PROJECTED DYNAMIC CONDITIONAL CORRELATIONS

Christian Brownlees[*]   Jordi Llorens[*]

PRELIMINARY AND INCOMPLETE
PLEASE DO NOT CITE WITHOUT PERMISSION
January 10, 2020

Abstract

The Dynamic Conditional Correlation (DCC) model is one of the leading approaches in the literature for modeling time-varying correlation matrices. In the DCC framework the conditional correlation matrix is modeled as a function of the so called pseudo-correlation matrix, a symmetric positive-definite proxy of the conditional correlation matrix that, however, is not guaranteed to have a unit diagonal. Conditional correlations are then obtained by appropriately rescaling this matrix. In this work we propose a novel DCC specification based on an alternative normalization of the pseudo-correlation matrix called Projected DCC (Pro-DCC). Rather than *rescaling*, we propose *projecting* the pseudo-correlation matrix onto the set of correlation matrices in order to obtain the correlation matrix closest to that pseudo-correlation matrix. A simulation study shows that projecting performs better than rescaling when the dimensionality of the correlation matrix is large and the degree of dependence is high. An empirical application to the constituents of the S&P 100 shows that the proposed methodology performs favourably to the standard DCC in an out-of-sample asset allocation exercise.

Keywords: Multivariate Volatility, DCC, Bregman projection, nearest-correlation matrix, Stein's loss.

1 Introduction

Estimating and forecasting the time-varying covariance matrix of asset returns is key for several applications in finance including asset allocation, risk management and systemic risk measurement. Over the years, the GARCH-DCC methodology of Engle (2002) has established itself as one of the leading paradigms in the literature due to its flexibility and ease of estimation (see also Engle and Sheppard, 2001). In a nutshell, the GARCH-DCC approach consists in modeling separately the conditional variances and the conditional correlation matrix. The conditional variances are modeled using GARCH whereas the conditional correlation matrix is modeled using the Dynamic Conditional Correlation (DCC) model. Recent research in the literature that is based on GARCH-DCC includes Engle, Ledoit, and Wolf (2019), Brownlees and Engle (2017) and De Nard, Ledoit, and Wolf (2018).

A key aspect of the DCC methodology is that the conditional correlation matrix is modeled as a function of the so called pseudo-correlation matrix. The pseudo-correlation matrix is a symmetric positive definite proxy of the conditional correlation matrix that, crucially, is not guaranteed to be a proper correlation matrix as it does not have a unit diagonal (almost surely). In order to obtain correlations, the pseudo-correlation matrix has to be appropriately normalized, and the standard strategy followed in the literature consists in rescaling this matrix (Engle, 2002; Tse and Tsui, 2002; Aielli, 2013). Engle (2009, Section 4.3) contains a discussion and a comparison of different rescaling approaches used in the literature. Despite the fact that rescaling is natural and commonly employed, it is unclear whether such an approach is in any sense optimal.

In this work we propose a modification of the standard DCC model based on an alternative normalization procedure of the pseudo-correlation matrix. Our modification consists in *projecting* the pseudo-correlation matrix onto the set of correlation matrices rather than *rescaling* it. In other words, we cast the normalization step of the pseudo-correlation matrix as a nearest-correlation matrix problem, that is the problem of finding the closest correlation matrix to a given pseudo-correlation matrix on the basis of an appropriate divergence function.

We begin this work by defining a class of projections for pseudo-correlation matrices. To do so, we first introduce the notion of Bregman divergence between two symmetric positive definite matrices (Bregman, 1967; Banerjee, Merugu, Dhillon, and Ghosh, 2005; Dhillon and Tropp, 2007; Patton, 2018). This family of divergences constitutes a rich collection of divergence functions that includes many familiar losses commonly encountered in the covariance estimation literature such as the Stein and square (Frobenius) losses (Stein, 1986; Dey and Srinivasan, 1985; Pourahmadi, 2013). We then define the projection of a pseudo-correlation matrix onto the set of correlation matrices as the correlation matrix that minimizes the Bregman matrix divergence with respect to that pseudo-correlation matrix. It is straightforward to establish that such a projection exists and is unique. Within this broad class of projections we focus in particular on the one implied by Stein's loss, which we name Stein's projection. Stein's loss is a natural loss function for covariance matrices that is related to the multivariate Gaussian log-likelihood, it is widely used (Ledoit and Wolf, 2018b), and it turns out that in our setting it delivers a fairly tractable projection. In fact, we derive a closed form expression to compute Stein projections in the two-dimensional case and an efficient iterative algorithm for the generic n-dimensional case.

We then introduce a novel DCC specification based on our pseudo-correlation matrix projection called Projected DCC (Pro-DCC). Simply put, the Pro-DCC corresponds to the classic DCC of Engle (2002) with the rescaling step of the pseudo-correlation matrix replaced by our proposed projection. In particular, in this work we focus on a version of the Pro-DCC based on Stein's projection. We acknowledge that alternative projections may be considered as well. In order to estimate the Pro-DCC we propose to follow the same multi-step procedure which is used to estimate other DCC-type models.

A simulation study is carried out to assess the performance of our projection-based methodology. We carry out two exercises. In the first exercise we simulate i.i.d. data from a multivariate Gaussian distribution with mean zero and covariance parameter given by a correlation matrix. We then estimate the correlation matrix of the simulated data by rescaling the sample covariance matrix (i.e. the sample correlation matrix) and by

projecting the sample covariance matrix onto the set of correlation matrices using Stein's projection. We find that the projection-based approach performs better than rescaling in terms of correlation estimation accuracy. Gains are greater when the dimensionality is larger and they are maximized when the cross-sectional dependence is high but not extreme. In the second exercise we compare the out-of-sample forecasting accuracy of DCC and Pro-DCC under misspecification, that is when the DGP differs from both models. In particular, here we consider the exact same set of DGPs as in the simulation study of Engle (2002). We find that Pro-DCC outperforms standard DCC for all the DGPs proposed in Engle (2002) except when the correlation process is constant.

A Global Minimum Variance Portfolio (GMVP) exercise using the constituents of the S&P 100 is used to measure the performance of Pro-DCC. The design of the exercise is close in spirit to the one of De Nard *et al.* (2018). We carry out two empirical exercises. In the first exercise we consider 1000 pairs of assets from the set of S&P 100 constituents and for each pair we carry out a GMVP forecasting exercise. We construct GMVP on the basis of the DCC, Pro-DCC as well as a DCC model in which we do not apply any normalization to the pseudo-correlation matrix (i.e. we take the correlation matrix equal to the pseudo-correlation matrix). Results show that all specifications have a similar out-of-sample performance. In the second exercise we carry out, again, a GMVP exercise using the entire set of assets. In this second exercise we consider the Pro-DCC, DCC, DCC with no rescaling and scalar VECH. We consider both the standard versions of these models as well as the versions based on nonlinear shrinkage. Results show that forecasts based on the standard and nonlinear shrinkage variant of the Pro-DCC achieve the best out-of-sample performance.

This paper is related to different strands of the literature. First, it is related to the literature on multivariate volatility models and the DCC. Important contributions in this area include Bollerslev (1990) Pakel, Shephard, Sheppard, and Engle (2018). Classic surveys of the literature on multivariate volatility modeling are Bauwens, Laurent, and Rombouts (2006) and Silvennoinen and Teräsvirta (2008). Second, it is related to the literature on matrix projections based on Bregman divergences and the nearest-correlation

matrix problem. Contributions in this area include the work of Higham (2002), Dhillon and Tropp (2007) and Kulis, Sustik, and Dhillon (2009).

The rest of the paper is structured as follows. Section 2 introduces the methodology. Section 3 contains the simulation study. Section 4 presents the empirical application. Section 5 concludes the paper. All proofs are collected in the Appendix.

2 Methodology

In this section we first concisely review the DCC model of Engle (2002) and we then introduce the Pro-DCC model.

2.1 The DCC Model

Let $r_t = (r_{1t}, \ldots, r_{nt})'$ denote an n-dimensional vector of log returns observed at time t, for t ranging from 1 to T. The key object of interest of this work is the conditional covariance matrix of returns given past information, that is $\boldsymbol{\Sigma}_t = \mathbb{Cov}_{t-1}(r_t)$. The GARCH-DCC framework is based on the following factorization of the conditional covariance matrix

$$\boldsymbol{\Sigma}_t = \mathbf{D}_t \, \mathbf{R}_t \, \mathbf{D}_t \,,$$

where $\mathbf{D}_t$ is a $n \times n$ diagonal matrix of conditional volatilities (standard deviations) and $\mathbf{R}_t$ is the $n \times n$ conditional correlation matrix.

The conditional volatility matrix $\mathbf{D}_t$ is modeled using some suitable GARCH specification. Assuming, for instance, GARCH(1,1) dynamics we have that the i-th diagonal element of $\mathbf{D}_t$, which we denote by d_{it}, is specified as

$$d_{it}^2 = \omega_i + a_i r_{it-1}^2 + b_i d_{it-1}^2 \,,$$

where ω_i, a_i and b_i are the GARCH(1,1) coefficients satisfying $\omega_i > 0$, $a_i > 0$, $b_i \geq 0$ and $a_i + b_i < 1$.

The conditional correlation matrix $\mathbf{R}_t$ is modeled using the DCC specification. The

DCC models the correlation process as a function of the so-called de-volatilized returns that are defined as $\epsilon_t = \mathbf{D}_t^{-1} r_t$. In the DCC framework the focus is on modeling the so-called pseudo-correlation matrix $\mathbf{Q}_t$ as

$$\mathbf{Q}_t = (1 - \alpha - \beta)\mathbf{C} + \alpha\,\epsilon_{t-1}\epsilon'_{t-1} + \beta\,\mathbf{Q}_{t-1} \tag{1}$$

where α and β are scalar parameters and $\mathbf{C}$ is an $n \times n$ positive definite matrix. It is straightforward to see by recursive substitution that

$$\mathbf{Q}_t = \frac{1 - \alpha - \beta}{1 - \beta}\mathbf{C} + \alpha\sum_{i=0}^{\infty} \beta^i \epsilon_{t-1-i}\epsilon'_{t-1-i} . \tag{2}$$

A crucial aspect of the DCC model on which we build upon in the next section is that the pseudo-correlation matrix is not guaranteed to be a correlation matrix. In particular, it is clear from (2) that $\mathbf{Q}_t$ is symmetric positive definite but (generally) not unit diagonal. Thus, an appropriate normalization step is required to obtain a correlation matrix. The standard approach consists of *rescaling* the pseudo-correlation matrix, that is

$$\mathbf{R}_t = \mathrm{diag}(\mathbf{Q}_t)^{-1/2}\,\mathbf{Q}_t\,\mathrm{diag}(\mathbf{Q}_t)^{-1/2} , \tag{3}$$

where for an $n \times n$ matrix $\mathbf{A}$, the notation $\mathrm{diag}(\mathbf{A})$ denotes the $n \times n$ diagonal matrix with the diagonal of $\mathbf{A}$.

The GARCH-DCC family of models is estimated using a multi-step procedure motivated by a QML argument. The first step consists of estimating the conditional standard deviation matrix $\mathbf{D}_t$ by estimating n univariate GARCH models. Next, the $\mathbf{C}$ matrix is estimated by correlation targeting as the sample second moment of the estimated standardized residuals, that is

$$\widehat{\mathbf{C}} = \frac{1}{T}\sum_{t=1}^{T} \hat{\epsilon}_t\hat{\epsilon}'_t ,$$

where $\hat{\epsilon}_{it} = r_{it}/\hat{\sigma}_{it}$ with $\hat{\sigma}_{it}$ being the estimated volatility of the first step. Last, the DCC parameters are obtained by maximizing the (Gaussian) quasi log-likelihood.

It is important to mention that the correlation targeting step has been subject to

some criticism. Albeit being intuitive, Engle's model formulation has some subtle issues first noted by Aielli (2013). The model "fixes the bug" of the DCC however the model dynamics are somehow not intuitive. Empirically, this model is found to perform similarly to the standard uncorrected DCC.

2.2   The Projected DCC Model

In this section we propose a novel DCC specification based on an alternative normalization procedure. Rather than *rescaling* the pseudo-correlation matrix as in equation (3) we propose *projecting* it onto the space of correlation matrices. In other words, we cast the problem of rescaling the pseudo-correlation matrix as a nearest-correlation matrix problem, that is finding the closest correlation matrix to a given pseudo-correlation matrix. In order to introduce our projection-based model some additional machinery is required.

We begin by introducing the notion of Bregman divergence for symmetric positive definite matrices, which is used in this work to measure nearness between two symmetric positive definite matrices.

Definition (Bregman Divergence). Let $\mathbb{S}_+^n$ be the set of $n \times n$ symmetric positive semi-definite matrices. Given a strictly convex and differentiable function $\phi: \mathbb{S}_+^n \to \mathbb{R}$, we define the Bregman matrix divergence $d_\phi : \mathbb{S}_+^n \times \mathrm{ri}(\mathbb{S}_+^n) \to [0, \infty)$ as

$$d_\phi(\mathbf{M}_1, \mathbf{M}_2) = \phi(\mathbf{M}_1) - \phi(\mathbf{M}_2) - \mathrm{tr}(\nabla\phi(\mathbf{M}_2)'(\mathbf{M}_1 - \mathbf{M}_2)) \ .$$

Note that this can be seen as the difference between the function ϕ evaluated at $\mathbf{M}_1$ and its first-order Taylor approximation around $\mathbf{M}_2$. Bregman divergences are a class of tractable divergences that enjoy a number of useful properties and are popular in the Machine Learning literature (Cesa-Bianchi and Lugosi, 2006). Bregman divergences are always positive, like distances, and are zero only when their arguments coincide. Unlike distances, they do not generally satisfy the triangle inequality and are only symmetric when ϕ is quadratic. Furthermore, they are always convex with respect to their first

7

argument and satisfy a generalized Pythagorean property (Dhillon and Tropp, 2007; Kulis *et al.*, 2009). Banerjee (2005) establishes the existence of a bijection between Bregman divergences and regular exponential families. If we set $\phi(\mathbf{M}) = -\log\det(\mathbf{M})$ then we have Stein's loss,

$$d_\phi(\mathbf{M}_1, \mathbf{M}_2) = \operatorname{tr}(\mathbf{M}_1\mathbf{M}_2^{-1}) - \ln\det(\mathbf{M}_1\mathbf{M}_2^{-1}) - n , \quad (4)$$

which can also be interpreted as the negative of the zero mean n-dimensional Gaussian log-likelihood (up to a constant). If we set $\phi(\mathbf{M}) = \|\mathbf{M}\|_F^2$ then we have the squared (Frobenius) loss

$$d_\phi(\mathbf{M}_1, \mathbf{M}_2) = \operatorname{tr}[(\mathbf{M}_1 - \mathbf{M}_2)^2] . \quad (5)$$

Finally, if we set $\phi(\mathbf{M}) = \operatorname{tr}(\mathbf{M}\log\mathbf{M})$ then we have the Von Neumann loss

$$d_\phi(\mathbf{M}_1, \mathbf{M}_2) = \operatorname{tr}(\mathbf{M}_1\log\mathbf{M}_1 - \mathbf{M}_1\log\mathbf{M}_2 - \mathbf{M}_1 + \mathbf{M}_2) . \quad (6)$$

We use Bregman divergences to introduce the following general class of projections for symmetric positive definite matrices onto the set of correlation matrices.

Lemma 1. *Let* $\mathbf{Q}$ *be a symmetric positive definite matrix and let* $\mathbb{C}^n$ *denote the set of correlation matrices. Define the Bregman projection* $P_\phi(\mathbf{Q})$ *as*

$$P_\phi(\mathbf{Q}) = \arg\min_{\mathbf{R}\in\mathbb{C}^n} d_\phi(\mathbf{R}, \mathbf{Q}) .$$

Then we have that $P_\phi(\mathbf{Q})$ *exists and is unique.*

We point out that existence and uniqueness of the Bregman projection follow from the fact that the set of correlation matrices is a convex set and that Bregman divergence is a convex function with respect to its first argument.

Finally, we introduce the Pro-DCC(1,1) as the

$$\mathbf{Q}_t = (1 - \alpha - \beta)\mathbf{C} + \alpha \epsilon_{t-1}\epsilon'_{t-1} + \beta \mathbf{Q}_{t-1}$$
$$\mathbf{R}_t = P_\phi(\mathbf{Q}_t) . \psi$$

In other words, the Pro-DCC replaces the rescaling equation of the DCC (3) with a projection. We point out that the Pro-DCC depends on a choice of an appropriate divergence function $\phi(\cdot)$. A natural choice for such a projection that also turns out to be computationally convenient is projection based on Stein's loss (cf (4)) and the Von Neumann. We call these projections, respectively, Stein's projection and Von Neumann projection for short.

A number of comments are in order. As discussed at the end of section 2.1, the original formulation of the DCC model was criticized in (Aielli, 2013) and a corrected version of the model was introduced. In particular, the correction involves replacing the standardized residuals ϵ_t by $\mathbf{Q}_t^* \epsilon_t$, where $\mathbf{Q}_t^* = \text{diag}(\mathbf{Q}_t)^{1/2}$. The mathematical justification of this correction is that the second moment of the innovations ϵ_t is in general not equal to the matrix $\mathbf{C}$, so $\mathbf{Q}_t$ cannot be thought of as a linear MGARCH process. Instead, in the corrected DCC specification, we have that $\mathbb{E}[\mathbf{Q}_t^* \epsilon_t \epsilon'_t \mathbf{Q}_t^*] = \mathbb{E}[\mathbf{Q}_t]$, and it is justified to estimate $\mathbf{C}$ via correlation targeting, i.e. $\hat{\mathbf{C}} = \frac{1}{T}\sum_{t=1}^T \mathbf{Q}_t^* \epsilon_t \epsilon'_t \mathbf{Q}_t^*$. Crucially, this holds because in the DCC model $\mathbb{E}_{t-1}[\epsilon_t \epsilon'_t] = \mathbf{R}_t = \text{diag}(\mathbf{Q}_t)^{-1/2}\mathbf{Q}_t \text{diag}(\mathbf{Q}_t)^{-1/2}$. Unfortunately, in the Pro-DCC specification the last identity does not hold, hence it is not justified to use the Aielli correction here. Therefore, we do not use the Aielli correction in the Pro-DCC methodology. As a sanity check, we carry out a simulated exercise where the DGP is Pro-DCC, and we observe that using Aielli's estimator produces virtually the same loss as using standard correlation targeting. We show this in Appendix B.

For high-dimensional settings, we propose to estimate the model by composite likelihood as in Pakel $et\ al.$ (2018). However, we note that in general, projecting the pseudo-correlation of any 2 assets i and j is not equal to the (i,j) entry of the projection of the entire matrix $P_\phi(\mathbf{Q}_t)$. This is because the projection takes into account the full correlation

structure and not just the correlation between assets i and j. An exception occurs when we use the Stein's projection of a pseudo-correlation matrix with a block-diagonal structure: in that case, the projection preserves the original structure and one can equivalently compute the projection block by block.

It is widely known that the sample correlation matrix performs poorly when the concentration ratio n/T is large -see Lecture 4 in Stein (1986). For that reason, we consider using a nonlinear shrinkage estimator to rectify the in-sample bias of the sample correlation as in Ledoit and Wolf (2018a).

2.2.1 Computing the Bregman Projection

In order to apply the Pro-DCC in practice it is key to be able to compute the projections in a fairly cheap way. We derive a closed-form expression for the projection in the 2 dimensional case for the Stein and von Neumann losses and we provide an efficient algorithm for the computation of the projection in the general n dimensional case for the Stein projection.

The following two lemmas derive the closed form of the projection.

Lemma 2. *Let* $\mathbf{Q}$ *be a* 2×2 *symmetric positive definite matrix. Consider the Bregman Projection of* $\mathbf{Q}$ *onto the set of correlation matrices under Stein's Loss. The unique minimizer of this problem is given by*

$$\hat{\rho} = \begin{cases} \dfrac{-\sqrt{1+4k^2}}{2k} & k \neq 0 \\ 0 & k = 0 \end{cases}$$

(7)

where $k = -\dfrac{q_{12}}{\det(\mathbf{Q})}$.

Lemma 3. *Let* $\mathbf{Q}$ *be a* 2×2 *symmetric positive definite matrix. Consider the Bregman Projection of* $\mathbf{Q}$ *onto the set of correlation matrices under the Von Neumann Divergence. Let* $\ln(\cdot)$ *denote the natural logarithm, and* $\log(\cdot)$ *the matrix logarithm.*[1] *Then, the unique*

---

[1] For symmetric positive definite matrices, the matrix logarithm is $\log \mathbf{Q} = \mathbf{U} \log \mathbf{\Lambda} U'$, where $U \mathbf{\Lambda} U'$ is the eigendecomposition of $\mathbf{Q}$ and $\log \mathbf{\Lambda}$ involves taking the natural logarithm of the eigenvalues.

minimizer of this problem is given by

$$\hat{\rho} = \tanh(k)$$

where k denotes the $(2,1)$-entry of $\log \mathbf{Q}$.[2]

It is important to emphasize that the optimal projection in these two cases looks very different from rescaling, thus implying that rescaling, at least as far as the Stein and Von Neumann divergences are concerned, is not optimal. We point out that we were not able to obtain informative closed form expressions for $n > 2$. However, computations show that in general the optimal projection of each correlation pair depends on the entire covariance matrix (not just that pair), implying, again, that rescaling cannot be optimal.

In the n-dimensional case we can derive an algorithm. Let $\mathbf{Q}$ be an $n \times n$ symmetric positive definite matrix and $d_\phi(\cdot, \cdot)$ be a matrix Bregman divergence with strictly convex function $\phi(\cdot)$.

Computing the Bregman projection $P_\phi(\mathbf{Q})$ is equivalent to solving the following optimization problem with n affine constraints (one for each diagonal element of $\mathbf{R}$):

$$\min_{\mathbf{R}} d_\phi(\mathbf{R}, \mathbf{Q}) \text{ subject to } \operatorname{tr}(\mathbf{R} e_i e_i') = 1 \text{ for all } i = 1, \ldots, n. \tag{8}$$

where e_i stands for the ith canonical basis vector.

Let $\mathbb{C}_i$ be the set of n-dimensional symmetric positive definite matrices whose i^{th} diagonal element is unity. Clearly, the set of correlation matrices $\mathbb{C}^n = \bigcap_{i=1}^n \mathbb{C}_i$.

To solve this problem, we use Bregman's *cyclic projections* method. This is an iterative algorithm in which one must project successively onto each basic constraint set $\mathbb{C}_i$ in the hope that the sequence of iterates will converge to the Bregman projection onto the

---

[2]Computations show that k has the following analytical expression $k = \frac{1}{1+\xi_1^2} \ln \lambda_1 + \frac{\xi_2}{1+\xi_2^2} \ln \lambda_2$ where

$$\lambda_i = \frac{1}{2}\left[q_{11} + q_{22} + (-1)^{i-1}\sqrt{(q_{11} - q_{22})^2 + 4q_{12}^2} \right]$$

$$\xi_i = -2q_{12} / \left(q_{11} - q_{22} + (-1)^i \sqrt{(q_{11} - q_{22})^2 + 4q_{12}^2} \right)$$

and $i = 1, 2$.

intersection $\mathbb{C}^n$.

Theorem 1 establishes the correctness of Bregman's algorithm. We refer to Dhillon and Tropp (2007) for a proof.

Theorem 1. *Suppose*

(i) ϕ is a convex function of Legendre type.

(ii) the domain of the conjugate function of ϕ is an open set.

(iii) $\mathbb{C}_1, \ldots, \mathbb{C}_n$ are affine spaces with intersection $\mathbb{C}^n$.

(iv) $\mathbb{C}^n \cap ri(dom(\phi))$ is nonempty.

(v) the control mapping $m: \mathbb{N} \to \{1, \ldots, n\}$ is a sequence that takes each output value an infinite number of times.

Define $P_{\phi,m(k)}(\mathbf{R}^{(k-1)})$ as the projection of $\mathbf{R}^{(k-1)}$ onto $\mathbb{C}_{m(k)}$. Choose $\mathbf{R}^{(0)}$ from $ri(dom(\phi))$, and form a sequence of iterates via successive Bregman projections

$$\mathbf{R}^{(k)} = P_{\phi,m(k)}(\mathbf{R}^{(k-1)})$$

Then the sequence of iterates $\{\mathbf{R}^{(k)}\}$ converges in norm to $P_\phi(\mathbf{R}^{(0)})$.

Lemma 4 establishes a closed-form formula for $P_{\phi,m(k)}(\mathbf{R}^{(k-1)})$ and is a special case of the derivation in Kulis *et al.* (2009) when $\phi(\cdot) = -\ln \det(\cdot)$.

Lemma 4. *Consider the setting in Theorem 1 and let $\phi(\cdot) = -\ln \det(\cdot)$. Then, for all $i \in \{1, \ldots, n\}$,*

$$P_{\phi,i}(\mathbf{R}^{(k-1)}) = \mathbf{R}^{(k-1)} + [\mathbf{R}_{ii}^{(k-1)}]^{-2} \left(1 - \mathbf{R}_{ii}^{(k-1)}\right) \mathbf{R}^{(k-1)} e_i e_i' \mathbf{R}^{(k-1)}$$

where $\mathbf{R}_{ii}^{(k)}$ stands for the ith diagonal element of the matrix $\mathbf{R}^{(k)}$.

We concisely describe this procedure in Algorithm 1. We point out that the algorithm has a complexity per iteration of $O(n^2)$.

Algorithm 1 Stein's Projection

Compute Stein's projection of a symmetric positive definite matrix $\mathbf{Q}$ onto the set of correlation matrices.

Input: A symmetric positive definite matrix $\mathbf{Q}$.

Initialization
Set $\mathbf{R}^{(0)} = \mathbf{Q}$.

Iterate until convergence
In the k-th iteration of the algorithm choose the i-th constraint as

$$i \psi= \arg \max_{s \in \{1,...,n\}} |1 - \mathbf{R}_{ss}^{(k-1)}| \, ,$$

and update the projection according to the formula

$$\mathbf{R}^{(k)} = \mathbf{R}^{(k-1)} + \left[\mathbf{R}_{ii}^{(k-1)}\right]^{-2} (1 - \mathbf{R}_{ii}^{(k-1)}) \mathbf{R}^{(k-1)} e_i e_i' \mathbf{R}^{(k-1)} \, ,$$

where e_i is defined as the ith canonical basis vector.

Covergence criteria
If $\max_s |1 - \mathbf{R}_{ss}^{(k)}| < \psi$tolerance then stop.

Output: The projected correlation matrix $\mathbf{R}^{(k)}$.

2.3   Rescaling vs Projecting a Pseudo-Correlation matrix

In the previous sub-section we argue that it is possible to project a pseudo-correlation matrix onto the set of correlation matrices with a very simple approach from a computational perspective. In this sub-section we show that the difference between rescaling and projecting is relevant enough in many cases.

For simplicity, consider a 2-dimensional pseudo-correlation matrix $\mathbf{Q}$ with diagonal elements q_{11}, q_{22} and off-diagonal element q_{12}. Simple algebra shows that when $q_{11}q_{22} = 1$, then the expression in equation 7 boils down to q_{12}, which trivially coincides with rescaling q_{12} by $\sqrt{q_{11}q_{22}}$. When $q_{11}q_{22} \neq 1$, this is generally not true, as it is shown in Figure 1. In the top-left panel we give an example of a combination of diagonal elements whose product is one, and observe that rescaling and projecting are equivalent. If the product is greater than 1, the projected correlation is below the rescaled one, and the difference increases as the product moves away from 1. The reverse pattern occurs when the product between the diagonal elements is lower than 1. We also note that the point at which the maximum difference occurs does not correspond to the same correlation level but is a

13

function of the product of the diagonal elements of $\mathbf{Q}$.

In Figure 1 we report only positive correlations, but a similar pattern emerges when they are negative (if the product of diagonal elements is greater than one, then the projected correlation is above the rescaled one, and it is below otherwise).
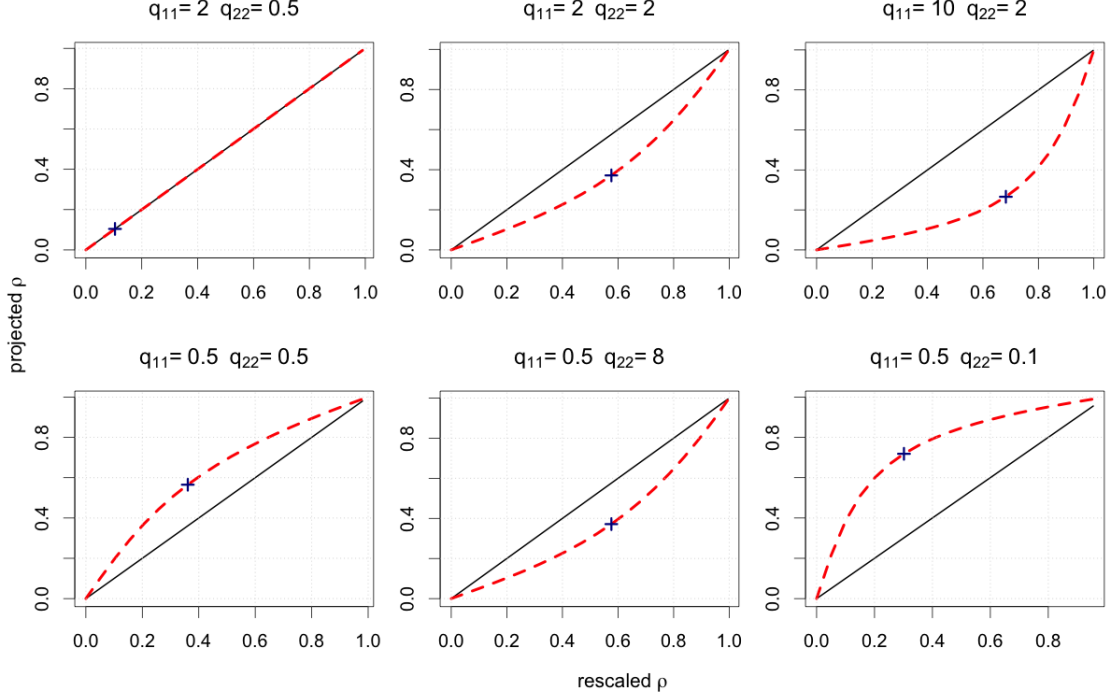

Figure 1: Projecting versus rescaling a pseudo-correlation matrix. Red: correlation computed using Stein's projection -see equation 7 - as a function of the rescaled correlation $q_{12}/\sqrt{q_{11}q_{22}}$. Black: 45 degree line. Cross: point of maximum difference between projecting and rescaling.

To further illustrate, consider a higher-dimensional setting. Assume that the n-dimensional pseudo-correlation matrix is given by

$$\mathbf{Q} = (1 - \kappa_2)\mathbf{R}(\kappa_1) + \kappa_2 vv',$$

where $\kappa_1, \kappa_2 \in (0, 1)$. The vector $v = (v_1, \ldots, v_n)'$ has a general entry $v_i = \sin(x_i)$, where $x_1 = \epsilon, x_n = 2\pi - \epsilon$, and $x_{i+1} = x_i + 2 \cdot \frac{\pi - \varepsilon}{n-1}$, for $\varepsilon > 0$. The first parameter κ_1 controls for the magnitude of the correlations and the second parameter κ_2 controls for the distance to an admissible correlation matrix $\mathbf{R}(\kappa_1)$, which is assumed to have a Toeplitz structure

14

- that is, its first row (or column) is given by the entries $1, \kappa_1, \kappa_1^2, \ldots, \kappa_1^{n-1}$. In Figure 2 we see that the difference -measured with the squared Frobenius norm 5 divided by n- between projecting and rescaling is monotonically increasing as $\mathbf{Q}$ moves away from the admissible correlation matrix $\mathbf{R}(\kappa_1)$, but it has an inverse-U shape with respect to the magnitude of the correlations. Moreover, the differences become more acute as the matrix dimension grows large.
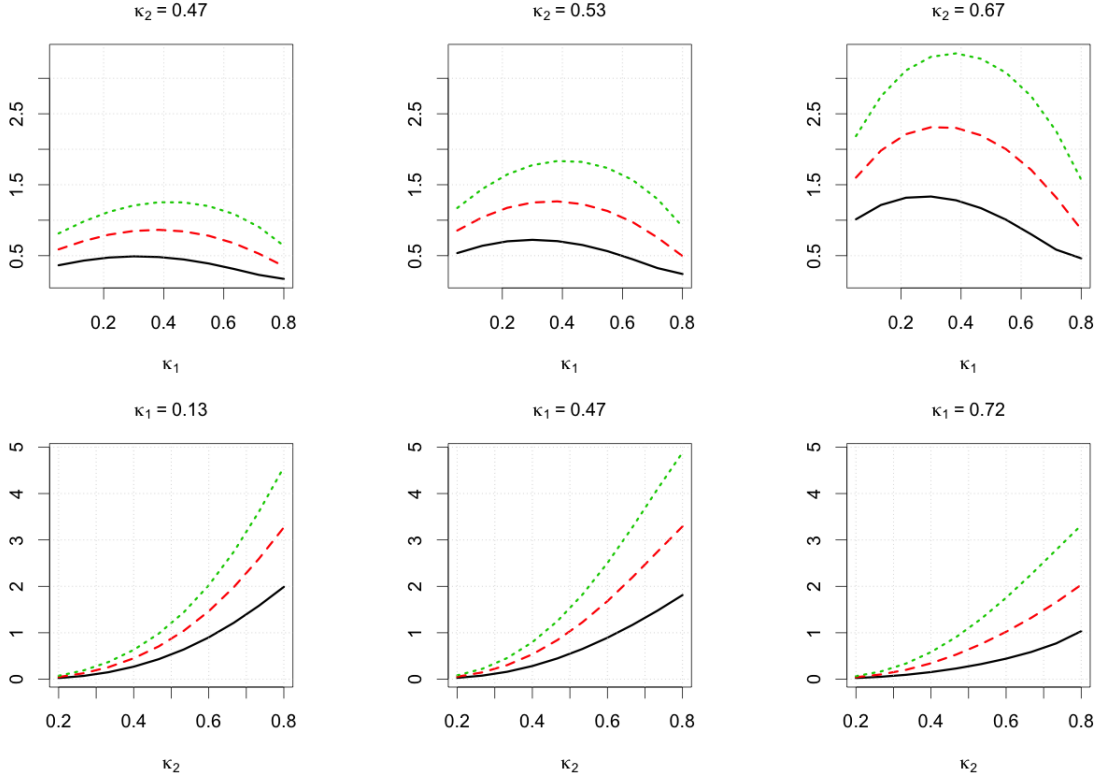

Figure 2: Projecting versus rescaling an n-dimensional pseudo-correlation matrix. Y-axis: $||\mathbf{R}_{resc} - \mathbf{R}_{proj}||_2^2/n = \operatorname{tr}((\mathbf{R}_{resc} - \mathbf{R}_{proj})^2)/n$, where $\mathbf{R}_{resc} = \operatorname{diag}(\mathbf{Q})^{-1/2}\mathbf{Q}\operatorname{diag}(\mathbf{Q})^{-1/2}$, and $\mathbf{R}_{proj} = P_{stein}(\mathbf{Q})$. Solid black line: $n = 50$. Red dashed line: $n = 75$. Green dotted line: $n = 100$. The pseudo-correlation matrix is $\mathbf{Q} = (1 - \kappa_2)\mathbf{R}(\kappa_1) + \kappa_2 v v^\top$ with $\kappa_1, \kappa_2 \in (0,1)$, v is an $n \times 1$ vector with i^{th} entry given by $v_i = \sin(x_i)$, where $x_{i+1} = x_i + 2(\pi - \varepsilon)/(n - 1), x_1 = \varepsilon, x_n = 2\pi - \varepsilon$. Finally, the notation $\mathbf{R}(\kappa_1)$ is used to denote a Toeplitz correlation matrix with parameter κ_1, i.e. with first row/column equal to $1, \kappa_1, \kappa_1^2, \ldots, \kappa_1^{n-1}$.

3   Simulation Study

3.1   Static estimators of multivariate Gaussian correlation matrix (known unit variance)

Let $r_t \sim \mathcal{N}(0, \mathbf{R})$, where $\mathbf{R}$ is an n-dimensional correlation matrix modelled with a single correlation parameter ρ, and consider a random sample (iid) of T observations from that distribution. Assume that population variances are known to be 1. In particular, we let $\mathbf{R} = (1 - \rho)\mathbf{I} + \rho \iota\iota'$, where ι is an n-dimensional vector of ones.

Let $\mathbf{S} = T^{-1} \sum_{t=1}^{T} r_t r_t'$ be the sample covariance matrix of the data, which is the MLE for the population *covariance*. It is crucial to note that *if the variances are unknown*, then the MLE for the *correlation* coefficient would be Pearson's sample correlation, i.e.

$$\hat{\mathbf{R}}^{(1)} = \operatorname{diag}(\mathbf{S})^{-1/2} \mathbf{S} \operatorname{diag}(\mathbf{S})^{-1/2} .$$

However, since the variances are constrained to be 1, the *constrained MLE* for the correlation matrix is not necessarily given by Pearson's sample correlation. In fact, Example 18.3 in Kendall and Stuart (1979) shows that the MLE in the bivariate Gaussian case is obtained by solving a cubic equation - which in large samples has only one real solution.

We consider two different candidates to estimate the correlation matrix $\mathbf{R}$. The first one consists in *rescaling* $\mathbf{S}$ to turn it into a correlation, whereas the second one *projects* $\mathbf{S}$ onto the correlation set using Stein's projection. From the example in section 2.2, we have that the zero-mean multivariate Gaussian density corresponds to Stein's loss, which is a Bregman divergence with $\phi(\cdot) = -\ln \det(\cdot)$. This means that computing the projection is indeed a very similar problem to finding the Gaussian MLE for the correlation matrix assuming unit variance. As discussed in the previous section, computing the projection involves solving a convex problem - as opposed to the MLE - and is much simpler to compute. We can write the proposed estimator as

$$\hat{\mathbf{R}}^{(2)} = P_\phi(\mathbf{S}) .$$

We evaluate the performance of our candidate estimators using a Montecarlo experiment. In each replication, we draw 500 observations from a multivariate Gaussian distribution with covariance parameter $\mathbf{R}$. We compute the loss of both candidate estimators with respect to the true correlation matrix. The following loss functions are used to assess the accuracy of each estimator:

$$\mathcal{L}_{Frob}(\hat{\mathbf{R}}, \mathbf{R}) = \sqrt{\frac{1}{n\psi} \text{tr}[(\hat{\mathbf{R}} - \mathbf{R})^2]}$$

$$\mathcal{L}_{MAE}(\hat{\mathbf{R}}, \mathbf{R}) = \frac{1}{n\psi} \sum_{i,j} |\hat{\mathbf{R}}_{ij} - \mathbf{R}_{ij}|$$

$$\mathcal{L}_{Stein}(\hat{\mathbf{R}}, \mathbf{R}) = \text{tr}(\hat{\mathbf{R}}\mathbf{R}^{-1}) - \ln\det(\hat{\mathbf{R}}\mathbf{R}^{-1}) - n$$

Note that we divide the squared Frobenius loss and the sum of absolute errors by $n\psi$ to establish a fair comparison as the dimension increases. We estimate $\mathbb{E}[\mathcal{L}]$ using the sample average of the losses obtained across Montecarlo replications.

We repeat the same exercise for different levels of the correlation parameter $\rho\psi$ as well as for the cross-sectional dimension n.

Figure 3 shows the excess loss in estimation accuracy which results from using sample correlation instead of projecting the sample covariance onto the correlation set with Stein's projection, where the truth is given by an equicorrelation matrix with parameter ρ. From this figure, we conclude that the gap between both methods increases with the dimensionality of the correlation matrix. In the equicorrelation model, we see that the gap is maximized at some intermediate value of the $\rho\psi$ parameter.
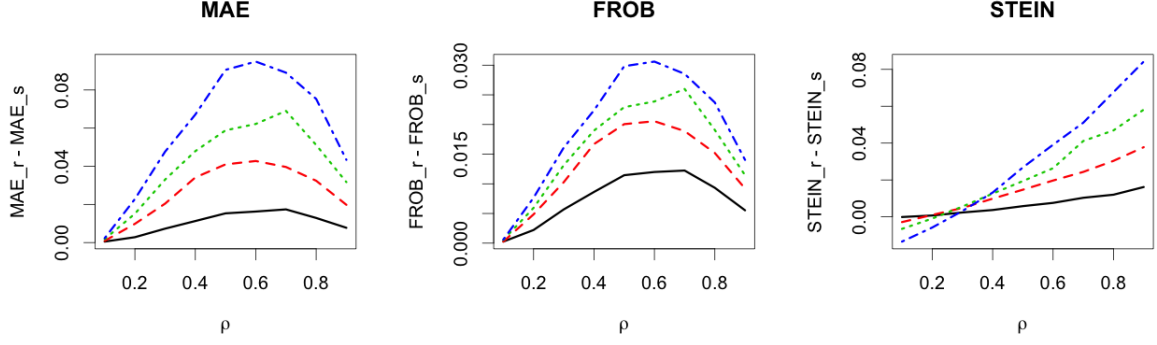
Figure 3: Static simulation study: Sample correlation versus Stein's projection of the sample covariance matrix onto the correlation set. From left to right panel, the y-axis shows $\mathbb{E}[\mathcal{L}_{MAE,resc}] - \mathbb{E}[\mathcal{L}_{MAE,proj}]$, $\mathbb{E}[\mathcal{L}_{Frob,resc}] - \mathbb{E}[\mathcal{L}_{Frob,proj}]$, and $\mathbb{E}[\mathcal{L}_{Stein,resc}] - \mathbb{E}[\mathcal{L}_{Stein,proj}]$, respectively. For each Montecarlo replication, we draw $T = 500$ observations from $\mathcal{N}(0, \mathbf{R})$, where $\mathbf{R} = (1 - \rho)\mathbf{I} + \rho\iota\iota'$. Black, red, green and blue lines correspond to $n = 10, 23, 36, 50$, respectively.

3.2 Estimation accuracy of competing dynamic correlation estimators

In this subsection we perform a Montecarlo simulation exercise with 500 replications to compare the estimation accuracy of Pro-DCC versus DCC.

Let $n = 2$. The return vector is generated as $r_t = \boldsymbol{\Sigma}_t^{1/2} z_t$, where $z_t \sim \mathcal{N}(0, \mathbf{I})$, and $\boldsymbol{\Sigma}_t = \mathbb{C}\text{ov}_{t-1}[r_t] = \mathbf{D}_t \mathbf{R}_t \mathbf{D}_t$, symmetric positive definite.

Clearly, $\epsilon_t \mid \mathcal{F}_{t-1} \sim \mathcal{N}\left(0, \begin{bmatrix} 1 & \rho_t \\ \rho_t & 1 \end{bmatrix}\right)$, where $\epsilon_t = \mathbf{D}_t^{-1} r_t$.

Consider the DGP's proposed in Engle (2002). The conditional variances follow the recursions

$$d_{1,t}^2 = .01 + .05 \, r_{1,t-1}^2 + .94 \, d_{1,t-1}^2$$

$$d_{2,t}^2 = .3 + .2 \, r_{2,t-1}^2 + .5 \, d_{2,t-1}^2$$

and we consider five different process for the conditional correlation: 1) Constant: $\rho_t = .9$, 2) Sine: $\rho_t = .5 + .4\cos(2\pi t/200)$, 3) Fast Sine: $\rho_t = .5 + .4\cos(2\pi t/20)$, 4) Step: $\rho_t = .9 - .5(t > 500)$ and 5) Ramp: $\rho_t = \mod(t/200)$.

18

Dynamic correlations for the SINE process are displayed in figure 4. We employ the following methodologies for estimation: 1) DCC: standard version of the Dynamic Conditional Correlation model, 2) Pro-DCC, the Projected DCC model presented in section 2.2, and 3) DCC-NoResc, a variant of the DCC model that omits the rescaling step, so it assumes that $\mathbf{R}_t = \mathbf{Q}_t$. For every replication, we draw $T\psi = 1500$ realizations from the DGP, and evaluate estimation accuracy using i) Stein's Loss as in 4, and ii) Root Mean Squared Error, which in practice we compute as $\sqrt{\frac{1}{T} \sum_{t=1}^{T} (\hat{\rho}_t - \rho_t)^2}$.

The distribution of excess losses across Montecarlo simulations is reported in figure 5. Finally, figure 6 reports the distribution of the estimated   parameter across Montecarlo replications for each methodology.

We draw the following conclusions from this simulated exercise: 1) Pro-DCC outperforms DCC for all DGP's proposed in Engle (2002). The only exception is the constant correlation process when we use Stein's loss. However, for this process, rescaling is not really relevant since neither DCC nor Pro-DCC are able to beat the DCC-NoResc benchmark, and 2) The estimated   parameter, which regulates the sensitivity of the conditional (pseudo) correlations to the innovations, is generally larger for Pro-DCC than for DCC. It has been documented that DCC usually underestimates this parameter. This would be less of a concern in our proposed algorithm. However, since these are actually different models and the relationship between both is not trivial, we should not claim that our model corrects this bias.
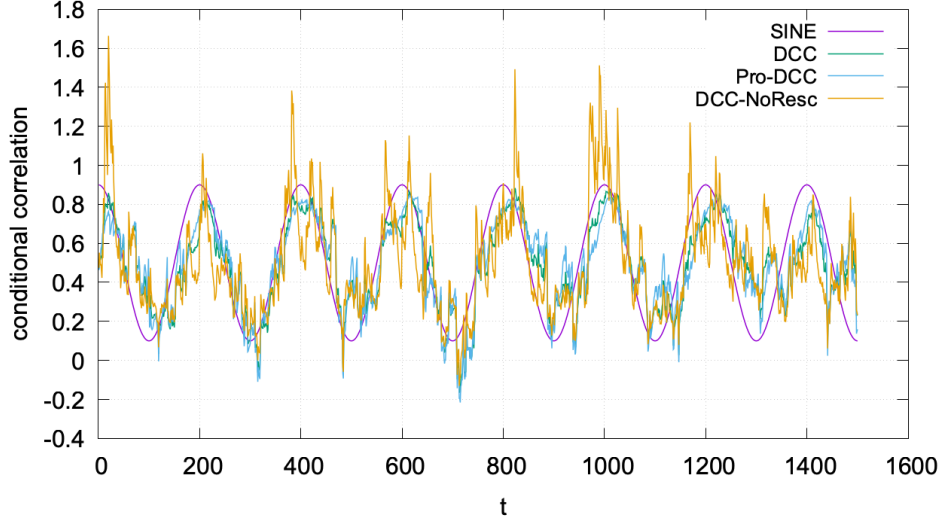
Figure 4: Simulation study: bivariate dynamic conditional correlations. Purple: true conditional correlation process (Sine). The remaining lines represent conditional correlation fitted values and one-step ahead forecasts for DCC (green), Pro-DCC (blue) and DCC-NoResc (dark orange).
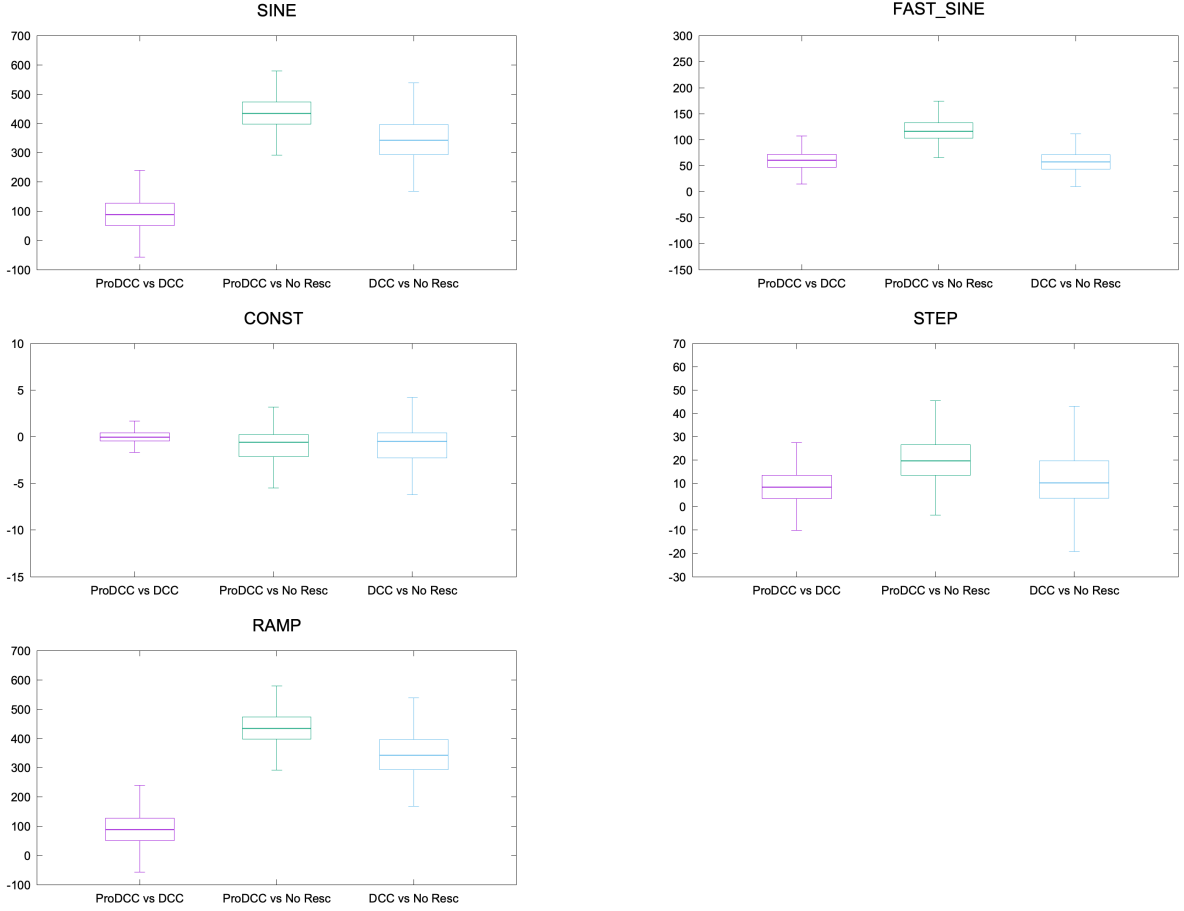

Figure 5: Simulation study: bivariate dynamic conditional correlations. Excess Stein's Loss (Cumulative) for different dynamic correlation processes. Purple: excess loss of DCC with respect to Pro-DCC. Green: excess loss of DCC-NoResc with respect to Pro-DCC. Blue: excess loss of DCC with respect to DCC-NoResc.
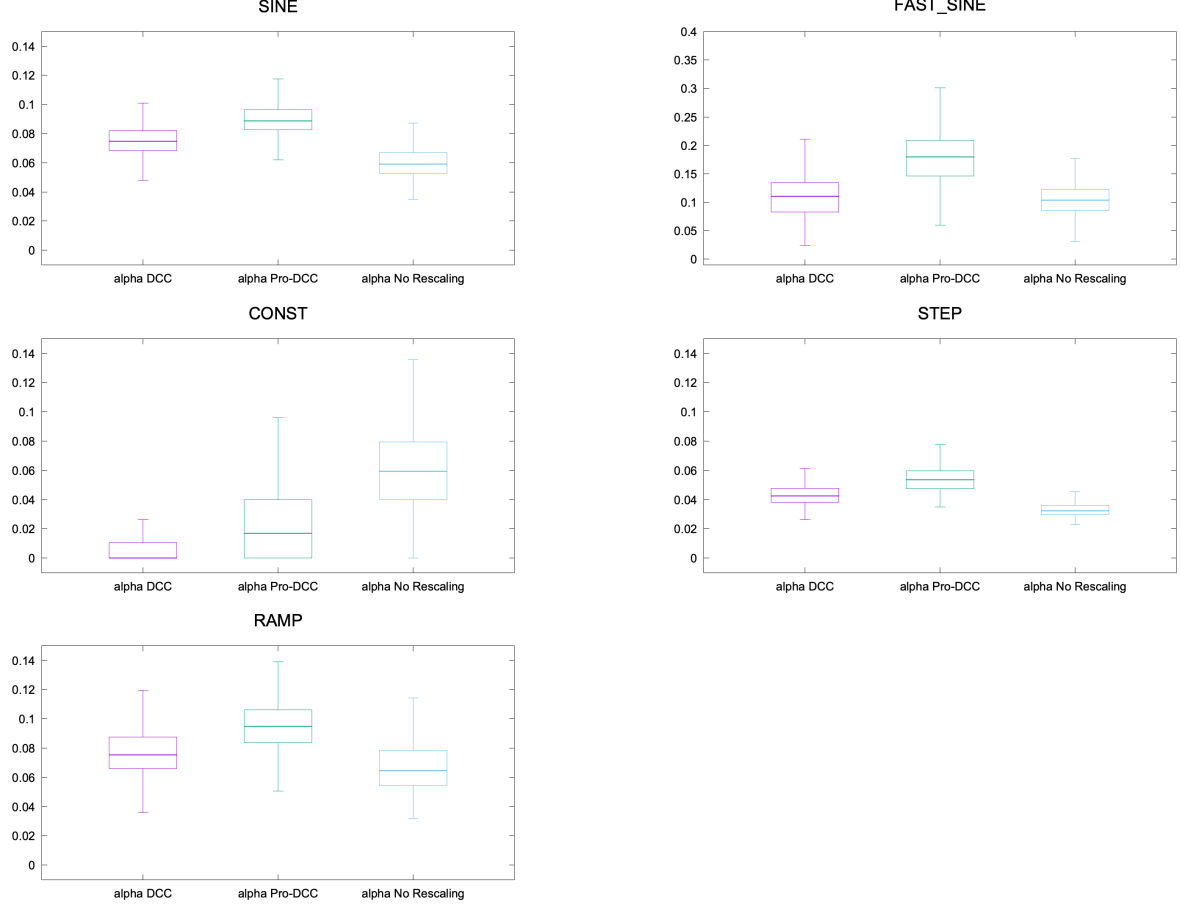
20

Figure 6: Simulation study: bivariate dynamic conditional correlations. Estimated parameter for different dynamic correlation processes. Purple: distribution of alpha estimates for DCC. Green: distribution of alpha estimates for Pro-DCC. Blue: distribution of alpha estimates for DCC-NoResc.

4 Empirical Application

4.1 Top-1000 pairs by correlation in S&P 100

We consider S&P 100 constituents' daily log returns ranging from 2007-01-01 to 2019-06-24 (source: Alpha Vantage). We carry out the exercise for 1000 pairs in the index, selected in decreasing order of their sample correlation coefficient across the whole period. We train Pro-DCC, DCC and DCC-NoResc (as in the previous section) until 2012-12-31 and use the remaining observations to test the performance of different covariance matrix estimators using the sample standard deviation of the Global Minimum Variance Portfolio (GMVP).

To compute the standard deviation of the Global Minimum Variance Portfolio (GMVP) for dimension n, we need the (dynamic) weights w_t that solve $\min_{w_t'\iota=1} w_t'\boldsymbol{\Sigma}_t w_t$, where ι is an n-dimensional vector of ones. It is easy to show that the minimizer is given by the following expression:

$$w_t^* = \frac{\boldsymbol{\Sigma}_t^{-1}\iota}{\iota'\boldsymbol{\Sigma}_t^{-1}\iota} \tag{9}$$

Clearly, the optimal weight vector in 9 is an infeasible quantity. Hence we estimate the standard deviation of the GMVP using weights that are based on a suitable estimator of $\boldsymbol{\Sigma}_t$:[3]

$$SD_{GMVP} = \sqrt{\tau^{-1}\sum_{t=T+}^{T+\tau}\left(\hat{w}_t^{*'}r_t r_t'\hat{w}_t^*\right)}$$

where $\hat{w}_t^* = \frac{\hat{\boldsymbol{\Sigma}}_{t|t-1}^{-1}\iota}{\iota'\hat{\boldsymbol{\Sigma}}_{t|t-1}^{-1}\iota}$.

From figures 7b and 7a we conclude with two main observations: 1) In terms of the standard deviation of the GMVP, it is not clear that any of the candidate estimators is able to outperform the equal-weighted portfolio (also known as "1/N"). In terms of the average GMVP standard deviation across pairs, the ranking is (from best to worst): 1/N, Pro-DCC, DCC, DCC-No Rescaling, Sample covariance. 2) Looking at the distribution of the estimated parameter across pairs suggests that $\psi_{Pro-DCC} > \psi_{DCC}$, same as in the simulation study.

4.2 Portfolio selection with all constituents in S&P 100

In this sub-section we test our algorithm in a higher-dimensional setting. We carry out a similar exercise as in the previous sub-section but for 86 (we have to discard 14 stocks for which the series is not long enough) instead of $n = 2$. We follow the common convention that 21 consecutive trading days constitute one "month". The training period ranges from 01/01/2007 to 31/12/2012. The out-of-sample period ranges from 01/01/2013 to 24/06/2019, resulting in a total of 77 months (1630 trading days). Following De Nard *et al.* (2018), we update portfolios every month and denote investment dates by $h =$

---

[3]Figures are reported in annualized terms. For daily log returns, this amounts to multiplying by $\sqrt{252}$.
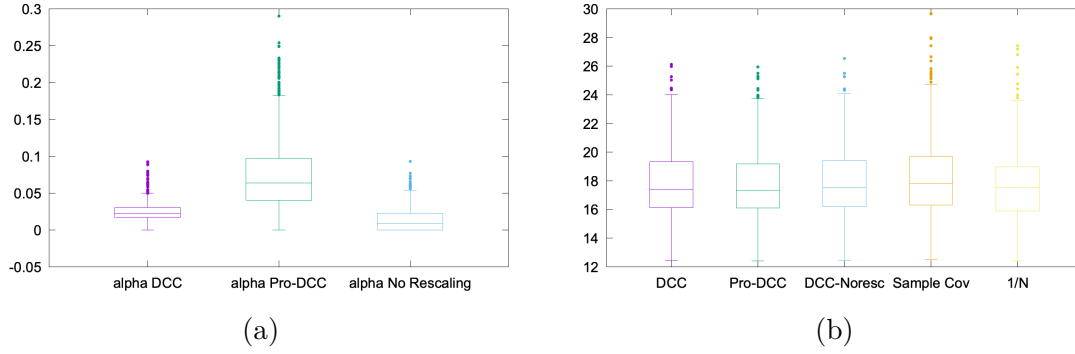
Figure 7: (a) Distribution (across 1000 different pairs from the S&P 100) of the estimated parameter for each methodology. (b) Distribution (across 1000 different pairs from the S&P 100) of the standard deviation of the Global Minimum Variance portfolio based on different candidate estimators of the conditional covariance matrix. The last two boxes correspond to static candidates for the purposes of benchmarking. Sample Cov stands for the unconditional sample covariance matrix, and $1/N$ refers to the equal-weighted portfolio, which in this case invests half of total wealth in each asset for each pair.

$1, \ldots, 77$. At any investment date h, a covariance matrix is filtered based on the estimated model.

	AV	SD	Sharpe	Turnover	Leverage	MaxWeight	MinWeight
ProDCC-NLS	10.54	11.36	0.93	0.07	0.51	0.54	-0.12
DCC-NLS	11.22	11.37	0.99	0.08	0.50	0.50	-0.14
ProDCC	10.53	11.44	0.92	0.08	0.50	0.57	-0.15
DCC	11.11	11.44	0.97	0.08	0.49	0.52	-0.16
DCCNoresc-NLS	9.93	11.45	0.87	0.07	0.50	0.51	-0.13
DCCNoresc	9.96	11.54	0.86	0.07	0.49	0.57	-0.15
NLS	10.66	11.69	0.91	0.00	0.48	0.22	-0.09
SampleCov	10.54	11.84	0.89	0.00	0.49	0.25	-0.10
ScalarVECH-NLS	15.48	12.09	1.28	0.07	0.47	0.48	-0.28
ScalarVECH	15.35	12.21	1.26	0.08	0.47	0.51	-0.30
1/N	13.04	12.60	1.03	0.00	0.00	0.01	0.01

Table 1: Empirical Application: Portfolio selection with all constituents in S&P 100. GMVP metrics for full out-of-sample period

We consider various covariance matrix estimators in this section. Since the dimensionality of the problem is considerably large, we also implement the analytical nonlinear shrinkage (NLS) methodology from Ledoit and Wolf (2018a) for each of them. The estimators are 1) DCC: the standard version of the DCC model, 2) Pro-DCC: the Projected DCC model based on Stein's projection as in section 2.2, 3) Scalar VECH: scalar-diagonal version of the VECH model, 4) DCC-Noresc: DCC model omitting the rescaling step, 5) Sample Cov: Sample covariance computed from training sample. This is a static

estimator, and 6) NLS: Analytical nonlinear shrinkage formula computed from training sample (also a static estimator). For benchmarking purposes, we also report metrics for the equal-weighted portfolio ("1/N").

We report the following three out-of-sample performance measures: 1) AV: the annualized out-of-sample average of portfolio returns, 2) SD: the annualized out-of-sample standard deviation of portfolio returns, and 3) Sharpe: the Sharpe ratio computed as AV / SD.[4]

The resulting portfolio metrics are presented in tables 1 and a breakdown by year can be found in table C1. These can be summarized as follows: a) Overall, Pro-DCC-NLS and Pro-DCC outperform all other candidate estimators in terms of both the standard deviation and the Sharpe ratio of the GMVP. b) The overall ranking in terms of the SD of the GMVP is 1) Pro-DCC-NLS, 2) DCC-NLS, 3) Pro-DCC, 4) DCC, 5) DCC-Noresc-NLS, 6) DCC-Noresc, 7) NLS, 8) Sample Cov, 9) Scalar VECH-NLS, 10) Scalar VECH, 11) 1/N. c) All dynamic estimators outperform static estimators (except for the scalar VECH). d) The (analytical) nonlinear shrinkage versions of all estimators considered in the exercise present superior performance than doing no shrinkage at all. Moreover, the relative ranking between models does not change when introducing the nonlinear shrinkage methodology, which suggests that it uniformly improves performance irrespective of the modeling choice for the time-varying covariance matrix.

5 Conclusions

In this paper we contribute to the multivariate GARCH-DCC literature with a novel DCC specification inspired by the literature on Bregman matrix projections and the nearest-correlation matrix problem. We demonstrate the benefits of using our proposed

---

[4]Additionally, the following portfolio metrics are computed:

- Turnover: $\frac{1}{N(\tau-1)} \sum_{t=T+2}^{T+\tau\psi} \sum_{i=1}^{N\psi} |w_{i,t\psi} - w_{i,t-1}|$.
- Proportion of leverage: $\frac{1}{N\tau\psi} \sum_{t=T+1}^{T+\tau\psi} \sum_{i=1}^{N} \mathbf{1}\{w_{i,t\psi} < 0\}$.
- Maximum weight: $\max_{i,t\psi} w_{i,t}$.
- Minimum weight: $\min_{i,t\psi} w_{i,t}$.

methodology (Pro-DCC) with respect to the standard GARCH-DCC model in a simulated exercise with 2 stocks. We also carry out a global minimum variance portfolio exercise based on historical data for 1000 pairs of stocks, and the differences in terms of out-of-sample performance turn out to be negligible. However, when we consider a setting where the cross-sectional dimension is considerably larger ($n = 86$), we find that the standard and nonlinear shrinkage versions of Pro-DCC outperform all other candidate estimators of the conditional covariance matrix in terms of the standard deviation and Sharpe ratio of the GMVP. As a secondary contribution, we illustrate on simulated iid data that when the DGP is a multivariate Gaussian process with unit variance, the estimator that results from projecting the sample covariance matrix onto the correlation set performs better than the sample correlation matrix in terms of Stein loss and -rather surprisingly-the excess portfolio variance of the GMVP.

A   Proofs

Proof of Lemma 1. See p. 206 in Bregman (1967). □

Proof of Lemma 2. Let $\mathbf{K} = \mathbf{Q}^{-1}$. Consider the $n = 2$ case:

$$\mathbf{R} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}, \quad \mathbf{Q} = \begin{bmatrix} q_{11} & q_{12} \\ q_{12} & q_{22} \end{bmatrix} \text{ and } \mathbf{K} = \begin{bmatrix} k_1 & k \\ k & k_2 \end{bmatrix}$$

Therefore,

$$\operatorname{tr}(\mathbf{RK}) = k_1 + 2k\rho + k_2, \text{ and } \det(\mathbf{RK}) = \det(\mathbf{K})\det(\mathbf{R}) = (k_1 k_2 - k^2)(1 - \rho^2)$$

and our minimization problem can be formulated as a univariate problem:

$$\min_{\rho} f(\rho) = \min_{\rho} k_1 + 2k\rho + k_2 - \ln(k_1 k_2 - k^2) - \ln(1 - \rho^2) - 2$$

Since the problem is convex and the domain of f is the open interval $(-1, 1)$, it suffices to take the first order condition and solve for ρ, which yields

$$\hat{\rho} = \begin{cases} \frac{1 - \sqrt{1+4k^2}}{2k} & k \neq 0 \\ 0 & k = 0 \end{cases}$$

where $k = -\frac{q_{12}}{\det(\mathbf{Q})}$. □

Proof of Lemma 3. Let $\mathbf{K} = \log \mathbf{Q}$. In the bivariate case, we have that

$$\operatorname{tr}(\mathbf{R}\log\mathbf{R}) - \operatorname{tr}(\mathbf{RK}) = \ln(1 - \rho^2) + \rho \ln\left(\frac{1+\rho}{1-\rho}\right) - 2k\rho + const = f(\rho)$$

which follows since the matrix logarithm of $\mathbf{R}$ is given by

$$
\log \mathbf{R} = \frac{1}{2} \left[\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \ln(1+\rho) & 0 \\ 0 & \ln(1-\rho) \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \right]
$$

Hence, the problem is equivalent to minimizing f_ψ with respect to ρ.[5] Since the problem is convex and the domain of f_ψ is the open interval $(-1, \psi)$, it suffices to take the first order condition and solve for ρ:

$$
\frac{-2\hat\rho\psi}{1-\hat\rho^2} + \ln\left(\frac{1+\hat\rho\psi}{1-\hat\rho\psi}\right)\left(1 + \hat\rho\psi\frac{1-\hat\rho\psi}{1+\hat\rho\psi}\frac{2}{(1-\hat\rho)^2}\right) - 2k\psi = 0
$$

$$
\implies \hat\rho\psi = \frac{e^{2k}-1}{e^{2k}+1} = \tanh(k)
$$

To find an analytical expression for k, let $\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}'$ be the eigendecomposition of $\mathbf{Q}$, where $\mathbf{V}$ is orthonormal. It is easy to verify that the eigenvalues of $\mathbf{Q}$ are given by

$$
\lambda_i = \frac{1}{2}\left[q_{11} + q_{22} + (-1)^{i-1}\sqrt{(q_{11}-q_{22})^2 + 4q_{12}^2} \right]
$$

where $i\psi = 1, 2$. Their corresponding eigenvectors are $v_i = [v_{i1}, v_{i2}]'$, where

$$
v_{i1} = -2q_{12}/\left(q_{11} - q_{22} + (-1)^i\sqrt{(q_{11}-q_{22})^2 + 4q_{12}^2} \right) v_{i2} := {}_i v_{i2}
$$

Imposing unit norm eigenvectors, we have that $v_{i2} = (1 + {}_i^2)^{-1/2}$. Hence, it is easy to see that the $(2,1)$ entry of the $\mathbf{K}$ matrix is given by

$$
k\psi = (\ln\lambda_1)v_{11}v_{12} + (\ln\lambda_2)v_{21}v_{22} = (\ln\lambda_1)\,{}_1 v_{12}^2 + (\ln\lambda_2)\,{}_2 v_{22}^2 = \frac{{}_1 \ln\lambda_1}{1 + {}_1^2} + \frac{{}_2 \ln\lambda_2}{1 + {}_2^2}
$$

$\square$

---

[5]Note that we can ignore the terms $\mathrm{tr}(\mathbf{R})$ and $\mathrm{tr}(\mathbf{Q})$ that appear in $d_\phi(\mathbf{R}, \mathbf{Q})$ as these do not depend on ρ.

Proof of Lemma 4. Let $\mathbf{R}^{(0)} = \mathbf{Q}$. Note that

$$P_{\phi,i}(\mathbf{R}^{(k-1)}) = \arg\min_{\mathbf{R}^{(k)} \in \mathbb{C}_i} d_\phi \left(\mathbf{R}^{(k)}, \mathbf{R}^{(k-1)} \right) ($$

The first order condition of the Lagrangian yields the following matrix update for $\mathbf{R}^{(k)}$:

$$\begin{cases} \nabla\phi(\mathbf{R}^{(k)}) = \nabla\phi(\mathbf{R}^{(k-1)}) + \quad e_i e_i' \\ \text{tr}(\mathbf{R}^{(k)} e_i e_i') = 1 \end{cases}$$

When $\phi(\cdot) = -\ln\det(\cdot)$, we have that $\nabla\phi(\mathbf{R}^{(k)}) = -[\mathbf{R}^{(k)}]^{-1}$, and the first equation of the system becomes

$$\mathbf{R}^{(k)} = \left([\mathbf{R}^{(k-1)}]^{-1} - \quad e_i e_i' \right)^{-1}$$

Using Sherman-Morrison's formula, we can re-write the first equation as

$$\mathbf{R}^{(k)} = \mathbf{R}^{(k-1)} + \frac{}{1 - \quad e_i' \mathbf{R}^{(k-1)} e_i} \mathbf{R}^{(k-1)} e_i e_i' \mathbf{R}^{(k-1)}$$

Note that $\text{tr}(\mathbf{R}^{(k-1)} e_i e_i') = e_i' \mathbf{R}^{(k-1)} e_i = \mathbf{R}_{ii}^{(k-1)}$. It follows that $\text{tr}(\mathbf{R}^{(k-1)} e_i e_i' \mathbf{R}^{(k-1)} e_i e_i') = (e_i' \mathbf{R}^{(k-1)} e_i)^2 = [\mathbf{R}_{ii}^{(k-1)}]^2$. Plugging the first equation in the second one and solving for we get

$$\text{tr}\left(\left[\mathbf{R}^{(k-1)} + \frac{}{1 - \quad \mathbf{R}_{ii}^{(k-1)}} \mathbf{R}^{(k-1)} e_i e_i' \mathbf{R}^{(k-1)} \right] e_i e_i' \right) = 1$$
$$= [\mathbf{R}_{ii}^{(k-1)}]^{-1} - 1$$

Replacing in the first equation of the system yields the desired result, since

$$\frac{}{1 - \quad \mathbf{R}_{ii}^{(k-1)}} = \frac{[\mathbf{R}_{ii}^{(k-1)}]^{-1} - 1}{1 - ([\mathbf{R}_{ii}^{(k-1)}]^{-1} - 1)\mathbf{R}_{ii}^{(k-1)}} = [\mathbf{R}_{ii}^{(k-1)}]^{-2} - [\mathbf{R}_{ii}^{(k-1)}]^{-1} = [\mathbf{R}_{ii}^{(k-1)}]^{-2} \left(1 - \mathbf{R}_{ii}^{(k-1)} \right) ($$

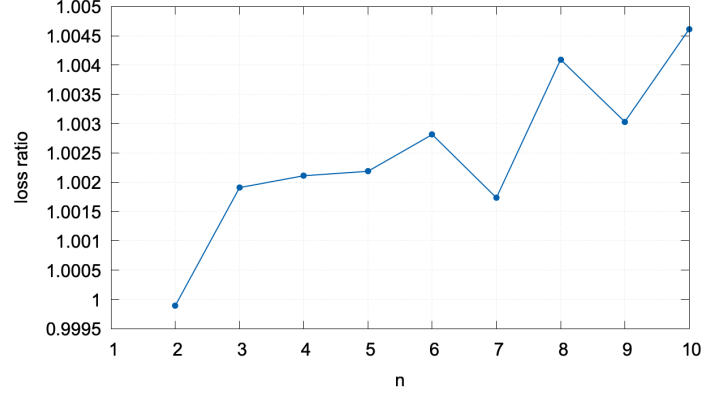$\square$

B Aielli's Critique

Figure B1: Covariance targeting vs Aielli correction in Pro-DCC. The ratio compares the loss of the sample second moment of the standardized residuals $\frac{1}{T}\sum_{t=1}^{T}\epsilon_t\epsilon_t'$ (covariance targeting) to the loss of Aielli's estimator, which is given by $\frac{1}{T}\sum_{t=1}^{T}\mathbf{Q}_t^*\epsilon_t\epsilon_t'\mathbf{Q}_t^*$. The loss of each estimator $\hat{\mathbf{C}}_s$ for simulation s is defined as $\frac{1}{S}\sum_{s=1}^{S}||\mathbf{C}-\hat{\mathbf{C}}_s||_F^2$, where S is the number of Montecarlo simulations, $\mathbf{C}$ is the true unconditional correlation matrix and $||\mathbf{A}||_F = \sqrt{\operatorname{tr}(\mathbf{A}'\mathbf{A})/n}$ for some $\mathbf{A} \in \mathbb{R}^{n\times n}$.

C Additional Tables

	AV	SD	Sharpe	Turnover	Leverage	MaxWeight	MinWeight
2013							
ScalarVECH-NLS	20.2645	9.6293	2.1045	0.0668	0.4678	0.4802	-0.1602
ScalarVECH	19.9699	9.7213	2.0542	0.0693	0.4650	0.5133	-0.1719
DCC-NLS	16.5642	10.0014	1.6562	0.0776	0.4992	0.5007	-0.1310
2014							
DCC-NLS	15.6294	9.3944	1.6637	0.0784	0.4882	0.3998	-0.1292
DCC	15.1084	9.4188	1.6041	0.0816	0.4863	0.3994	-0.1369
DCCNoresc	14.4356	9.4522	1.5272	0.0688	0.4983	0.3685	-0.1050
2015							
NLS	3.8565	14.3740	0.2683	0.0000	0.4767	0.2208	-0.0856
SampleCov	3.6372	14.4431	0.2518	0.0000	0.4884	0.2455	-0.1024
ScalarVECH-NLS	9.3066	14.9324	0.6232	0.0934	0.4747	0.4692	-0.2772
2016							
ProDCC-NLS	19.9057	10.1197	1.9670	0.0717	0.5310	0.3640	-0.1215
DCC-NLS	19.6058	10.1252	1.9363	0.0747	0.5134	0.3912	-0.1039
ProDCC	20.9271	10.1287	2.0661	0.0749	0.5233	0.3641	-0.1344
2017							
EqualWeighted	17.2063	6.4566	2.6649	0.0000	0.0000	0.0116	0.0116
DCCNoresc-NLS	14.3169	6.8502	2.0900	0.0549	0.5078	0.2802	-0.0933
DCC-NLS	12.2975	6.8950	1.7835	0.0590	0.4883	0.2863	-0.1176
2018							
DCC-NLS	-6.2292	14.3685	-0.4335	0.0726	0.4853	0.3062	-0.1044
DCC	-6.6338	14.4764	-0.4582	0.0750	0.4852	0.3161	-0.1144
ProDCC-NLS	-5.5740	14.5076	-0.3842	0.0681	0.4900	0.3159	-0.1018
2019							
DCCNoresc-NLS	17.9405	10.5199	1.7054	0.0647	0.4874	0.2800	-0.0863
ProDCC-NLS	13.0457	10.5955	1.2312	0.0753	0.5030	0.3159	-0.0796
DCCNoresc	18.1649	10.5963	1.7143	0.0681	0.4813	0.2929	-0.0921

Table C1: GMVP metrics by year (showing top 3 by SD)

References

Aielli, G. P. (2013). Dynamic conditional correlation: On properties and estimation. *Journal of Business and Economic Statistics*, **31**.

Banerjee, A., Merugu, S., Dhillon, I. S., and Ghosh, J. (2005). Clustering with bregman divergences. *Journal of Machine Learning Research*, **6**, 1705–1749.

Bauwens, L., Laurent, S., and Rombouts, J. (2006). Multivariate garch models: a survey. *Journal of Applied Econometrics*, **21**, 79–109.

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

Bregman, L. M. (1967). The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Computational Mathematics and Mathematical Physics*, **7**(3), 200 – 217.

Brownlees, C. and Engle, R. (2017). SRISK: A Conditional Capital Shortfall Measure of Systemic Risk. *Review of Financial Studies*, **30**(1), 48–79.

Cesa-Bianchi, N. and Lugosi, G. (2006). *Prediction, Learning, and Games*. Cambridge University Press, New York, NY, USA.

De Nard, G., Ledoit, O., and Wolf, M. (2018). Factor models for portfolio selection in large dimensions: The good, the better and the ugly. *Journal of Financial Econometrics*. Forthcoming.

Dey, D. K. and Srinivasan, C. (1985). Estimation of a covariance matrix under stein's loss. *The Annals of Statistics*, **13**(4), 1581–1591.

Dhillon, S. I. and Tropp, A. J. (2007). Matrix nearness problems with bregman divergences. *SIAM Journal on Matrix Analysis and Applications*, **29**.

Engle, R. (2002). Dynamic conditional correlation. *Journal of Business and Economic Statistics*, **20**(3), 339–350.

Engle, R. (2009). *Anticipating correlations: a new paradigm for risk management*. Princeton University Press.

Engle, R. and Colacito, R. (2006). Testing and valuing dynamic correlations for asset allocation. *Journal of Business and Economic Statistics*, **24**, 238–253.

Engle, R. and Sheppard, K. (2001). Theoretical and empirical properties of dynamic conditional correlation multivariate garch. Technical report, NBER. Working Paper No. W8554.

Engle, R. F., Ledoit, O., and Wolf, M. (2019). Large dynamic covariance matrices. *Journal of Business and Economic Statistics*, **37**(2), 363–375.

Higham, N. J. (2002). Computing the nearest correlation matrix—a problem from finance. *IMA Journal of Numerical Analysis*, **22**(3), 329–343.

Kendall, S. and Stuart, A. (1979). The advanced theory of statistics (4ed.), volume 2.

Kulis, B., Sustik, M. A., and Dhillon, I. S. (2009). Low-rank kernel learning with bregman matrix divergences. *J. Mach. Learn. Res.*, **10**, 341–376.

Ledoit, O. and Wolf, M. (2004). Honey, i shrunk the sample covariance matrix. *The Journal of Portfolio Management*, **30**(4), 110–119.

Ledoit, O. and Wolf, M. (2018a). Analytical nonlinear shrinkage of large-dimensional covariance matrices. *ECON-Working Papers 264, Department of Economics - University of Zurich*.

Ledoit, O. and Wolf, M. (2018b). Optimal estimation of a large-dimensional covariance matrix under stein's loss. *ECON-Working Papers No. 122, Department of Economics - University of Zurich*.

Pakel, C., Shephard, N., Sheppard, K. K., and Engle, R. F. (2018). Fitting vast dimensional time-varying covariance models. *NYU Working Paper No. FIN-08-009*.

Patton, A. (2018). Comparing possibly misspecified forecasts. *Journal of Business and Economic Statistics*. Forthcoming.

Pourahmadi, M. (2013). *High-Dimensional Covariance Estimation*. John Wiley & Sons, Inc.

Silvennoinen, A. and Teräsvirta, T. (2008). Multivariate garch models. SSE/EFI Working Paper Series in Economics and Finance 669, Stockholm School of Economics.

Stein, C. (1986). Lectures on the theory of estimation of many parameters. *Journal of Soviet Mathematics*, **34**(1), 1373.

Tse, L. D. Y. K. and Tsui, A. K. C. (2002). Evaluating the hedging performance of the constant-correlation garch model. *Applied Financial Economics*, **12**, 791–798.