfolio returns.

5.3.1 Model confidence set based on correlation loss functions

We compare the models under consideration by computing the model confidence set (MCS) of [Hans 03, Hans 11] using the following loss function constructed using the GARCH standardized returns in (5.3) and the conditional correlation matrices implied by the different models:

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

where $\rho_{ij,t}$ is the (i,j)-entry of the model dependent conditional correlation matrix R_t of the standardized returns introduced in (3.4). The loss function is based on the fact that under the correct specification (3.1), the GARCH standardized returns

$$\boldsymbol{\varepsilon}_t := D_t^{-1/2}\mathbf{r}_t \tag{5.3}$$

have the correlation matrix R_t defined in (3.4). The results are provided in Tables 6.8 and 6.9. Each table shows, for a given number of assets, the order of exclusion of models from the confidence set: 6 means that the corresponding model was the first to be excluded (with its p-value underneath), 1 means that the corresponding model was the last. The MCS at the significance level of 5% is identified by the set of bold red figures, and at 10% by union of the bold red and black figures.

5.3.2 Tests based on the use of 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 variance, where H_t is the relevant conditional covariance matrix of $\mathbf{r}_t$ (see (3.1)). We construct two kinds of portfolios:

- **The minimum variance portfolio (MVP)**: it is defined by choosing a weight vector $\mathbf{w}_t$ that minimizes $\mathbf{w}^\top H_t \mathbf{w}$ subjected to the constraint $\sum_{i=1}^n w_{t,i} = 1$. The solution of this problem is given by
$$\mathbf{w}_t = \frac{1}{\mathbf{i}_n^\top H_t^{-1} \mathbf{i}_n} H_t^{-1} \mathbf{i}_n. \tag{5.4}$$
We use this expression to construct the sequence of variance minimizing portfolios associated to each of the different models under test.
- **The equally weighted portfolio (EWP)**: in this case $\mathbf{w}_t := \mathbf{i}_n/n$ for each date t.

Portfolio tests construction. The first three tests that we consider are based on the observation that under a correct specification of the type (3.1), the standardized portfolio return series
$$y_t := \frac{\mathbf{w}_t^\top \mathbf{r}_t}{\sqrt{\mathbf{w}_t^\top H_t \mathbf{w}_t}} \tag{5.5}$$
has unconditional variance equal to one. Using the following tests, we can assess the validity of different hypotheses for the series
$$\widehat{y}_t := \frac{\mathbf{w}_t^\top \mathbf{r}_t}{\sqrt{\mathbf{w}_t^\top \widehat{H}_t \mathbf{w}_t}}, \tag{5.6}$$
constructed using the one-step ahead forecast of the conditional covariance matrices $\widehat{H}_t$ implied by each of the estimated models under consideration.

Engle-Colacito regression test [Engl 06]: it is constructed by estimating the regression
$$\widehat{y}_t - 1 = \lambda + u_t, \qquad t \in \{T_{est}+1, \ldots, T\}, \tag{5.7}$$
where u_t is the error term and λ is the intercept coefficient. The test assesses the null hypothesis that λ is equal to 0 using a heteroskedasticity and autocorrealtion consistent (HAC) t-statistic ([Andr 91]). The results are presented in the tables 6.10-6.13 that contain the HAC t-statistics and their p-values.

Model confidence set (MCS) based on the predictive ability for squared portfolio returns: we evaluate the one-step ahead predictive ability of the models under consideration by computing model confidence sets using the loss function
$$d_t := \left((\mathbf{w}_t^\top \mathbf{r}_t)^2 - \mathbf{w}_t^\top \widehat{H}_t \mathbf{w}_t \right)^2. \tag{5.8}$$
The results of this procedure are provided in Tables 6.14-6.17. The information that they contain is organized in the same way that was already described at the end of Section 5.3.1.

VaR backtesting and the dynamic quantile test (HIT test) [Engl 04]: the Value at Risk $\mathrm{VaR}_\alpha(t)$ at confidence level α of a given portfolio return ($p_t = \mathbf{w}_t^\top \mathbf{r}_t$) distribution at time t is its α-quantile. When the returns distribution is forecasted using a given model, the backtesting of the dynamical structure of the VaR violations, that is, the occurrences for which $\mathbf{w}_t^\top \mathbf{r}_t < \mathrm{VaR}_\alpha(t)$, can be used as a tool to assess its pertinence. We define the binary random variable HIT that counts the VaR violations as
$$\mathrm{HIT}_\alpha(t) := \mathbf{1}_{\{\mathbf{w}_t^\top \mathbf{r}_t < \mathrm{VaR}_\alpha(t)\}}.$$
Under a correct model specification, $E\left[\mathrm{HIT}_\alpha(t)\right] = \alpha$ and the random variables $\{\mathrm{HIT}_\alpha(t)\}_{t\in\{T_{est}+1,\ldots,T\}}$ are serially independent and independent of other elements in the conditioning information set, like the quantity $\mathrm{VaR}_\alpha(t)$. We test these features using the HIT-test [Engl 04], which is a F-test on the regression
$$\mathrm{HIT}_\alpha(t) - \overline{\mathrm{HIT}}_\alpha = \lambda + \beta_1 \mathrm{HIT}_\alpha(t-1) + \cdots + \beta_l \mathrm{HIT}_\alpha(t-l) + \beta_{l+1} \mathrm{VaR}_\alpha(t) + u_t, \tag{5.9}$$

where $\overline{\text{HIT}}_\alpha = \frac{1}{T_{out}} \sum_{t=T_{est}+1}^{T} \text{HIT}_\alpha(t)$ and l is an exogenous number of lags. The null hypothesis is that all the regression coefficients, including the intercept, are equal to zero. In order to compute $\text{VaR}_\alpha(t)$, we use the fact that the forecasted returns $\widehat{p}_t$ are Gaussian with zero mean and estimated variance $\widehat{\sigma}^2_{p,t} = \mathbf{w}_t^\top \widehat{H_t} \mathbf{w}_t$ and hence $\text{VaR}_\alpha(t) = \Phi^{-1}(\alpha)\widehat{\sigma}_{p,t}$, where Φ is the cumulative density function of the standard normal distribution.

The results of this test with lag order $l = 5$ for the two datasets and portfolios are described in Tables 6.18-6.21. For each portfolio cardinality, the tables contain the hit averages $\overline{\text{HIT}}_\alpha$ for the confidence levels 1, 5, and 10%, as well as the corresponding p-values of the F-test on the regression (5.9).

5.4 Results of the tests on the empirical performance of the competing models

All the tests presented in the previous subsection are carried out in several dimensions ranging from 5 up to 30 in such a way that for the n-dimensional case we have picked the first n assets in the DJIA datasets arranged in alphabetical order. The following conclusions about the results of the tests that we conducted can be drawn:

1. **MCS for the correlation of the standardized returns (Tables 6.8-6.9)**: the first observation is that the size of the MCS goes down when the dimension increases. In both periods, the MCS contains 5 or 6 models for dimensions up to 15. In Period I, for dimensions from 20 to 30, the MCS contains a single model; in Period II, it contains one model for $n = 20$, and two models for $n = 25$ and 30. Those models are always Almon or Almon shuffle, except in Period I for $n = 20$ where it is the scalar model. The Almon shuffle model performs relatively well in all dimensions in Period II (with a score of 8), distantly followed by the Almon and rank one deficient models (scores 20), the scalar model having only the fourth score (22). In Period I, the scalar model is the winner (with a score of 10), closely followed by the Almon model (score 13), and then the rank one deficient model (score 18).

2. **Tests based on the use of portfolio returns.**

   - **Engle-Colacito regression test (Tables 6.10-6.13)**: the test results are the same in most cases across models for each dimension, that is, rejections (or non-rejections) at the level of 5% occur for all models. Exceptions occur in Period I: for dimension 25 (MVP and EWP), where no rejection happens only for the scalar and both Almon models (also rank one deficient in EWP); for dimension 10 (EWP), where only Hadamard, rank one deficient, and Almon shuffle models are not rejected. In Period II, the exceptions are for dimension 25, MVP (no rejection for scalar and Almon shuffle), EWP (no rejection for the Hadamard model). However, when differences of this kind happen, the p-values are rather close to 5%.
   - **MCS for the predictive ability of squared portfolio returns (Tables 6.14-6.17)**: the results do not favor systematically a particular model. For minimum variance portfolios, the model confidence sets at the confidence level of 90% (95% in a few cases) include all the models in both periods. The reported model scores are very similar in Period I, while in Period II they are more favorable to the rank one deficient model. For equally weighted portfolios, the model confidence sets do not include all six models; they only include the scalar model in dimensions up to 20.
   - **VaR backtesting and the dynamic quantile test (HIT test, Tables 6.18-6.21)**: for each dimension, the percentages of VaR violations (at 1, 5, and 10% confidence) and the p-values of the F-tests of independence are in most cases similar across models. For equally weighted portfolios, the number of VaR violations are not very good across the board, see

e.g. the results for $n = 30$ in Tables 6.19 and 6.21. We have also carried these tests out using the filtered historical simulation procedure [Baro 98] that constructs the forecasted returns distribution and the corresponding VaRs by bootstrapping the estimation residuals; the number of VaR violations obtained with this method are much closer to their nominal levels for all models.

The results of the three tests based on portfolio returns clearly indicate that the different DCC models considered in this paper have more or less the same performance in terms of out-of-sample specification tests when using this approach. No model dominates systematically the others, and no model is systematically dominated. This may be due to the fact that the different DCC models, and consequently these tests results, rely on the same univariate GARCH models for the conditional variances. In the context of portfolio performances, the correlations are often of secondary importance with respect to the volatilities of the different assets. Consequently, the test results are probably mostly influenced by the common GARCH components of the different models and therefore, unless the DCC models under consideration produce very different correlation forecasts, it is not surprising that the test based on portfolio returns do not let us discriminate clearly between the different parameterizations of the correlation dynamics.

On the contrary, the tests behind the MCS based on the correlation of the standardized returns do not depend directly on the univariate GARCH equations (though they depend on them indirectly, through the standardization of the returns) and, as we saw, favors clearly the most parsimonious models (Almon, Almon shuffle, scalar and rank one deficient).

In any case, we recommend that the choice of the optimal model is done *a posteriori* since the grounds for this decision depend on the dataset, time period, and dimensionality of the application of interest. It is clear that all these factors influence significantly the performance of the DCC models and that a careful empirical study has to be conducted in order to select one of them.

6 Conclusions

In this paper we used several parameterizations for the Dynamic Conditional Correlation (DCC) family of models which go beyond the standard scalar case that appears in most applications. We adapted an optimization technique based on Bregman divergences designed to effectively perform the two-stages quasi-maximum likelihood estimation of these models by handling all the linear and non-linear parameter constraints that need to be imposed in order to ensure the stationarity of the processes and the positive semidefiniteness of the conditional covariance matrices.

More explicitly, the DCC parameterizations that we work with are the scalar, Hadamard, and rank deficient ones already proposed in the literature. We also introduce new parameterizations of the DCC model, that we call Almon DCC and Almon shuffle DCC. They are particular cases of the rank one deficient model. The number of parameters of the latter increases linearly with the dimension of the return process, whereas the corresponding number in the Almon specifications is independent of that dimension. Even though this property also applies to the scalar model, the Almon version is more flexible as it does not impose that the same parameter values determine the dynamic pattern of the conditional correlations. We provided all the mathematical and algorithmic details needed to implement the proposed optimization method for each specification of the DCC model. Computer codes will be made available to the community.

We carried out an extensive empirical study using two datasets based on the thirty constituents of the Dow Jones Industrial Average Index considered during two periods that differ as to the presence or not of the financial crisis events of 2008-09 in the sample used for estimation. The proposed estimation procedure was applied to all the DCC models under study in various dimensions up to thirty. Moreover, we compared the empirical in- and out-of-sample performances of the different types of parameterizations

in the forecasting of conditional correlations using a variety of statistical specification tests. Our results provide substantial evidence that some non-scalar DCC parameterizations, in particular the Almon (shuffle) one and rank one deficient, are worth using in the volatility modeling of asset returns of dimension up to thirty, and perhaps more, though this remains an open issue.

Two extensions of all this work are on our research agenda. In the first one, we shall develop and apply the method of composite quasi-maximum likelihood estimation to the non-scalar DCC models. [Engl 08b] have found that this method reduces the bias in the estimation of the parameters of the scalar DCC model, so we shall be interested to know whether this result applies also to non-scalar models. The second extension will consist in applying the optimization tools of this paper to various non-scalar diagonal VEC (DVEC) models and compare the performance of DCC and DVEC models.

A possible further development of the research reported in this paper is to apply the same tools to the cDCC model of Aielli [Aiel 13]. This extension will not be trivial, because the cDCC model does not lend itself to a targeting procedure of the constant term matrix S of the dynamic equation of the model. Instead, a profiling method must be used, as discussed in [Aiel 13]. The profiling implies that at every evaluation of the objective function in the optimization method, the matrix S is updated through the intrinsic parameters that determine A and B. The consequences of this are that the mathematical derivations needed to implement the constrained optimization algorithm will be more tedious, and estimation will be more computer intensive. It is an open question whether the cDCC model would then be applicable in dimension as high as thirty.

Appendix: Matrix expressions of the recursions in Proposition 4.1

The proposition below provides explicit matrix expressions for the operator recursions (4.5)-(4.10) given in Proposition 4.1. These expressions are needed to compute iteratively the gradient of the log-likelihood function. The proof is provided in the Technical Appendix A.4. We refer the reader to Section 2 for the definitions of the operators and other mathematical objects in the proposition.

Proposition 6.1 *Consider* $\boldsymbol{\theta} := (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$, $\boldsymbol{\Theta}(\boldsymbol{\theta}) := (A(\boldsymbol{\theta}_1), B(\boldsymbol{\theta}_2))$, *and the differential operators* $T^*_{\boldsymbol{\Theta}} H_t$, $T^*_{\boldsymbol{\Theta}} Q_t : \mathbb{S}_n \times \mathbb{S}_n \longrightarrow \mathbb{S}_n \times \mathbb{S}_n$, *which for every component* Θ *of* $\boldsymbol{\Theta}$ *are defined by the relations* (4.5)-(4.7). *Let* $\mathcal{A}^t_Q, \mathcal{B}^t_Q, \mathcal{C}^t_\Theta : \mathbb{R}^N \longrightarrow \mathbb{R}^N$ *with* $N := \frac{1}{2} n\,(n+1)$ *be the linear maps, defined as*

$$\mathcal{A}^t_Q := \text{vech} \circ T^*_A Q_t \circ \text{math}, \tag{6.1}$$

$$\mathcal{B}^t_Q := \text{vech} \circ T^*_B Q_t \circ \text{math}, \tag{6.2}$$

$$\mathcal{C}^t_\Theta := \text{vech} \circ T^*_\Theta H_t \circ \text{math}, \tag{6.3}$$

and let $A_t, B_t, C_t \in \mathbb{S}_N$ *be the matrices associated to the operators* (6.1), (6.2), *and* (6.3), *respectively. Then, the matrices* $\{A_t\}_{t\in\{1,\ldots,T\}}$, $\{B_t\}_{t\in\{1,\ldots,T\}}$, *and* $\{C_t\}_{t\in\{1,\ldots,T\}}$ *are determined by the recursions*

$$A_t = \text{diag}\left(\text{vech}\left(\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}^\top_{t-1} - S\right)\right) + A_{t-1}\text{diag}\left(\text{vech}\left(B\right)\right), \tag{6.4}$$

$$B_t = \text{diag}\left(\text{vech}\left(Q_{t-1} - S\right)\right) + B_{t-1}\text{diag}\left(\text{vech}\left(B\right)\right), \tag{6.5}$$

$$\begin{aligned} C_t = \Theta_t\Big\{ & L_n\left(\left(Q^{*-1}_t D_t\right) \otimes \left(Q^{*-1}_t D_t\right)\right) - \frac{1}{2} P^d_n L_n((Q^{*-1}_t Q_t Q^{*-1}_t D_t) \otimes (Q^{*-2}_t D_t) \\ & + (Q^{*-1}_t D_t) \otimes (Q^{*-2}_t Q_t Q^{*-1}_t D_t))\Big\} \cdot D_n \;\text{ with }\; \Theta_t = \{A_t, B_t\}, \end{aligned} \tag{6.6}$$

and the initial values $A_0 = B_0 = C_0 = 0$, *where* L_n, D_n, *and* P^d_n *are the elimination, duplication, and diagonalization matrices in dimension* n, *respectively, defined in* (2.7)-(2.9).
Finally, consider the map $\mathcal{G}_{\boldsymbol{\theta}} : \mathbb{R}^N \times \mathbb{R}^N \longrightarrow \mathcal{P} \times \mathcal{P}$ *defined by*

$$\mathcal{G}_{\boldsymbol{\theta}} := T^*_{\boldsymbol{\theta}} \boldsymbol{\Theta} \circ \text{math}, \tag{6.7}$$

where $T^{*}_{\boldsymbol{\theta}}\boldsymbol{\Theta} : \mathbb{S}_n \times \mathbb{S}_n \longrightarrow \mathcal{P} \times \mathcal{P}$ *is the adjoint of the map* $T_{\boldsymbol{\theta}}\boldsymbol{\Theta} : \mathcal{P} \times \mathcal{P} \longrightarrow \mathbb{S}_n \times \mathbb{S}_n$ *and* $\mathcal{P} \times \mathcal{P}$ *the intrinsic* $\boldsymbol{\theta}$ *parameter space that depends on the mapping* $\boldsymbol{\Theta}(\boldsymbol{\theta})$ *associated to each of the model specifications considered in Proposition 4.1. Then:*

(i) ***The Hadamard family:***

$$\begin{aligned} &\mathcal{G}_{\boldsymbol{\theta}} : \mathbb{R}^N \times \mathbb{R}^N \longrightarrow \mathbb{R}^N \times \mathbb{R}^N \\ &\mathcal{G}_{\boldsymbol{\theta}}(\mathbf{v}_1, \mathbf{v}_2) = 2\left(\mathbf{v}_1 - \frac{1}{2}\text{vech}(\text{Diag}(\text{math}(\mathbf{v}_1))), \mathbf{v}_2 - \frac{1}{2}\text{vech}(\text{Diag}(\text{math}(\mathbf{v}_2)))\right). \end{aligned} \tag{6.8}$$

(ii) ***The rank deficient family:***

$$\begin{aligned} &\mathcal{G}_{\boldsymbol{\theta}} : \mathbb{R}^N \times \mathbb{R}^N \longrightarrow \mathbb{R}^{N^*} \times \mathbb{R}^{N^*} \\ &\mathcal{G}_{\boldsymbol{\theta}}(\mathbf{v}_1, \mathbf{v}_2) = 2\ (\text{vec}_r(\text{math}(\mathbf{v}_1)\text{mat}_r(\boldsymbol{\theta}_1)), \text{vec}_r(\text{math}(\mathbf{v}_2)\text{mat}_r(\boldsymbol{\theta}_2))), \end{aligned} \tag{6.9}$$

with $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \in \mathbb{R}^{N^*} \times \mathbb{R}^{N^*}$*, and where* $r \leq n \in \mathbb{N}$*,* $N^* := nr - \frac{1}{2}r(r-1)$*.*

(iii) ***The Almon family:***

$$\begin{aligned} &\mathcal{G}_{\boldsymbol{\theta}} : \mathbb{R}^N \times \mathbb{R}^N \longrightarrow \mathbb{R}^3 \times \mathbb{R}^3 \\ &\mathcal{G}_{\boldsymbol{\theta}}(\mathbf{v}_1, \mathbf{v}_2) = 2\ (K_{\boldsymbol{\theta}_1}^{\top}\text{math}(\mathbf{v}_1)\text{alm}_n(\boldsymbol{\theta}_1), K_{\boldsymbol{\theta}_2}^{\top}\text{math}(\mathbf{v}_2)\text{alm}_n(\boldsymbol{\theta}_2)), \end{aligned} \tag{6.10}$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \in \mathbb{R}^3 \times \mathbb{R}^3$*, and* $K_{\boldsymbol{\theta}_i}$ *(for* $i = 1$ *and* 2*) has been defined in Proposition 4.1.*

(iv) ***The scalar family:***

$$\begin{aligned} &\mathcal{G}_{\boldsymbol{\theta}} : \mathbb{R}^N \times \mathbb{R}^N \longrightarrow \mathbb{R} \times \mathbb{R} \\ &\mathcal{G}_{\boldsymbol{\theta}}(\mathbf{v}_1, \mathbf{v}_2) = \left(\langle \text{math}(\mathbf{v}_1), \mathbf{i}_n\mathbf{i}_n^{\top}\rangle, \langle \text{math}(\mathbf{v}_2), \mathbf{i}_n\mathbf{i}_n^{\top}\rangle\right). \end{aligned} \tag{6.11}$$

Proposition 6.1 provides a complete computational recipe for the calculation of the gradient of the log-likelihood function associated to each of the DCC model parameterizations under study. We note that this result could be easily extended to any other model prescription by simply writing down the associated operator $\mathcal{G}_{\boldsymbol{\theta}} : \mathbb{R}^N \times \mathbb{R}^N \longrightarrow \mathcal{P} \times \mathcal{P}$ in (6.7).

References

[Aiel 13] G. P. Aielli. "Dynamic Conditional Correlation: on Properties and Estimation". *Journal of Business & Economic Statistics*, Vol. 31, No. 3, pp. 282–299, July 2013.

[Alex 98] C. Alexander. "Orthogonal GARCH". In: C. Alexander, Ed., *Mastering Risk*, pp. 21–38, Financial Times-Prentice Hall, 1998.

[Almo 65] S. Almon. "The distributed lag between capital appropriations and expenditures". *Econometrica*, Vol. 33, No. 1, pp. 178–196, 1965.

[Andr 91] D. W. K. Andrews. "Heteroskedasticity and autocorrelation consistent covariance matrix estimation". *Econometrica*, Vol. 59, No. 3, pp. 817–858, 1991.

[Bapa 97] R. B. Bapat and T. E. S. Raghavan. *Nonnegative Matrices and Applications*. Cambridge University Press, 1997.

[Baro 98] G. Barone-Adesi, F. Bourgoin, and K. Giannopoulos. "Don't look back". *Risk*, Vol. 11, No. 8, pp. 100–103, 1998.

[Bauw 13] L. Bauwens and E. Otranto. "Modeling the dependence of conditional correlations on volatility". 2013.

[Bill 03] M. Billio, M. Caporin, and M. Gobbo. "Block dynamic conditional correlation multivariate GARCH models". 2003.

[Bill 05] M. Billio and M. Caporin. "Multivariate Markov switching dynamic conditional correlation GARCH representations for contagion analysis". *Statistical Methods & Applications*, Vol. 14, pp. 145–161, 2005.

[Bill 06] M. Billio, M. Caporin, and M. Gobbo. "Flexible Dynamic Conditional Correlation multivariate GARCH models for asset allocation". *Applied Financial Economics Letters*, Vol. 2, No. 2, pp. 123–130, 2006.

[Boll 86] T. Bollerslev. "Generalized autoregressive conditional heteroskedasticity". *Journal of Econometrics*, Vol. 31, No. 3, pp. 307–327, 1986.

[Boll 90] T. Bollerslev. "Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model". *Review of Economics and Statistics*, Vol. 72, No. 3, pp. 498–505, 1990.

[Breg 67] L. Bregman. "The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming". *Zh. vychisl. mat. Mat. Fiz.*, Vol. 7, No. 3, pp. 620–631, 1967.

[Capo 12] M. Caporin and M. McAleer. "Do we really need both BEKK and DCC? A tale of two multivariate GARCH models.". *Journal of Economic Surveys*, Vol. 26, No. 4, pp. 736–751, Sep. 2012.

[Capp 06] L. Cappiello, R. F. Engle, and K. K. Sheppard. "Asymmetric dynamics in the correlations of global equity and bond returns". *Journal of Financial Economics*, Vol. 4, No. 4, pp. 537–572, 2006.

[Chre 13] S. Chrétien and J.-P. Ortega. "Multivariate GARCH estimation via a Bregman-proximal trust-region method". *Computational Statistics and Data Analysis*, Vol. To appear, 2013.

[Cola 11] R. Colacito, R. F. Engle, and E. Ghysels. "A component model for dynamic correlations". *Journal of Econometrics*, Vol. 164, pp. 45–59, 2011.

[Dhil 07] I. S. Dhillon and J. A. Tropp. "Matrix nearness problems with Bregman divergences". *SIAM J. Matrix Anal. Appl.*, Vol. 29, No. 4, pp. 1120–1146, 2007.

[Engl 02] R. F. Engle. "Dynamic conditional correlation -a simple class of multivariate GARCH models". *Journal of Business and Economic Statistics*, Vol. 20, pp. 339–350, 2002.

[Engl 04] R. F. Engle and S. Manganelli. "Conditional autoregressive value at risk by regression quantiles". *Journal of Business & Economic Statistics*, Vol. 22, No. 4, pp. 367–381, Oct. 2004.

[Engl 06] R. F. Engle and R. Colacito. "Testing and Valuing Dynamic Correlations for Asset Allocation". *Journal of Business & Economic Statistics*, Vol. 24, No. 2, pp. 238–253, 2006.

[Engl 08a] R. F. Engle. "High dimensional dynamic correlations". In: J. L. Castle and N. Shephard, Eds., *The Methodology and Practice of Econometrics: Papers in Honour of David F Hendry*, Oxford University Press, Oxford, 2008.

[Engl 08b] R. F. Engle, N. Shephard, and K. K. Sheppard. "Fitting Vast Dimensional Time-Varying Covariance Models". 2008.

[Engl 09] R. Engle. *Anticipating Correlations*. Princeton University Press, Princeton, NJ, 2009.

[Engl 82] R. F. Engle. "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation". *Econometrica*, Vol. 50, No. 4, pp. pp. 987–1007, 1982.

[Fern 05] M. Fernandes, B. de Sa Mota, and G. Rocha. "A multivariate conditional autoregressive range model". *Economics Letters*, Vol. 86, pp. 435–440, 2005.

[Hafn 09] C. M. Hafner and P. H. Franses. "A generalized Dynamic Conditional Correlation model: simulation and application to many assets". *Econometric Reviews*, Vol. 28, No. 6, pp. 612–631, 2009.

[Hans 03] P. R. Hansen, A. Lunde, and J. M. Nason. "Choosing the best volatility models: the model confidence set approach". *Oxford Bulletin of Economics and Statistics*, Vol. 65, No. s1, pp. 839–861, Dec. 2003.

[Hans 11] P. R. Hansen, A. Lunde, and J. M. Nason. "The model confidence set". *Econometrica*, Vol. 79, No. 2, pp. 453–497, 2011.

[Horn 94] R. A. Horn and C. R. Johnson. *Topics in matrix analysis*. Cambridge University Press, Cambridge, 1994.

[Jame 61] W. James and C. Stein. "Estimation with quadratic loss". In: *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I*, pp. 361–379, Univ. California Press, Berkeley, Calif., 1961.

[Kuli 09a] B. Kulis, M. A. Sustik, and I. S. Dhillon. "Low-rank kernel learning with Bregman matrix divergences". *J. Mach. Learn. Res.*, Vol. 10, pp. 341–376, 2009.

[Kuli 09b] B. Kulis, S. S. Sustik, and I. S. Dhillon. "Convex perturbations for scalable semidefinite programming". In: *Proceedings of the 12th International Conference on Artificial Intelligence and Statistics (AISTATS) 2009*, pp. 296–303, 2009.

[Lutk 05] H. Lütkepohl. *New Introduction to Multiple Time Series Analysis*. Springer-Verlag, Berlin, 2005.

[Nour 14] D. Noureldin, N. Shephard, and K. Sheppard. "Multivariate Rotated ARCH models". *Journal of Econometrics*, Vol. 179, No. 1, pp. 16–30, 2014.

[Pell 06] D. Pelletier. "Regime switching for dynamic correlations". *Journal of Econometrics*, Vol. 131, No. 1-2, pp. 445–473, 2006.

[Shar 64] W. F. Sharpe. "Capital asset prices: a theory of market equilibrium under conditions of risk". *The Journal of Finance*, Vol. 19, No. 3, pp. 425–442, 1964.

[Tse 02] Y. K. Tse and A. K. C. Tsui. "A multivariate GARCH with time-varying correlations". *Journal of Business and Economic Statistics*, Vol. 20, pp. 351–362, 2002.

	n	5	10	15	20	25	30
Hadamard	$n(n+1)$	30	110	240	420	650	930
Rank Deficient ($r=2$)	$2(2n-1)$	18	38	58	78	98	118
Rank Deficient ($r=1$)	$2n$	10	20	30	40	50	60
Almon	6	6	6	6	6	6	6
Scalar	2	2	2	2	2	2	2

Table 6.1: Number of parameters needed for the different DCC model parameterizations as a function of the process dimension n. These numbers represent exclusively the parameters obtained in the second stage estimation that we study in detail in this paper.

Period II. Relative computation time/Relative number of gradient calls

n		Almon	Almon Shuffle	Rank Deficient ($r=1$)	Rank Deficient ($r=2$)	Hadamard
5	time	0.6599	0.7783	0.7828	6.5984	1.0486
	grad calls	0.6613	0.7742	0.6452	3.1935	0.6774
10	time	2.5637	5.4408	1.7209	28.3579	2.8449
	grad calls	0.7500	1.2333	0.9333	3.2667	0.9000
15	time	3.0451	16.9112	6.7186	14.4835	2.6498
	grad calls	1.1190	4.8095	1.7619	2.5952	1.0952
20	time	1.1361	0.9467	2.7956	2.6302	0.6662
	grad calls	1.2449	1.0816	1.8163	1.8571	0.7959
25	time	1.2971	1.8973	2.0043	2.6563	0.5554
	grad calls	1.4286	1.7959	1.7143	2.1837	0.8163
30	time	1.8630	1.9192	1.9769	1.9992	0.3627
	grad calls	1.9792	2.3750	2.0625	1.8958	0.7083

Table 6.2: Time and number of gradient calls necessary to carry out the estimation of the various DCC models under consideration in different dimensions when using the Period II sample. The figures provided are relative to the computational effort necessary to estimate the scalar model in the corresponding dimension.

DJIA dataset. Period I.

n		Scalar	Almon	Almon Shuffle	Rank Deficient ($r=1$)	Rank Deficient ($r=2$)	Hadamard
5	-log$L$	-13.9889	-13.9892	-13.9912	-13.9926	-13.9932	**-13.9945**
	AIC$^{\text{rank}}$	-27.9665^{3}	-27.9644^{4}	-27.9683^{2}	**-27.9685**^{1}	-27.9643^{5}	-27.9591^{6}
	LR p-value	—	0.7736	**0.0085**	**0.0049**	0.0603	0.2032
10	-log$L$	-28.3038	-28.3042	-28.3038	-28.3102	-28.3130	**-28.3163**
	AIC$^{\text{rank}}$	-56.5862^{2}	-56.5843^{3}	-56.5835^{4}	**-56.5870**^{1}	-56.5807^{5}	-56.5392^{6}
	LR p-value	—	0.6537	1.0000	**0.0034**	**0.0191**	0.9933
15	-log$L$	-42.0486	-42.0489	-42.0510	-42.0592	-42.0670	**-42.0725**
	AIC$^{\text{rank}}$	-84.0659^{3}	-84.0637^{5}	-84.0680^{2}	**-84.0684**^{1}	-84.0653^{4}	-83.9550^{6}
	LR p-value	—	0.8443	**0.0067**	**0.0001**	**0.0000**	1.0000
20	-log$L$	-56.6593	-56.6597	-56.6613	-56.6742	-56.6855	**-56.6951**
	AIC$^{\text{rank}}$	-113.2773^{4}	-113.2753^{5}	-113.2786^{3}	**-113.2818**^{1}	-113.2790^{2}	-113.0702^{6}
	LR p-value	—	0.7419	**0.0183**	**0.0000**	**0.0000**	1.0000
25	-log$L$	-71.1817	-71.1819	-71.1835	-71.1985	-71.2108	**-71.2296**
	AIC$^{\text{rank}}$	-142.3120^{3}	-142.3098^{4}	-142.3129^{2}	**-142.3137**^{1}	-142.3062^{5}	-141.9760^{6}
	LR p-value	—	0.8475	**0.0295**	**0.0000**	**0.0000**	1.0000
30	-log$L$	-85.9232	-85.9253	-85.9267	-85.9523	-85.9599	**-86.0091**
	AIC$^{\text{rank}}$	-171.7851^{4}	-171.7866^{3}	-171.7894^{2}	**-171.8047**^{1}	-171.7811^{5}	-171.3381^{6}
	LR p-value	—	**0.0139**	**0.0004**	**0.0000**	**0.0000**	1.0000
	AIC score$^{\text{rank}}$	19^{3}	24^{4}	15^{2}	**6**^{1}	26^{5}	36^{6}

Table 6.3: Normalized values of minus the log-likelihood function (-log$L$), associated AIC statistics, and p-values of the likelihood ratio (LR) tests that take as null model the scalar one and all the others as alternatives. The smallest values of minus the log-likelihood function are displayed in black bold and the p-values smaller than 0.05 are shown in bold. Exponents on the AIC row indicate the rank of the model from 6 (the worse) to 1 (the best). The "AIC score" row at the bottom contains aggregated ranks by models. Figures in red point to the model that exhibits the lowest AIC value.

DJIA dataset. Period II.

n		Scalar	Almon	Almon Shuffle	Rank Deficient ($r=1$)	Rank Deficient ($r=2$)	Hadamard
5	-log$L$	-13.4686	-13.4690	-13.4713	-13.4728	-13.4735	**-13.4736**
	AIC$^{\text{rank}}$	-26.9258^{3}	-26.9239^{5}	-26.9286^{2}	**-26.9289**^{1}	-26.9250^{4}	-26.9173^{6}
	LR p-value	—	0.6671	**0.0024**	**0.0013**	**0.0205**	0.3437
10	-log$L$	-27.8050	-27.8059	-27.8084	-27.8149	-27.8178	**-27.8186**
	AIC$^{\text{rank}}$	-55.5888^{4}	-55.5878^{5}	-55.5928^{2}	**-55.5964**^{1}	-55.5902^{3}	-55.5439^{6}
	LR p-value	—	0.2675	**0.0005**	**0.0000**	**0.0001**	0.9735
15	-log$L$	-41.8992	-41.9006	-41.9026	-41.9164	-41.9235	**-41.9266**
	AIC$^{\text{rank}}$	-83.7671^{5}	-83.7671^{4}	-83.7712^{3}	**-83.7829**^{1}	-83.7784^{2}	-83.6633^{6}
	LR p-value	—	0.0888	**0.0004**	**0.0000**	**0.0000**	0.9999
20	-log$L$	-56.4619	-56.4629	-56.4662	-56.4860	-56.4969	**-56.5068**
	AIC$^{\text{rank}}$	-112.8824^{4}	-112.8819^{5}	-112.8885^{3}	**-112.9053**^{1}	-112.9018^{2}	-112.6935^{6}
	LR p-value	—	0.1750	**0.0000**	**0.0000**	**0.0000**	1.0000
25	-log$L$	-71.0132	-71.0148	-71.0172	-71.0357	-71.0523	**-71.0724**
	AIC$^{\text{rank}}$	-141.9750^{5}	-141.9757^{4}	-141.9805^{3}	-141.9880^{2}	**-141.9893**^{1}	-141.6614^{6}
	LR p-value	—	**0.0417**	**0.0001**	**0.0000**	**0.0000**	1.0000
30	-log$L$	-85.9028	-85.9063	-85.9065	-85.9347	-85.9565	**-85.9935**
	AIC$^{\text{rank}}$	-171.7443^{5}	-171.7487^{4}	-171.7490^{3}	-171.7694^{2}	**-171.7743**^{1}	-171.3070^{6}
	LR p-value	—	**0.0003**	**0.0002**	**0.0000**	**0.0000**	1.0000
	AIC score$^{\text{rank}}$	26^{4}	27^{5}	16^{3}	**8**^{1}	13^{2}	36^{6}

Table 6.4: Normalized values of minus the log-likelihood function (-log$L$), associated AIC statistics, and p-values of the likelihood ratio (LR) tests that take as null model the scalar one and all the others as alternatives. See caption of Table 6.3 for an explanation of the table entries.

DJIA dataset. Period I. Matrix A

n		McGyver	Scalar	Hadamard	Rank Deficient ($r=1$)	Rank Deficient ($r=2$)	Almon	Almon Shuffle
5	mean (median)	0.0111 (0.0077)	0.0040	0.0083	0.0030	0.0042	0.0048	0.0046
	std	0.0120	-	0.0078	0.0060	0.0071	0.0015	0.0069
	min	0.0000	0.0040	-0.0012	-0.0018	-0.0014	0.0018	0.0001
	max	0.0559	0.0040	0.0243	0.0172	0.0216	0.0070	0.0296
10	mean (median)	0.0097 (0.0069)	0.0046	0.0056	0.0076	0.0044	0.0050	0.0045
	std	0.0102	-	0.0046	0.0055	0.0037	0.0011	0.0016
	min	0.0000	0.0046	-0.0021	0.0024	-0.0015	0.0035	0.0025
	max	0.0466	0.0046	0.0279	0.0437	0.0193	0.0083	0.0101
15	mean (median)	0.0095 (0.0064)	0.0036	0.0038	0.0039	0.0032	0.0036	0.0036
	std	0.0105	-	0.0038	0.0026	0.0025	0.0004	0.0015
	min	0.0000	0.0036	-0.0030	0.0005	-0.0034	0.0024	0.0014
	max	0.0559	0.0036	0.0237	0.0172	0.0146	0.0042	0.0088
20	mean (median)	0.0088 (0.0059)	0.0026	0.0025	0.0029	0.0025	0.0026	0.0026
	std	0.0100	-	0.0024	0.0015	0.0020	0.0001	0.0006
	min	0.0000	0.0026	-0.0025	0.0006	-0.0034	0.0023	0.0009
	max	0.0559	0.0026	0.0128	0.0088	0.0100	0.0029	0.0035
25	mean (median)	0.0092 (0.0061)	0.0024	0.0023	0.0033	0.0029	0.0023	0.0024
	std	0.0099	-	0.0023	0.0020	0.0021	0.0000	0.0004
	min	0.0000	0.0024	-0.0037	0.0004	-0.0036	0.0023	0.0015
	max	0.0559	0.0024	0.0139	0.0162	0.0141	0.0024	0.0033
30	mean (median)	0.0091 (0.0060)	0.0021	0.0019	0.0028	0.0024	0.0021	0.0022
	std	0.0100	-	0.0023	0.0018	0.0012	0.0001	0.0006
	min	0.0000	0.0021	-0.0047	-0.0003	0.0005	0.0018	0.0010
	max	0.0578	0.0021	0.0127	0.0100	0.0090	0.0023	0.0036

DJIA dataset. Period I. Matrix B

n		McGyver	Scalar	Hadamard	Rank Deficient ($r=1$)	Rank Deficient ($r=2$)	Almon	Almon Shuffle
5	mean (median)	0.7731 (0.9484)	0.9764	0.9354	0.9873	0.9774	0.9692	0.9696
	std	0.3311	-	0.0439	0.0115	0.0106	0.0107	0.0088
	min	0.0000	0.9764	0.8700	0.9576	0.9560	0.9539	0.9584
	max	0.9974	0.9764	0.9803	0.9994	0.9990	0.9943	0.9921
10	mean (median)	0.8623 (0.9698)	0.9801	0.9745	0.8791	0.9804	0.9769	0.9797
	std	0.2744	-	0.0081	0.1978	0.0094	0.0069	0.0097
	min	0.0001	0.9801	0.9580	0.2466	0.9600	0.9606	0.9509
	max	0.9974	0.9801	0.9917	0.9955	0.9971	0.9877	0.9921
15	mean (median)	0.8707 (0.9652)	0.9805	0.9781	0.9701	0.9810	0.9799	0.9793
	std	0.2372	-	0.0069	0.0230	0.0130	0.0014	0.0106
	min	0.0000	0.9805	0.9607	0.8915	0.9305	0.9777	0.9434
	max	0.9974	0.9805	0.9935	0.9972	0.9965	0.9835	0.9933
20	mean (median)	0.8514 (0.9656)	0.9871	0.9862	0.9756	0.9854	0.9865	0.9865
	std	0.2717	-	0.0030	0.0247	0.0080	0.0012	0.0029
	min	0.0000	0.9871	0.9795	0.8382	0.9580	0.9847	0.9780
	max	0.9977	0.9871	0.9935	0.9975	0.9970	0.9907	0.9913
25	mean (median)	0.8398 (0.9610)	0.9857	0.9851	0.9589	0.9748	0.9861	0.9847
	std	0.2741	-	0.0027	0.0436	0.0200	0.0014	0.0028
	min	0.0000	0.9857	0.9792	0.7953	0.8685	0.9833	0.9766
	max	0.9977	0.9857	0.9926	0.9970	0.9970	0.9900	0.9892
30	mean (median)	0.8364 (0.9617)	0.9864	0.9857	0.9689	0.9779	0.9866	0.9841
	std	0.2781	-	0.0022	0.0272	0.0159	0.0026	0.0047
	min	0.0000	0.9864	0.9795	0.8425	0.9033	0.9828	0.9674
	max	0.9977	0.9864	0.9932	0.9971	0.9958	0.9960	0.9911

Table 6.5: Period I. Estimated parameter matrices A and B of (3.5). Mean, median, standard deviation, minimum and maximum of the entries of the estimated matrices are reported. "McGyver" stands for the method with this name introduced in [Engl 08a] based on the use of bivariate scalar DCC models; in this case the median value of estimates are reported between parentheses.

DJIA dataset. Period II. Matrix A

n		McGyver	Scalar	Hadamard	Rank Deficient ($r=1$)	Rank Deficient ($r=2$)	Almon	Almon Shuffle
5	mean (median)	0.0103 (0.0086)	0.0063	0.0093	0.0071	0.0092	0.0064	0.0065
	std	0.0090	-	0.0102	0.0054	0.0106	0.0039	0.0042
	min	0.0000	0.0063	0.0004	0.0021	-0.0000	0.0026	0.0026
	max	0.0432	0.0063	0.0493	0.0276	0.0520	0.0194	0.0201
10	mean (median)	0.0118 (0.0101)	0.0060	0.0066	0.0065	0.0063	0.0060	0.0062
	std	0.0089	-	0.0036	0.0025	0.0028	0.0003	0.0013
	min	0.0000	0.0060	-0.0015	0.0025	0.0012	0.0056	0.0045
	max	0.0432	0.0060	0.0189	0.0124	0.0122	0.0071	0.0112
15	mean (median)	0.0101 (0.0079)	0.0044	0.0047	0.0048	0.0045	0.0044	0.0045
	std	0.0082	-	0.0038	0.0029	0.0035	0.0008	0.0014
	min	0.0000	0.0044	-0.0032	0.0005	-0.0022	0.0021	0.0026
	max	0.0432	0.0044	0.0243	0.0193	0.0197	0.0055	0.0100
20	mean (median)	0.0099 (0.0074)	0.0034	0.0033	0.0037	0.0031	0.0035	0.0035
	std	0.0089	-	0.0026	0.0022	0.0023	0.0005	0.0005
	min	0.0000	0.0034	-0.0039	0.0004	-0.0030	0.0027	0.0021
	max	0.0535	0.0034	0.0162	0.0144	0.0110	0.0052	0.0042
25	mean (median)	0.0093 (0.0069)	0.0031	0.0029	0.0032	0.0030	0.0031	0.0031
	std	0.0093	-	0.0023	0.0018	0.0017	0.0005	0.0003
	min	0.0000	0.0031	-0.0027	0.0008	-0.0010	0.0022	0.0021
	max	0.0545	0.0031	0.0155	0.0130	0.0098	0.0046	0.0035
30	mean (median)	0.0093 (0.0070)	0.0027	0.0025	0.0027	0.0025	0.0026	0.0027
	std	0.0092	-	0.0023	0.0019	0.0021	0.0005	0.0002
	min	0.0000	0.0027	-0.0040	0.0004	-0.0021	0.0020	0.0020
	max	0.0592	0.0027	0.0138	0.0112	0.0131	0.0047	0.0031

DJIA dataset. Period II. Matrix B

n		McGyver	Scalar	Hadamard	Rank Deficient ($r=1$)	Rank Deficient ($r=2$)	Almon	Almon Shuffle
5	mean (median)	0.8748 (0.9673)	0.9815	0.9728	0.9694	0.9725	0.9819	0.9788
	std	0.2433	-	0.0121	0.0170	0.0126	0.0106	0.0057
	min	0.0006	0.9815	0.9461	0.9265	0.9436	0.9550	0.9672
	max	0.9945	0.9815	0.9928	0.9944	0.9937	0.9956	0.9881
10	mean	0.8748 (0.9647)	0.9801	0.9742	0.9667	0.9733	0.9785	0.9772
	std	0.2617	-	0.0088	0.0228	0.0148	0.0051	0.0022
	min	0.0000	0.9801	0.9478	0.8785	0.9197	0.9641	0.9743
	max	0.9949	0.9801	0.9928	0.9962	0.9962	0.9865	0.9842
15	mean (median)	0.8874 (0.9658)	0.9827	0.9797	0.9737	0.9788	0.9822	0.9815
	std	0.2309	-	0.0070	0.0158	0.0141	0.0009	0.0049
	min	0.0000	0.9827	0.9626	0.9258	0.9297	0.9808	0.9653
	max	0.9975	0.9827	0.9942	0.9984	0.9970	0.9851	0.9878
20	mean (median)	0.8769 (0.9672)	0.9866	0.9854	0.9806	0.9857	0.9862	0.9851
	std	0.2415	-	0.0038	0.0104	0.0073	0.0013	0.0029
	min	0.0000	0.9866	0.9758	0.9434	0.9601	0.9840	0.9805
	max	0.9975	0.9866	0.9951	0.9985	0.9985	0.9900	0.9942
25	mean (median)	0.8752 (0.9724)	0.9862	0.9857	0.9816	0.9848	0.9860	0.9853
	std	0.2490	-	0.0026	0.0094	0.0076	0.0007	0.0028
	min	0.0000	0.9862	0.9791	0.9465	0.9485	0.9851	0.9812
	max	0.9975	0.9862	0.9925	0.9964	0.9977	0.9881	0.9950
30	mean (median)	0.8776 (0.9718)	0.9866	0.9858	0.9836	0.9844	0.9867	0.9858
	std	0.2391	-	0.0020	0.0099	0.0089	0.0021	0.0029
	min	0.0000	0.9866	0.9781	0.9383	0.9420	0.9839	0.9815
	max	0.9975	0.9866	0.9920	0.9985	0.9974	0.9943	0.9964

Table 6.6: Period II. Estimated parameter matrices A and B of (3.5). Mean, median, standard deviation, minimum and maximum of the entries of the estimated matrices are reported. "McGyver" stands for the method with this name introduced in [Engl 08a] based on the use of bivariate scalar DCC models; in this case the median value of estimates are reported between parentheses.

DJIA dataset. Period I

n	a_1	a_2	a_3	b_1	b_2	b_3
5	-0.97647	-0.02374	0.01028	0.01344	-0.02047	0.00293
10	-0.94870	-0.00176	0.00065	-0.00758	0.00247	-0.00042
15	-0.96304	0.00090	0.00019	-0.01338	0.00346	-0.00030
20	-0.97398	0.00448	-0.00015	-0.00434	-0.00001	-0.00002
25	-0.96169	0.00099	-0.00001	-0.01266	0.00112	-0.00004
30	-0.96930	0.00124	-0.00001	-0.01275	0.00121	-0.00004

DJIA dataset. Period II

n	a_1	a_2	a_3	b_1	b_2	b_3
5	-0.78996	-0.07976	0.00909	-0.02519	0.01330	-0.00232
10	-0.87945	-0.01598	0.00117	-0.00609	-0.00197	0.00014
15	-0.90004	-0.01347	0.00090	-0.00926	0.00122	-0.00012
20	-0.93593	0.00045	-0.00007	-0.00144	-0.00152	0.00007
25	-0.94422	0.00077	-0.00005	-0.00140	-0.00112	0.00004
30	-0.95261	0.00113	-0.00004	-0.00088	-0.00095	0.00003

Table 6.7: Estimates of the parameters of the Almon functions $(\text{alm}_n(\mathbf{a}))_i = a_1 + \exp(a_2 i + a_3 i^2)$ and $(\text{alm}_n(\mathbf{b}))_i = b_1 + \exp(b_2 i + b_3 i^2)$ for the Almon shuffle models. The symbol i represents the entry number and n is the dimension of the model.

DJIA dataset. Period I. MCS for the correlation of the standardized returns.

n		Scalar	Hadamard	Rank Deficient ($r=1$)	Rank Deficient ($r=2$)	Almon	Almon Shuffle
5	Position	**1**	**6**	**3**	**5**	**2**	**4**
	p-value	1.000	0.066	0.170	0.170	0.170	0.170
10	Position	**2**	**6**	**1**	**3**	**4**	**5**
	p-value	0.108	0.085	1.000	0.108	0.108	0.108
15	Position	**2**	6	**1**	**4**	**3**	**5**
	p-value	0.683	0.001	1.000	0.159	0.159	0.159
20	Position	**1**	6	3	5	2	4
	p-value	1.000	0.001	0.025	0.025	0.025	0.025
25	Position	2	6	5	3	**1**	4
	p-value	0.004	0.000	0.004	0.004	1.000	0.004
30	Position	2	6	5	3	**1**	4
	p-value	0.000	0.000	0.000	0.000	1.000	0.000
	Score	10	36	18	23	13	26

Table 6.8: Model confidence sets (MCS) constructed using the loss function (5.2) based on the correlation of the standardized returns defined in (5.3). For each model and dimension, the integer value in the first row indicates the order of elimination of the model from the MCS (6 stands for the first eliminated model, 5 for the second eliminated model, and so on). In the second row we report the p-value of the test leading to the decision of eliminating the given model from the MCS. The set of integer values printed in bold red identifies the MCS at the confidence level of 95%. The union of integer values printed in bold black and bold red identifies the MCS at the confidence level of 90%. The score of each model in the last row is the sum of the integer values of the six dimensions.

DJIA dataset. Period II. MCS for the correlation of the standardized returns.

n		Scalar	Hadamard	Rank Deficient ($r=1$)	Rank Deficient ($r=2$)	Almon	Almon Shuffle
5	Position	5	**4**	**1**	**3**	6	**2**
	p-value	0.008	0.073	1.000	0.228	0.005	0.228
10	Position	**5**	**3**	**1**	**4**	**6**	**2**
	p-value	0.190	0.433	1.000	0.260	0.190	0.433
15	Position	**3**	**5**	**4**	**6**	**2**	**1**
	p-value	0.242	0.127	0.242	0.061	0.811	1.000
20	Position	3	6	5	4	2	**1**
	p-value	0.003	0.003	0.003	0.003	0.026	1.000
25	Position	3	6	5	4	**2**	**1**
	p-value	0.002	0.000	0.001	0.002	0.074	1.000
30	Position	3	6	4	5	**2**	**1**
	p-value	0.006	0.000	0.006	0.001	0.394	1.000
	Score	22	30	20	26	20	8

Table 6.9: Model confidence sets (MCS) constructed using the loss function (5.2) based on the correlation of the standardized returns defined in (5.3). See caption of Table 6.8 for an explanation of the table entries.

DJIA dataset. Period I. Engle-Colacito regression test for the minimum variance portfolio returns.

n		Scalar	Hadamard	Rank Deficient ($r=1$)	Rank Deficient ($r=2$)	Almon	Almon Shuffle
5	t-stats	**3.76***	4.04*	4.13*	4.09*	3.81*	3.98*
	p-value	0.00	0.00	0.00	0.00	0.00	0.00
10	t-stats	1.73	1.80	1.67	1.90	1.74	**1.65**
	p-value	0.08	0.07	0.10	0.06	0.08	0.10
15	t-stats	-0.55	-0.16	**0.05**	0.26	-0.45	-0.16
	p-value	0.58	0.87	0.96	0.80	0.65	0.87
20	t-stats	-0.51	**-0.02**	0.44	0.43	-0.43	-0.22
	p-value	0.61	0.99	0.66	0.67	0.66	0.82
25	t-stats	**1.47**	2.17*	2.48*	2.15*	1.56	1.96
	p-value	0.14	0.03	0.01	0.03	0.12	0.05
30	t-stats	**2.70***	3.73*	3.85*	3.60*	2.97*	3.43*
	p-value	0.01	0.00	0.00	0.00	0.00	0.00

Table 6.10: Results of the Engle-Colacito regression test. The t-stat values refer to the intercept λ of the Engle-Colacito regression (5.7) obtained using a HAC estimator. The p-value can be used to test the null hypothesis that $\lambda = 0$. The symbol * (respectively **) indicates rejection at the 5% (respectively 1%) significance level. Values corresponding to models that exhibit the maximum p-value for a given portfolio cardinality are printed in red.

DJIA dataset. Period I. Engle-Colacito regression test for the equally weighted portfolio returns.

n		Scalar	Hadamard	Rank Deficient ($r=1$)	Rank Deficient ($r=2$)	Almon	Almon Shuffle
5	t-stats	**2.72***	2.80*	2.87*	2.78*	2.77*	2.82*
	p-value	0.01	0.01	0.00	0.01	0.01	0.00
10	t-stats	1.99*	1.91	**1.84**	2.01*	2.06*	1.91
	p-value	0.05	0.06	0.07	0.04	0.04	0.06
15	t-stats	**-1.21**	-1.57	-1.24	-1.44	-1.22	-1.29
	p-value	0.23	0.12	0.22	0.15	0.22	0.20
20	t-stats	-2.32*	-2.89*	**-2.16***	-2.84*	-2.29*	-2.28*
	p-value	0.02	0.00	0.03	0.00	0.02	0.02
25	t-stats	-1.94	-2.39*	**-1.68**	-2.36*	-1.86	-1.76
	p-value	0.05	0.02	0.09	0.02	0.06	0.08
30	t-stats	-3.45*	-3.88*	-3.10*	**-3.01***	-3.18*	-3.12*
	p-value	0.00	0.00	0.00	0.00	0.00	0.00

Table 6.11: Results of the Engle-Colacito regression test. The t-stat values refer to the intercept λ of the Engle-Colacito regression (5.7) obtained using a HAC estimator. The p-value can be used to test the null hypothesis that $\lambda = 0$. The symbol * (respectively **) indicates rejection at the 5% (respectively 1%) significance level. Values corresponding to models that exhibit the maximum p-value for a given portfolio cardinality are printed in red.

DJIA dataset. Period II. Engle-Colacito regression test for the minimum variance portfolio returns.

n		Scalar	Hadamard	Rank Deficient ($r=1$)	Rank Deficient ($r=2$)	Almon	Almon Shuffle
5	t-stats	-3.81*	-3.77*	-3.58*	-3.70*	-3.87*	**-3.49***
	p-value	0.00	0.00	0.00	0.00	0.00	0.00
10	t-stats	0.39	0.30	0.38	**0.25**	0.33	0.29
	p-value	0.70	0.76	0.71	0.81	0.74	0.77
15	t-stats	**0.38**	0.66	0.40	0.49	0.43	0.41
	p-value	0.71	0.51	0.69	0.63	0.67	0.68
20	t-stats	0.58	1.01	0.72	0.82	0.62	**0.50**
	p-value	0.56	0.31	0.47	0.41	0.54	0.62
25	t-stats	1.96	2.77*	2.29*	2.31*	2.06*	**1.83**
	p-value	0.05	0.01	0.02	0.02	0.04	0.07
30	t-stats	2.58*	3.63*	2.73*	3.04*	2.73*	**2.35***
	p-value	0.01	0.00	0.01	0.00	0.01	0.02

Table 6.12: Results of the Engle-Colacito regression test. The t-stat values refer to the intercept λ of the Engle-Colacito regression (5.7) obtained using a HAC estimator. The p-value can be used to test the null hypothesis that $\lambda = 0$. The symbol * (respectively **) indicates rejection at the 5% (respectively 1%) significance level. Values corresponding to models that exhibit the maximum p-value for a given portfolio cardinality are printed in red.

DJIA dataset. Period II. Engle-Colacito regression test for the equally weighted portfolio returns.

n		Scalar	Hadamard	Rank Deficient ($r=1$)	Rank Deficient ($r=2$)	Almon	Almon Shuffle
5	t-stats	-2.42*	-2.48*	-2.32*	-2.41*	-2.54*	**-2.23***
	p-value	0.02	0.01	0.02	0.02	0.01	0.03
10	t-stats	**-2.00***	-2.30*	-2.08*	-2.36*	-2.13*	-2.32*
	p-value	0.05	0.02	0.04	0.02	0.03	0.02
15	t-stats	-1.70	-1.62	-1.89	**-1.60**	-1.71	-1.80
	p-value	0.09	0.11	0.06	0.11	0.09	0.07
20	t-stats	-2.84*	-2.57*	-2.63*	**-2.50***	-2.82*	-3.20*
	p-value	0.00	0.01	0.01	0.01	0.00	0.00
25	t-stats	-2.31*	**-1.87**	-1.98*	-2.03*	-2.25*	-2.63*
	p-value	0.02	0.06	0.05	0.04	0.02	0.01
30	t-stats	-2.97*	-2.82*	-2.87*	**-2.68***	-2.86*	-3.33*
	p-value	0.00	0.00	0.00	0.01	0.00	0.00

Table 6.13: Results of the Engle-Colacito regression test. The t-stat values refer to the intercept λ of the Engle-Colacito regression (5.7) obtained using a HAC estimator. The p-value can be used to test the null hypothesis that $\lambda = 0$. The symbol * (respectively **) indicates rejection at the 5% (respectively 1%) significance level. Values corresponding to models that exhibit the maximum p-value for a given portfolio cardinality are printed in red.

DJIA dataset. Period I. MCS of EPA for the minimum variance portfolio squared returns.

n		Scalar	Hadamard	Rank Deficient ($r=1$)	Rank Deficient ($r=2$)	Almon	Almon Shuffle
5	Position	**3**	**2**	**4**	**1**	**5**	**6**
	p-value	0.856	0.856	0.810	1.000	0.810	0.810
10	Position	**4**	**2**	**1**	**3**	**5**	**6**
	p-value	0.335	0.579	1.000	0.335	0.335	0.270
15	Position	**5**	**4**	**1**	**3**	**6**	**2**
	p-value	0.374	0.588	1.000	0.600	0.290	0.600
20	Position	**2**	**5**	**4**	**6**	**3**	**1**
	p-value	0.206	0.206	0.206	0.206	0.206	1.000
25	Position	**3**	**6**	**2**	**5**	**4**	**1**
	p-value	0.803	0.465	0.910	0.803	0.803	1.000
30	Position	**1**	**6**	**5**	**4**	**3**	**2**
	p-value	1.000	0.383	0.407	0.407	0.407	0.638
	Score	18	25	17	22	26	18

Table 6.14: Model confidence sets based on the predictive ability for squared portfolio returns using the loss function defined in (5.8). See the caption of Table 6.8 for an explanation of the table entries.

DJIA dataset. Period I. MCS of EPA for the equally weighted portfolio squared returns.

n		Scalar	Hadamard	Rank Deficient ($r=1$)	Rank Deficient ($r=2$)	Almon	Almon Shuffle
5	Position	**2**	**5**	**4**	6	**1**	**3**
	p-value	0.296	0.296	0.296	0.002	1.000	0.296
10	Position	**3**	**4**	6	**2**	**1**	5
	p-value	0.071	0.071	0.001	0.345	1.000	0.001
15	Position	**1**	6	**2**	5	**3**	**4**
	p-value	1.000	0.005	0.685	0.005	0.326	0.199
20	Position	**2**	6	**1**	5	**3**	**4**
	p-value	0.367	0.006	1.000	0.011	0.367	0.158
25	Position	4	5	**1**	6	**2**	3
	p-value	0.005	0.004	1.000	0.004	0.061	0.049
30	Position	5	6	**2**	**1**	3	4
	p-value	0.002	0.002	0.136	1.000	0.018	0.018
	Score	17	32	16	25	13	23

Table 6.15: Model confidence sets based on the predictive ability for squared portfolio returns using the loss function defined in (5.8). See the caption of Table 6.8 for an explanation of the table entries.

DJIA dataset. Period II. MCS of EPA for the minimum variance portfolio squared returns.

n		Scalar	Hadamard	Rank Deficient ($r=1$)	Rank Deficient ($r=2$)	Almon	Almon Shuffle
5	Position	**2**	**6**	**1**	**3**	**5**	**4**
	p-value	0.562	0.214	1.000	0.562	0.214	0.562
10	Position	**4**	**6**	**3**	**5**	**2**	**1**
	p-value	0.379	0.071	0.379	0.379	0.379	1.000
15	Position	**5**	**3**	**1**	**4**	**2**	**6**
	p-value	0.368	0.676	1.000	0.676	0.676	0.368
20	Position	**4**	**2**	**1**	**5**	**3**	**6**
	p-value	0.339	0.932	1.000	0.339	0.836	0.339
25	Position	**4**	**6**	**1**	**3**	**2**	**5**
	p-value	0.618	0.100	1.000	0.618	0.618	0.618
30	Position	**3**	**6**	**1**	**5**	**2**	**4**
	p-value	0.544	0.054	1.000	0.312	0.808	0.544
	Score	22	29	8	25	16	26

Table 6.16: Model confidence sets based on the predictive ability for squared portfolio returns using the loss function defined in (5.8). See the caption of Table 6.8 for an explanation of the table entries.

DJIA dataset. Period II. MCS of EPA for the equally weighted portfolio squared returns.

n		Scalar	Hadamard	Rank Deficient ($r=1$)	Rank Deficient ($r=2$)	Almon	Almon Shuffle
5	Position	4	5	2	3	6	**1**
	p-value	0.000	0.000	0.029	0.000	0.000	1.000
10	Position	**1**	4	**2**	6	3	5
	p-value	1.000	0.000	0.438	0.000	0.000	0.000
15	Position	**3**	**2**	6	**1**	**4**	5
	p-value	0.486	0.772	0.001	1.000	0.484	0.003
20	Position	**5**	**2**	**3**	**1**	**4**	6
	p-value	0.061	0.374	0.374	1.000	0.061	0.000
25	Position	5	**1**	**2**	3	4	6
	p-value	0.000	1.000	0.674	0.000	0.000	0.000
30	Position	5	3	4	**1**	2	6
	p-value	0.000	0.036	0.002	1.000	0.036	0.000
	Score	23	17	19	15	23	29

Table 6.17: Results of the model confidence set of EPA for equally weighted portfolio squared returns. The loss function is defined in (5.8). See caption of Table 6.8 for an explanation of the table entries.

DJIA dataset. Period I. HIT test of the minimum variance portfolio returns.

n		Scalar	Hadamard	Rank Deficient ($r=1$)	Rank Deficient ($r=2$)	Almon	Almon Shuffle
5	1%	**2.54**	3.07	3.07	2.94	2.67	2.80
	p-value	0.7895	0.9538	**0.9549**	0.9046	0.8707	0.9205
	5%	**6.81***	7.21*	6.94*	6.94*	6.94*	**6.81***
	p-value	0.0170	0.0108	0.0112	0.0112	**0.0223**	0.0172
	10%	11.08**	11.48**	11.75*	11.48*	**10.95***	11.62
	p-value	0.0085	0.0075	0.0211	0.0295	0.0362	**0.0811**
10	1%	**1.60**	1.87	1.74	1.74	**1.60**	**1.60**
	p-value	0.5135	0.7272	**0.9820**	0.5750	0.5126	0.5120
	5%	4.54	**4.67**	4.54	**4.67**	4.54	4.54
	p-value	0.6107	0.5839	**0.7906**	0.6146	0.6113	0.6114
	10%	9.35	8.95	8.68	**9.48**	9.21	9.21
	p-value	0.1459	0.0816	0.1247	0.0738	0.1585	**0.1588**
15	1%	**0.93**	1.07	1.20	1.20	**0.93**	**0.93**
	p-value	0.9965	0.1357	0.0845	0.2323	**0.9968**	0.9968
	5%	4.14	4.41	4.41	**4.54**	4.14	4.14
	p-value	0.2980	0.4356	0.4524	**0.4987**	0.2977	0.2992
	10%	7.61	8.14	8.68	**8.81**	7.88	8.54
	p-value	0.4368	0.6871	**0.8502**	0.5405	0.4536	0.8441
20	1%	**1.20**	1.47	1.47	1.34	1.34	1.47
	p-value	0.2141	0.9301	0.5153	**0.9826**	0.3886	0.5197
	5%	3.47	3.87	**4.14**	4.01	3.47	3.60
	p-value	0.7794	0.7344	0.7765	**0.7980**	0.7794	0.7789
	10%	7.48	8.28	**8.41**	**8.41**	7.61	7.74
	p-value	0.9621	0.7913	0.9138	0.7433	**0.9693**	0.9448
25	1%	**1.34**	1.47	1.87	1.74*	**1.34**	1.60*
	p-value	0.4985	**0.9939**	0.0772	0.0350	0.4985	0.0105
	5%	4.67	4.81	**5.07**	5.21	4.67	4.81
	p-value	0.6049	0.4531	**0.7543**	0.4600	0.6049	0.5051
	10%	9.35	10.68	10.55	**9.88**	9.48	9.75
	p-value	0.7228	0.9699	0.9094	**0.9739**	0.5810	0.7412
30	1%	**1.74**	2.54	2.80	2.40	1.87	2.27
	p-value	0.7757	0.9153	0.9129	0.9175	0.8351	**0.9213**
	5%	**5.74**	6.81	7.34	7.08	6.01	6.54
	p-value	0.9235	0.9682	**0.9890**	0.9019	0.9568	0.6951
	10%	**10.28**	12.28	12.02	11.21	10.68	10.81
	p-value	0.8853	0.6322	**0.9880**	0.7384	0.9599	0.9487

Table 6.18: Results of the HIT test. For each assets cardinality n and model we report the average number of VaR violations when this risk measure is computed at the 1, 5, and 10% confidence levels. The percentage printed in red corresponds to the model that yields the closest number to the specified confidence level. The p-values correspond to the F-test on the HIT regression (5.9) with five lags. The highest p-values are marked in bold. When they imply that the null hypothesis of independence is rejected at the 5% (respectively 1%) level, the corresponding average number of VaR violations is marked with * (respectively **).

DJIA dataset. Period I. HIT test of the equally weighted portfolio returns.

n		Scalar	Hadamard	Rank Deficient ($r=1$)	Rank Deficient ($r=2$)	Almon	Almon Shuffle
5	1%	2.27	2.27	2.27	2.27	**2.14**	2.27
	p-value	0.5905	0.6204	**0.6210**	0.6204	0.4791	0.6198
	5%	6.28	6.54	6.54	6.54	**6.01**	6.54
	p-value	**0.7918**	0.6796	0.6793	0.6776	0.5968	0.6796
	10%	**11.88**	**11.88**	**11.88**	**11.88**	**11.88**	12.02
	p-value	**0.3977**	0.2019	0.2019	0.2017	0.3792	0.2615
10	1%	1.47	**1.34**	1.47	**1.34**	1.47	1.47
	p-value	0.9944	0.9936	**0.9944**	0.9931	0.9943	0.9944
	5%	6.28	**6.14**	6.28	**6.14**	6.28	6.28
	p-value	0.0708	**0.1067**	0.0717	0.1066	0.0703	0.0704
	10%	9.48*	**9.61***	9.48*	**9.61***	9.48*	9.35
	p-value	0.0390	0.0333	0.0403	0.0309	0.0382	**0.2340**
15	1%	**1.07**	**1.07**	**1.07**	**1.07**	**1.07**	**1.07**
	p-value	**0.9961**	0.9960	0.9952	0.9952	0.9959	0.9956
	5%	**3.87**	3.60	**3.87**	3.60	**3.87**	**3.87**
	p-value	**0.2759**	0.2613	0.2753	0.2594	0.2758	0.2756
	10%	**6.81**	6.54	6.41	6.41	6.68	6.54
	p-value	**0.1950**	0.1605	0.0975	0.0969	0.1506	0.1652
20	1%	1.20	**0.93***	1.20	**0.93***	1.20	1.20
	p-value	0.3060	0.0453	0.3059	0.0452	0.3061	**0.3066**
	5%	**2.67**	2.54	**2.67**	**2.67**	**2.67**	**2.67**
	p-value	0.3405	0.2177	0.3399	0.3356	**0.3410**	0.3409
	10%	**5.34**	5.21	**5.34**	5.21	**5.34**	**5.34**
	p-value	0.7912	0.7699	0.4568	0.7709	**0.7925**	0.4519
25	1%	**0.80**	0.53	**0.80**	0.53	**0.80**	0.67
	p-value	0.9998	0.9253	0.9998	0.9184	**0.9998**	0.9992
	5%	3.34	3.20	**3.60**	3.07	3.47	3.47
	p-value	0.8904	0.9086	0.7351	**0.9182**	0.8690	0.8687
	10%	6.14	6.01*	**6.28**	6.01*	6.14*	6.14*
	p-value	0.0501	0.0350	**0.0719**	0.0349	0.0499	0.0497
30	1%	0.80	**0.93**	**0.93**	**0.93**	0.80	0.80
	p-value	**0.9954**	0.9759	0.9752	0.9778	0.9952	0.9951
	5%	2.94	2.94	**3.07**	**3.07**	2.94	**3.07**
	p-value	0.4821	0.4812	0.5690	**0.5697**	0.4807	0.5674
	10%	6.01	5.87	6.14	**6.28**	6.14	**6.28**
	p-value	0.6916	**0.9274**	0.7334	0.6628	0.7295	0.6599

Table 6.19: Results of the HIT test. For each assets cardinality n and model we report the average number of VaR violations when this risk measure is computed at the 1, 5, and 10% confidence levels. The percentage printed in red corresponds to the model that yields the closest number to the specified confidence level. The p-values correspond to the F-test on the HIT regression (5.9) with five lags. The highest p-values are marked in bold. When they imply that the null hypothesis of independence is rejected at the 5% (respectively 1%) level, the corresponding average number of VaR violations is marked with * (respectively **).

DJIA dataset. Period II. HIT test of the minimum variance portfolio returns.

n		Scalar	Hadamard	Rank Deficient ($r=1$)	Rank Deficient ($r=2$)	Almon	Almon Shuffle
5	1%	0.40	0.40	**0.53**	0.40	**0.53**	**0.53**
	p-value	1.0000	**1.0000**	0.9999	1.0000	1.0000	0.9999
	5%	3.60	3.47	3.60	3.60	3.60	**3.74**
	p-value	0.3355	0.2632	0.3367	0.3349	0.3337	**0.5322**
	10%	**8.95**	8.54	8.54	8.54	8.81	8.68
	p-value	**0.5005**	0.2296	0.2278	0.2292	0.4985	0.4310
10	1%	1.60	**1.47**	**1.47**	**1.47**	**1.47**	**1.47**
	p-value	0.9916	0.9924	0.9940	0.9933	**0.9955**	0.9946
	5%	**4.81**	**4.81**	**4.81**	**4.81**	**4.81**	4.67
	p-value	0.7859	0.7784	0.7780	0.7850	0.7894	**0.9189**
	10%	7.34	7.34	**7.74**	7.34	7.48	7.34
	p-value	**0.8484**	0.8423	0.7044	0.7633	0.8345	0.8480
15	1%	**1.07**	1.60	1.34	1.47	1.34	1.20*
	p-value	0.0994	0.1511	**0.3995**	0.1273	0.0579	0.0105
	5%	**4.94**	4.54	4.67	4.81	5.07	4.81
	p-value	**0.8766**	0.6078	0.6573	0.5982	0.8753	0.8643
	10%	8.54	9.08	8.68	**9.21**	8.54	8.81
	p-value	0.9149	**0.9735**	0.9310	0.9350	0.9061	0.9493
20	1%	**1.34***	1.47	**1.34***	**1.34***	**1.34***	**1.34***
	p-value	0.0314	**0.0939**	0.0295	0.0332	0.0305	0.0317
	5%	**4.94**	4.81	**4.94**	4.81	5.07	4.81
	p-value	0.9438	0.9355	0.9262	0.9370	**0.9476**	0.9107
	10%	8.81	10.28	9.61	**10.01**	8.81	8.81
	p-value	0.9216	0.9227	0.9930	**0.9932**	0.9157	0.9228
25	1%	**1.07**	1.60	1.47	1.47	1.20	**1.07**
	p-value	**0.9990**	0.2535	0.5740	0.1508	0.3132	0.9990
	5%	4.67	5.34	**5.07**	4.81	4.81	4.67
	p-value	0.8538	0.3596	**0.9060**	0.4802	0.8694	0.8544
	10%	11.21	12.15	11.62	11.35	11.48	**10.95**
	p-value	0.9361	0.9297	**0.9829**	0.9757	0.9566	0.9548
30	1%	**1.07**	1.60	1.20	1.34	1.20	**1.07**
	p-value	**0.9997**	0.6447	0.2330	0.4011	0.2327	0.9997
	5%	**4.94**	5.87	5.21	5.61	5.07	5.07
	p-value	0.3227	0.2609	0.7368	**0.8300**	0.3966	0.4063
	10%	10.95	12.42	11.75	11.88	11.62	**10.81**
	p-value	0.5710	**0.8997**	0.8530	0.8707	0.7243	0.6226

Table 6.20: Results of the HIT test. For each assets cardinality n and model we report the average number of VaR violations when this risk measure is computed at the 1, 5, and 10% confidence levels. The percentage printed in red corresponds to the model that yields the closest number to the specified confidence level. The p-values correspond to the F-test on the HIT regression (5.9) with five lags. The highest p-values are marked in bold. When they imply that the null hypothesis of independence is rejected at the 5% (respectively 1%) level, the corresponding average number of VaR violations is marked with * (respectively **).

DJIA dataset. Period II. HIT test of the equally weighted portfolio returns.

n		Scalar	Hadamard	Rank Deficient ($r=1$)	Rank Deficient ($r=2$)	Almon	Almon Shuffle
5	1%	0.53	**0.67**	**0.67**	**0.67**	0.53	**0.67**
	p-value	0.9864	0.9986	**0.9987**	0.9985	0.9870	0.9981
	5%	**3.20**	**3.20**	**3.20**	**3.20**	**3.20**	**3.20**
	p-value	0.0982	0.0985	0.0963	0.0986	**0.0995**	0.0975
	10%	**8.28**	8.14	8.14	8.14	8.14	**8.28**
	p-value	0.9162	0.9483	**0.9487**	0.9481	0.9483	0.7887
10	1%	**1.34**	**1.34**	**1.34**	**1.34**	**1.34**	**1.34**
	p-value	**0.9959**	0.9949	0.9948	0.9951	0.9956	0.9944
	5%	**4.01**	**4.01**	**4.01**	**4.01**	**4.01**	**4.01**
	p-value	**0.4566**	0.4355	0.4183	0.4350	0.4543	0.4417
	10%	**6.94**	**6.94**	**6.94**	6.68	**6.94**	6.81
	p-value	0.2370	**0.3799**	0.3659	0.1903	0.2331	0.2884
15	1%	**2.00**	**2.00**	**2.00**	**2.00**	**2.00**	**2.00**
	p-value	0.6435	0.6407	0.6429	0.6388	**0.6465**	0.6439
	5%	4.01	**4.14**	4.01	**4.14**	**4.14**	4.01
	p-value	0.8840	0.9467	0.8897	**0.9521**	0.9361	0.8837
	10%	7.48	7.88	7.74	**8.01**	7.61	7.48
	p-value	0.9053	0.8020	0.8267	0.6327	0.9059	**0.9059**
20	1%	1.34	1.34	1.34	1.47	1.34	**1.20**
	p-value	0.9964	0.9961	0.9957	0.9899	**0.9967**	0.9902
	5%	**3.87**	3.74	**3.87**	3.74	**3.87**	**3.87**
	p-value	0.9354	**0.9414**	0.9368	0.9412	0.9358	0.9351
	10%	7.61	**7.74**	**7.74**	**7.74**	7.61	7.34
	p-value	0.3273	0.3136	0.2648	0.2452	**0.3279**	0.2836
25	1%	**1.07**	1.20	1.34	1.20	1.20	**1.07**
	p-value	0.9959	0.9988	0.9972	0.9985	**0.9989**	0.9952
	5%	4.27	**4.41**	4.14	4.27	4.27	4.14
	p-value	0.6704	0.6125	0.6940	0.6678	0.6694	**0.7031**
	10%	7.08	**7.34**	**7.34**	7.08	7.21	6.81
	p-value	0.6365	0.5347	0.3536	0.6351	0.5319	**0.7141**
30	1%	1.07	1.07	1.07	1.07	1.07	**0.93**
	p-value	0.9796	0.9768	0.9752	0.9844	0.9732	**0.9966**
	5%	3.74	3.74	**3.87**	**3.87**	**3.87**	3.60
	p-value	**0.9449**	0.9423	0.9347	0.9348	0.9348	0.8676
	10%	**6.68**	6.54	6.41	6.54	6.54	6.54
	p-value	0.8804	0.9154	0.8955	**0.9172**	0.9106	0.8993

Table 6.21: Results of the HIT test. For each assets cardinality n and model we report the average number of VaR violations when this risk measure is computed at the 1, 5, and 10% confidence levels. The percentage printed in red corresponds to the model that yields the closest number to the specified confidence level. The p-values correspond to the F-test on the HIT regression (5.9) with five lags. The highest p-values are marked in bold. When they imply that the null hypothesis of independence is rejected at the 5% (respectively 1%) level, the corresponding average number of VaR violations is marked with * (respectively **).

Technical appendices to the paper "Estimation and empirical performance of non-scalar dynamic conditional correlation models"

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

This document contains three technical appendices to the paper [Bauw 14] that are meant for online publication, as a companion to the paper.

[1]Université catholique de Louvain, CORE, Voie du Roman Pays 34 L1.03.01, B-1348 Louvain-La-Neuve. Belgium; University of Johannesburg, Department of Economics, Johannesburg, South Africa. 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: LB 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". LG acknowledges financial support from the Faculty for the Future Program of the Schlumberger Foundation. LG and JPO acknowledge partial financial support of the Région de Franche-Comté (Convention 2013C-5493).

A Proofs of Propositions

A.1 Proof of Proposition 2.1

In order to prove the Proposition 2.1, we introduce the auxiliary operator σ.

The operator σ: Let $r \leq n \in \mathbb{N}$, $N^* = nr - \frac{1}{2}r\,(r-1)$, and

$$S = \{(i,j)\,|i \in \{1,\ldots,n\}\,, j \in \{1,\ldots,r\}\,, i \geq j\}\,.$$

Given a matrix $A \in \mathbb{L}_{n,r}$ and $\mathbf{v} = \mathrm{vec}_r\,(A) \in \mathbb{R}^{N^*}$ the operator $\sigma : S \longrightarrow \{1,\ldots,N^*\}$ assigns to the position of the entry (i,j), $i \geq j$, of the matrix A the position of the corresponding element of $\mathbf{v}$ in the vec_r representation. We refer to the inverse of this operator as $\sigma^{-1} : \{1,\ldots,N^*\} \longrightarrow S$.

Proof of Proposition 2.1: In order to prove (2.13), we use the following chain of equalities:

$$\begin{aligned}
\langle A, \mathrm{mat}_r\,(\mathbf{v})\rangle &= \mathrm{tr}\left(A^\top \cdot \mathrm{mat}_r\,(\mathbf{v})\right) = \sum_{i=1}^{n}\sum_{j=1}^{r} A_{ij}(\mathrm{mat}_r\,(\mathbf{v}))_{ij} = \sum_{i=r}^{n}\sum_{j=1}^{r} A_{ij}(\mathrm{mat}_r\,(\mathbf{v}))_{ij} \\
&\quad + \sum_{i=1}^{r-1}\sum_{j=1}^{r} A_{ij}(\mathrm{mat}_r\,(\mathbf{v}))_{ij} = \sum_{i=r}^{n}\sum_{j=1}^{r} A_{ij}(\mathrm{mat}_r\,(\mathbf{v}))_{ij} + \sum_{i=1}^{r-1}\sum_{j=1}^{i} A_{ij}(\mathrm{mat}_r\,(\mathbf{v}))_{ij} \\
&\quad + \sum_{i=1}^{r-1}\sum_{j=i+1}^{r} A_{ij}(\mathrm{mat}_r\,(\mathbf{v}))_{ij} = \sum_{i=r}^{n}\sum_{j=1}^{r} A_{ij}(\mathrm{mat}_r\,(\mathbf{v}))_{ij} + \sum_{i=1}^{r-1}\sum_{j=1}^{i} A_{ij}(\mathrm{mat}_r\,(\mathbf{v}))_{ij} \\
&= \sum_{i=r}^{n}\sum_{j=1}^{r} A_{ij} v_{\sigma(i,j)} + \sum_{i=1}^{r-1}\sum_{j=1}^{i} A_{ij} v_{\sigma(i,j)} = \sum_{q=1}^{N^*} A_{\sigma^{-1}(q)} v_q = \langle \mathrm{vec}_r\,(A)\,, \mathbf{v}\rangle.
\end{aligned}$$

Since $A \in \mathbb{L}_{n,r}$ and $\mathbf{v} \in \mathbb{R}^{N^*}$ in these equalities are arbitrary, the identity $\langle A, \mathrm{mat}_r\,(\mathbf{v})\rangle = \langle \mathrm{vec}_r\,(A)\,, \mathbf{v}\rangle$ ensures that $\langle \mathrm{mat}_r^*\,(A)\,, \mathbf{v}\rangle = \langle \mathrm{vec}_r\,(A)\,, \mathbf{v}\rangle$, which yields (2.13). In order to prove the relation (2.14) we write down:

$$\begin{aligned}
\langle \mathbf{v}, \mathrm{vec}_r\,(A)\rangle &= \mathrm{tr}\left(\mathbf{v}^\top \mathrm{vec}_r\,(A)\right) = \sum_{i=1}^{N^*} v_i(\mathrm{vec}_r\,(A))_i = \sum_{i=1}^{n}\sum_{j=1}^{r} v_{\sigma(i,j)}(\mathrm{vec}_r\,(A))_{\sigma(i,j)} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{r}(\mathrm{mat}_r\,(\mathbf{v}))_{ij} A_{ij} = \mathrm{tr}((\mathrm{mat}_r(\mathbf{v})^\top \cdot A) = \langle \mathrm{mat}_r\,(\mathbf{v})\,, A\rangle,
\end{aligned}$$

which yields $\langle \mathrm{vec}_r^*\,(\mathbf{v})\,, A\rangle = \langle \mathrm{mat}_r\,(\mathbf{v})\,, A\rangle$ and hence proves (2.14).

A.2 Proof of Proposition 3.1

We start by showing that the map Ψ is injective. Let $A, B \in \mathbb{L}_{n,m}^+$ be such that $[A] = [B]$. This implies the existence of an element $O \in \mathbb{O}(r)$ such that $AO^{-1} = B$ or, equivalently, $OA^\top = B^\top$. This equality can be written in terms of matrix entries as

$$O\left(\begin{array}{ccccc|c} a_{11} & a_{21} & a_{31} & \cdots & a_{r1} & \\ 0 & a_{22} & a_{32} & \cdots & a_{r2} & \\ 0 & 0 & a_{33} & \cdots & a_{r3} & \\ \vdots & \vdots & \vdots & \ddots & \vdots & \widetilde{A} \\ 0 & 0 & 0 & \cdots & a_{rr} & \end{array}\right) = \left(\begin{array}{ccccc|c} b_{11} & b_{21} & b_{31} & \cdots & b_{r1} & \\ 0 & b_{22} & b_{32} & \cdots & b_{r2} & \\ 0 & 0 & b_{33} & \cdots & b_{r3} & \\ \vdots & \vdots & \vdots & \ddots & \vdots & \widetilde{B} \\ 0 & 0 & 0 & \cdots & b_{rr} & \end{array}\right), \tag{A.1}$$

with $\widetilde{A}, \widetilde{B} \in \mathbb{M}_{r,n-r}$. We now proceed recursively by analyzing the n different equations included in (A.1). First of all, since $A, B \in \mathbb{L}^+_{n,m}$, then $a_{11}, b_{11} > 0$ and hence the equation $O\,(a_{11}, 0, \cdots, 0)^\top = (b_{11}, 0, \ldots, 0)^\top$ implies that $a_{11} = b_{11}$ and that $O \in \mathbb{O}(r)$ belongs to the subgroup of $\mathbb{O}(r)$ isomorphic to $\mathbb{O}(r-1)$ that leaves invariant the vectors in $\text{span}\{\mathbf{e}_r^{(1)}\}$. This statement, together with the second equation included in (A.1), that is, $O\,(a_{21}, a_{22}, 0, \ldots, 0)^\top = (b_{21}, b_{22}, 0, \ldots, 0)^\top$ imply that $a_{21} = b_{21}$ and that $O\,(0, a_{22}, 0, \ldots, 0)^\top = (0, b_{22}, 0, \ldots, 0)^\top$. Since $a_{22}, b_{22} > 0$ then $a_{22} = b_{22}$ necessarily and we can conclude that $O \in \mathbb{O}(r)$ belongs to the subgroup of $\mathbb{O}(r)$ isomorphic to $\mathbb{O}(r-2)$ that leaves invariant the vectors in $\text{span}\{\mathbf{e}_r^{(1)}, \mathbf{e}_r^{(2)}\}$. We are able to conclude by repeating this procedure r times that $O = \mathbf{I}_r$ necessarily and that $A = B$, as required.

In order to show that Ψ is also surjective, we have to prove that for any $[B] \in \mathbb{M}_{n,r}/\mathbb{O}(r)$, there exists $A \in \mathbb{L}^+_{n,m}$ such that $\Psi(A) = [B]$. Let $B \in \mathbb{M}_{n,r}$ be an arbitrary element in the orbit $[B]$ such that

$$B^\top = \left(\begin{array}{ccccc|c} b_{11} & b_{21} & b_{31} & \cdots & b_{r1} & \\ b_{12} & b_{22} & b_{32} & \cdots & b_{r2} & \\ \vdots & \vdots & \vdots & \ddots & \vdots & \widetilde{B} \\ b_{1r} & b_{2r} & b_{3r} & \cdots & b_{rr} & \end{array}\right),$$

with $\widetilde{B} \in \mathbb{M}_{r,n-r}$. Let $O_1 \in \mathbb{O}(r)$ be such that $O_1\,(b_{11}, b_{12}, \ldots, b_{1r})^\top = \left(b^1_{11}, 0, \ldots, 0\right)^\top$, for some $b^1_{11} > 0$ and let

$$\left(\begin{array}{ccccc|c} b^1_{11} & b^1_{21} & b^1_{31} & \cdots & b^1_{r1} & \\ 0 & b^1_{22} & b^1_{32} & \cdots & b^1_{r2} & \\ \vdots & \vdots & \vdots & \ddots & \vdots & \widetilde{B}^1 \\ 0 & b^1_{2r} & b^1_{3r} & \cdots & b^1_{rr} & \end{array}\right) := O_1 B^\top.$$

Consider now another element $O_2 \in \mathbb{O}(r)$ that leaves invariant the vectors in $\text{span}\{\mathbf{e}_r^{(1)}\}$ and such that $O_2\left(b^1_{21}, b^1_{22}, \ldots, b^1_{2r}\right)^\top = \left(b^1_{21}, b^2_{22}, 0, \ldots, 0\right)^\top$, for some $b^2_{22} > 0$, and let

$$\left(\begin{array}{ccccc|c} b^1_{11} & b^1_{21} & b^1_{31} & \cdots & b^1_{r1} & \\ 0 & b^2_{22} & b^2_{32} & \cdots & b^2_{r2} & \\ 0 & 0 & b^2_{33} & \cdots & b^2_{r3} & \\ \vdots & \vdots & \vdots & \ddots & \vdots & \widetilde{B}^2 \\ 0 & 0 & b^2_{3r} & \cdots & b^2_{rr} & \end{array}\right) := O_2 O_1 B^\top.$$

If we iterate r times this construction we obtain r elements $O_1, O_2, \ldots, O_r \in \mathbb{O}(r)$ such that the matrix A defined by $A^\top := O_r O_{r-1} \cdots O_2 O_1 B^\top$ belongs to $\mathbb{L}^+_{n,m}$ and since by construction $[A] = [B]$ we have that $\Psi(A) = [A] = [B]$, and the result follows.

A.3 Proof of Proposition 4.1

We first compute the differential of the log-likelihood function in (4.2). Indeed, for any $\boldsymbol{\delta\theta} \in T_{\boldsymbol{\theta}}\mathcal{P} \times T_{\boldsymbol{\theta}}\mathcal{P}$:

$$\begin{aligned} d_{\boldsymbol{\theta}} l_t \cdot \boldsymbol{\delta\theta} &= d_{H_t} l_t\left(H_t\left(\boldsymbol{\Theta}\left(\boldsymbol{\theta}\right)\right)\right) \cdot T_{\boldsymbol{\Theta}} H_t\left(\boldsymbol{\Theta}\left(\boldsymbol{\theta}\right)\right) \cdot T_{\boldsymbol{\theta}}\boldsymbol{\Theta}\left(\boldsymbol{\theta}\right) \cdot \boldsymbol{\delta\theta} \\ &= \langle \nabla_{H_t} l_t, T_{\boldsymbol{\Theta}} H_t \cdot T_{\boldsymbol{\theta}}\boldsymbol{\Theta} \cdot \boldsymbol{\delta\theta} \rangle = \langle T^*_{\boldsymbol{\theta}}\boldsymbol{\Theta} \cdot T^*_{\boldsymbol{\Theta}} H_t \cdot \nabla_{H_t} l_t, \boldsymbol{\delta\theta} \rangle, \end{aligned}$$

where we used the chain rule on the function $\boldsymbol{\Theta} = \boldsymbol{\Theta}(\boldsymbol{\theta})$. This relation shows that

$$\nabla_{\boldsymbol{\theta}} l_t\left(\boldsymbol{\theta}; \mathbf{r}_t\right) = T^*_{\boldsymbol{\theta}}\boldsymbol{\Theta} \cdot T^*_{\boldsymbol{\Theta}} H_t \cdot \nabla_{H_t} l_t\left(\boldsymbol{\theta}; \mathbf{r}_t\right),$$

which yields (4.3). The proof of (4.4) is standard and can be found, for instance, in Apendix 7.6 of [Chre 13]. In order to prove (4.5) we first consider the tangent map $T_{\boldsymbol{\Theta}}H_t : \mathbb{S}_n \times \mathbb{S}_n \longrightarrow \mathbb{S}_n \times \mathbb{S}_n$ and derive the expression that determines it for every component Θ of $\boldsymbol{\Theta}$. First,

$$\begin{aligned} T_\Theta H_t \cdot \delta\Theta &= D_t(T_\Theta(Q_t^{*-1}) \cdot \delta\Theta)Q_tQ_t^{*-1}D_t + D_tQ_t^{*-1}(T_\Theta Q_t \cdot \delta\Theta)Q_t^{*-1}D_t \\ &\quad + D_tQ_t^{*-1}Q_t(T_\Theta(Q_t^{*-1}) \cdot \delta\Theta)D_t, \quad \delta\Theta \in \mathbb{S}_n. \end{aligned} \tag{A.2}$$

Now, since

$$T_\Theta Q_t^* \cdot \delta\Theta = \frac{1}{2}Q_t^{*-1}\mathrm{Diag}\left(T_\Theta Q_t \cdot \delta\Theta\right),$$

the following relation holds true

$$T_\Theta(Q_t^{*-1}) \cdot \delta\Theta = -\frac{1}{2}Q_t^{*-2}\mathrm{Diag}(T_\Theta Q_t \cdot \delta\Theta)Q_t^{*-1}. \tag{A.3}$$

When we substitute the expression (A.3) into the relation (A.2) we obtain an explicit expression for the map $T_{\boldsymbol{\Theta}}H_t : \mathbb{S}_n \times \mathbb{S}_n \longrightarrow \mathbb{S}_n \times \mathbb{S}_n$. We now compute its adjoint $T^*_{\boldsymbol{\Theta}}H_t : \mathbb{S}_n \times \mathbb{S}_n \longrightarrow \mathbb{S}_n \times \mathbb{S}_n$ by dualizing (A.2) with respect to the Frobenius inner product. Indeed, let $\delta\Theta, \Delta \in \mathbb{S}_n$ arbitrary, then:

$$\begin{aligned} \langle T^*_\Theta H_t \cdot \Delta, \delta\Theta\rangle &= \langle \Delta, T_\Theta H_t \cdot \delta\Theta\rangle = \langle \Delta, D_t(T_\Theta(Q_t^{*-1}) \cdot \delta\Theta)Q_tQ_t^{*-1}D_t\rangle + \langle \Delta, D_tQ_t^{*-1}(T_\Theta Q_t \cdot \delta\Theta)Q_t^{*-1}D_t\rangle \\ &\quad + \langle \Delta, D_tQ_t^{*-1}Q_t\left(T_\Theta\left(Q_t^{*-1}\right) \cdot \delta\Theta\right)D_t\rangle = \langle T^*_\Theta\left(Q_t^{*-1}\right)\left(D_t\Delta D_tQ_t^{*-1}Q_t\right), \delta\Theta\rangle \\ &\quad + \langle T^*_\Theta Q_t\left(Q_t^{*-1}D_t\Delta D_tQ_t^{*-1}\right), \delta\Theta\rangle + \langle T^*_\Theta Q_t^{*-1}\left(Q_tQ_t^{*-1}D_t\Delta D_t\right), \delta\Theta\rangle, \end{aligned}$$

and hence

$$T^*_\Theta H_t \cdot \Delta = T^*_\Theta Q_t\left(Q_t^{*-1}D_t\Delta D_tQ_t^{*-1}\right) + T^*_\Theta\left(Q_t^{*-1}\right)\left[D_t\Delta D_tQ_t^{*-1}Q_t + Q_tQ_t^{*-1}D_t\Delta D_t\right]. \tag{A.4}$$

In order to explicitly write down the second summand of this relation, we dualize (A.3). For any $\delta\Theta, \Delta \in \mathbb{S}_n$ we obtain:

$$\begin{aligned} \langle T^*_\Theta(Q_t^{*-1})\Delta, \delta\Theta\rangle &= -\frac{1}{2}\mathrm{tr}(\Delta^\top Q_t^{*-2}\mathrm{Diag}(T_\Theta Q_t \cdot \delta\Theta)Q_t^{*-1}) = -\frac{1}{2}\mathrm{tr}(Q_t^{*-1}\Delta^\top Q_t^{*-2}\mathrm{Diag}(T_\Theta Q_t \cdot \delta\Theta)) \\ &= -\frac{1}{2}\langle Q_t^{*-2}\Delta Q_t^{*-1}, \mathrm{Diag}(T_\Theta Q_t \cdot \delta\Theta)\rangle = -\frac{1}{2}\langle \mathrm{Diag}(Q_t^{*-2}\Delta Q_t^{*-1}), \mathrm{Diag}(T_\Theta Q_t \cdot \delta\Theta)\rangle \\ &= -\frac{1}{2}\langle T^*_\Theta Q_t\mathrm{Diag}(Q_t^{*-2}\Delta Q_t^{*-1}), \delta\Theta\rangle, \end{aligned}$$

which immediately yields

$$T^*_\Theta\left(Q_t^{*-1}\right) \cdot \Delta = -\frac{1}{2}T^*_\Theta Q_t\mathrm{Diag}\left(Q_t^{*-2}\Delta Q_t^{*-1}\right). \tag{A.5}$$

Substituting (A.5) into (A.4) we obtain:

$$T^*_\Theta H_t \cdot \Delta = T^*_\Theta Q_t\left[Q_t^{*-1}D_t\Delta D_tQ_t^{*-1} - \frac{1}{2}\mathrm{Diag}\left(Q_t^{*-2}\left(D_t\Delta D_tQ_t^{*-1}Q_t + Q_tQ_t^{*-1}D_t\Delta D_t\right)Q_t^{*-1}\right)\right],$$

which proves (4.5) in the statement of the Proposition.

In order to prove the relations (4.6) and (4.7) we start by differentiating (3.5). For arbitrary $\delta A, \delta B \in \mathbb{S}_n$ we obtain:

$$\begin{aligned} T_AQ_t \cdot \delta A &= -\delta A \odot S + \delta A \odot \left(\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}^\top\right) + B \odot (T_AQ_{t-1} \cdot \delta A) \\ &= \delta A \odot \left(\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}^\top - S\right) + B \odot (T_AQ_{t-1} \cdot \delta A), \end{aligned} \tag{A.6}$$

$$\begin{aligned} T_BQ_t \cdot \delta B &= -\delta B \odot S + \delta B \odot Q_{t-1} + B \odot (T_BQ_{t-1} \cdot \delta B) \\ &= \delta B \odot (Q_{t-1} - S) + B \odot (T_BQ_{t-1} \cdot \delta B). \end{aligned} \tag{A.7}$$

We now compute the corresponding adjoints with respect to the Frobenius inner product. Indeed, for any $\delta A, \Delta \in \mathbb{S}_n$ we have

$$\begin{aligned}\langle T_A^* Q_t \cdot \Delta, \delta A\rangle &= \operatorname{tr}\left[\Delta^\top \left(\delta A \odot \left(\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}^\top - S\right)\right)\right] + \operatorname{tr}\left[\Delta^\top \left(B \odot \left(T_A Q_{t-1} \cdot \delta A\right)\right)\right] \\ &= \operatorname{tr}\left[\delta A^\top \left(\Delta \odot \left(\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}^\top - S\right)\right)\right] + \operatorname{tr}\left[\left(T_A Q_{t-1} \cdot \delta A\right)^\top \left(\Delta \odot B\right)\right] \\ &= \left\langle \Delta \odot \left(\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}^\top - S\right), \delta A\right\rangle + \left\langle T_A^* Q_{t-1}\left(\Delta \odot B\right), \delta A\right\rangle,\end{aligned}$$

where we used the Hadamard product trace property (2.4). This expression yields

$$T_A^* Q_t \cdot \Delta = \Delta \odot \left(\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}^\top - S\right) + T_A^* Q_{t-1}\left(\Delta \odot B\right)$$

which proves (4.6). We analogously show (4.7) by dualizing (A.7). Let $\delta B, \Delta \in \mathbb{S}_n$ be arbitrary; the following relation holds true:

$$\langle T_B^* Q_t \cdot \Delta, \delta B\rangle = \operatorname{tr}\left[\Delta^\top \left(\delta B \odot \left(Q_{t-1} - S\right)\right)\right] = \operatorname{tr}\langle \Delta \odot \left(Q_{t-1} - S\right), \delta B\rangle + \left\langle T_B^* Q_{t-1}\left(\Delta \odot B\right), \delta B\right\rangle,$$

which is equivalent to $T_B^* Q_t \cdot \Delta = \Delta \odot \left(Q_{t-1} - S\right) + T_B^* Q_{t-1}\left(\Delta \odot B\right)$.

We conclude by proving the relations (4.8)-(4.11) that are obtained out of the dualization of the tangent map $T_{\boldsymbol{\theta}}\boldsymbol{\Theta}$ of the relation $\boldsymbol{\Theta}\left(\boldsymbol{\theta}\right)$ that provides the connection between the variables $\boldsymbol{\Theta}$ and the intrinsic parameter space $\boldsymbol{\theta}$ for each of the families considered in Section 3.

(i) ***The Hadamard family:*** let $\boldsymbol{\delta a}, \boldsymbol{\delta b} \in \mathbb{R}^N$, then the map $T_{\boldsymbol{\theta}}\boldsymbol{\Theta} : \mathbb{R}^N \times \mathbb{R}^N \longrightarrow \mathbb{S}_n \times \mathbb{S}_n$ is given by

$$T_{\boldsymbol{\theta}}\boldsymbol{\Theta} \cdot (\boldsymbol{\delta a}, \boldsymbol{\delta b}) = \left(\operatorname{math}\left(\boldsymbol{\delta a}\right), \operatorname{math}\left(\boldsymbol{\delta b}\right)\right).$$

Dualizing this relation, we obtain for any $\Delta_1, \Delta_2 \in \mathbb{S}_n$:

$$\langle T_{\boldsymbol{\theta}}^*\boldsymbol{\Theta}(\Delta_1, \Delta_2), (\boldsymbol{\delta a}, \boldsymbol{\delta b})\rangle = \langle (\Delta_1, \Delta_2), (\operatorname{math}(\boldsymbol{\delta a}), \operatorname{math}(\boldsymbol{\delta b}))\rangle = \langle (\operatorname{math}^*(\Delta_1), \operatorname{math}^*(\Delta_2), (\boldsymbol{\delta a}, \boldsymbol{\delta b}))\rangle,$$

and hence

$$T_{\boldsymbol{\theta}}^*\boldsymbol{\Theta}(\Delta_1, \Delta_2) = (\operatorname{math}^*(\Delta_1), \operatorname{math}^*(\Delta_2)),$$

which yields (4.8).

(ii) ***The rank deficient family:*** let $\boldsymbol{\delta a}, \boldsymbol{\delta b} \in \mathbb{R}^{N^*}$, the tangent map $T_{\boldsymbol{\theta}}\boldsymbol{\Theta} : \mathbb{R}^{N^*} \times \mathbb{R}^{N^*} \longrightarrow \mathbb{S}_n \times \mathbb{S}_n$ is given by

$$\begin{aligned}T_{\boldsymbol{\theta}}\boldsymbol{\Theta} \cdot (\boldsymbol{\delta a}, \boldsymbol{\delta b}) = \big(&\operatorname{mat}_r(\boldsymbol{\delta a})(\operatorname{mat}_r(\boldsymbol{a}))^\top + \operatorname{mat}_r(\boldsymbol{a})(\operatorname{mat}_r(\boldsymbol{\delta a}))^\top, \\ &\operatorname{mat}_r(\boldsymbol{\delta b})(\operatorname{mat}_r(\boldsymbol{b}))^\top + \operatorname{mat}_r(\boldsymbol{b})(\operatorname{mat}_r(\boldsymbol{\delta b}))^\top\big).\end{aligned}$$

We now dualize this relation in order to prove (4.9). For $\Delta_1, \Delta_2 \in \mathbb{S}_n$ arbitrary we write

$$\begin{aligned}\langle T_{\boldsymbol{\theta}}^*\boldsymbol{\Theta}(\Delta_1, \Delta_2), (\boldsymbol{\delta a}, \boldsymbol{\delta b})\rangle &= \langle (\Delta_1, \Delta_2), T_{\boldsymbol{\theta}}\boldsymbol{\Theta} \cdot (\boldsymbol{\delta a}, \boldsymbol{\delta b})\rangle \\ &= \langle (\Delta_1, \Delta_2), (\operatorname{mat}_r(\boldsymbol{\delta a})(\operatorname{mat}_r(\boldsymbol{a}))^\top, \operatorname{mat}_r(\boldsymbol{\delta b})(\operatorname{mat}_r(\boldsymbol{b}))^\top)\rangle \\ &\quad + \langle (\Delta_1, \Delta_2), (\operatorname{mat}_r(\boldsymbol{a})(\operatorname{mat}_r(\boldsymbol{\delta a}))^\top, \operatorname{mat}_r(\boldsymbol{b})(\operatorname{mat}_r(\boldsymbol{\delta b}))^\top)\rangle \\ &= 2\langle (\Delta_1 \operatorname{mat}_r(\boldsymbol{a}), \Delta_2 \operatorname{mat}_r(\boldsymbol{b})), (\operatorname{mat}_r(\boldsymbol{\delta a}), \operatorname{mat}_r(\boldsymbol{\delta b}))\rangle \\ &= 2\langle (\operatorname{mat}_r^*(\Delta_1 \operatorname{mat}_r(\boldsymbol{a})), \operatorname{mat}_r^*(\Delta_2 \operatorname{mat}_r(\boldsymbol{b})), (\boldsymbol{\delta a}, \boldsymbol{\delta b}))\rangle. \qquad \text{(A.8)}\end{aligned}$$

Recall that by Proposition 2.13 the operator mat_r^* equals vec_r and hence the relation (A.8) yields

$$T_{\boldsymbol{\theta}}^*\boldsymbol{\Theta}(\Delta_1, \Delta_2) = 2(\operatorname{vec}_r(\Delta_1 \operatorname{mat}_r(\boldsymbol{a})), \operatorname{vec}_r(\Delta_2 \operatorname{mat}_r(\boldsymbol{b}))),$$

and proves (4.9), as required.

(iii) ***The Almon family:*** let $\boldsymbol{\delta a}, \boldsymbol{\delta b} \in \mathbb{R}^3$, then the map $T_{\boldsymbol{\theta}}\boldsymbol{\Theta} : \mathbb{R}^3 \times \mathbb{R}^3 \longrightarrow \mathbb{S}_n \times \mathbb{S}_n$ is determined by

$$T_{\boldsymbol{\theta}}\boldsymbol{\Theta} \cdot (\boldsymbol{\delta a}, \boldsymbol{\delta b}) = \Big((K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a})(\mathrm{alm}_n(\boldsymbol{a}))^\top + \mathrm{alm}_n(\boldsymbol{a})(K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a})^\top, \\ (K_{\boldsymbol{b}} \cdot \boldsymbol{\delta b})(\mathrm{alm}_n(\boldsymbol{b}))^\top + \mathrm{alm}_n(\boldsymbol{b})(K_{\boldsymbol{b}} \cdot \boldsymbol{\delta b})^\top \Big), \tag{A.9}$$

where we used the expression (2.16) for the tangent map $T_{\boldsymbol{\theta}}\mathrm{alm}_n : \mathbb{R}^3 \longrightarrow \mathbb{R}^n$. In order to prove (4.10) we dualize the relation (A.9). For $\Delta_1, \Delta_2 \in \mathbb{S}_n$ arbitrary we compute

$$\begin{aligned} \langle T^*_{\boldsymbol{\theta}}\boldsymbol{\Theta}(\Delta_1, \Delta_2), (\boldsymbol{\delta a}, \boldsymbol{\delta b}) \rangle &= \langle (\Delta_1, \Delta_2), T_{\boldsymbol{\theta}}\boldsymbol{\Theta} \cdot (\boldsymbol{\delta a}, \boldsymbol{\delta b}) \rangle \\ &= \langle (\Delta_1, \Delta_2), ((K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a})(\mathrm{alm}_n(\boldsymbol{a}))^\top, (K_{\boldsymbol{b}} \cdot \boldsymbol{\delta b})(\mathrm{alm}_n(\boldsymbol{b}))^\top) \rangle \\ &\quad + \langle (\Delta_1, \Delta_2), (\mathrm{alm}_n(\boldsymbol{a})(K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a})^\top, \mathrm{alm}_n(\boldsymbol{b})(K_{\boldsymbol{b}} \cdot \boldsymbol{\delta b})^\top) \rangle \\ &= 2\langle (\Delta_1 \mathrm{alm}_n(\boldsymbol{a}), \Delta_2 \mathrm{alm}_n(\boldsymbol{b})), (K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a}, K_{\boldsymbol{b}} \cdot \boldsymbol{\delta b}) \rangle \\ &= 2\langle \big(K_{\boldsymbol{a}}^\top \cdot \Delta_1 \cdot \mathrm{alm}_n(\boldsymbol{a}), K_{\boldsymbol{b}}^\top \cdot \Delta_2 \cdot \mathrm{alm}_n(\boldsymbol{b})\big), (\boldsymbol{\delta a}, \boldsymbol{\delta b}) \rangle, \end{aligned}$$

which yields (4.10), as required. ■

(iv) ***The scalar family:*** let $\delta a, \delta b \in \mathbb{R}$. The map $T_{\boldsymbol{\theta}}\boldsymbol{\Theta} : \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{S}_n \times \mathbb{S}_n$ is given by

$$T_{\boldsymbol{\theta}}\boldsymbol{\Theta} \cdot (\delta a, \delta b) = \big(\delta a \mathbf{i}_n \mathbf{i}_n^\top, \delta b \mathbf{i}_n \mathbf{i}_n^\top\big).$$

Hence, for $\Delta_1, \Delta_2 \in \mathbb{S}_n$ arbitrary we have

$$\begin{aligned} \langle T^*_{\boldsymbol{\theta}}\boldsymbol{\Theta}(\Delta_1, \Delta_2), (\delta a, \delta b) \rangle &= \langle (\Delta_1, \Delta_2), T_{\boldsymbol{\theta}}\boldsymbol{\Theta}(\delta a, \delta b) \rangle \\ &= \langle (\Delta_1, \Delta_2), \big(\delta a \mathbf{i}_n \mathbf{i}_n^\top, \delta b \mathbf{i}_n \mathbf{i}_n^\top\big) \rangle = \langle \big(\langle \Delta_1, \mathbf{i}_n \mathbf{i}_n^\top \rangle, \langle \Delta_2, \mathbf{i}_n \mathbf{i}_n^\top \rangle\big), (\delta a, \delta b) \rangle, \end{aligned}$$

and consequently,

$$T^*_{\boldsymbol{\theta}}\boldsymbol{\Theta}(\Delta_1, \Delta_2) = \big(\langle \Delta_1, \mathbf{i}_n \mathbf{i}_n^\top \rangle, \langle \Delta_2, \mathbf{i}_n \mathbf{i}_n^\top \rangle\big),$$

which proves (4.11), as required.

A.4 Proof of Proposition 6.1

We start by proving (6.4). Let $\Delta := \mathrm{math}(\mathbf{v})$, with $\mathbf{v} \in \mathbb{R}^N$ arbitrary. Then by (4.6), recalling that the operator math is the inverse of the operator vech, and using the expression (2.3) for the Hadamard product of two vectors, we obtain

$$\begin{aligned} A_t \cdot \mathbf{v} &= \mathrm{vech}\big(\mathrm{math}(\mathbf{v}) \odot \big(\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}^\top - S\big)\big) + \mathrm{vech}\, T^*_A Q_{t-1}\, \mathrm{math}\, \mathrm{vech}(\mathrm{math}(\mathbf{v}) \odot B) \\ &= \mathrm{vech}\big(\mathrm{math}(\mathbf{v}) \odot \big(\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}^\top - S\big)\big) + A_{t-1}\, \mathrm{vech}(\mathrm{math}(\mathbf{v}) \odot B) \\ &= \mathbf{v} \odot \mathrm{vech}\big(\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}^\top - S\big) + A_{t-1}(\mathbf{v} \odot \mathrm{vech}(B)) \\ &= \mathrm{diag}\big(\mathrm{vech}\big(\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}^\top - S\big)\big) \cdot \mathbf{v} + A_{t-1}\mathrm{diag}(\mathrm{vech}(B)) \cdot \mathbf{v}, \end{aligned}$$

which yields (6.4) as required. Analogously, by (4.7):

$$\begin{aligned} B_t \cdot \mathbf{v} &= \mathrm{vech}(\mathrm{math}(\mathbf{v}) \odot (Q_{t-1} - S)) + \mathrm{vech}\, T^*_B Q_{t-1}\, \mathrm{math}\, \mathrm{vech}(\mathrm{math}(\mathbf{v}) \odot B) \\ &= \mathbf{v} \odot \mathrm{vech}(Q_{t-1} - S) + B_{t-1}(\mathbf{v} \odot \mathrm{vech}(B)) = \mathrm{diag}(\mathrm{vech}(Q_{t-1} - S)) \cdot \mathbf{v} + A_{t-1}\mathrm{diag}(\mathrm{vech}(B)) \cdot \mathbf{v}, \end{aligned}$$

and hence (6.5) follows.

We now prove (6.6) out of (4.5) and apply the relations (2.7)-(2.10) between the vech, math, vec, and Diag:

$$\begin{aligned}
C_t \cdot \mathbf{v} &= \operatorname{vech} T_\Theta^* Q_t \operatorname{math} \operatorname{vech}\left(Q_t^{*-1} D_t \operatorname{math}(\mathbf{v})\, D_t Q_t^{*-1}\right) \\
&\quad -\frac{1}{2}\operatorname{vech} T_\Theta^* Q_t \operatorname{math} \operatorname{vech} \operatorname{Diag}(Q_t^{*-2} D_t \operatorname{math}(\mathbf{v})\, D_t Q_t^{*-1} Q_t Q_t^{*-1} + Q_t^{*-2} Q_t Q_t^{*-1} D_t \operatorname{math}(\mathbf{v})\, D_t Q_t^{*-1}) \\
&= \Theta_t L_n \operatorname{vec}\left(Q_t^{*-1} D_t \operatorname{math}(\mathbf{v})\, D_t Q_t^{*-1}\right) - \frac{1}{2}\Theta_t L_n \operatorname{vec} \operatorname{Diag}(Q_t^{*-2} D_t \operatorname{math}(\mathbf{v})\, D_t Q_t^{*-1} Q_t Q_t^{*-1} \\
&\quad + Q_t^{*-2} Q_t Q_t^{*-1} D_t \operatorname{math}(\mathbf{v})\, D_t Q_t^{*-1}) = \Theta_t L_n \left(\left(Q_t^{*-1} D_t\right) \otimes \left(Q_t^{*-1} D_t\right)\right) D_n \cdot \mathbf{v} \\
&\quad -\frac{1}{2}\Theta_t L_n \operatorname{vec} \operatorname{math} P_n^d L_n \operatorname{vec}(Q_t^{*-2} D_t \operatorname{math}(\mathbf{v})\, D_t Q_t^{*-1} Q_t Q_t^{*-1} + Q_t^{*-2} Q_t Q_t^{*-1} D_t \operatorname{math}(\mathbf{v})\, D_t Q_t^{*-1}) \\
&= \Theta_t L_n \left(\left(Q_t^{*-1} D_t\right) \otimes \left(Q_t^{*-1} D_t\right)\right) D_n \cdot \mathbf{v} - \frac{1}{2}\Theta_t P_n^d L_n(\left(Q_t^{*-1} Q_t Q_t^{*-1} D_t\right) \otimes \left(Q_t^{*-2} D_t\right) \\
&\quad + \left(Q_t^{*-1} D_t\right) \otimes \left(Q_t^{*-2} Q_t Q_t^{*-1} D_t\right)) D_n \cdot \mathbf{v},
\end{aligned}$$

with $\Theta_t = \{A_t, B_t\}$, which proves (6.6) as required. Finally, the relations (6.8)-(6.11) are obtained from (4.8)-(4.11) by straightforward substitution. ■.

B Implementation of the constrained optimization in the estimation of DCC models

In the following paragraphs we provide explicit expressions for the Bregman divergences, the local penalized model, its gradient and Jacobian, for the rank deficient DCC models, the Almon DCC and the scalar DCC model specifications that we presented in Section 3. The corresponding results for the Hadamard DCC model are provided in Subsection 4.3 of the paper. The detailed derivations of these results are contained in Appendix C.

B.1 The rank deficient DCC model

As explained in Section 3.2, in this case the parameter matrices $A, B \in \mathbb{S}_n$ have a common prescribed rank $r \in \{1, \ldots, n-1\}$, that is, $\operatorname{rank}(A) = \operatorname{rank}(B) = r$. The intrinsic parameterization is provided by two vectors $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^{N^*}$, $N^* = nr - \frac{1}{2}r(r-1)$ using the operator $\operatorname{mat}_r : \mathbb{R}^{N^*} \longrightarrow \mathbb{L}_{n,r}$, presented in Subsection 2.2 and setting: $A := \operatorname{mat}_r(\boldsymbol{a}) \operatorname{mat}_r(\boldsymbol{a})^\top$, $B := \operatorname{mat}_r(\boldsymbol{b}) \operatorname{mat}_r(\boldsymbol{b})^\top$. Let $\boldsymbol{\theta} := (\boldsymbol{a}, \boldsymbol{b}) \in \mathbb{R}^{N^*} \times \mathbb{R}^{N^*}$; the intrinsic parameter subspace $\mathcal{P}$ is $\mathbb{R}^{N^*}$ of dimension N^*. The dynamics of the conditional correlation matrix process $\{Q_t\}$ for the rank deficient family is given in (3.13) and $\boldsymbol{\theta}$ is subjected to the identification, stationarity, and positivity constraints (3.14)-(3.16).

Constraints and the local model. We group these constraints according to the classification in Subsection 4.2.2 which yields two linear, two non-linear, and one positive definiteness constraint. In order to specify them let $\{i_1, \ldots, i_r\}$ with $i_j = n(j-1) + \frac{1}{2}j(3-j)$, $j \in \{1, \ldots, r\}$, be the entries spelled out in (2.12) of any vector in $\mathbb{R}^{N^*}$ that amount to the main diagonal of the corresponding matrix in $\mathbb{L}_{n,r}$ via the mat_r representation. Let $C_{\boldsymbol{a}}^{(1)}, C_{\boldsymbol{b}}^{(2)} \in \mathbb{M}_{r,N^*}$ be the matrices that have as rows the canonical

unit vectors $\{\mathbf{e}_{N^*}^{(i_1)},\ldots,\mathbf{e}_{N^*}^{(i_r)}\}$ with a minus sign in the front. Then:

$$L_1(\boldsymbol{a}) := -C_{\boldsymbol{a}}^{(1)}\boldsymbol{a} > \mathbf{0}_r, \tag{B.1}$$

$$L_2(\boldsymbol{b}) := -C_{\boldsymbol{b}}^{(2)}\boldsymbol{b} > \mathbf{0}_r, \tag{B.2}$$

$$N_1(\boldsymbol{\theta}) := \mathbf{i}_N - \text{vech}(\text{mat}_r(\boldsymbol{a})\,\text{mat}_r(\boldsymbol{a})^\top + \text{mat}_r(\boldsymbol{b})\,\text{mat}_r(\boldsymbol{b})^\top) > \mathbf{0}_N, \tag{B.3}$$

$$N_2(\boldsymbol{\theta}) := \mathbf{i}_N + \text{vech}(\text{mat}_r(\boldsymbol{a})\,\text{mat}_r(\boldsymbol{a})^\top + \text{mat}_r(\boldsymbol{b})\,\text{mat}_r(\boldsymbol{b})^\top) > \mathbf{0}_N, \tag{B.4}$$

$$M(\boldsymbol{\theta}) := (\mathbf{i}_n\mathbf{i}_n^\top - \text{mat}_r(\boldsymbol{a})\,\text{mat}_r(\boldsymbol{a})^\top - \text{mat}_r(\boldsymbol{b})\,\text{mat}_r(\boldsymbol{b})^\top) \odot S \succ 0, \tag{B.5}$$

where $L_1(\boldsymbol{a}), L_2(\boldsymbol{b}) \in \mathbb{R}^r$, $N_1(\boldsymbol{\theta}), N_2(\boldsymbol{\theta}) \in \mathbb{R}^N$, and $M(\boldsymbol{\theta}) \in \mathbb{S}_n$. The local penalized model for the optimization problem (4.17) in this case is given by

$$\begin{aligned}\tilde{f}^{(k)}(\boldsymbol{\theta}) =& f(\boldsymbol{\theta}^{(k)}) + \nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta}^{(k)})(\boldsymbol{\theta}-\boldsymbol{\theta}^{(k)}) + \frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}^{(k)})^\top H^{(k)}(\boldsymbol{\theta}-\boldsymbol{\theta}^{(k)}) \\ &+ L_1 D_M(\boldsymbol{\theta},\boldsymbol{\theta}^{(k)}) + \sum_{j=1}^{2} L_2^j \mathbf{i}_N^\top D_{N_j}(\boldsymbol{\theta},\boldsymbol{\theta}^{(k)}) + \sum_{j=1}^{2} L_3^j \mathbf{i}_r^\top D_{L_j}(\boldsymbol{\theta},\boldsymbol{\theta}^{(k)}),\end{aligned} \tag{B.6}$$

where L_1, $\mathbf{L}_2 := \left(L_2^1, L_2^2\right)^\top$, $\mathbf{L}_3 := \left(L_3^1, L_3^2\right)^\top$ specify the penalization strengths, $f(\boldsymbol{\theta}^{(k)})$ is minus the log-likelihood function evaluated at (4.1), $\nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta}^{(k)})$ is its gradient, that is $\nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta}^{(k)}) = -\nabla_{\boldsymbol{\theta}} \text{log} L(\boldsymbol{\theta}^{(k)})$ which is determined by relations (4.3)-(4.7), and (4.9) in Proposition 4.1. Finally, $H^{(k)}$ is its Hessian computed at the point $\boldsymbol{\theta}^{(k)}$. In relation (B.6) the Bregman divergences associated to each of the constraints (B.1)-(B.5) are easily obtained from (4.14)-(4.16):

$$D_{L_1}^j(\boldsymbol{\theta},\boldsymbol{\theta}^{(k)}) = \frac{a_{i_j}}{a_{i_j}^{(k)}} - \log\frac{a_{i_j}}{a_{i_j}^{(k)}} - 1, \quad i_j = n(j-1) + \frac{1}{2}j(3-j), \quad j \in \{1,\ldots,r\}, \tag{B.7}$$

$$D_{L_2}^j(\boldsymbol{\theta},\boldsymbol{\theta}^{(k)}) = \frac{b_{i_j}}{b_{i_j}^{(k)}} - \log\frac{b_{i_j}}{b_{i_j}^{(k)}} - 1, \quad i_j = n(j-1) + \frac{1}{2}j(3-j), \quad j \in \{1,\ldots,r\}, \tag{B.8}$$

$$D_{N_1}^i(\boldsymbol{\theta},\boldsymbol{\theta}^{(k)}) = \frac{(N_1(\boldsymbol{\theta}))_i}{(N_1(\boldsymbol{\theta}^{(k)}))_i} - \log\frac{(N_1(\boldsymbol{\theta}))_i}{(N_1(\boldsymbol{\theta}^{(k)}))_i} - 1, \quad i \in \{1,\ldots,N\}, \tag{B.9}$$

$$D_{N_2}^i(\boldsymbol{\theta},\boldsymbol{\theta}^{(k)}) = \frac{(N_2(\boldsymbol{\theta}))_i}{(N_2(\boldsymbol{\theta}^{(k)}))_i} - \log\frac{(N_2(\boldsymbol{\theta}))_i}{(N_2(\boldsymbol{\theta}^{(k)}))_i} - 1, \quad i \in \{1,\ldots,N\}, \tag{B.10}$$

$$D_M(\boldsymbol{\theta},\boldsymbol{\theta}^{(k)}) = \text{tr}(M(\boldsymbol{\theta})\cdot M(\boldsymbol{\theta}^{(k)})^{-1}) - \log\det(M(\boldsymbol{\theta})\cdot M(\boldsymbol{\theta}^{(k)})^{-1}) - n. \tag{B.11}$$

Gradient of the local model. A straightforward computation provided in Technical Appendix C.2 gives the following expressions for the components of the gradient of the local model (B.6):

$$\begin{aligned}&\nabla_{\boldsymbol{a}}\tilde{f}^{(k)}(\boldsymbol{\theta}) = \nabla_{\boldsymbol{a}} f(\boldsymbol{\theta}^{(k)}) + H_{\boldsymbol{a}}^{(k)}(\boldsymbol{a}-\boldsymbol{a}^{(k)}) + 2L_1\text{vec}_r(((M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1})\odot S)\text{mat}_r(\boldsymbol{a})) \\ &+ 2\sum_{i=1}^{N}\left(L_2^1\left[\frac{1}{(N_1(\boldsymbol{\theta}))_i} - \frac{1}{(N_1(\boldsymbol{\theta}^{(k)}))_i}\right] - L_2^2\left[\frac{1}{(N_2(\boldsymbol{\theta}))_i} - \frac{1}{(N_2(\boldsymbol{\theta}^{(k)}))_i}\right]\right)\cdot\text{vec}_r(\text{vech}^*(\mathbf{e}_N^{(i)})\cdot\text{mat}_r(\boldsymbol{a})) \\ &- L_3^1\sum_{j=1}^{r}\left[\frac{1}{a_{i_j}} - \frac{1}{a_{i_j}^{(k)}}\right]\cdot\mathbf{e}_{N^*}^{(i_j)}, \quad i_j = n(j-1)+\frac{1}{2}j(3-j), \quad j\in\{1,\ldots,r\}.\end{aligned} \tag{B.12}$$

$$\nabla_{\boldsymbol{b}}\tilde{f}^{(k)}(\boldsymbol{\theta}) = \nabla_{\boldsymbol{b}}f(\boldsymbol{\theta}^{(k)}) + H_{\boldsymbol{b}}^{(k)}(\boldsymbol{b}-\boldsymbol{b}^{(k)}) + 2L_1(\mathrm{vec}_r(((M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1}) \odot S)\mathrm{mat}_r(\boldsymbol{b}))$$
$$+ 2\sum_{i=1}^{N}\left(L_2^1\left[\frac{1}{(N_1(\boldsymbol{\theta}))_i} - \frac{1}{(N_1(\boldsymbol{\theta}^{(k)}))_i}\right] - L_2^2\left[\frac{1}{(N_2(\boldsymbol{\theta}))_i} - \frac{1}{(N_2(\boldsymbol{\theta}^{(k)}))_i}\right]\right) \cdot \mathrm{vec}_r(\mathrm{vech}^*(\mathbf{e}_N^{(i)}) \cdot \mathrm{mat}_r(\boldsymbol{b}))$$
$$- L_3^2\sum_{j=1}^{r}\left[\frac{1}{b_{i_j}} - \frac{1}{b_{i_j}^{(k)}}\right] \cdot \mathbf{e}_{N^*}^{(i_j)}, \quad i_j = n(j-1) + \frac{1}{2}j(3-j), \quad j \in \{1,\ldots,r\}. \tag{B.13}$$

where $\nabla_{\boldsymbol{a}}f(\boldsymbol{\theta}^{(k)})$ and $\nabla_{\boldsymbol{b}}f(\boldsymbol{\theta}^{(k)})$ are the components of the gradient of minus the log-likelihood function computed at the point $\boldsymbol{\theta}^{(k)}$; $\mathbf{e}_{N^*}^{(i)} \in \mathbb{R}^{N^*}$, $i \in \{1,\ldots,N^*\}$ are the canonical unit vectors.

Jacobian of the local model. We use the general relation (4.24) to determine the tangent map to $\nabla_{\boldsymbol{\theta}}\tilde{f}^{(k)}(\boldsymbol{\theta})$ for any $\boldsymbol{\delta\theta} := (\boldsymbol{\delta a}, \boldsymbol{\delta b}) \in \mathbb{R}^{N^*} \times \mathbb{R}^{N^*}$:

$$T_{\boldsymbol{\theta}}\nabla_{\boldsymbol{a}}\tilde{f}^{(k)}(\boldsymbol{\delta a}, \boldsymbol{\delta b}) = H_{\boldsymbol{a}}^{(k)}\boldsymbol{\delta a} + 2L_1\mathrm{vec}_r(((M(\boldsymbol{\theta})^{-1}(W \odot S)M(\boldsymbol{\theta})^{-1}) \odot S)\mathrm{mat}_r(\boldsymbol{a})$$
$$+ ((M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1}) \odot S)\mathrm{mat}_r(\boldsymbol{\delta a})) + 2\sum_{i=1}^{N}\left\{\left(L_2^1\left[\frac{1}{(N_1(\boldsymbol{\theta}))_i} - \frac{1}{(N_1(\boldsymbol{\theta}^{(k)}))_i}\right]\right.\right.$$
$$\left.- L_2^2\left[\frac{1}{(N_2(\boldsymbol{\theta}))_i} - \frac{1}{(N_2(\boldsymbol{\theta}^{(k)}))_i}\right]\right) \cdot \mathrm{vec}_r(\mathrm{vech}^*(\mathbf{e}_N^{(i)}) \cdot \mathrm{mat}_r(\boldsymbol{\delta a}))$$
$$\left.+ (\mathrm{vech}(W))_i \cdot \left(\frac{L_2^1}{(N_1(\boldsymbol{\theta}))_i^2} + \frac{L_2^2}{(N_2(\boldsymbol{\theta}))_i^2}\right) \cdot \mathrm{vec}_r(\mathrm{vech}^*(\mathbf{e}_N^{(i)}) \cdot \mathrm{mat}_r(\boldsymbol{a}))\right\}$$
$$+ L_3^1\sum_{j=1}^{r}\frac{\delta a_{i_j}}{a_{i_j}^2} \cdot \mathbf{e}_{N^*}^{(i_j)}, \quad i_j = n(j-1) + \frac{1}{2}j(3-j). \tag{B.14}$$

$$T_{\boldsymbol{\theta}}\nabla_{\boldsymbol{b}}\tilde{f}^{(k)}(\boldsymbol{\delta a}, \boldsymbol{\delta b}) = H_{\boldsymbol{b}}^{(k)}\boldsymbol{\delta b} + 2L_1\mathrm{vec}_r(((M(\boldsymbol{\theta})^{-1}(W \odot S)M(\boldsymbol{\theta})^{-1}) \odot S)\mathrm{mat}_r(\boldsymbol{b})$$
$$+ ((M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1}) \odot S)\mathrm{mat}_r(\boldsymbol{\delta b})) + 2\sum_{i=1}^{N}\left\{\left(L_2^1\left[\frac{1}{(N_1(\boldsymbol{\theta}))_i} - \frac{1}{(N_1(\boldsymbol{\theta}^{(k)}))_i}\right]\right.\right.$$
$$\left.- L_2^2\left[\frac{1}{(N_2(\boldsymbol{\theta}))_i} - \frac{1}{(N_2(\boldsymbol{\theta}^{(k)}))_i}\right]\right) \cdot \mathrm{vec}_r(\mathrm{vech}^*(\mathbf{e}_N^{(i)}) \cdot \mathrm{mat}_r(\boldsymbol{\delta b}))$$
$$\left.+ (\mathrm{vech}(W))_i \cdot \left(\frac{L_2^1}{(N_1(\boldsymbol{\theta}))_i^2} + \frac{L_2^2}{(N_2(\boldsymbol{\theta}))_i^2}\right) \cdot \mathrm{vec}_r(\mathrm{vech}^*(\mathbf{e}_N^{(i)}) \cdot \mathrm{mat}_r(\boldsymbol{b}))\right\}$$
$$+ L_3^1\sum_{j=1}^{r}\frac{\delta b_{i_j}}{b_{i_j}^2} \cdot \mathbf{e}_{N^*}^{(i_j)}, \quad i_j = n(j-1) + \frac{1}{2}j(3-j). \tag{B.15}$$

with $W := \mathrm{mat}_r(\boldsymbol{\delta a})\mathrm{mat}_r(\boldsymbol{a})^\top + \mathrm{mat}_r(\boldsymbol{a})\mathrm{mat}_r(\boldsymbol{\delta a})^\top + \mathrm{mat}_r(\boldsymbol{\delta b})\mathrm{mat}_r(\boldsymbol{b})^\top + \mathrm{mat}_r(\boldsymbol{b})\mathrm{mat}_r(\boldsymbol{\delta b})^\top$.

The (1,1) and (1, 2) blocks of the Jacobian matrix can be obtained from (B.14) by taking $\boldsymbol{\delta\theta}$ of the form $(\boldsymbol{\delta a}, \mathbf{0})$, $(\mathbf{0}, \boldsymbol{\delta b})$. Analogously, the (2,1) and (2, 2) blocks of the Jacobian are prescribed by (B.15) by setting $\boldsymbol{\delta\theta}$ as $(\boldsymbol{\delta a}, \mathbf{0})$, $(\mathbf{0}, \boldsymbol{\delta b})$, respectively.

B.2 The Almon DCC model

In this subsection we provide the detailed description of the Bregman divergences based optimization algorithm for the new Almon DCC model that we introduced in Section 3.3. This model can be seen

as a particular case of the rank deficient DCC model with parameter matrices $A, B \in \mathbb{S}_n$ in (3.5) of rank $r = 1$. The matrices are intrinsically parametrized by two vectors $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^3$ via the Almon lag operator $\text{alm}_n : \mathbb{R}^3 \longrightarrow \mathbb{R}^n$ (see Subsection 2.2 for the definition and properties of the operator) by defining $A := \text{alm}_n(\boldsymbol{a})\ \text{alm}_n(\boldsymbol{a})^\top$ and $B := \text{alm}_n(\boldsymbol{b})\ \text{alm}_n(\boldsymbol{b})^\top$. Let $\boldsymbol{\theta} := (\boldsymbol{a}, \boldsymbol{b}) \in \mathbb{R}^3 \times \mathbb{R}^3$; the intrinsic parameter subspace $\mathcal{P}$ is $\mathbb{R}^3$ of dimension $P = 3$. The DCC model in the Almon specification is given in (3.17). The parameter constraints that are imposed in order to ensure that the process admits a stationary solution and that the resulting conditional correlation and hence the conditional covariance matrices are positive definite, are prescribed by the relations (3.18)-(3.20).

Constraints and the local model. We group the parameter constraints following the classification suggested in Subsection 4.2.2. This leads to the following set of four nonlinear positivity and one positive definiteness constraint:

$$N_1(\boldsymbol{a}) := (\text{alm}_n(\boldsymbol{a}))_1 = -\mathbf{e}_n^{(1)T} \cdot \text{alm}_n(\boldsymbol{a}) > 0, \tag{B.16}$$

$$N_2(\boldsymbol{b}) := (\text{alm}_n(\boldsymbol{b}))_1 = -\mathbf{e}_n^{(1)T} \cdot \text{alm}_n(\boldsymbol{b}) > 0, \tag{B.17}$$

$$N_3(\boldsymbol{\theta}) := \mathbf{i}_N - \text{vech}(\text{alm}_n(\boldsymbol{a})\,\text{alm}_n(\boldsymbol{a})^\top + \text{alm}_n(\boldsymbol{b})\,\text{alm}_n(\boldsymbol{b})^\top) > \mathbf{0}_N, \tag{B.18}$$

$$N_4(\boldsymbol{\theta}) := \mathbf{i}_N + \text{vech}(\text{alm}_n(\boldsymbol{a})\,\text{alm}_n(\boldsymbol{a})^\top + \text{alm}_n(\boldsymbol{b})\,\text{alm}_n(\boldsymbol{b})^\top) > \mathbf{0}_N, \tag{B.19}$$

$$M(\boldsymbol{\theta}) := (\mathbf{i}_n\mathbf{i}_n^\top - \text{alm}_n(\boldsymbol{a})\,\text{alm}_n(\boldsymbol{a})^\top - \text{alm}_n(\boldsymbol{b})\,\text{alm}_n(\boldsymbol{b})^\top) \odot S \succ 0, \tag{B.20}$$

where $N_1(\boldsymbol{a}), N_2(\boldsymbol{b}) \in \mathbb{R}$, $N_3(\boldsymbol{\theta}), N_4(\boldsymbol{\theta}) \in \mathbb{R}^N$, $M(\boldsymbol{\theta}) \in \mathbb{S}_n$, and $\mathbf{e}_n^{(1)} \in \mathbb{R}^n$ is the canonical unit vector. In the case of the Almon family of models the local penalized model for the optimization problem (4.17) with incorporated Bregman divergences associated to the constraints (B.16)-(B.20) is given by

$$\begin{aligned}\tilde{f}^{(k)}(\boldsymbol{\theta}) =& f(\boldsymbol{\theta}^{(k)}) + \nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta}^{(k)})(\boldsymbol{\theta} - \boldsymbol{\theta}^{(k)}) + \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}^{(k)})^\top H^{(k)}(\boldsymbol{\theta} - \boldsymbol{\theta}^{(k)}) \\ &+ L_1 D_M(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) + \sum_{j=1}^{4} L_2^j \mathbf{i}_{m_j}^\top D_{N_j}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}),\end{aligned} \tag{B.21}$$

where L_1, $\mathbf{L}_2 := \left(L_2^1, \ldots, L_2^4\right)^\top$ specify the penalization strengths and $m_j = \dim\left\{D_{N_j}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})\right\}$, $j \in \{1, \ldots, 4\}$. In this relation $f(\boldsymbol{\theta}^{(k)})$ is minus the log-likelihood function evaluated at (4.1), $\nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta}^{(k)})$ is its gradient, that is $\nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta}^{(k)}) = -\nabla_{\boldsymbol{\theta}} \text{log} L(\boldsymbol{\theta}^{(k)})$ which is determined by relations (4.3)-(4.7), and (4.10) in Proposition 4.1; $H^{(k)}$ is the Hessian of the minus log-likelihood computed at the point $\boldsymbol{\theta}^{(k)}$; $D_{N_j}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$, $j = \{1, \ldots, 4\}$ and $D_M(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$ are the Bregman divergences associated to the constraints (B.16)-(B.19) and (B.20), respectively. The expressions of these divergences by (4.15) and (4.14) can be written as

$$D_{N_1}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = \frac{(\text{alm}_n(\boldsymbol{a}))_1}{(\text{alm}_n(\boldsymbol{a})^{(k)})_1} - \text{log}\frac{(\text{alm}_n(\boldsymbol{a}))_1}{(\text{alm}_n(\boldsymbol{a})^{(k)})_1} - 1, \tag{B.22}$$

$$D_{N_2}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = \frac{(\text{alm}_n(\boldsymbol{b}))_1}{(\text{alm}_n(\boldsymbol{b})^{(k)})_1} - \text{log}\frac{(\text{alm}_n(\boldsymbol{b}))_1}{(\text{alm}_n(\boldsymbol{b})^{(k)})_1} - 1, \tag{B.23}$$

$$D^i_{N_j}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = \frac{(N_j(\boldsymbol{\theta}))_i}{(N_j(\boldsymbol{\theta}^{(k)}))_i} - \text{log}\frac{(N_j(\boldsymbol{\theta}))_i}{(N_j(\boldsymbol{\theta}^{(k)}))_i} - 1, \;\; i \in \{1, \ldots, N\}, \;\; j \in \{3, 4\}, \tag{B.24}$$

$$D_M(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = \text{tr}(M(\boldsymbol{a}, \boldsymbol{b}) \cdot M(\boldsymbol{a}^{(k)}, \boldsymbol{b}^{(k)})^{-1}) - \text{log}\ \det(M(\boldsymbol{a}, \boldsymbol{b}) \cdot M(\boldsymbol{a}^{(k)}, \boldsymbol{b}^{(k)})^{-1}) - n. \tag{B.25}$$

Gradient of the local model. By means of a computation provided in Technical Appendix C.3, we obtain the expressions for the components of the gradient of the local model (B.21), namely:

$$\begin{aligned}
&\nabla_{\boldsymbol{a}}\tilde{f}^{(k)}(\boldsymbol{\theta}) = \nabla_{\boldsymbol{a}} f(\boldsymbol{\theta}^{(k)}) + H_{\boldsymbol{a}}^{(k)}(\boldsymbol{a}-\boldsymbol{a}^{(k)}) \\
&+ 2L_1 K_{\boldsymbol{a}}^{\top}((M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1}) \odot S) \cdot \mathrm{alm}_n(\boldsymbol{a}) - L_2^1 \left[\frac{1}{(\mathrm{alm}_n(\boldsymbol{a}))_1} - \frac{1}{(\mathrm{alm}_n(\boldsymbol{a})^{(k)})_1}\right]\left(K_{\boldsymbol{a}}^{\top}\cdot \mathbf{e}_n^{(1)}\right) \\
&+ 2\sum_{i=1}^{N}\left(L_2^3\left[\frac{1}{(N_2(\boldsymbol{\theta}))_i} - \frac{1}{(N_2(\boldsymbol{\theta}^{(k)}))_i}\right] - L_2^4\left[\frac{1}{(N_3(\boldsymbol{\theta}))_i} - \frac{1}{(N_3(\boldsymbol{\theta}^{(k)}))_i}\right]\right)(K_{\boldsymbol{a}}\cdot \mathrm{vech}^*(\mathbf{e}_N^{(i)})\cdot \mathrm{alm}_n(\boldsymbol{a})),
\end{aligned} \tag{B.26}$$

$$\begin{aligned}
&\nabla_{\boldsymbol{b}}\tilde{f}^{(k)}(\boldsymbol{\theta}) = \nabla_{\boldsymbol{b}} f(\boldsymbol{\theta}^{(k)}) + H_{\boldsymbol{b}}^{(k)}(\boldsymbol{b}-\boldsymbol{b}^{(k)}) \\
&+ 2L_1 K_{\boldsymbol{b}}^{\top}((M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1}) \odot S) \cdot \mathrm{alm}_n(\boldsymbol{b}) - L_2^2 \left[\frac{1}{(\mathrm{alm}_n(\boldsymbol{b}))_1} - \frac{1}{(\mathrm{alm}_n(\boldsymbol{b})^{(k)})_1}\right]\left(K_{\boldsymbol{b}}^{\top}\cdot \mathbf{e}_n^{(1)}\right) \\
&+ 2\sum_{i=1}^{N}\left(L_2^3\left[\frac{1}{(N_2(\boldsymbol{\theta}))_i} - \frac{1}{(N_2(\boldsymbol{\theta}^{(k)}))_i}\right] - L_2^4\left[\frac{1}{(N_3(\boldsymbol{\theta}))_i} - \frac{1}{(N_3(\boldsymbol{\theta}^{(k)}))_i}\right]\right)(K_{\boldsymbol{b}}\cdot \mathrm{vech}^*(\mathbf{e}_N^{(i)})\cdot \mathrm{alm}_n(\boldsymbol{b})),
\end{aligned} \tag{B.27}$$

where $\nabla_{\boldsymbol{a}} f(\boldsymbol{\theta}^{(k)})$ and $\nabla_{\boldsymbol{b}} f(\boldsymbol{\theta}^{(k)})$ are the components of the gradient of minus the log-likelihood function computed at the point $\boldsymbol{\theta}^{(k)}$. Recall that $K_{\boldsymbol{a}}, K_{\boldsymbol{b}} \in \mathbb{M}_{n,3}$ are the matrices introduced in (2.16), $\mathbf{e}_{N^*}^{(i)} \in \mathbb{R}^{N^*}$, $i \in \{1,\dots,N^*\}$ and $\mathbf{e}_N^{(i)}$, $i \in \{1,\dots,N\}$ are the canonical unit vectors introduced in Subsection 2.1.

Jacobian of the local model. We use the expression (4.24) to determine the tangent map to $\nabla_{\boldsymbol{\theta}}\tilde{f}^{(k)}(\boldsymbol{\theta})$ for any $\boldsymbol{\delta\theta} := (\boldsymbol{\delta a}, \boldsymbol{\delta b}) \in \mathbb{R}^3 \times \mathbb{R}^3$. Computations detailed in Technical Appendix C.3 provide the following results:

$$\begin{aligned}
&T_{\boldsymbol{\theta}}\nabla_{\boldsymbol{a}}\tilde{f}^{(k)}(\boldsymbol{\delta a}, \boldsymbol{\delta b}) = H_{\boldsymbol{a}}^{(k)}\boldsymbol{\delta a} + 2L_1\big\{(K_{\boldsymbol{aa}}\cdot\boldsymbol{\delta a})^{\top}\cdot((M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1})\odot S)\cdot \mathrm{alm}_n(\boldsymbol{a}) \\
&\quad + K_{\boldsymbol{a}}\cdot((M(\boldsymbol{\theta})^{-1}(W\odot S)M(\boldsymbol{\theta})^{-1})\odot S)\cdot\mathrm{alm}_n(\boldsymbol{a}) + K_{\boldsymbol{a}}^{\top}\cdot((M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1})\odot S)\cdot(K_{\boldsymbol{a}}\cdot\boldsymbol{\delta a})\big\} \\
&\quad + L_2^1\left\{\frac{(K_{\boldsymbol{a}}\cdot\boldsymbol{\delta a})_1}{(\mathrm{alm}_n(\boldsymbol{a}))_1^2}\cdot(K_{\boldsymbol{a}}^{\top}\cdot\mathbf{e}_n^{(1)}) - \left[\frac{1}{(\mathrm{alm}_n(\boldsymbol{a}))_1} - \frac{1}{(\mathrm{alm}_n(\boldsymbol{a}^{(k)}))_1}\right]((K_{\boldsymbol{aa}}\cdot\boldsymbol{\delta a})^{\top}\cdot\mathbf{e}_n^{(1)})\right\} \qquad \text{(B.28)} \\
&\quad + 2\sum_{i=1}^{N}\left\{\left(L_2^3\left[\frac{1}{(N_2(\boldsymbol{\theta}))_i} - \frac{1}{(N_2(\boldsymbol{\theta}^{(k)}))_i}\right] - L_2^4\left[\frac{1}{(N_3(\boldsymbol{\theta}))_i} - \frac{1}{(N_3(\boldsymbol{\theta}^{(k)}))_i}\right]\right)\cdot((K_{\boldsymbol{aa}}\cdot\boldsymbol{\delta a})^{\top}\cdot\mathrm{vech}^*(\mathbf{e}_N^{(i)})\cdot\mathrm{alm}_n(\boldsymbol{a})\right. \\
&\quad \left. + K_{\boldsymbol{a}}^{\top}\cdot\mathrm{vech}^*(\mathbf{e}_N^{(i)})\cdot(K_{\boldsymbol{a}}\cdot\boldsymbol{\delta a}) + (\mathrm{vech}(W))_i\cdot\left[\frac{L_2^3}{(N_2(\boldsymbol{\theta}))_i^2} + \frac{L_2^4}{(N_3(\boldsymbol{\theta}))_i^2}\right]\cdot(K_{\boldsymbol{a}}^{\top}\cdot\mathrm{vech}^*(\mathbf{e}_N^{(i)})\cdot\mathrm{alm}_n(\boldsymbol{a}))\right\},
\end{aligned} \tag{B.29}$$

$$\begin{aligned}
&T_{\boldsymbol{\theta}}\nabla_{\boldsymbol{b}}\tilde{f}^{(k)}(\boldsymbol{\delta a}, \boldsymbol{\delta b}) = H_{\boldsymbol{b}}^{(k)}\boldsymbol{\delta b} + 2L_1\big\{(K_{\boldsymbol{bb}}\cdot\boldsymbol{\delta b})^{\top}\cdot((M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1})\odot S)\cdot \mathrm{alm}_n(\boldsymbol{b}) \\
&\quad + K_{\boldsymbol{b}}\cdot((M(\boldsymbol{\theta})^{-1}(W\odot S)M(\boldsymbol{\theta})^{-1})\odot S)\cdot\mathrm{alm}_n(\boldsymbol{b}) + K_{\boldsymbol{b}}^{\top}\cdot((M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1})\odot S)\cdot(K_{\boldsymbol{b}}\cdot\boldsymbol{\delta b})\big\} \\
&\quad + L_2^2\left\{\frac{(K_{\boldsymbol{b}}\cdot\boldsymbol{\delta b})_1}{(\mathrm{alm}_n(\boldsymbol{b}))_1^2}\cdot(K_{\boldsymbol{b}}^{\top}\cdot\mathbf{e}_n^{(1)}) - \left[\frac{1}{(\mathrm{alm}_n(\boldsymbol{b}))_1} - \frac{1}{(\mathrm{alm}_n(\boldsymbol{b}^{(k)}))_1}\right]((K_{\boldsymbol{bb}}\cdot\boldsymbol{\delta b})^{\top}\cdot\mathbf{e}_n^{(1)})\right\}
\end{aligned}$$

$$+2\sum_{i=1}^{N}\left\{\left(L_2^3\left[\frac{1}{(N_2(\boldsymbol{\theta}))_i}-\frac{1}{(N_2(\boldsymbol{\theta}^{(k)}))_i}\right]-L_2^4\left[\frac{1}{(N_3(\boldsymbol{\theta}))_i}-\frac{1}{(N_3(\boldsymbol{\theta}^{(k)}))_i}\right]\right)\cdot((K_{\boldsymbol{bb}}\cdot\boldsymbol{\delta b})^\top\cdot\mathrm{vech}^*(\mathbf{e}_N^{(i)})\cdot\mathrm{alm}_n(\boldsymbol{b})$$
$$+K_{\boldsymbol{b}}^\top\cdot\mathrm{vech}^*(\mathbf{e}_N^{(i)})\cdot(K_{\boldsymbol{b}}\cdot\boldsymbol{\delta b})+(\mathrm{vech}(W))_i\cdot\left[\frac{L_2^3}{(N_2(\boldsymbol{\theta}))_i^2}+\frac{L_2^4}{(N_3(\boldsymbol{\theta}))_i^2}\right]\cdot(K_{\boldsymbol{b}}^\top\cdot\mathrm{vech}^*(\mathbf{e}_N^{(i)})\cdot\mathrm{alm}_n(\boldsymbol{b}))\Big\}, \tag{B.30}$$

with

$$W:=(K_{\boldsymbol{a}}\cdot\boldsymbol{\delta a})\cdot\mathrm{alm}_n(\boldsymbol{a})^\top+\mathrm{alm}_n(\boldsymbol{a})\cdot(K_{\boldsymbol{a}}\cdot\boldsymbol{\delta a})^\top+(K_{\boldsymbol{b}}\cdot\boldsymbol{\delta b})\cdot\mathrm{alm}_n(\boldsymbol{b})^\top+\mathrm{alm}_n(\boldsymbol{b})\cdot(K_{\boldsymbol{b}}\cdot\boldsymbol{\delta b})^\top,$$
$$K_{\boldsymbol{aa}}\cdot\boldsymbol{\delta a}:=(\mathbf{0}_n\mid\mathbf{k}_n^1\odot(K_{\boldsymbol{a}}^0\cdot\boldsymbol{\delta a})\mid\mathbf{k}_n^2\odot(K_{\boldsymbol{a}}^0\cdot\boldsymbol{\delta a})),$$
$$K_{\boldsymbol{bb}}\cdot\boldsymbol{\delta b}:=(\mathbf{0}_n\mid\mathbf{k}_n^1\odot(K_{\boldsymbol{b}}^0\cdot\boldsymbol{\delta b})\mid\mathbf{k}_n^2\odot(K_{\boldsymbol{b}}^0\cdot\boldsymbol{\delta b})),$$
$$K_{\boldsymbol{a}}^0:=\left(\mathbf{0}_n\mid\mathbf{k}_n^1\odot\mathrm{alm}_n(\bar{\boldsymbol{a}})\mid\mathbf{k}_n^2\odot\mathrm{alm}_n(\bar{\boldsymbol{a}})\right),$$
$$K_{\boldsymbol{b}}^0:=\left(\mathbf{0}_n\mid\mathbf{k}_n^1\odot\mathrm{alm}_n(\bar{\boldsymbol{b}})\mid\mathbf{k}_n^2\odot\mathrm{alm}_n(\bar{\boldsymbol{b}})\right).$$

The expressions in (B.29)-(B.30), despite their apparent complexity, are explicit and provide the tool to compute the blocks of the Jacobian matrix. The (1,1) and (1, 2) blocks of the Jacobian matrix can be obtained from (B.29) by taking $\boldsymbol{\delta\theta}$ of the form $(\boldsymbol{\delta a},\mathbf{0})$, $(\mathbf{0},\boldsymbol{\delta b})$. Analogously, the (2,1) and (2, 2) blocks of the Jacobian are prescribed by (B.30) by setting $\boldsymbol{\delta\theta}$ as $(\boldsymbol{\delta a},\mathbf{0})$, $(\mathbf{0},\boldsymbol{\delta b})$, respectively.

B.3 The scalar DCC model

In this case the parameter matrices $A,B\in\mathbb{S}_n$ in (3.5) 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 parameter subspace $\mathcal{P}$ is in this case $\mathbb{R}$ and its dimension $P=1$; we denote $\boldsymbol{\theta}:=(a,b)\in\mathbb{R}\times\mathbb{R}$. The scalar DCC model specification together with the associated stationarity and positivity constraints are provided in (3.21)-(3.23).

Constraints and the local model. In this case, the necessary constraints can be reduced to three linear positivity constraints, namely:

$$L_1(\boldsymbol{\theta}):=1-C_a^{(1)}a-C_b^{(1)}b>0, \tag{B.31}$$
$$L_2(a):=-C_a^{(2)}a\geq 0, \tag{B.32}$$
$$L_3(b):=-C_b^{(3)}b\geq 0, \tag{B.33}$$

where $L_1(\boldsymbol{\theta}),L_2(a),L_3(b)\in\mathbb{R}$, $C_a^{(1)}=C_b^{(1)}=1$, and $C_a^{(2)}=C_b^{(3)}=-1$. The local penalized model for the optimization problem (4.17) in this case is given by

$$\tilde{f}^{(k)}(\boldsymbol{\theta})=f(\boldsymbol{\theta}^{(k)})+\nabla_{\boldsymbol{\theta}}f(\boldsymbol{\theta}^{(k)})(\boldsymbol{\theta}-\boldsymbol{\theta}^{(k)})+\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}^{(k)})^\top H^{(k)}(\boldsymbol{\theta}-\boldsymbol{\theta}^{(k)})+\sum_{j=1}^{3}L_1^jD_{L_j}(\boldsymbol{\theta},\boldsymbol{\theta}^{(k)}), \tag{B.34}$$

where the vector $\mathbf{L}_1:=\left(L_1^1,L_1^2,L_1^3\right)^\top$ contains the penalization strengths, $f(\boldsymbol{\theta}^{(k)})$ is minus the log-likelihood function in (4.1), $\nabla_{\boldsymbol{\theta}}f(\boldsymbol{\theta}^{(k)})=-\nabla_{\boldsymbol{\theta}}\log L(\boldsymbol{\theta}^{(k)}$ is its gradient which for the scalar family is determined by relations (4.3)-(4.7), and (4.11) in Proposition 4.1, and $H^{(k)}$ is its Hessian computed at the point $\boldsymbol{\theta}^{(k)}$. The Bregman divergences associated to each constraint (B.31)-(B.33) and obtained

from (4.16) take the following expressions:

$$D_{L_1}(\boldsymbol{\theta},\boldsymbol{\theta}^{(k)}) = \frac{1-a-b}{1-a^{(k)}-b^{(k)}} - \log\frac{1-a-b}{1-a^{(k)}-b^{(k)}} - 1, \tag{B.35}$$

$$D_{L_2}(\boldsymbol{\theta},\boldsymbol{\theta}^{(k)}) = \frac{a}{a^{(k)}} - \log\frac{a}{a^{(k)}} - 1, \tag{B.36}$$

$$D_{L_3}(\boldsymbol{\theta},\boldsymbol{\theta}^{(k)}) = \frac{b}{b^{(k)}} - \log\frac{b}{b^{(k)}} - 1. \tag{B.37}$$

Gradient of the local model. A straightforward computation yields:

$$\nabla_a \tilde{f}^{(k)}(\boldsymbol{\theta}) = \nabla_a f(\boldsymbol{\theta}^{(k)}) + H_a^{(k)}(a-a^{(k)}) + L_1^1\left[\frac{1}{1-a-b} - \frac{1}{1-a^{(k)}-b^{(k)}}\right] - L_1^2\left[\frac{1}{a} - \frac{1}{a^{(k)}}\right], \tag{B.38}$$

$$\nabla_b \tilde{f}^{(k)}(\boldsymbol{\theta}) = \nabla_b f(\boldsymbol{\theta}^{(k)}) + H_b^{(k)}(b-b^{(k)}) + L_1^1\left[\frac{1}{1-a-b} - \frac{1}{1-a^{(k)}-b^{(k)}}\right] - L_1^3\left[\frac{1}{b} - \frac{1}{b^{(k)}}\right], \tag{B.39}$$

where $\nabla_a f(\boldsymbol{\theta}^{(k)})$, $\nabla_b f(\boldsymbol{\theta}^{(k)})$ are the components of the gradient of minus the log-likelihood function $\nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta}^{(k)}) = -\nabla_{\boldsymbol{\theta}} \log L(\boldsymbol{\theta}^{(k)}, \mathbf{r})$ determined by the expressions (4.3)-(4.7) and by relation (4.11).

Jacobian of the local model. A straightforward computation yields for any $(\delta a, \delta b) \in \mathbb{R}\times\mathbb{R}$ the following expressions:

$$T_{\boldsymbol{\theta}} \nabla_a \tilde{f}^{(k)}(\delta a, \delta b) = H_a^{(k)}\delta a + L_1^1 \frac{\delta a + \delta b}{(1-a-b)^2} + L_1^2 \frac{\delta a}{a^2}, \tag{B.40}$$

$$T_{\boldsymbol{\theta}} \nabla_b \tilde{f}^{(k)}(\delta a, \delta b) = H_b^{(k)}\delta b + L_1^1 \frac{\delta b + \delta b}{(1-a-b)^2} + L_1^3 \frac{\delta b}{b^2}. \tag{B.41}$$

The (1,1) and (1, 2) blocks of the Jacobian matrix can be obtained from (B.40) by simply taking $(\delta a, \delta b)$ of the form $(\delta a, 0)$, $(0, \delta b)$, and the (2,1) and (2, 2) blocks of the Jacobian matrix are prescribed by (B.41) by setting $(\delta a, \delta b)$ as $(\delta a, 0)$, $(0, \delta b)$, respectively.

C Detailed computations for the implementation of the constrained optimization in the estimation of non-scalar DCC models

We provide here detailed derivations of the results in Subsection 4.3 of the paper and in Appendix B.

C.1 The Hadamard DCC model

The gradient of the local model. The computation of $\nabla_{\boldsymbol{\theta}} \tilde{f}^{(k)}(\boldsymbol{\theta})$ requires the expressions of the gradients of each of the divergences in (4.31)-(4.35).

We start with the Bregman divergences $D_{L_1}(\boldsymbol{\theta},\boldsymbol{\theta}^{(k)}), D_{L_2}(\boldsymbol{\theta},\boldsymbol{\theta}^{(k)}) \in \mathbb{R}^N$ associated to the linear constraints (4.25) and (4.26), respectively. It is easy to check by (4.23) that for each component $\theta \in \mathbb{R}^N$ of $\boldsymbol{\theta}$, the gradients of these divergences are determined by

$$\nabla_\theta D^i_{L_1}(\boldsymbol{\theta},\boldsymbol{\theta}^{(k)}) = \left[\frac{1}{1-a_i-b_i} - \frac{1}{1-a_i^{(k)}-b_i^{(k)}}\right]\mathbf{e}_N^{(i)}, \quad i \in \{1,\dots,N\}, \tag{C.1}$$

$$\nabla_\theta D^i_{L_2}(\boldsymbol{\theta},\boldsymbol{\theta}^{(k)}) = -\left[\frac{1}{1+a_i+b_i} - \frac{1}{1+a_i^{(k)}+b_i^{(k)}}\right]\mathbf{e}_N^{(i)}, \quad i \in \{1,\dots,N\}, \tag{C.2}$$

where $\mathbf{e}_N^{(i)} \in \mathbb{R}^N$ are the canonical unit vectors defined by $\mathbf{e}_N^{(i)} = (\delta_{ij})_{j\in\{1,\dots,N\}}$.

Regarding the Bregman divergences $D_{M_1}(\boldsymbol{a}, \boldsymbol{a}^{(k)})$, $D_{M_2}(\boldsymbol{b}, \boldsymbol{b}^{(k)}) \in \mathbb{S}_n$ in (4.33)-(4.34) related to the positive semidefiniteness constraints (4.27) and (4.28), we have

$$\nabla_{\boldsymbol{a}} D_{M_1}(\boldsymbol{a}, \boldsymbol{a}^{(k)}) = -\mathrm{math}^*(M_1(\boldsymbol{a})^{-1} - M_1(\boldsymbol{a}^{(k)})^{-1}), \tag{C.3}$$

$$\nabla_{\boldsymbol{b}} D_{M_2}(\boldsymbol{b}, \boldsymbol{b}^{(k)}) = -\mathrm{math}^*(M_2(\boldsymbol{b})^{-1} - M_2(\boldsymbol{b}^{(k)})^{-1}), \tag{C.4}$$

where $\mathrm{math}^* : \mathbb{S}_n \longrightarrow \mathbb{R}^N$ is the adjoint map of the operator $\mathrm{math} : \mathbb{R}^N \longrightarrow \mathbb{S}_n$, both introduced in Subsection 2.2.

We now proceed with the expression for the gradient of the divergence $D_{M_3}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) \in \mathbb{S}_n$ defined in (4.35) and related to the positive definiteness constraint (4.29). By relation (4.21) it can be written down as

$$\nabla_{\boldsymbol{\theta}} D_{M_3}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = (\mathrm{math}^*((M_3(\boldsymbol{\theta})^{-1} - M_3(\boldsymbol{\theta}^{(k)})^{-1}) \odot S), \mathrm{math}^*((M_3(\boldsymbol{\theta})^{-1} - M_3(\boldsymbol{\theta}^{(k)})^{-1}) \odot S)), \tag{C.5}$$

where we just used that the tangent map $T_{\boldsymbol{\theta}} M_3 : \mathbb{R}^N \times \mathbb{R}^N \longrightarrow \mathbb{S}_n$ for any $\boldsymbol{\delta\theta} := (\boldsymbol{\delta a}, \boldsymbol{\delta b}) \in \mathbb{R}^N \times \mathbb{R}^N$ is given by

$$T_{\boldsymbol{\theta}} M_3 \cdot \boldsymbol{\delta\theta} = -\mathrm{math}(\boldsymbol{\delta a} + \boldsymbol{\delta b}) \odot S. \tag{C.6}$$

Substitution of (C.1)-(C.5) into (4.20) gives the components (4.36)-(4.37) of the gradient of the local model (4.30) for the Hadamard family, as required.

The Jacobian of the local model. The only new element required in the expression (4.24) are the adjoints $T^*_{\boldsymbol{\theta}} M_j : \mathbb{S}_n \longrightarrow \mathbb{R}^N \times \mathbb{R}^N$ of the tangent maps $T_{\boldsymbol{\theta}} M_j : \mathbb{R}^N \times \mathbb{R}^N \longrightarrow \mathbb{S}_n$, $j = 1, 2, 3$. In the case of $j = 1$ and $j = 2$, these adjoints are given by math^*. We now compute the expression for the adjoint map $T^*_{\boldsymbol{\theta}} M_3 : \mathbb{S}_n \longrightarrow \mathbb{R}^N \times \mathbb{R}^N$ by dualizing (C.6) for $\Delta \in \mathbb{S}_n$ arbitrary and obtain

$$\begin{aligned}
\langle T^*_{\boldsymbol{\theta}} M_3 \cdot \Delta, (\boldsymbol{\delta a}, \boldsymbol{\delta b})\rangle &= \langle \Delta, T_{\boldsymbol{\theta}} M_3 (\boldsymbol{\delta a}, \boldsymbol{\delta b})\rangle = -\mathrm{tr}\left(\Delta((\mathrm{math}(\boldsymbol{\delta a} + \boldsymbol{\delta b})) \odot S)\right) = -\mathrm{tr}\left((\mathrm{math}(\boldsymbol{\delta a}) \odot S)\Delta\right) \\
&\quad -\mathrm{tr}\left((\mathrm{math}(\boldsymbol{\delta b}) \odot S)\Delta\right) = -\mathrm{tr}\left((\Delta \odot S)\ \mathrm{math}(\boldsymbol{\delta a})\right) - \mathrm{tr}\left((\Delta \odot S)\ \mathrm{math}(\boldsymbol{\delta b})\right) \\
&= -\langle \Delta \odot S, \mathrm{math}(\boldsymbol{\delta a})\rangle - \langle \Delta \odot S, \mathrm{math}(\boldsymbol{\delta b})\rangle = -\langle \mathrm{math}^*(\Delta \odot S), \boldsymbol{\delta a}\rangle \\
&\quad -\langle \mathrm{math}^*(\Delta \odot S), \boldsymbol{\delta b}\rangle = \langle (\mathrm{math}^*(\Delta \odot S), \mathrm{math}^*(\Delta \odot S)), (\boldsymbol{\delta a}, \boldsymbol{\delta b})\rangle,
\end{aligned}$$

where we used the Hadamard product trace property (2.4) which immediately yields

$$T^*_{\boldsymbol{\theta}} M_3 \cdot \Delta = -\left(\mathrm{math}^*(\Delta \odot S), \mathrm{math}^*(\Delta \odot S)\right). \tag{C.7}$$

The relations (4.38)-(4.39) hence follow from (4.24), as required.

C.2 The rank deficient DCC model

Gradient of the local model. The computation of the gradient $\nabla \tilde{f}(\boldsymbol{\theta})$ requires the expressions for the gradients of the components of each of the Bregman divergences defined in (B.7)-(B.11).

Regarding the divergences (B.7)-(B.8) associated with the linear positivity constraints we write

$$\nabla_{\boldsymbol{\theta}} D^j_{L_1}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = -\left[\frac{1}{a_{i_j}} - \frac{1}{a^{(k)}_{i_j}}\right] \mathbf{e}^{(i_j)}_{N^*}, \quad i_j = n(j-1) + \frac{1}{2}j(3-j), \quad j \in \{1, \dots, r\}, \tag{C.8}$$

$$\nabla_{\boldsymbol{\theta}} D^j_{L_2}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = -\left[\frac{1}{b_{i_j}} - \frac{1}{b^{(k)}_{i_j}}\right] \mathbf{e}^{(i_j)}_{N^*}, \quad i_j = n(j-1) + \frac{1}{2}j(3-j), \quad j \in \{1, \dots, r\}, \tag{C.9}$$

where we recall that $\mathbf{e}_{N^*}^{(i_j)} \in \mathbb{R}^{N^*}$ is the canonical unit vector whose i_j-th entry equals one.

We now carry out the computation of the gradients of the Bregman divergences (B.9)-(B.10) related to the nonlinear positivity constraints (B.3)-(B.4) using the relation (4.22). In order to provide explicit expressions for those gradients we first compute the tangent maps $T_{\boldsymbol{\theta}}N_1 : \mathbb{R}^{N^*} \times \mathbb{R}^{N^*} \longrightarrow \mathbb{R}^N$ and $T_{\boldsymbol{\theta}}N_2 : \mathbb{R}^{N^*} \times \mathbb{R}^{N^*} \longrightarrow \mathbb{R}^N$:

$$T_{\boldsymbol{\theta}}N_1 \cdot \boldsymbol{\delta\theta} = -\mathrm{vech}(\mathrm{mat}_r(\boldsymbol{\delta a})\mathrm{mat}_r(\boldsymbol{a})^\top + \mathrm{mat}_r(\boldsymbol{a})\mathrm{mat}_r(\boldsymbol{\delta a})^\top + \mathrm{mat}_r(\boldsymbol{\delta b})\mathrm{mat}_r(\boldsymbol{b})^\top + \mathrm{mat}_r(\boldsymbol{b})\mathrm{mat}_r(\boldsymbol{\delta b})^\top), \tag{C.10}$$

$$T_{\boldsymbol{\theta}}N_2 \cdot \boldsymbol{\delta\theta} = \mathrm{vech}(\mathrm{mat}_r(\boldsymbol{\delta a})\mathrm{mat}_r(\boldsymbol{a})^\top + \mathrm{mat}_r(\boldsymbol{a})\mathrm{mat}_r(\boldsymbol{\delta a})^\top + \mathrm{mat}_r(\boldsymbol{\delta b})\mathrm{mat}_r(\boldsymbol{b})^\top + \mathrm{mat}_r(\boldsymbol{b})\mathrm{mat}_r(\boldsymbol{\delta b})^\top), \tag{C.11}$$

with $\boldsymbol{\delta\theta} := (\boldsymbol{\delta a}, \boldsymbol{\delta b}) \in \mathbb{R}^{N^*} \times \mathbb{R}^{N^*}$. We determine their adjoints $T^*_{\boldsymbol{\theta}}N_1 : \mathbb{R}^N \longrightarrow \mathbb{R}^{N^*} \times \mathbb{R}^{N^*}$ and $T^*_{\boldsymbol{\theta}}N_2 : \mathbb{R}^N \longrightarrow \mathbb{R}^{N^*} \times \mathbb{R}^{N^*}$ by dualizing (C.10)-(C.2). Indeed, for $\mathbf{v} \in \mathbb{R}^N$ arbitrary, we have

$$\begin{aligned}
\langle T^*_{\boldsymbol{\theta}}N_1 \cdot \mathbf{v}, (\boldsymbol{\delta a}, \boldsymbol{\delta b})\rangle =& \langle \mathbf{v}, T_{\boldsymbol{\theta}}N_1\,(\boldsymbol{\delta a}, \boldsymbol{\delta b})\rangle = -\langle \mathbf{v}, \mathrm{vech}(\mathrm{mat}_r(\boldsymbol{\delta a})\mathrm{mat}_r(\boldsymbol{a})^\top + \mathrm{mat}_r(\boldsymbol{a})\mathrm{mat}_r(\boldsymbol{\delta a})^\top \\
&+ \mathrm{mat}_r(\boldsymbol{\delta b})\mathrm{mat}_r(\boldsymbol{b})^\top + \mathrm{mat}_r(\boldsymbol{b})\mathrm{mat}_r(\boldsymbol{\delta b})^\top)\rangle = -\langle \mathbf{v}, \mathrm{vech}(\mathrm{mat}_r(\boldsymbol{\delta a})\mathrm{mat}_r(\boldsymbol{a})^\top)\rangle \\
&- \langle \mathbf{v}, \mathrm{vech}(\mathrm{mat}_r(\boldsymbol{a})\mathrm{mat}_r(\boldsymbol{\delta a})^\top)\rangle - \langle \mathbf{v}, \mathrm{vech}(\mathrm{mat}_r(\boldsymbol{\delta b})\mathrm{mat}_r(\boldsymbol{b})^\top)\rangle \\
&- \langle \mathbf{v}, \mathrm{vech}(\mathrm{mat}_r(\boldsymbol{b})\mathrm{mat}_r(\boldsymbol{\delta b})^\top)\rangle = -\langle \mathrm{vech}^*(\mathbf{v})\mathrm{mat}_r(\boldsymbol{a}), \mathrm{mat}_r(\boldsymbol{\delta a})\rangle \\
&- \langle \mathrm{mat}_r(\boldsymbol{a})^\top \mathrm{vech}^*(\mathbf{v}), \mathrm{mat}_r(\boldsymbol{\delta a})^\top\rangle - \langle \mathrm{vech}^*(\mathbf{v})\mathrm{mat}_r(\boldsymbol{b}), \mathrm{mat}_r(\boldsymbol{\delta b})\rangle \\
&- \langle \mathrm{mat}_r(\boldsymbol{b})^\top \mathrm{vech}^*(\mathbf{v}), \mathrm{mat}_r(\boldsymbol{\delta b})^\top\rangle = -2\langle \mathrm{mat}_r^*(\mathrm{vech}^*(\mathbf{v})\mathrm{mat}_r(\boldsymbol{a})), \boldsymbol{\delta a}\rangle \\
&- 2\langle \mathrm{mat}_r^*(\mathrm{vech}^*(\mathbf{v})\mathrm{mat}_r(\boldsymbol{b})), \boldsymbol{\delta b}\rangle \\
=& -2\langle (\mathrm{vec}_r(\mathrm{vech}^*(\mathbf{v})\mathrm{mat}_r(\boldsymbol{a})), \mathrm{vec}_r(\mathrm{vech}^*(\mathbf{v})\mathrm{mat}_r(\boldsymbol{b})), (\boldsymbol{\delta a}, \boldsymbol{\delta b}))\rangle,
\end{aligned}$$

where we used that by Proposition 2.1 mat_r^* equals vec_r. We hence obtain that

$$T^*_{\boldsymbol{\theta}}N_1 \cdot \mathbf{v} = -2(\mathrm{vec}_r(\mathrm{vech}^*(\mathbf{v})\mathrm{mat}_r(\boldsymbol{a})), \mathrm{vec}_r(\mathrm{vech}^*(\mathbf{v})\mathrm{mat}_r(\boldsymbol{b})). \tag{C.12}$$

Analogously, it can be easily checked that:

$$T^*_{\boldsymbol{\theta}}N_2 \cdot \mathbf{v} = 2(\mathrm{vec}_r(\mathrm{vech}^*(\mathbf{v})\mathrm{mat}_r(\boldsymbol{a})), \mathrm{vec}_r(\mathrm{vech}^*(\mathbf{v})\mathrm{mat}_r(\boldsymbol{b})). \tag{C.13}$$

Consequently, the gradients of the components of the Bregman divergences $D^i_{N_1}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}), D^i_{N_2}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$, $i \in \{1, \ldots, N\}$ are given by the relations

$$\nabla_{\boldsymbol{\theta}} D^i_{N_1}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = 2\left[\frac{1}{(N_1(\boldsymbol{\theta}))_i} - \frac{1}{(N_1(\boldsymbol{\theta}^{(k)}))_i}\right](\mathrm{vec}_r(\mathrm{vech}^*(\mathbf{e}_N^{(i)}) \cdot \mathrm{mat}_r(\boldsymbol{a}), \mathrm{vec}_r(\mathrm{vech}^*(\mathbf{e}_N^{(i)}) \cdot \mathrm{mat}_r(\boldsymbol{b})), \tag{C.14}$$

$$\nabla_{\boldsymbol{\theta}} D^i_{N_2}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = -2\left[\frac{1}{(N_2(\boldsymbol{\theta}))_i} - \frac{1}{(N_2(\boldsymbol{\theta}^{(k)}))_i}\right](\mathrm{vec}_r(\mathrm{vech}^*(\mathbf{e}_N^{(i)}) \cdot \mathrm{mat}_r(\boldsymbol{a}), \mathrm{vec}_r(\mathrm{vech}^*(\mathbf{e}_N^{(i)}) \cdot \mathrm{mat}_r(\boldsymbol{b})), \tag{C.15}$$

where $\mathbf{e}_N^{(i)} \in \mathbb{R}^N$, $i \in \{1, \ldots, N\}$ are the canonical unit vectors.

We now proceed with the positive definite constraint (B.5), for which we recall that

$$M(\boldsymbol{\theta}) = (\mathbf{i}_n\mathbf{i}_n^\top - \mathrm{mat}_r(\boldsymbol{a})\,\mathrm{mat}_r(\boldsymbol{a})^\top - \mathrm{mat}_r(\boldsymbol{b})\,\mathrm{mat}_r(\boldsymbol{b})^\top) \odot S.$$

In order to provide the gradient for the Bregman divergence associated to the constraint (B.5), we refer to the relation (4.21), whose explicit formulation requires the expression of the adjoint map

$T^*_{\boldsymbol{\theta}}M : \mathbb{S}_n \longrightarrow \mathbb{R}^{N^*} \times \mathbb{R}^{N^*}$. We hence first determine the corresponding tangent map $T_{\boldsymbol{\theta}}M : \mathbb{R}^{N^*} \times \mathbb{R}^{N^*} \longrightarrow \mathbb{S}_n$. Indeed, for any $\boldsymbol{\delta\theta} = (\boldsymbol{\delta a}, \boldsymbol{\delta b}) \in \mathbb{R}^{N^*} \times \mathbb{R}^{N^*}$:

$$T_{\boldsymbol{\theta}}M \cdot \boldsymbol{\delta\theta} = -(\mathrm{mat}_r(\boldsymbol{\delta a})\mathrm{mat}_r(\boldsymbol{a})^\top + \mathrm{mat}_r(\boldsymbol{a})\mathrm{mat}_r(\boldsymbol{\delta a})^\top + \mathrm{mat}_r(\boldsymbol{\delta b})\mathrm{mat}_r(\boldsymbol{b})^\top + \mathrm{mat}_r(\boldsymbol{b})\mathrm{mat}_r(\boldsymbol{\delta b})^\top) \odot S.$$

We dualize this expression in order to obtain the required adjoint, that is, for arbitrary $\Delta \in \mathbb{S}_n$:

$$\begin{aligned}
&\langle T^*_{\boldsymbol{\theta}}M \cdot \Delta, (\boldsymbol{\delta a}, \boldsymbol{\delta b})\rangle = \langle \Delta, T_{\boldsymbol{\theta}}M \cdot (\boldsymbol{\delta a}, \boldsymbol{\delta b})\rangle \\
&= -\langle \Delta, (\mathrm{mat}_r(\boldsymbol{\delta a})\mathrm{mat}_r(\boldsymbol{a})^\top + \mathrm{mat}_r(\boldsymbol{a})\mathrm{mat}_r(\boldsymbol{\delta a})^\top + \mathrm{mat}_r(\boldsymbol{\delta b})\mathrm{mat}_r(\boldsymbol{b})^\top + \mathrm{mat}_r(\boldsymbol{b})\mathrm{mat}_r(\boldsymbol{\delta b})^\top) \odot S\rangle \\
&= -\langle \Delta, (\mathrm{mat}_r(\boldsymbol{\delta a})\mathrm{mat}_r(\boldsymbol{a})^\top) \odot S\rangle - \langle \Delta, (\mathrm{mat}_r(\boldsymbol{a})\mathrm{mat}_r(\boldsymbol{\delta a})^\top) \odot S\rangle - \langle \Delta, (\mathrm{mat}_r(\boldsymbol{\delta b})\mathrm{mat}_r(\boldsymbol{b})^\top) \odot S\rangle \\
&\quad - \langle \Delta, (\mathrm{mat}_r(\boldsymbol{b})\mathrm{mat}_r(\boldsymbol{\delta b})^\top) \odot S\rangle = -\langle (\Delta \odot S)\mathrm{mat}_r(\boldsymbol{a}), \mathrm{mat}_r(\boldsymbol{\delta a})\rangle - \langle \mathrm{mat}_r(\boldsymbol{a})^\top(\Delta \odot S), \mathrm{mat}_r(\boldsymbol{\delta a})^\top\rangle \\
&\quad - \langle (\Delta \odot S)\mathrm{mat}_r(\boldsymbol{b}), \mathrm{mat}_r(\boldsymbol{\delta b})\rangle - \langle \mathrm{mat}_r(\boldsymbol{b})^\top(\Delta \odot S), \mathrm{mat}_r(\boldsymbol{\delta b})^\top\rangle = -2\ \langle \mathrm{mat}^*_r((\Delta \odot S)\mathrm{mat}_r(\boldsymbol{a})), \boldsymbol{\delta a}\rangle \\
&\quad - 2\ \langle \mathrm{mat}^*_r((\Delta \odot S)\mathrm{mat}_r(\boldsymbol{b})), \boldsymbol{\delta b}\rangle = -2\langle (\mathrm{mat}^*_r((\Delta \odot S)\mathrm{mat}_r(\boldsymbol{a})), \mathrm{mat}^*_r((\Delta \odot S)\mathrm{mat}_r(\boldsymbol{b}))), (\boldsymbol{\delta a}, \boldsymbol{\delta b})\rangle.
\end{aligned}$$

Recall that by Proposition 2.1, mat^*_r equals vec_r and we hence have

$$T^*_{\boldsymbol{\theta}}M \cdot \Delta = -2\ (\mathrm{vec}_r((\Delta \odot S)\mathrm{mat}_r(\boldsymbol{a})), \mathrm{vec}_r((\Delta \odot S)\mathrm{mat}_r(\boldsymbol{b}))) \in \mathbb{R}^{N^*} \times \mathbb{R}^{N^*}. \tag{C.16}$$

If we substitute (C.16) into (4.21) we obtain the following expression for the gradient of the Bregman divergence $D_M(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$:

$$\begin{aligned}
\nabla_{\boldsymbol{\theta}} D_M(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = -T^*_{\boldsymbol{\theta}}M(M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1}) = 2\ (&\mathrm{vec}_r(((M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1}) \odot S)\ \mathrm{mat}_r(\boldsymbol{a})), \\
&\mathrm{vec}_r(((M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1}) \odot S)\ \mathrm{mat}_r(\boldsymbol{b}))),
\end{aligned} \tag{C.17}$$

Finally, the relations (C.8), (C.14), (C.15), and (C.17) lead to expressions (B.12)-(B.13) for the components of the gradient for the local model (B.6) in the case of rank deficient family.

Jacobian of the local model. We use the general expression (4.24) to determine the tangent map of $\nabla_{\boldsymbol{\theta}}\tilde{f}^{(k)}(\boldsymbol{\theta})$ for any $\boldsymbol{\delta\theta} := (\boldsymbol{\delta a}, \boldsymbol{\delta b})$. The only ingredient that needs to be specified in order to apply (4.24) is the tangent map of the gradient $\nabla_{\boldsymbol{\theta}} D_M(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$ (C.16) of the Bregman divergence (B.11) associated to the positive definite constraint (B.5). A straightforward computation yields

$$\begin{aligned}
&T_{\boldsymbol{\theta}}(\nabla_{\boldsymbol{\theta}} D_M(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})) \cdot \boldsymbol{\delta\theta} = -T_{\boldsymbol{\theta}}(T^*_{\boldsymbol{\theta}}M(M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1})) \cdot \boldsymbol{\delta\theta} \\
&= 2\ T_{\boldsymbol{\theta}}\left(\mathrm{vec}_r(((M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1}) \odot S)\mathrm{mat}_r(\boldsymbol{a})), \mathrm{vec}_r(((M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1}) \odot S)\mathrm{mat}_r(\boldsymbol{b}))\right) \cdot \boldsymbol{\delta\theta} \\
&= -2(\mathrm{vec}_r(((M(\boldsymbol{\theta})^{-1}(T_{\boldsymbol{\theta}}M \cdot \boldsymbol{\delta\theta})M(\boldsymbol{\theta})^{-1}) \odot S)\mathrm{mat}_r(\boldsymbol{a}) - ((M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1}) \odot S)\mathrm{mat}_r(\boldsymbol{\delta a})), \\
&\qquad \mathrm{vec}_r(((M(\boldsymbol{\theta})^{-1}(T_{\boldsymbol{\theta}}M \cdot \boldsymbol{\delta\theta})M(\boldsymbol{\theta})^{-1}) \odot S)\mathrm{mat}_r(\boldsymbol{b}) - ((M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1}) \odot S)\mathrm{mat}_r(\boldsymbol{\delta b}))).
\end{aligned} \tag{C.18}$$

The substitution of the expressions for the tangent maps and their adjoints provided in this appendix into (4.24) yields the relations (B.14)-(B.15) that can be used to obtain the blocks of the Jacobian matrix for the local model (B.6) in the case of the rank deficient DCC specification.

C.3 The Almon DCC model

Gradient of the local model. We determine the explicit expression for the gradient of the local model which requires computation of the gradients of each of the Bregman divergences presented

in (B.22)-(B.25). For the divergences associated with the nonlinear positivity constraints we use the relation (4.22) and for $D_{N_1}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$ and $D_{N_2}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$ related to the inequalities (B.16)-(B.17) we write

$$\nabla_{\boldsymbol{\theta}} D_{N_1}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = -\left[\frac{1}{(\mathrm{alm}_n(\boldsymbol{a}))_1} - \frac{1}{(\mathrm{alm}_n(\boldsymbol{a}^{(k)}))_1}\right]\left(K_{\boldsymbol{a}}^{\top} \cdot \mathbf{e}_n^{(1)}\right), \tag{C.19}$$

$$\nabla_{\boldsymbol{\theta}} D_{N_2}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = -\left[\frac{1}{(\mathrm{alm}_n(\boldsymbol{b}))_1} - \frac{1}{(\mathrm{alm}_n(\boldsymbol{b}^{(k)}))_1}\right]\left(K_{\boldsymbol{b}}^{\top} \cdot \mathbf{e}_n^{(1)}\right), \tag{C.20}$$

where we used the definition (2.17) of the adjoint $T^*_{\mathbf{v}}\mathrm{alm}_n : \mathbb{R}^n \longrightarrow \mathbb{R}^3$ of the tangent map $T_{\mathbf{v}}\mathrm{alm}_n : \mathbb{R}^3 \longrightarrow \mathbb{R}^n$; recall that $K_{\boldsymbol{a}}, K_{\boldsymbol{b}} \in \mathbb{M}_{n,3}$ are the matrices introduced in (2.16) and $\mathbf{e}_n^{(1)} \in \mathbb{R}^n$ is the canonical unit vector.

We now carry out the computation of the gradients for the components $D^i_{N_2}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}), D^i_{N_3}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$, $i \in \{1, \ldots, N\}$ of the Bregman divergences associated to the nonlinear positivity constraints in (B.18) and (B.19), respectively. For that we first need to introduce the maps $T_{\boldsymbol{\theta}}N_2 : \mathbb{R}^{N^*} \times \mathbb{R}^{N^*} \longrightarrow \mathbb{R}^N$ and $T_{\boldsymbol{\theta}}N_3 : \mathbb{R}^{N^*} \times \mathbb{R}^{N^*} \longrightarrow \mathbb{R}^N$. Let $\boldsymbol{\delta\theta} := (\boldsymbol{\delta a}, \boldsymbol{\delta b}) \in \mathbb{R}^3 \times \mathbb{R}^3$, then the following relations hold true:

$$\begin{aligned} T_{\boldsymbol{\theta}}N_2 \cdot \boldsymbol{\delta\theta} = &-\mathrm{vech}((K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a}) \cdot \mathrm{alm}_n(\boldsymbol{a})^{\top} + \mathrm{alm}_n(\boldsymbol{a}) \cdot (K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a})^{\top} \\ &+ (K_{\boldsymbol{b}} \cdot \boldsymbol{\delta b}) \cdot \mathrm{alm}_n(\boldsymbol{b})^{\top} + \mathrm{alm}_n(\boldsymbol{b}) \cdot (K_{\boldsymbol{b}} \cdot \boldsymbol{\delta b})^{\top}), \end{aligned} \tag{C.21}$$

$$\begin{aligned} T_{\boldsymbol{\theta}}N_3 \cdot \boldsymbol{\delta\theta} = &\mathrm{vech}((K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a}) \cdot \mathrm{alm}_n(\boldsymbol{a})^{\top} + \mathrm{alm}_n(\boldsymbol{a}) \cdot (K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a})^{\top} \\ &+ (K_{\boldsymbol{b}} \cdot \boldsymbol{\delta b}) \cdot \mathrm{alm}_n(\boldsymbol{b})^{\top} + \mathrm{alm}_n(\boldsymbol{b}) \cdot (K_{\boldsymbol{b}} \cdot \boldsymbol{\delta b})^{\top}). \end{aligned} \tag{C.22}$$

In order to use (4.22), we need to determine the adjoint maps $T^*_{\boldsymbol{\theta}}N_1 : \mathbb{R}^N \longrightarrow \mathbb{R}^{N^*} \times \mathbb{R}^{N^*}$ and $T^*_{\boldsymbol{\theta}}N_2 : \mathbb{R}^N \longrightarrow \mathbb{R}^{N^*} \times \mathbb{R}^{N^*}$ by dualizing the relations (C.21) and (C.22). For $\mathbf{v} \in \mathbb{R}^N$ arbitrary we compute

$$\begin{aligned} &\langle T^*_{\boldsymbol{\theta}}N_2 \cdot \mathbf{v}, (\boldsymbol{\delta a}, \boldsymbol{\delta b})\rangle = \langle \mathbf{v}, T_{\boldsymbol{\theta}}N_2(\boldsymbol{\delta a}, \boldsymbol{\delta b})\rangle = -\langle \mathbf{v}, \mathrm{vech}((K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a}) \cdot \mathrm{alm}_n(\boldsymbol{a})^{\top} + \mathrm{alm}_n(\boldsymbol{a}) \cdot (K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a})^{\top} \\ &\quad + (K_{\boldsymbol{b}} \cdot \boldsymbol{\delta b}) \cdot \mathrm{alm}_n(\boldsymbol{b})^{\top} + \mathrm{alm}_n(\boldsymbol{b}) \cdot (K_{\boldsymbol{b}} \cdot \boldsymbol{\delta b})^{\top})\rangle = -\langle \mathbf{v}, \mathrm{vech}((K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a}) \cdot \mathrm{alm}_n(\boldsymbol{a})^{\top})\rangle \\ &\quad - \langle \mathbf{v}, \mathrm{vech}(\mathrm{alm}_n(\boldsymbol{a}) \cdot (K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a})^{\top})\rangle - \langle \mathbf{v}, \mathrm{vech}((K_{\boldsymbol{b}} \cdot \boldsymbol{\delta b}) \cdot \mathrm{alm}_n(\boldsymbol{b})^{\top})\rangle - \langle \mathbf{v}, \mathrm{vech}(\mathrm{alm}_n(\boldsymbol{b}) \cdot (K_{\boldsymbol{b}} \cdot \boldsymbol{\delta b})^{\top})\rangle \\ &= -\langle \mathrm{vech}^*(\mathbf{v}) \cdot \mathrm{alm}_n(\boldsymbol{a}), K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a}\rangle - \langle \mathrm{alm}_n(\boldsymbol{a})^{\top} \cdot \mathrm{vech}^*(\mathbf{v}), (K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a})^{\top}\rangle - \langle \mathrm{vech}^*(\mathbf{v}) \cdot \mathrm{alm}_n(\boldsymbol{b}), K_{\boldsymbol{b}} \cdot \boldsymbol{\delta b}\rangle \\ &\quad - \langle \mathrm{alm}_n(\boldsymbol{b})^{\top} \cdot \mathrm{vech}^*(\mathbf{v}), (K_{\boldsymbol{b}} \cdot \boldsymbol{\delta b})^{\top}\rangle = -2\langle K_{\boldsymbol{a}}^{\top} \cdot \mathrm{vech}^*(\mathbf{v}) \cdot \mathrm{alm}_n(\boldsymbol{a}), \boldsymbol{\delta a}\rangle - 2\langle K_{\boldsymbol{b}}^{\top} \cdot \mathrm{vech}^*(\mathbf{v}) \cdot \mathrm{alm}_n(\boldsymbol{b}), \boldsymbol{\delta b}\rangle \\ &= -2\langle (K_{\boldsymbol{a}}^{\top} \cdot \mathrm{vech}^*(\mathbf{v}) \cdot \mathrm{alm}_n(\boldsymbol{a}), K_{\boldsymbol{b}}^{\top} \cdot \mathrm{vech}^*(\mathbf{v}) \cdot \mathrm{alm}_n(\boldsymbol{b})), (\boldsymbol{\delta a}, \boldsymbol{\delta b})\rangle, \end{aligned}$$

which immediately yields

$$T^*_{\boldsymbol{\theta}}N_2 \cdot \mathbf{v} = -2(K_{\boldsymbol{a}}^{\top} \cdot \mathrm{vech}^*(\mathbf{v}) \cdot \mathrm{alm}_n(\boldsymbol{a}), K_{\boldsymbol{b}}^{\top} \cdot \mathrm{vech}^*(\mathbf{v}) \cdot \mathrm{alm}_n(\boldsymbol{b})),$$

and, analogously, it can be easily checked that

$$T^*_{\boldsymbol{\theta}}N_3 \cdot \mathbf{v} = 2(K_{\boldsymbol{a}}^{\top} \cdot \mathrm{vech}^*(\mathbf{v}) \cdot \mathrm{alm}_n(\boldsymbol{a}), K_{\boldsymbol{b}}^{\top} \cdot \mathrm{vech}^*(\mathbf{v}) \cdot \mathrm{alm}_n(\boldsymbol{b})).$$

Hence we obtain the following expression for the gradients of divergences $D_{N_3}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$ and $D_{N_4}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$:

$$\begin{aligned} \nabla_{\boldsymbol{\theta}} D^i_{N_j}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = 2 \cdot (-1)^j \cdot \Bigg(&\left[\frac{1}{(N_j(\boldsymbol{\theta}))_i} - \frac{1}{(N_j(\boldsymbol{\theta}^{(k)}))_i}\right] (K_{\boldsymbol{a}}^{\top} \cdot \mathrm{vech}^*(\mathbf{e}_N^{(i)}) \cdot \mathrm{alm}_n(\boldsymbol{a})), \\ &\left[\frac{1}{(N_j(\boldsymbol{\theta}))_i} - \frac{1}{(N_j(\boldsymbol{\theta}^{(k)}))_i}\right] (K_{\boldsymbol{b}}^{\top} \cdot \mathrm{vech}^*(\mathbf{e}_N^{(i)}) \cdot \mathrm{alm}_n(\boldsymbol{b}))\Bigg), \quad j \in \{3, 4\}. \end{aligned} \tag{C.23}$$

We now proceed with the positive definite constraint (B.5), for which we recall that

$$M(\boldsymbol{\theta}) = (\mathbf{i}_n\mathbf{i}_n^\top - \mathrm{alm}_n(\boldsymbol{a})\,\mathrm{alm}_n(\boldsymbol{a})^\top - \mathrm{alm}_n(\boldsymbol{b})\,\mathrm{alm}_n(\boldsymbol{b})^\top) \odot S.$$

In order to provide the gradient for the Bregman divergences associated to the constraint (B.5), we refer to the relation (4.21), whose explicit formulation requires the expression of the adjoint map $T^*_{\boldsymbol{\theta}}M : \mathbb{S}_n \longrightarrow \mathbb{R}^{N^*} \times \mathbb{R}^{N^*}$. We hence first determine the corresponding tangent map $T_{\boldsymbol{\theta}}M : \mathbb{R}^{N^*} \times \mathbb{R}^{N^*} \longrightarrow \mathbb{S}_n$ for any $\boldsymbol{\delta\theta} = (\boldsymbol{\delta a}, \boldsymbol{\delta b}) \in \mathbb{R}^{N^*} \times \mathbb{R}^{N^*}$ in a straightforward manner

$$T_{\boldsymbol{\theta}}M \cdot \boldsymbol{\delta\theta} = -((K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a}) \cdot \mathrm{alm}_n(\boldsymbol{a})^\top + \mathrm{alm}_n(\boldsymbol{a}) \cdot (K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a})^\top + (K_{\boldsymbol{b}} \cdot \boldsymbol{\delta b}) \cdot \mathrm{alm}_n(\boldsymbol{b})^\top + \mathrm{alm}_n(\boldsymbol{b}) \cdot (K_{\boldsymbol{b}} \cdot \boldsymbol{\delta b})^\top) \odot S$$

and dualize it in order to obtain the required adjoint, that is, for arbitrary $\Delta \in \mathbb{S}_n$, we compute

$$\begin{aligned}
&\langle T^*_{\boldsymbol{\theta}}M \cdot \Delta, (\boldsymbol{\delta a}, \boldsymbol{\delta b})\rangle = \langle \Delta, T_{\boldsymbol{\theta}}M \cdot (\boldsymbol{\delta a}, \boldsymbol{\delta b})\rangle \\
&= -\langle \Delta, ((K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a}) \cdot \mathrm{alm}_n(\boldsymbol{a})^\top + \mathrm{alm}_n(\boldsymbol{a}) \cdot (K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a})^\top + (K_{\boldsymbol{b}} \cdot \boldsymbol{\delta b}) \cdot \mathrm{alm}_n(\boldsymbol{b})^\top + \mathrm{alm}_n(\boldsymbol{b}) \cdot (K_{\boldsymbol{b}} \cdot \boldsymbol{\delta b})^\top) \odot S\rangle \\
&= -\langle \Delta, ((K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a}) \cdot \mathrm{alm}_n(\boldsymbol{a})^\top) \odot S\rangle - \langle \Delta, (\mathrm{alm}_n(\boldsymbol{a}) \cdot (K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a})^\top) \odot S\rangle - \langle \Delta, ((K_{\boldsymbol{b}} \cdot \boldsymbol{\delta b}) \cdot \mathrm{alm}_n(\boldsymbol{b})^\top) \odot S\rangle \\
&\quad - \langle \Delta, (\mathrm{alm}_n(\boldsymbol{b}) \cdot (K_{\boldsymbol{b}} \cdot \boldsymbol{\delta b})^\top) \odot S\rangle = -\langle (\Delta \odot S)\mathrm{alm}_n(\boldsymbol{a}), K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a}\rangle - \langle \mathrm{alm}_n(\boldsymbol{a})^\top(\Delta \odot S), (K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a})^\top\rangle \\
&\quad - \langle (\Delta \odot S)\mathrm{alm}_n(\boldsymbol{b}), K_{\boldsymbol{b}} \cdot \boldsymbol{\delta b}\rangle - \langle \mathrm{alm}_n(\boldsymbol{b})^\top(\Delta \odot S), (K_{\boldsymbol{b}} \cdot \boldsymbol{\delta b})^\top\rangle = -2\,\langle K_{\boldsymbol{a}}^\top(\Delta \odot S)\mathrm{alm}_n(\boldsymbol{a}), \boldsymbol{\delta a}\rangle \\
&\quad - 2\,\langle K_{\boldsymbol{b}}^\top(\Delta \odot S)\mathrm{alm}_n(\boldsymbol{b}), \boldsymbol{\delta b}\rangle = -2\langle (K_{\boldsymbol{a}}^\top(\Delta \odot S)\mathrm{alm}_n(\boldsymbol{a}), K_{\boldsymbol{b}}^\top(\Delta \odot S)\mathrm{alm}_n(\boldsymbol{b})), (\boldsymbol{\delta a}, \boldsymbol{\delta b})\rangle.
\end{aligned}$$

This relation is equivalent to

$$T^*_{\boldsymbol{\theta}}M \cdot \Delta = -2\,(K_{\boldsymbol{a}}^\top(\Delta \odot S)\mathrm{alm}_n(\boldsymbol{a}), K_{\boldsymbol{b}}^\top(\Delta \odot S)\mathrm{alm}_n(\boldsymbol{b})) \in \mathbb{R}^3 \times \mathbb{R}^3. \tag{C.24}$$

Directly substituting (C.24) into (4.21) gives the following expression for the required gradient of the Bregman divergence $D_M(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$

$$\begin{aligned}
\nabla_{\boldsymbol{\theta}} D_M(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = -T^*_{\boldsymbol{\theta}}M(M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1}) = 2\,(K_{\boldsymbol{a}}^\top((M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1}) \odot S)\,\mathrm{alm}_n(\boldsymbol{a}), \\
K_{\boldsymbol{b}}^\top((M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1}) \odot S)\,\mathrm{alm}_n(\boldsymbol{b})).
\end{aligned} \tag{C.25}$$

Finally, the relations (C.19), (C.20), (C.23), and (C.25) lead to expressions (B.26)-(B.27) for the components of the gradient of the local model (B.21) in the case of the Almon family of DCC models.

Jacobian of the local model. Again we use the general expression (4.24) to determine the tangent map of $\nabla_{\boldsymbol{\theta}}\tilde{f}^{(k)}(\boldsymbol{\theta})$ for any $\boldsymbol{\delta\theta} := (\boldsymbol{\delta a}, \boldsymbol{\delta b}) \in \mathbb{R}^3 \times \mathbb{R}^3$. In order to apply (4.24), we only need to specify the tangent map of the gradient $\nabla_{\boldsymbol{\theta}} D_M(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$ (C.25) of the Bregman divergence (B.25) associated to the positive definite constraint (B.20). A straightforward computation yields

$$\begin{aligned}
&T_{\boldsymbol{\theta}}(\nabla_{\boldsymbol{\theta}} D_M(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})) \cdot \boldsymbol{\delta\theta} = -T_{\boldsymbol{\theta}}(T^*_{\boldsymbol{\theta}}M(M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1})) \cdot \boldsymbol{\delta\theta} \\
&=2\,T_{\boldsymbol{\theta}}(K_{\boldsymbol{a}}^\top \cdot ((M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1}) \odot S) \cdot \mathrm{alm}_n(\boldsymbol{a}), K_{\boldsymbol{b}}^\top \cdot ((M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1}) \odot S) \cdot \mathrm{alm}_n(\boldsymbol{b})) \\
&=2\big((K_{\boldsymbol{aa}} \cdot \boldsymbol{\delta a})^\top \cdot ((M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1}) \odot S) \cdot \mathrm{alm}_n(\boldsymbol{a}) - K_{\boldsymbol{a}} \cdot ((M(\boldsymbol{\theta})^{-1}(T_{\boldsymbol{\theta}}M \cdot \boldsymbol{\delta\theta})M(\boldsymbol{\theta})^{-1}) \odot S) \\
&\quad \cdot \mathrm{alm}_n(\boldsymbol{a}) + K_{\boldsymbol{a}}^\top \cdot ((M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1}) \odot S) \cdot (K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a}), (K_{\boldsymbol{bb}} \cdot \boldsymbol{\delta b})^\top \cdot ((M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1}) \odot S) \\
&\quad \cdot \mathrm{alm}_n(\boldsymbol{b}) + K_{\boldsymbol{b}} \cdot ((-M(\boldsymbol{\theta})^{-1}(T_{\boldsymbol{\theta}}M \cdot \boldsymbol{\delta\theta})M(\boldsymbol{\theta})^{-1}) \odot S) \cdot \mathrm{alm}_n(\boldsymbol{b}) + K_{\boldsymbol{b}}^\top \cdot ((M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1}) \odot S) \\
&\quad \cdot (K_{\boldsymbol{b}} \cdot \boldsymbol{\delta b})\big),
\end{aligned} \tag{C.26}$$

where

$$
\begin{aligned}
K_{\boldsymbol{aa}} \cdot \boldsymbol{\delta a} &:= (\mathbf{0}_n \mid \mathbf{k}_n^1 \odot (K_{\boldsymbol{a}}^0 \cdot \boldsymbol{\delta a}) \mid \mathbf{k}_n^2 \odot (K_{\boldsymbol{a}}^0 \cdot \boldsymbol{\delta a})),\\
K_{\boldsymbol{bb}} \cdot \boldsymbol{\delta b} &:= (\mathbf{0}_n \mid \mathbf{k}_n^1 \odot (K_{\boldsymbol{b}}^0 \cdot \boldsymbol{\delta b}) \mid \mathbf{k}_n^2 \odot (K_{\boldsymbol{b}}^0 \cdot \boldsymbol{\delta b})),\\
K_{\boldsymbol{a}}^0 &:= \left(\mathbf{0}_n \mid \mathbf{k}_n^1 \odot \mathrm{alm}_n(\bar{\boldsymbol{a}}) \mid \mathbf{k}_n^2 \odot \mathrm{alm}_n(\bar{\boldsymbol{a}})\right),\\
K_{\boldsymbol{b}}^0 &:= \left(\mathbf{0}_n \mid \mathbf{k}_n^1 \odot \mathrm{alm}_n(\bar{\boldsymbol{b}}) \mid \mathbf{k}_n^2 \odot \mathrm{alm}_n(\bar{\boldsymbol{b}})\right).
\end{aligned}
$$

The substitution of the expressions for the tangent maps and their adjoints provided in this appendix into (4.24) yields the relations (B.29)-(B.30) that can be used to obtain the blocks of the Jacobian matrix for the local model (B.21) in the case of the Almon DCC specification.

References

[Bauw 14] L. Bauwens, L. Grigoryeva, and J.-P. Ortega. "Estimation and empirical performance of non-scalar dynamic conditional correlation models". 2014.

[Chre 13] S. Chrétien and J.-P. Ortega. "Multivariate GARCH estimation via a Bregman-proximal trust-region method". *Computational Statistics and Data Analysis*, Vol. To appear, 2013.

Recent titles

CORE Discussion Papers

2013/47 Helmuth CREMER, Firouz GAHVARI and Pierre PESTIEAU. Uncertain altruism and the provision of long term care.

2013/48 Claire DUJARDIN, Vincent LORANT and Isabelle THOMAS. Self-assessed health of elderly people in Brussels: does the built environment matter?

2013/49 Marc FLEURBAEY, Marie-Louise LEROUX, Pierre PESTIEAU and Grégory PONTHIERE. Fair retirement under risky lifetime.

2013/50 Manuel FÖRSTER, Ana MAULEON and Vincent VANNETELBOSCH. Trust and manipulation in social networks.

2013/51 Anthony PAPAVASILIOU, Yi HE and Alva SVOBODA. Self-commitment of combined cycle units under electricity price uncertainty.

2013/52 Ana MAULEON, Elena MOLIS, Vincent VANNETELBOSCH and Wouter VERGOTE. Dominance invariant one-to-one matching problems.

2013/53 Jean GABSZEWICZ and Skerdilajda ZANAJ. (Un)stable vertical collusive agreements.

2013/54 François MANIQUET and Massimo MORELLI. Approval quorums dominate participation quorums.

2013/55 Mélanie LEFÈVRE and Joe THARAKAN. Intermediaries, transport costs and interlinked transactions.

2013/56 Gautier M. KRINGS, Jean-François CARPANTIER and Jean-Charles DELVENNE. Trade integration and the trade imbalances in the European Union: a network perspective.

2013/57 Philip USHCHEV, Igor SLOEV and Jacques-François THISSE. Do we go shopping downtown or in the 'burbs'? Why not both?

2013/58 Mathieu PARENTI. Large and small firms in a global market: David vs. Goliath.

2013/59 Paul BELLEFLAMME and Francis BLOCH. Dynamic protection of innovations through patents and trade secrets.

2013/60 Christian HAEDO and Michel MOUCHART. Specialized agglomerations with areal data: model and detection.

2013/61 Julien MARTIN and Florian MAYNERIS. High-end variety exporters defying distance: micro facts and macroeconomic implications.

2013/62 Luca G. DEIDDA and Dimitri PAOLINI. Wage premia, education race, and supply of educated workers.

2013/63 Laurence A. WOLSEY and Hande YAMAN. Continuous knapsack sets with divisible capacities.

2013/64 Francesco DI COMITE, Jacques-François THISSE and Hylke VANDENBUSSCHE. Verti-zontal differentiation in export markets.

2013/65 Carl GAIGNÉ, Stéphane RIOU and Jacques-François THISSE. How to make the metropolitan area work? Neither big government, nor laissez-faire.

2013/66 Yu. NESTEROV and Vladimir SHIKHMAN. Algorithmic models of market equilibrium.

2013/67 Cristina PARDO-GARCIA and Jose J. SEMPERE-MONERRIS. Equilibrium mergers in a composite good industry with efficiencies.

2013/68 Federica RUSSO, Michel MOUCHART and Guillaume WUNSCH. Confounding and control in a multivariate system. An issue in causal attribution.

2013/69 Marco DI SUMMA. The convex hull of the all-different system with the inclusion property: a simple proof.

2013/70 Philippe DE DONDER and Pierre PESTIEAU. Lobbying, family concerns and the lack of political support for estate taxation.

2013/71 Alexander OSHARIN, Jacques-François THISSE, Philip USHCHEV and Valery VERBUS. Monopolistic competition and income dispersion.

2013/72 N. Baris VARDAR. Imperfect resource substitution and optimal transition to clean technologies.

2013/73 Alejandro LAMAS and Philippe CHEVALIER. Jumping the hurdles for collaboration: fairness in operations pooling in the absence of transfer payments.

Recent titles

CORE Discussion Papers - continued

2013/74 Mehdi MADANI and Mathieu VAN VYVE. A new formulation of the European day-ahead electricity market problem and its algorithmic consequences.
2014/1 Erik SCHOKKAERT and Tom TRUYTS. Preferences for redistribution and social structure.
2014/2 Maarten VAN DIJCK and Tom TRUYTS. The agricultural invasion and the political economy of agricultural trade policy in Belgium, 1875-1900.
2014/3 Ana MAULEON, Nils ROEHL and Vincent VANNETELBOSCH. Constitutions and social networks.
2014/4 Nicolas CARAYOL, Rémy DELILLE and Vincent VANNETELBOSCH. Allocating value among farsighted players in network formation.
2014/5 Yu. NESTEROV and Vladimir SHIKHMAN. Convergent subgradient methods for nonsmooth convex minimization.
2014/6 Yuri YATSENKO, Natali HRITONENKO and Thierry BRECHET. Modeling of enrironmental adaptation versus pollution mitigation.
2014/7 Sanjeeb DASH, Oktay GÜNLÜK and Laurence A. WOLSEY. The continuous knapsack set.
2014/8 Simon BUCKLE, Mirabelle MUÛLS, Joerg LEIB and Thierry BRECHET. Prospects for Paris 2015: do major emitters want the same climate.
2014/9 Lionel ARTIGE, Antoine DEDRY and Pierre PESTIEAU. Social security and economic integration.
2014/10 Mikhail ISKAKOV, Alexey ISKAKOV and Alexey ZAKHAROV. Equilibria in secure strategies in the Tullock contest.
2014/11 Helmuth CREMER and Pierre PESTIEAU. Means-tested long term care and family transfers.
2014/12 Luc BAUWENS, Lyudmila GRIGORYEVA and Juan-Pablo ORTEGA. Estimation and empirical performance of non-scalar dynamic conditional correlation models.

Books

V. GINSBURGH and S. WEBER (2011), *How many languages make sense? The economics of linguistic diversity*. Princeton University Press.
I. THOMAS, D. VANNESTE and X. QUERRIAU (2011), *Atlas de Belgique – Tome 4 Habitat*. Academia Press.
W. GAERTNER and E. SCHOKKAERT (2012), *Empirical social choice*. Cambridge University Press.
L. BAUWENS, Ch. HAFNER and S. LAURENT (2012), *Handbook of volatility models and their applications*. Wiley.
J-C. PRAGER and J. THISSE (2012), *Economic geography and the unequal development of regions*. Routledge.
M. FLEURBAEY and F. MANIQUET (2012), *Equality of opportunity: the economics of responsibility*. World Scientific.
J. HINDRIKS (2012), *Gestion publique*. De Boeck.
M. FUJITA and J.F. THISSE (2013), *Economics of agglomeration: cities, industrial location, and globalization.* (2nd edition). Cambridge University Press.
J. HINDRIKS and G.D. MYLES (2013). *Intermediate public economics.* (2nd edition). MIT Press.
J. HINDRIKS, G.D. MYLES and N. HASHIMZADE (2013). *Solutions manual to accompany intermediate public economics.* (2nd edition). MIT Press.

CORE Lecture Series

R. AMIR (2002), Supermodularity and complementarity in economics.
R. WEISMANTEL (2006), Lectures on mixed nonlinear programming.
A. SHAPIRO (2010), Stochastic programming: modeling and theory.