International Journal of Forecasting 39 (2023) 1761–1776

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Projected Dynamic Conditional Correlations

Jordi Llorens-Terrazas [a,*], Christian Brownlees [a,b]

[a] Department of Economics and Business, Universitat Pompeu Fabra, 08005, Barcelona, Spain
[b] Barcelona School of Economics, Spain

ARTICLE INFO

Keywords:
Multivariate volatility
DCC
Bregman projection
Nearest-correlation matrix
Stein's loss

ABSTRACT

We propose a novel specification of the Dynamic Conditional Correlation (DCC) model based on an alternative normalization of the pseudo-correlation matrix called Projected DCC (Pro-DCC). Our modification consists in *projecting*, rather than *rescaling*, the pseudo-correlation matrix onto the set of correlation matrices in order to obtain a well defined conditional correlation matrix. A simulation study shows that projecting performs better than rescaling when the dimensionality of the correlation matrix is large. An empirical application to the constituents of the S&P 100 shows that the proposed methodology performs favorably to the standard DCC in an out-of-sample asset allocation exercise.

© 2022 The Author(s). Published by Elsevier B.V. on behalf of International Institute of Forecasters. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

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 & 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 Brownlees and Engle (2017), De Nard, Ledoit, and Wolf (2021), Engle, Ledoit, and Wolf (2019) and Van Os and Van Dijk (2021).

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 (Aielli, 2013; Engle, 2002; Tse & Tsui, 2002). 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 for symmetric

* Corresponding author.
E-mail address: jordi.llorens@upf.edu (J. Llorens-Terrazas).
URL: https://sites.google.com/barcelonagse.eu/jordi-llorens-terrazas/ (J. Llorens-Terrazas).

https://doi.org/10.1016/j.ijforecast.2022.06.003
0169-2070/© 2022 The Author(s). Published by Elsevier B.V. on behalf of International Institute of Forecasters. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

positive definite matrices (Banerjee, Merugu, Dhillon, & Ghosh, 2005; Bregman, 1967; Dhillon & Tropp, 2007; Patton, 2020), which is used in this work to measure nearness between two symmetric positive definite matrices. 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 (Dey & Srinivasan, 1985; Pourahmadi, 2013; Stein, 1986). In addition, this class of loss functions has been the focus of attention in the financial econometrics literature in the context of ranking multivariate volatility models by their forecasting performance (Laurent, Rombouts, & Violante, 2013; Patton, 2020). In particular, the former paper establishes under mild assumptions that consistent volatility forecast rankings using conditionally unbiased proxies are obtained if and only if the loss function is of the Bregman type.

We 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 Wishart log-density –or equivalently, the zero mean multivariate Gaussian log-likelihood with respect to the covariance parameter–, it is widely used (Ledoit & Wolf, 2018), and it guarantees to deliver a positive definite projection. Moreover, 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 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 main 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 and that gains are larger in higher dimensional systems. In the second exercise we compare the estimation accuracy of DCC and Pro-DCC under misspecification, that is when the DGP differs from both models. In particular, we consider a dynamic equicorrelation matrix model (Engle & Kelly, 2012) in which the dynamic correlation evolves according to the cosine function, in the spirit of one of the DGPs considered in the simulation exercise of Engle (2002). We find that Pro-DCC outperforms standard DCC and that the gains increase with the dimensionality of the system and degree of cross-sectional dependence.

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. (2021). We compare Pro-DCC to DCC, and we consider both the standard versions of these models as well as versions that rely on nonlinear shrinkage (Ledoit & Wolf, 2020) for covariance targeting. Results show that forecasts based on the standard and nonlinear shrinkage variant of the Pro-DCC achieve the best out-of-sample performance. For completeness, we also consider GMVPs with exposure constraints in order to understand if the advantage of Pro-DCC is due to shrinkage (Fan, Zhang, & Yu, 2012; Jagannathan & Ma, 2003). We find that adding 1-norm constraints substantially improves performance for both the DCC and Pro-DCC, hence suggesting that Pro-DCC performs favorably even after controlling for shrinkage.

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, besides the one we have already mentioned, include Bollerslev (1990) and 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 financial econometrics literature on large dimensional covariance estimation for asset allocation, which include the contributions of, among others, Hautsch, Kyj, and Malec (2015), Hautsch and Voigt (2019) and De Nard et al. (2021). Last, 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 Dhillon and Tropp (2007), Higham (2002) 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 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 $\boldsymbol{\Sigma}_t = \mathbb{C}\text{ov}_{t-1}(r_t)$. 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.

In the GARCH-DCC framework, the conditional volatility matrix $\mathbf{D}_t$ is typically modeled using some appropriate GARCH specification. Assuming, for instance, GJR-GARCH(1,1) dynamics we have that the ith diagonal element of $\mathbf{D}_t$, which we denote by d_{it}, is specified as

$$d_{it}^2 = \omega_i + a_i\, r_{it-1}^2 + \gamma_i\, \mathbb{I}_{t-1}^{-} r_{it-1}^2 + b_i\, d_{it-1}^2 \,,$$

where ω_i, a_i, γ_i and b_i are the GJR-GARCH(1,1) coefficients satisfying $\omega_i > 0$, $a_i > 0$, $b_i \geq 0$ and $a_i + \gamma_i/2 + b_i < 1$.

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 model the conditional correlation matrix is determined by the so-called pseudo-correlation matrix $\mathbf{Q}_t$ which evolves according to the equation

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

where α and β are scalar parameters that satisfy $\alpha > 0$, $\beta > 0, \alpha + \beta < 1$ 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 covariance targeting using 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 of the de-volatilized returns (see Engle, 2002, for details).

We remark that, albeit being intuitive, the estimation strategy put forward by Engle (2002) has some consistency issues first noted by Aielli (2013). These have motivated Aielli (2013) to introduce a "corrected" version of the model called Corrected DCC (CDCC). However, 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 Eq. (3) we propose *projecting* it onto the space of correlation matrices. In other words, we cast the problem of normalizing 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 real matrices.

Definition 1 (*Bregman Divergence*)**.** Given a strictly convex and differentiable function ϕ of Legendre type,[1] we define the Bregman matrix divergence as

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

for any two real matrices $\mathbf{M}_1$ and $\mathbf{M}_2$.

Bregman divergences 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 & Lugosi, 2006). Bregman divergences are always positive, like distances, and are zero only when their arguments coincide. Unlike distances, they are not necessarily symmetric and they do not necessarily satisfy the triangle inequality. Furthermore, they are always convex with respect to their first argument and satisfy a generalized Pythagorean property (Dhillon & Tropp, 2007; Kulis et al., 2009). Finally, Banerjee et al. (2005) establishes the existence of a bijection between Bregman divergences and regular exponential families. In this paper the discrepancy between a correlation matrix $\mathbf{R}$ and a pseudo-correlation matrix $\mathbf{Q}$ is measured by the divergence $d_\phi(\mathbf{R}, \mathbf{Q})$.

Depending on the choice of the function ϕ, we obtain a number of well-known loss functions for covariance matrices. If we set $\phi(\mathbf{M}) = -\ln\det(\mathbf{M})$ then we have Stein's loss,

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

where $\ln(\cdot)$ denotes the natural logarithm. This divergence can also be interpreted as the negative of the n-dimensional Wishart log-density (up to a constant) or, equivalently, the zero mean multivariate Gaussian log-likelihood with respect to the covariance parameter. If we set $\phi(\mathbf{M}) = \mathrm{tr}(\mathbf{M}\log\mathbf{M} - \mathbf{M})$ then we have the Von Neumann loss

$$d_\phi(\mathbf{M}_1, \mathbf{M}_2) = \mathrm{tr}(\mathbf{M}_1\log\mathbf{M}_1 - \mathbf{M}_1\log\mathbf{M}_2 - \mathbf{M}_1 + \mathbf{M}_2) \,, \tag{5}$$

where $\log(\cdot)$ denotes the matrix logarithm.[2] Finally, if we set $\phi(\mathbf{M}) = \|\mathbf{M}\|_F^2$ then we have the squared Frobenius loss.

[1] A function is of Legendre type if it is essentially smooth and essentially strictly convex (Bauschke & Borwein, 1997).

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

Let $\mathbb{S}^n_+$ ($\mathbb{S}^n_{++}$) be the set of n-dimensional symmetric positive semidefinite (positive definite) matrices. We use Bregman divergences to introduce the following general class of projections of symmetric positive definite matrices onto the set of correlation matrices –which is understood as the set of n-dimensional symmetric positive definite matrices with unit diagonal.

Lemma 1. *Let $\mathbf{Q} \in \mathbb{S}^n_{++}$ and let $\mathbb{C}^n$ denote the set of correlation matrices. Furthermore, assume that ϕ is a closed convex proper function of Legendre type and differentiable on $\mathrm{int}(\mathrm{dom}(\phi)) = \mathbb{S}^n_{++}$. Define the Bregman projection*

$$P_\phi(\mathbf{Q}) = \arg\min_{\mathbf{R}\in\mathbb{C}^n} d_\phi(\mathbf{R}, \mathbf{Q}). \tag{6}$$

Then we have that there exists a unique $P_\phi(\mathbf{Q}) \in \mathbb{S}^n_{++}$.

A few remarks are in order. First, we emphasize that the projection depends on the choice of the function ϕ. A natural choice that also turns out to be computationally convenient is to define a projection based on Stein's loss defined in (4) and the Von Neumann loss defined in (5). We call these projections, respectively, Stein's projection and Von Neumann projection for short.

Second, we point out that existence and uniqueness of the Bregman projection hold because (*i*) $\mathbb{S}^n_{++} \cap \mathbb{C}^n = \mathbb{C}^n \neq \emptyset$, (*ii*) ϕ is Legendre and (*iii*) $\mathrm{int}(\mathrm{dom}(\phi)) = \mathbb{S}^n_{++}$. The Stein and Von Neumann losses satisfy requirements (*ii*) and (*iii*), hence it follows from Bauschke and Borwein (1997) that the projection exists in $\mathbb{S}^n_{++}$ and is unique. However, the Frobenius loss does not satisfy property (*iii*) over the set of positive semidefinite matrices. We remark that a Frobenius projection can be uniquely defined, but it would not necessarily preserve positive definiteness.

Third, in the case of the Stein's loss we have that the projection is related to constrained maximum likelihood estimation (MLE) of the covariance of the zero-mean multivariate Gaussian with unit diagonal. In particular, the constrained MLE would be given by $\arg\min_{\mathbf{R}\in\mathbb{C}^n} d_\phi(\mathbf{Q}, \mathbf{R})$ where $\mathbf{Q}$ is the sample covariance matrix. Note that in general this differs from Pearson's sample correlation. In fact, in Example 18.3 of Kendall and Stuart (1979) it is shown that the constrained MLE for a bivariate Gaussian distribution is obtained by solving a cubic equation – which in large samples has only one real solution. In higher dimensions, finding such MLE is computationally burdensome. On the contrary, projecting the sample covariance under Stein's loss involves solving a convex problem (as opposed to the MLE) and the solution is easily obtained with a much lower computational burden.

Finally, we introduce the Pro-DCC(1,1), that is

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

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 ϕ.

A number of comments are in order. First, we remark that the projection yields the closest correlation matrix with respect to the loss induced by ϕ, which does not automatically imply any optimality properties for the purposes of forecasting or minimizing other relevant classes of losses such as the GMVP. Nevertheless, Sections 3 and 4 –which provide both simulation and empirical evidence that the projection performs favorably to standard rescaling– shed some light on this point. Second, as discussed at the end of Section 2.1, the original formulation of the DCC model was criticized by Aielli (2013) and a corrected version of the model was introduced. We remark that we do not use the Aielli correction in the Pro-DCC methodology.

For large dimensional models, we propose estimating the model by composite likelihood as in Pakel et al. (2018). However, we note that for the Stein and Von Neumann cases, 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 for this loss function the projection takes into account the full correlation structure and not just the correlation between assets i and j. In practice, the resulting composite likelihood estimates of the dynamic parameters do not vary substantially from their full likelihood counterparts, so this is a minor concern.

Last, we remark that, as it is widely known, 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 (2020).

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 computationally 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*

$$\hat{\rho} = \begin{cases} \frac{1-\sqrt{1+4k_s^2}}{2k_s} & k_s \neq 0 \\ 0 & k_s = 0 \end{cases}, \tag{7}$$

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

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. Then, the unique minimizer of this problem is given by*

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

where k_v denotes the off-diagonal entry of $\log \mathbf{Q}$.[3]

It is important to emphasize that the optimal projection in these two cases looks different from rescaling, thus implying that rescaling, at least as far as the Stein and Von Neumann divergences are concerned, is not optimal.

In the n-dimensional case we can derive an algorithm. 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 $\mathbf{R}$):

$$\min_{\mathbf{R}\in\mathbb{S}^n_{++}} d_\phi(\mathbf{R},\mathbf{Q}) \text{ subject to } \mathbf{R}_{ii}=1 \text{ for all } i=1,\dots,n. \quad (8)$$

$\mathbf{R}_{ii}$ stands for the ith diagonal element of the matrix $\mathbf{R}$.

To solve this problem, we use Bregman's *cyclic projections* method. Let $\mathbb{C}_i$ be the set of n-dimensional symmetric positive definite matrices whose ith diagonal element is unity. Clearly, the set of correlation matrices $\mathbb{C}^n = \bigcap_{i=1}^n \mathbb{C}_i$. Bregman's cyclic projections method is an iterative algorithm in which one projects successively onto each basic constraint set $\mathbb{C}_i$ until the sequence of iterates converges to the Bregman projection onto the intersection $\mathbb{C}^n$. Theorem 1 establishes that this algorithm is asymptotically valid. We refer to Dhillon and Tropp (2007) and the references therein for a proof.

Theorem 1. *Suppose*

1. *ϕ is a closed convex proper function of Legendre type such that* $\mathrm{int}(\mathrm{dom}(\phi)) = \mathbb{S}^n_{++}$.
2. *$\{\mathbb{C}_i\}_{i=1}^n$ are the sets of n-dimensional symmetric positive definite matrices with unit ith diagonal entry.*
3. *the control mapping $m : \mathbb{N} \to \{1,\dots,n\}$ is a sequence that takes each output value an infinite number of times.*

For $k = 1,2,\dots$, define $P_{\phi,m(k)}(\mathbf{R}^{(k-1)})$ as the Bregman projection of $\mathbf{R}^{(k-1)}$ onto $\mathbb{C}_{m(k)}$. Choose $\mathbf{R}^{(0)} = \mathbf{Q} \in \mathbb{S}^n_{++}$, 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 spectral norm to $P_\phi(\mathbf{Q})$.

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 $\phi(\mathbf{M}) = -\ln\det(\mathbf{M})$.

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

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

where e_i denotes the ith canonical basis vector.

We concisely describe this procedure in Algorithm 1. We point out that the algorithm has a complexity of $O(n^2)$ per iteration. We remark that Algorithm 1 is an iterative procedure that relies on a tolerance parameter to determine convergence. In practice, in the empirical application and simulations we have found that a tolerance value of 10^{-6} is sufficiently accurate.

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

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

and update the projection according to the formula

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

where e_i is defined as the i^{th} canonical basis vector.

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

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

2.3. Discussion

2.3.1. Projecting vs rescaling a pseudo-correlation matrix

In this section we show that the difference between rescaling and projecting can be relevant enough in many cases.

We first consider the difference between rescaling and projecting in a 2-dimensional setting. We denote by q_{11}, q_{22} the diagonal elements of the pseudo-correlation matrix $\mathbf{Q}$ and denote by q_{12} its off-diagonal element. Simple algebra shows that when $q_{11}q_{22} = 1$, then the expression in Eq. (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 Fig. 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 is larger than 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 function of the product of the diagonal elements of $\mathbf{Q}$. Note that in Fig. 1 we report only positive correlations. An analogous 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).

[3] Straightforward computations show that k_v has the following analytical expression $k_v = \frac{\psi_1 \ln\lambda_1}{1+\psi_1^2} + \frac{\psi_2 \ln\lambda_2}{1+\psi_2^2}$ where

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

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

and $i = 1, 2$.

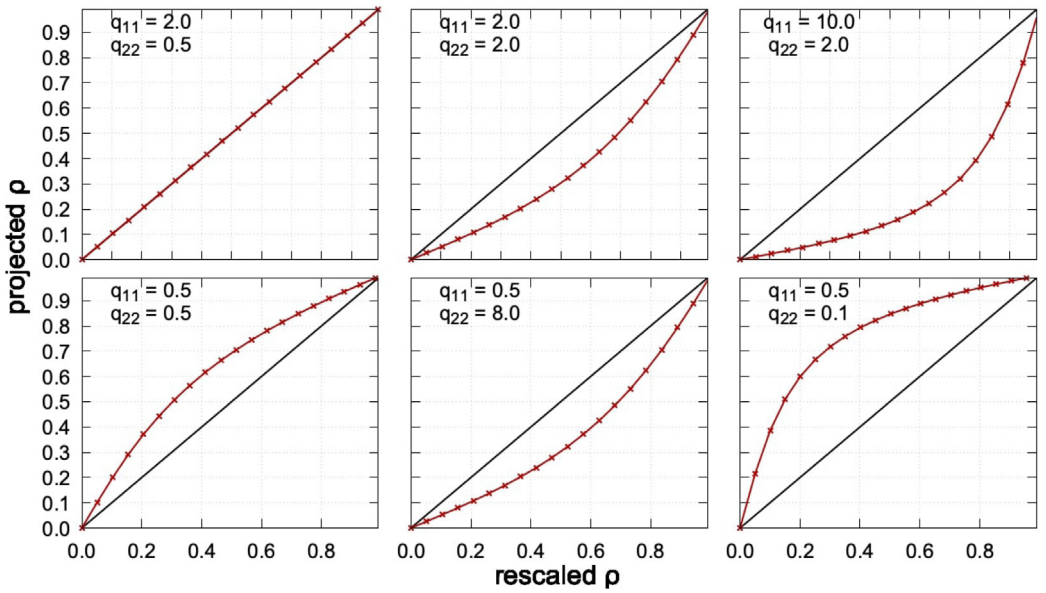

Fig. 1. Projecting versus rescaling a pseudo-correlation matrix. Red: correlation computed using Stein's projection – see Eq. (7) – as a function of the rescaled correlation $q_{12}/\sqrt{q_{11}q_{22}}$. Black: 45 degree line. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Next, we illustrate the difference between rescaling and projecting in a large dimensional setting. Assume that the n-dimensional pseudo-correlation matrix is given by

$$\mathbf{Q} = \mathrm{Diag}(1+\xi+v)^{1/2}\mathbf{R}(\kappa_1)\,\mathrm{Diag}(1+\xi+v)^{1/2}\ ,$$

where $\kappa_1 \in (0,1)$, $\xi > 0$, $\mathbf{R}(\kappa_1)$ is an $n \times n$ matrix with Toeplitz structure, v is an n-dimensional vector, and $\mathrm{Diag}(1+\xi+v)$ denotes the $n \times n$ diagonal matrix with diagonal given by $1+\xi+v$. In particular, we have that the (i,j)-th element of $\mathbf{R}(\kappa_1)$ is $\kappa_1^{|i-j|}$. The $v = (v_1, \ldots, v_n)'$ vector has its ith element equal to $\sin(x_i)$, where $x_1 = \varepsilon$, $x_n = 2\pi - \varepsilon$, and $x_{i+1} = x_i + 2 \cdot \frac{\pi-\varepsilon}{n-1}$, for $\varepsilon > 0$. In Fig. 2 we see that the difference – measured with the squared Frobenius norm divided by n – between projecting and rescaling has an inverse-U shape with respect to the magnitude of the correlations. Importantly, the figure also shows that differences become more pronounced with the matrix dimension.

2.3.2. Projecting as shrinking

This sub-section explores the relationship between projecting and shrinkage. The main message is that there are some cases where the projection may be interpreted as a methodology that shrinks the eigenvalues of a given pseudo-correlation matrix compared to the rescaled version. However, this does not hold in general for all possible pseudo-correlation matrices. Hence, interpreting the projection as shrinkage is not entirely straightforward. To see this, consider first the case where $n = 2$ with the Stein's loss. From Laurent et al. (2013), we have that when $\phi(\mathbf{M}) = -\ln\det(\mathbf{M})$, then

$$d_\phi(\mathbf{R},\mathbf{Q}) = \sum_{i,j=1}^{n} \frac{\lambda_i}{\mu_j}(v_i'u_j)^2 - \sum_{i=1}^{n} \ln\frac{\lambda_i}{\mu_i} - n\ , \tag{9}$$

where λ_i, v_i and μ_j, u_j are the i and jth eigenvalues/vectors of $\mathbf{R}$ and $\mathbf{Q}$, respectively. If $u_i = v_i$ for all $i = 1, 2$, and since $\lambda_1 + \lambda_2 = 2$ with $\lambda_1 > \lambda_2 > 0$, then we can re-parameterize the matrix nearness problem as a function of λ_1:

$$\min_{\mathbf{R}\in\mathbb{C}^n} d_\phi(\mathbf{R},\mathbf{Q})$$
$$\iff \min_{\lambda_1\in(0,2)} \left(\frac{1}{\mu_1} - \frac{1}{\mu_2}\right)\lambda_1 - \ln\lambda_1 - \ln(2-\lambda_1),$$

with unique solution given by

$$\lambda_1 = 1 - \frac{1}{g} + \sqrt{1+\frac{1}{g^2}}\ , \qquad g := \frac{1}{\mu_2} - \frac{1}{\mu_1}\ ,$$

which is to be compared against $2\mu_1/\mathrm{tr}(\mathbf{Q})$, i.e. the eigenvalue obtained via rescaling. After some algebra, it can be shown that $\lambda_1 < 2\mu_1/\mathrm{tr}(\mathbf{Q})$ whenever $\mathrm{tr}(\mathbf{Q}) > 2$, and viceversa. Therefore, even in this simplified scenario we can see that the projection methodology may have an effect similar to shrinking the eigenvalues of the rescaled $\mathbf{Q}$, but it may as well have the opposite effect.

Clearly, the analysis becomes more complex when $n > 2$ since normalizing $\mathbf{Q}$ into a correlation matrix involves not just a change in the eigenspectrum but also rotating the eigenvectors. Under the same setup of Section 2.3.1, Fig. 3 illustrates the difference between projecting $\mathbf{Q}$ versus shrinking and rescaling $\mathbf{Q}$ via the analytical nonlinear shrinkage formula of Ledoit and Wolf (2020) for different values of the concentration ratio n/T. The figure shows that the gap widens with the degree of dependence and the dimension –as in the previous sub-section– as well as with the concentration ratio n/T, hence suggesting that projecting and shrinking are essentially different operations.

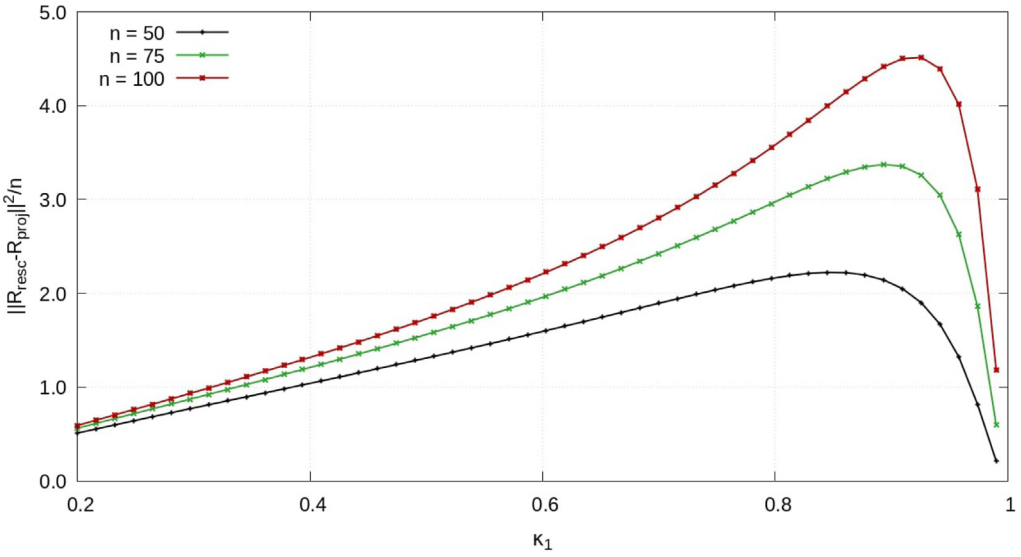

Fig. 2. Projecting versus rescaling an n-dimensional pseudo-correlation matrix. The y-axis represents $\|\mathbf{R}_{resc} - \mathbf{R}_{proj}\|_2^2/n = \text{tr}((\mathbf{R}_{resc} - \mathbf{R}_{proj})^2)/n$, where $\mathbf{R}_{resc} = \text{diag}(\mathbf{Q})^{-1/2}\mathbf{Q}\,\text{diag}(\mathbf{Q})^{-1/2}$, and $\mathbf{R}_{proj} = P_{STEIN}(\mathbf{Q})$. The pseudo-correlation matrix is $\mathbf{Q} = \text{Diag}(1+\xi+v)^{1/2}\mathbf{R}(\kappa_1)\,\text{Diag}(1+\xi+v)^{1/2}$, with $\kappa_1 \in (0, 1)$, $\xi = 0.05$, v is an $n \times 1$ vector with ith entry given by $v_i = \sin(x_i)$, where $x_{i+1} = x_i + 2(\pi - \varepsilon)/(n-1), x_1 = \varepsilon, x_n = 2\pi - \varepsilon$. 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 $1, \kappa_1, \kappa_1^2, \ldots, \kappa_1^{n-1}$.

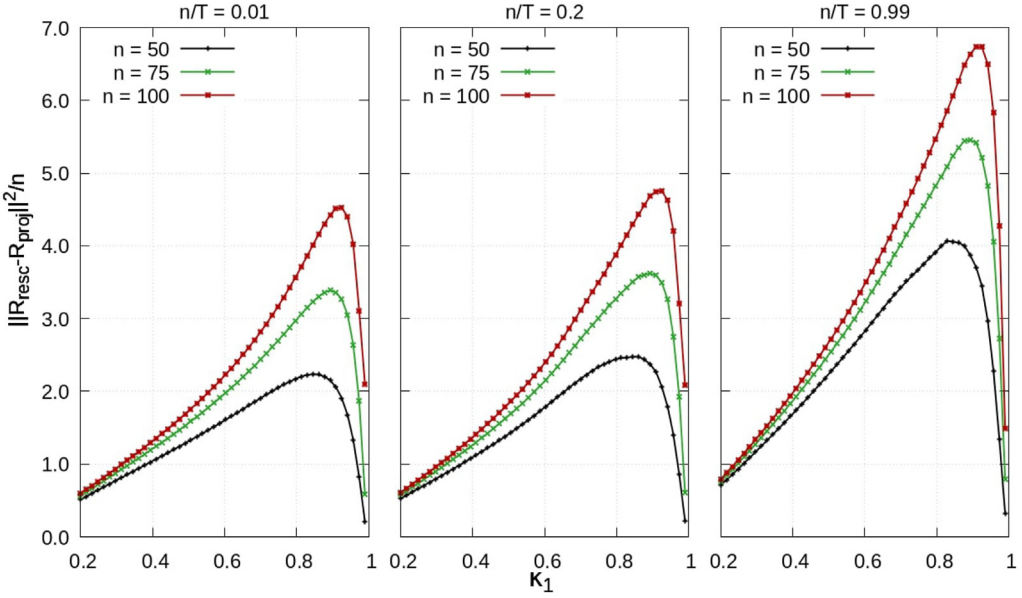

Fig. 3. Projecting versus shrinking and rescaling an n-dimensional pseudo-correlation matrix. The y-axis represents $\|\mathbf{R}_{s-r} - \mathbf{R}_{proj}\|_2^2/n = \text{tr}((\mathbf{R}_{s-r} - \mathbf{R}_{proj})^2)/n$, where $\mathbf{R}_{s-r} = \text{diag}(\tilde{\mathbf{Q}})^{-1/2}\tilde{\mathbf{Q}}\,\text{diag}(\tilde{\mathbf{Q}})^{-1/2}$, $\mathbf{R}_{proj} = P_{STEIN}(\mathbf{Q})$, and $\tilde{\mathbf{Q}}$ is obtained after applying the nonlinear shrinkage estimator of Ledoit and Wolf (2020) to $\mathbf{Q}$ with bandwidth $h = T^{-1/3}$, where T is chosen to match the desired concentration ratio n/T. The pseudo-correlation matrix is $\mathbf{Q} = \text{Diag}(1+\xi+v)^{1/2}\mathbf{R}(\kappa_1)\,\text{Diag}(1+\xi+v)^{1/2}$, with $\kappa_1 \in (0, 1)$, $\xi = 0.05$, v is an $n \times 1$ vector with ith entry given by $v_i = \sin(x_i)$, where $x_{i+1} = x_i + 2(\pi - \varepsilon)/(n-1), x_1 = \varepsilon, x_n = 2\pi - \varepsilon$. 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 $1, \kappa_1, \kappa_1^2, \ldots, \kappa_1^{n-1}$.

2.3.3. Measurement error

In the context of the DCC model, the observed pseudo-correlation matrix $\mathbf{Q}$ is estimated. It is natural to consider the impact of measurement error on the projection methodology as opposed to rescaling. The analysis of this question is inspired by Laurent et al. (2013). Let $\tilde{\mathbf{Q}}$ denote the estimated value of $\mathbf{Q}$, and denote $\tilde{\mathbf{R}}$ and $\tilde{\mathbf{P}}$ the rescaled/projected $\tilde{\mathbf{Q}}$. The structure of Bregman divergences allows us to write the average discrepancy between rescaling and projecting as

$$\mathbb{E}[d_\phi(\tilde{\mathbf{R}}, \mathbf{Q}) - d_\phi(\tilde{\mathbf{P}}, \mathbf{Q})]$$

$$= \mathbb{E}[\phi(\tilde{\mathbf{R}}) - \phi(\tilde{\mathbf{P}})] - \mathbb{E}\,\mathrm{tr}(\nabla\phi(\mathbf{Q})(\tilde{\mathbf{P}} - \tilde{\mathbf{R}}))\ .$$

It follows that the measurement error does not impact the proximity rankings of $\tilde{\mathbf{R}}$ and $\tilde{\mathbf{P}}$ provided that this discrepancy is positive. Although it seems challenging to derive general conclusions for general Legendre functions ϕ, if we consider $\phi(\mathbf{M}) = -\ln\det(\mathbf{M})$, then the proximity rankings are not affected by measurement error whenever the trace of $\mathbf{Q}$ is large enough and the eigenvalues of $\tilde{\mathbf{R}}$ are more dispersed than those of $\tilde{\mathbf{P}}$.

3. Simulation study

In this section we carry out different Monte Carlo exercises to analyse the performance of Stein's projection and the Pro-DCC for correlation modeling.

3.1. Static correlations

In the Monte Carlo exercise of this section we study the performance of our proposed projection in a static environment. We carry out a Monte Carlo exercise designed as follows. We simulate $T = 500$ i.i.d. random draws from an equicorrelation process given by

$$r \sim \mathcal{N}(0, \mathbf{R}),\ \text{where } \mathbf{R} = (1-\rho)\mathbf{I} + \rho\iota\iota',$$

where ι is an n-dimensional vector of ones, $\rho \in (0, 1)$ and $\mathbf{I}$ is the $n \times n$ identity matrix. The focus of the exercise lies in the estimation of the population correlation matrix $\mathbf{R}$. We consider two competing estimators. The first candidate estimator is Pearson's sample correlation matrix, that is

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

where $\mathbf{S}$ denote the sample covariance matrix $\mathbf{S} = \frac{1}{T}\sum_{t=1}^{T} r_t r_t'$. The second estimator is the projected sample covariance matrix of the data, defined as $\hat{\mathbf{R}}^{(2)} = P_{STEIN}(\mathbf{S})$.

In each replication, we compute the loss of both estimators with respect to the true correlation matrix. The losses under consideration are the Frobenius and the MAE. These are defined as follows:

$$\mathcal{L}_{Frob}(\hat{\mathbf{R}}, \mathbf{R}) = \sqrt{\frac{1}{n}\,\mathrm{tr}[(\hat{\mathbf{R}} - \mathbf{R})^2]} \tag{10}$$

$$\mathcal{L}_{MAE}(\hat{\mathbf{R}}, \mathbf{R}) = \frac{1}{n}\sum_{i,j} |\hat{\mathbf{R}}_{ij} - \mathbf{R}_{ij}| \tag{11}$$

Note that we divide the squared Frobenius loss and the sum of absolute errors by n to establish a fair comparison as the dimension increases. We estimate $\mathbb{E}[\mathcal{L}]$ using the sample average of the losses obtained across 1000 Monte Carlo replications, and repeat the same exercise for different levels of the correlation parameter ρ as well as the cross-sectional dimension n. In particular, we consider all combinations of ρ in $\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$ and n in $\{10, 23, 36, 50\}$. We remark that the concentration ratio is rather low (it only reaches a maximum of 0.1) since this exercise is based on the sample covariance, which is optimal under fixed-n large-T asymptotics but performs poorly in finite samples when the concentration ratio n/T is large. In unreported Monte Carlo exercises we have also implemented the projection of the non-linear shrinkage estimator (Ledoit & Wolf, 2020) of the covariance matrix onto the correlation set and compared them to the sample correlation. The results were also favorable to our proposed methodology.

Fig. 4 reports the excess loss of the sample correlation matrix with respect to our proposed estimator. From this figure, it is clear that the gap between both methods is positive and increases with the dimensionality of the correlation matrix. We also observe that the gap is maximized at some intermediate value of the ρ parameter between zero and one.

To conclude this sub-section, note that the results are congruent with our findings from Section 2.3.1. The main message is that care must be taken in choosing an appropriate way of normalizing a positive definite matrix to a correlation. Rescaling is one (obvious) way of doing so, but it may not be necessarily optimal.

3.2. Dynamic correlations

In the Monte Carlo exercise of this section we study the performance of our proposed projection in a dynamic environment. We carry out a Monte Carlo exercise designed as follows. We consider a dynamic equicorrelation model where the scalar correlation parameter is governed by a cosine wave, in the spirit of Engle and Kelly (2012) and one of the simulated DGPs considered in Engle (2002). More precisely, we consider the DGP given by the equations

$$\epsilon_t = \mathbf{R}_t^{1/2} z_t,\ \text{where } z_t \sim \mathcal{N}(0, \mathbf{I}),\ \text{for } t = 1, \ldots, T\ ,$$
$$\mathbf{R}_t = (1-\rho_t)\mathbf{I} + \rho_t\iota\iota'\ ,$$
$$\rho_t = \rho + 0.1\cos(2\pi t/200)\ .$$

In each replication, we draw $T = 1000$ observations from the DGP described above for different levels of the ambient dimension $n = 50, 100, 200$ and $\rho = 0.4, 0.5, 0.6, 0.7, 0.8$.

To predict the sequence of matrices $\mathbf{R}_t$, we consider the DCC and Pro-DCC models. As customary, the parameters of the two specifications are estimated by maximizing the Quasi Log-Likelihood of the data with respect to the correlation matrix parameter, where the intercept matrix $\mathbf{C}$ is estimated by covariance targeting.

A few remarks are in order. In this exercise, we abstract from estimation of dynamic volatilities since the DCC and Pro-DCC methodologies employ the same univariate GARCH strategy to estimate the conditional volatilities. The exercise thus focuses on the main source of the difference between both strategies, that is the correlation modeling step. Similar to the previous sub-section, we employ the sample covariance for the covariance targeting step, but in unreported exercises we have also implemented the non-linear shrinkage estimator and the results were also favourable to our proposed methodology.

In order to evaluate the precision of the correlation forecasts with respect to the true correlation process $\mathbf{R}_t$ we use again the Frobenius and MAE loss functions given by Eqs. (10) and (11). For each replication of the exercise

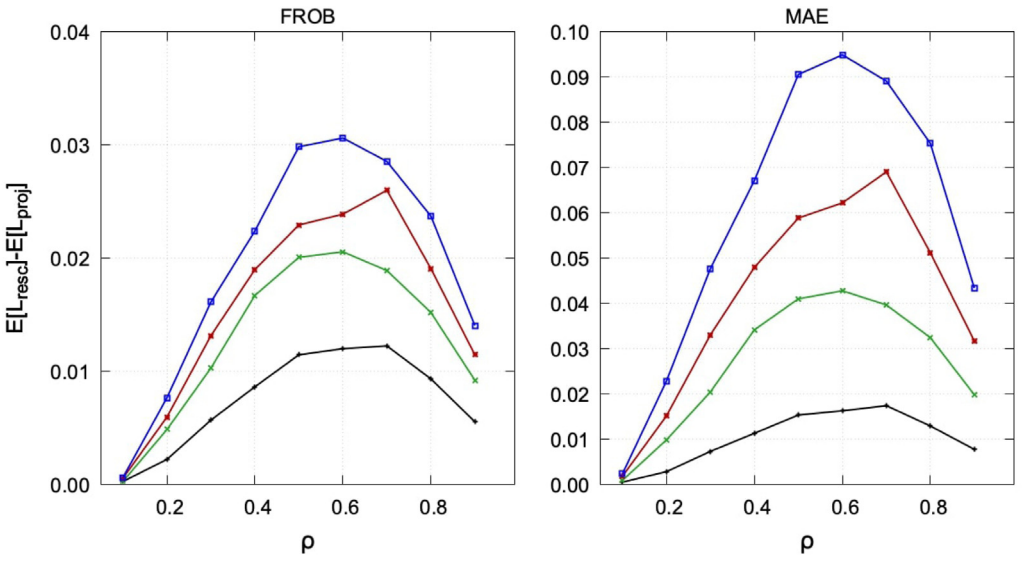

Fig. 4. Static simulation study: Sample correlation versus Stein's projection of the sample covariance matrix onto the correlation set. The y-axis shows $\mathbb{E}[\mathcal{L}_{Frob,resc}] - \mathbb{E}[\mathcal{L}_{Frob,proj}]$ and $\mathbb{E}[\mathcal{L}_{MAE,resc}] - \mathbb{E}[\mathcal{L}_{MAE,proj}]$. 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, green, red and blue lines correspond to $n = 10, 23, 36, 50$, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

we compute the average losses across all the time periods $t = 1, \ldots, T$ and we then average again these across 1000 replications of the Monte Carlo exercise.

The correlation plot in Fig. 5 illustrates how the Pro-DCC correlation delivers a performance improvement over the DCC. In Fig. 6 we report the results of the exercise for different values of n and ρ. Overall, results show that the Pro-DCC performs systematically better than rescaling and that the gains of the projections are larger when the dimension of the system is larger and when the degree of correlation is higher. These results are robust to the choice of the amplitude and period of the cosine wave. We remark that here we have focused on the positive range of the correlations to avoid issues related to the positive definiteness of the equicorrelation matrix since the corresponding lower bound for the ρ_t parameter varies with the dimension n.

4. Empirical application

In this Section we carry out an out-of-sample asset allocation exercise using the constituents of the S&P 100 to assess the benefits of the Pro-DCC for forecasting. The exercise design is close in spirit to the one in De Nard et al. (2021). An important difference between the latter and our paper is that we use publicly available daily data from Alpha Vantage (closing stock prices adjusted for dividends and splits) and that the focus is exclusively on the S&P 100. Hence, the investment universe is fixed rather than time-varying and smaller than the one considered in that paper.

We consider the $n = 86$ constituents of the S&P 100 that have continuously been trading between 2011-01-01 and 2019-06-30 (2136 trading days). We transform adjusted close price data into log-returns, namely $r_{it} = 100 \times \log(P_{it}/P_{it-1})$ for $i = 1, \ldots, n$. The exercise consists in constructing the global minimum variance portfolio (GVMP) once a month on the basis of different covariance forecasts and to then measure the accuracy of the different covariance forecasts on the basis of asset allocation metrics.[4] It should be emphasized that despite the fact that portfolio rebalancing occurs on a monthly basis, all models are estimated using daily data. The GMVP is defined as $r_{\mathsf{GMVP}\,t} = w_t^{*'} r_t$ where

$$w_t^* = \arg\min_{w_t'\iota=1} w_t' \boldsymbol{\Sigma}_t w_t \ ,$$

with ι denoting an n-dimensional vector of ones. As it is well known, the minimizer of this optimization problem is given by

$$w_t^* = \frac{\boldsymbol{\Sigma}_t^{-1}\iota}{\iota'\boldsymbol{\Sigma}_t^{-1}\iota} \ . \tag{12}$$

The GMVP has become a fairly standard metric to evaluate covariance forecasts. In particular, the appeal of the GMVP lies in the fact that this asset allocation strategy only depends on covariance forecasts and in particular it does not rely on forecasts of the expected returns. As an additional exercise, we consider the GMVP with exposure constraints. In the literature, it is generally documented that adding exposure constraints improves the minimum-variance portfolio allocations and has an interpretation in terms of shrinkage estimation (Fan et al., 2012; Jagannathan & Ma, 2003). The GMVP with 1-norm constraints is formulated as

$$w_t^* = \arg\min_{w_t'\iota=1, \|w_t\|_1 \le \gamma} w_t' \boldsymbol{\Sigma}_t w_t, \tag{13}$$

[4] We follow the common convention that 21 consecutive trading days constitute one "month".

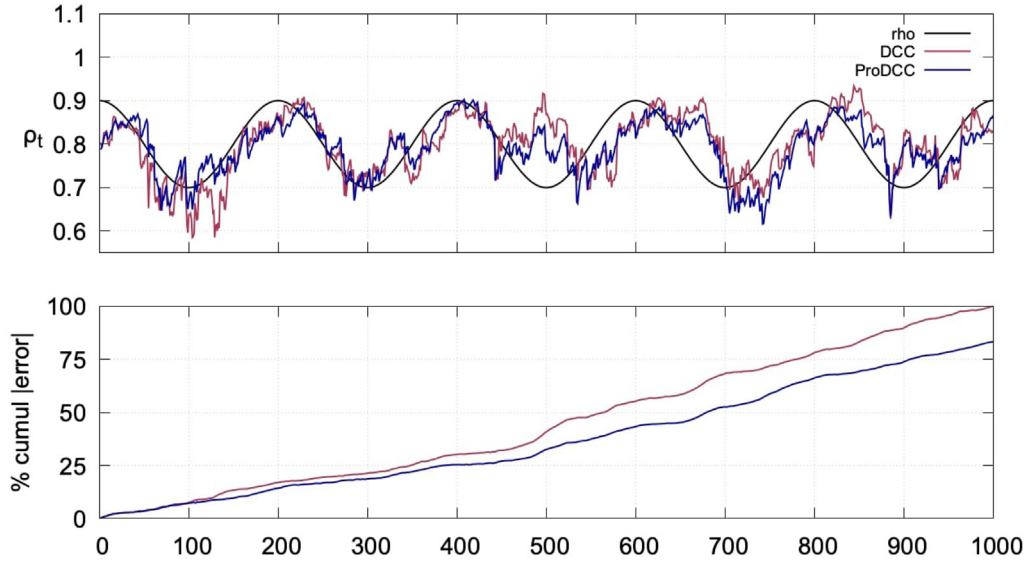

Fig. 5. Dynamic simulation study: Top panel: Dynamic correlations for the true DGP (black), versus the corresponding estimates from DCC (red) and Pro-DCC (blue), where the true DGP is given by $\rho_t = \rho + .1\cos(2\pi t/200)$. Bottom panel: cumulative sum of absolute error for each methodology (same color legend), divided by the sum of absolute errors from the DCC model times 100.

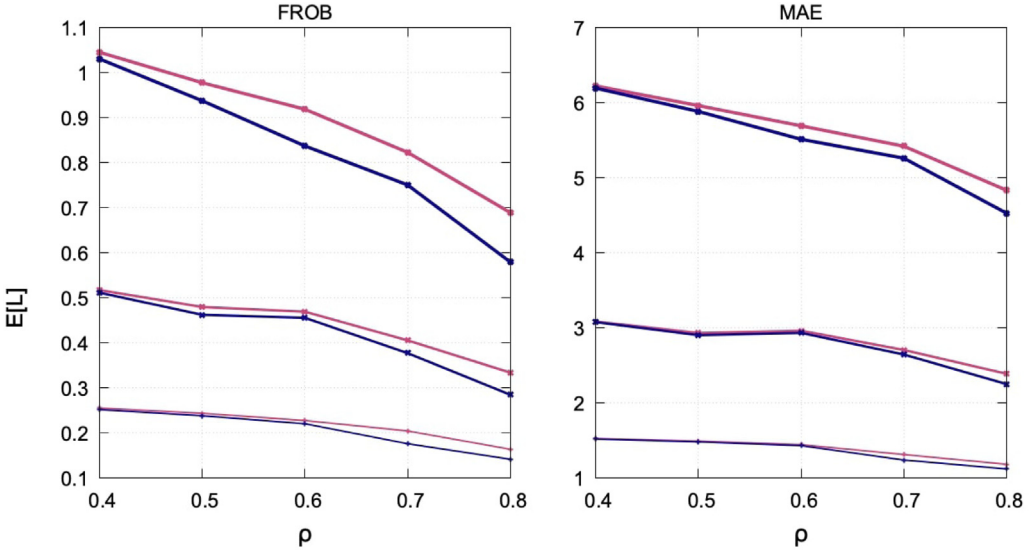

Fig. 6. Dynamic simulation study: DCC's rescaling (red) versus Pro-DCC's projection (blue) based on Stein's loss for dimensions $n = 50$, (bottom pair), $n = 100$ (middle pair) and $n = 200$ (top pair). The y-axis shows the average Frobenius (left panel) and MAE (left panel) losses across Montecarlo simulations, where each loss is computed as the average loss across $t = 1, \ldots, T$. The x-axis shows the magnitude of the parameter ρ, which captures the level of cross-sectional dependence of the process.

for some $\gamma > 0$. We consider the well-known choices $\gamma = 1$ (the no-short sale portfolio) and $\gamma = 2$ (the 150/50 portfolio, i.e. the portfolio that allows a maximum of 50% of short positions). In order to understand whether projections represent an improvement beyond shrinkage, we compute $\gamma_t = \sum_{i=1}^{n} |w_{i,t}|$ for each t for the portfolio based on the Projected DCC and add the portfolio constraints of the standard DCC with exposure constraint $\gamma = \gamma_t$.[5]

To compute the GMVP from the data, we train some suitable dynamic covariance model in-sample and compute one-step ahead covariance forecasts out-of-sample on a rolling basis, keeping the parameters fixed to their in-sample estimates. The portfolio is updated every month to mitigate the costs of rebalancing. We note that in De Nard et al. (2021) the approach to forecasting is slightly different, as they compute the average of k-step ahead forecasts for k ranging from 1 to 21. However, in order to compute those forecasts, it is assumed that $\mathbb{E}[\mathbf{R}_{t+k}|\mathcal{F}_t] = \mathbb{E}[\mathbf{Q}_{t+k}|\mathcal{F}_t]$, which is an approximation. In this work we

[5] We thank one of the referees for the suggestion.

refrain from following this strategy and we focus on one-step ahead forecasts only in order to make the comparison between DCC and Pro-DCC more transparent.

We consider different covariance forecasting strategies to construct the GMVP. We entertain: DCC, the standard version of the DCC model; Pro-DCC, the Projected DCC model based on Stein's projection; RiskMetrics, the RM2006 methodology of Zumbach (2007), and RollCov, Rolling sample covariance computed with a window of six months. For the GARCH-DCC/Pro-DCC specifications, the marginal distribution for each asset is assumed to be GARCH(1,1). For benchmarking purposes, we also report metrics for the equal-weighted portfolio ("1/N").

Moreover, since the dimensionality of the problem is fairly large, the DCC and Pro-DCC models are implemented using standard covariance targeting as well as covariance targeting based on the analytical nonlinear shrinkage (NLS) methodology from Ledoit and Wolf (2020). Also, DCC models are estimated using the composite likelihood approach with contiguous pairs for the dynamic correlation models as in Pakel et al. (2018), which significantly reduces the computational burden of estimation. We evaluate the performance of these different forecasting strategies using the following three out-of-sample performance measures: AV, the annualized out-of-sample average of the GMVP returns; SD, the annualized out-of-sample standard deviation of the GMVP returns, and Sharpe, the Sharpe ratio computed as AV/SD.[6] For instance, the annualized volatility of the portfolio is given by

$$SD_{\text{GMVP}} = \sqrt{252} \times \sqrt{\tau^{-1} \sum_{t=T+1}^{T+\tau} r^2_{\text{GMVP}\,t}} \,,$$

where τ is the length of the out-of-sample period and T is the length of the in-sample period.

The resulting portfolio metrics are presented in Table 1 and results using different split dates for the in-sample and out-of-sample periods can be found in Tables 2 and 3. Results from tests of the difference in portfolio variances using HAC inference as in Ledoit and Wolf (2008) are shown in Tables 4–6. The findings of the exercise can be summarised as follows. First, overall, we have that Pro-DCC-NLS and Pro-DCC-SC outperform all other candidate estimators in terms of both the standard deviation and the Sharpe ratio of the GMVP. The (analytical) nonlinear shrinkage versions of all estimators considered in the exercise present superior performance than doing no shrinkage. Second, 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. Third, the empirical results show that the 1-norm constrained portfolios (denoted with γ in Tables 1–3) substantially improve performance for both the DCC and Pro-DCC. In a nutshell, these results suggest that projecting, rescaling and shrinking are essentially different operations, and that the advantage of Pro-DCC versus DCC persists even after controlling for shrinkage. Incidentally we note that the equal-weighted portfolio ("1/N") in this sample performs relatively well in comparison to other studies – despite being the worst performing portfolio in terms of standard deviation. However, it must be noted that for larger cross sections and/or periods of distress like the 2008 global financial crisis, the results from the 1/N portfolio deteriorate significantly – see for example De Nard et al. (2021) and Engle et al. (2019).

4.1. Alternative parameterization of the Global Minimum Variance Portfolio

We provide a new approach to analyze the weights given by the Global Minimum Variance Portfolio. We define $\mathbf{K}_t := \boldsymbol{\Sigma}_t^{-1}$ and $\boldsymbol{\Omega}_t := \mathbf{R}_t^{-1}$, and note that the $n \times n$ matrix of partial correlations can be written as

$$\boldsymbol{\varrho}_t = -\operatorname{diag}(\boldsymbol{\Omega}_t)^{-1/2} \boldsymbol{\Omega}_t \operatorname{diag}(\boldsymbol{\Omega}_t)^{-1/2} \,.$$

Defining $v_{i,t} := \frac{\sqrt{\Omega_{ii,t}}}{d_{i,t}}$, after straightforward computations we have that

$$w_t^* = \frac{\mathbf{K}_t \iota}{\iota' \mathbf{K}_t \iota} = \frac{v_t \odot \boldsymbol{\varrho}_t v_t}{\iota'(v_t \odot \boldsymbol{\varrho}_t v_t)}, \tag{14}$$

where $v_t = (v_{1,t}, \ldots, v_{n,t})'$ and $\Omega_{ii,t}$ is the ith diagonal element of $\boldsymbol{\Omega}_t$. This means that the sign and magnitude of the GMVP weights depend on 3 factors: (1) $d_{i,t}$, the volatility of the ith asset, which always shrinks the exposure to 0. (2) $s_{i,t} := \frac{1}{\sqrt{\Omega_{ii,t}}}$ or in words, the "spillover" effect which is proportional to the determinant of the conditional correlation matrix $\mathbf{R}_t$ given by all assets excluding i. The higher the $s_{i,t}$, the stronger is the shrinkage towards 0. (3) $\varrho_{ij,t}$: partial correlation between assets i and j, which measures the direct effects between assets i and j conditional on all other assets in the portfolio. Note that if the ith asset is directly and strongly connected to the rest of the system, the weight is likely to be negative (i.e. increased exposure in short position).

The parameterization allows us to deepen our understanding of the GMVP weight vector. In Fig. 7 we can visualize the alternative parameterization of the weight vector for a given date. For instance, if we focus on asset number 20, we observe that the weight implied by the DCC model is larger than the one implied by the Pro-DCC model (the actual weights on the given date are 7.5% vs 1.8%, respectively). From the second column we learn that if all assets are independent of each other, then the weight should be the lowest since this asset is the most volatile at the given date. Both models convey that partial correlation effects for this asset are on the lower end of the distribution, which makes the asset more attractive

[6] Additionally, the following portfolio metrics are computed:
- Turnover: $\frac{1}{n(H-1)} \sum_{h=1}^{H} \|\hat{w}_{h+1} - \hat{w}_h^{hold}\|_1$, where H is the total number of months out-of-sample, and w_h^{hold} is the weight vector right before the next monthly update, that is, taking into account the price evolution of each asset during the month.
- Proportion of leverage: $\frac{1}{nH} \sum_{h=1}^{H} \sum_{i=1}^{n} \mathbf{1}\{w_{i,h} < 0\}$.
- Gamma: $\frac{1}{H} \sum_{h=1}^{H} \sum_{i=1}^{n} |w_{i,h}|$.
- Maximum weight: $\max_{i,h} w_{i,h}$.
- Minimum weight: $\min_{i,h} w_{i,h}$.

Table 1
Portfolio selection with all constituents in S&P 100. Performance metrics for the out-of-sample period, which ranges from 2015-12-31 to 2019-06-30. AV: annualized average portfolio return. SD: annualized volatility of portfolio returns. Sharpe: AV/SD. Turnover: $\frac{1}{n(H-1)}\sum_{h=1}^{H}\|\hat{w}_{h+1}-\hat{w}_h^{hold}\|_1$, where H is the total number of months out-of-sample, and w_h^{hold} is the weight vector right before the next monthly update, that is, taking into account the price evolution of each asset during the month. Leverage: $\frac{1}{nH}\sum_{h=1}^{H}\sum_{i=1}^{n}\mathbf{1}\{w_{i,h}<0\}$. Gamma: $\frac{1}{H}\sum_{h=1}^{H}\sum_{i=1}^{n}|w_{i,h}|$. Maximum weight: $\max_{i,h} w_{i,h}$. Minimum weight: $\min_{i,h} w_{i,h}$. The annualized volatility of the best portfolio is highlighted in boldface.

	AV	SD	Sharpe	Turnover	Leverage	Gamma	MaxWeight	MinWeight
DCC-SC	7.880	12.489	0.631	0.065	0.493	2.945	0.680	−0.125
Pro-DCC-SC	10.228	11.918	0.858	0.093	0.488	2.935	0.439	−0.125
DCC-SC($\gamma=\gamma_t$)	7.833	12.478	0.628	0.293	0.490	2.897	0.679	−0.125
DCC-SC($\gamma=1$)	7.880	11.989	0.657	0.019	0.019	1.001	0.911	−0.004
Pro-DCC-SC($\gamma=1$)	9.133	**11.427**	0.799	0.028	0.019	1.001	0.592	−0.004
DCC-SC($\gamma=2$)	8.975	12.133	0.740	0.081	0.447	2.000	0.678	−0.093
Pro-DCC-SC($\gamma=2$)	11.245	11.676	0.963	0.036	0.451	2.000	0.441	−0.116
DCC-NLS	7.745	12.362	0.626	0.071	0.492	2.756	0.685	−0.097
Pro-DCC-NLS	10.139	11.823	0.858	0.050	0.490	2.733	0.432	−0.107
RiskMetrics	8.387	12.251	0.685	0.173	0.428	4.724	0.466	−0.418
RollCov	14.931	19.799	0.754	0.559	0.463	7.787	1.017	−0.795
1/N	10.440	12.426	0.840	0.016	0.000	1.000	0.012	0.012

Table 2
Portfolio selection with all constituents in S&P 100. Performance metrics for the out-of-sample period, which ranges from 2016-12-31 to 2019-06-30. AV: annualized average portfolio return. SD: annualized volatility of portfolio returns. Sharpe: AV/SD. Turnover: $\frac{1}{n(H-1)}\sum_{h=1}^{H}\|\hat{w}_{h+1}-\hat{w}_h^{hold}\|_1$, where H is the total number of months out-of-sample, and w_h^{hold} is the weight vector right before the next monthly update, that is, taking into account the price evolution of each asset during the month. Leverage: $\frac{1}{nH}\sum_{h=1}^{H}\sum_{i=1}^{n}\mathbf{1}\{w_{i,h}<0\}$. Gamma: $\frac{1}{H}\sum_{h=1}^{H}\sum_{i=1}^{n}|w_{i,h}|$. Maximum weight: $\max_{i,h} w_{i,h}$. Minimum weight: $\min_{i,h} w_{i,h}$. The annualized volatility of the best portfolio is highlighted in boldface.

	AV	SD	Sharpe	Turnover	Leverage	Gamma	MaxWeight	MinWeight
DCC-SC	4.122	12.439	0.331	0.051	0.491	2.749	0.558	−0.107
Pro-DCC-SC	5.881	12.112	0.486	0.058	0.489	2.691	0.481	−0.104
DCC-SC($\gamma=\gamma_t$)	4.275	12.406	0.345	0.044	0.489	2.669	0.567	−0.107
DCC-SC($\gamma=1$)	10.284	11.878	0.866	0.013	0.023	1.001	0.803	−0.004
Pro-DCC-SC($\gamma=1$)	9.196	**11.454**	0.803	0.013	0.021	1.000	0.681	−0.003
DCC-SC($\gamma=2$)	6.635	12.207	0.544	0.039	0.462	2.000	0.578	−0.087
Pro-DCC-SC($\gamma=2$)	7.588	11.884	0.638	0.050	0.473	2.000	0.487	−0.099
DCC-NLS	4.295	12.348	0.348	0.038	0.494	2.627	0.567	−0.100
Pro-DCC-NLS	5.960	12.017	0.496	0.055	0.495	2.567	0.492	−0.092
RiskMetrics	1.229	12.100	0.102	0.125	0.416	4.329	0.388	−0.368
RollCov	1.591	15.945	0.100	0.253	0.455	7.048	0.815	−0.606
1/N	9.519	12.126	0.785	0.013	0.000	1.000	0.012	0.012

Table 3
Portfolio selection with all constituents in S&P 100. Performance metrics for the out-of-sample period, which ranges from 2017-12-31 to 2019-06-30. AV: annualized average portfolio return. SD: annualized volatility of portfolio returns. Sharpe: AV / SD. Turnover: $\frac{1}{n(H-1)}\sum_{h=1}^{H}\|\hat{w}_{h+1}-\hat{w}_h^{hold}\|_1$, where H is the total number of months out-of-sample, and w_h^{hold} is the weight vector right before the next monthly update, that is, taking into account the price evolution of each asset during the month. Leverage: $\frac{1}{nH}\sum_{h=1}^{H}\sum_{i=1}^{n}\mathbf{1}\{w_{i,h}<0\}$. Gamma: $\frac{1}{H}\sum_{h=1}^{H}\sum_{i=1}^{n}|w_{i,h}|$. Maximum weight: $\max_{i,h} w_{i,h}$. Minimum weight: $\min_{i,h} w_{i,h}$. The annualized volatility of the best portfolio is highlighted in boldface.

	AV	SD	Sharpe	Turnover	Leverage	Gamma	MaxWeight	MinWeight
DCC-SC	0.186	14.672	0.013	0.069	0.499	2.606	0.392	−0.136
Pro-DCC-SC	2.198	14.443	0.152	0.064	0.500	2.483	0.336	−0.096
DCC-SC($\gamma=\gamma_t$)	0.165	14.623	0.011	0.136	0.493	2.468	0.388	−0.132
DCC-SC($\gamma=1$)	1.858	13.755	0.135	0.015	0.026	1.001	0.599	−0.004
Pro-DCC-SC($\gamma=1$)	2.111	**13.442**	0.157	0.043	0.023	1.001	0.503	−0.002
DCC-SC($\gamma=2$)	1.065	14.427	0.074	0.065	0.466	2.000	0.414	−0.120
Pro-DCC-SC($\gamma=2$)	3.095	14.172	0.218	0.072	0.467	2.000	0.342	−0.085
DCC-NLS	0.105	14.581	0.007	0.066	0.503	2.511	0.395	−0.129
Pro-DCC-NLS	2.019	14.350	0.141	0.052	0.499	2.384	0.337	−0.088
RiskMetrics	−4.857	14.514	−0.335	0.159	0.426	4.776	0.386	−0.362
RollCov	−9.864	18.875	−0.523	0.232	0.462	7.715	0.814	−0.569
1/N	4.461	14.745	0.303	0.016	0.000	1.000	0.012	0.012

Table 4
Difference in variance test (Ledoit & Wolf, 2011) for different candidate portfolios. Split Date: 2015-12-31. Every cell computes the difference in standard deviation from the column portfolio minus the standard deviation of the row portfolio. Figures are reported in annualized terms i.e. multiplying by $\sqrt{252}$).

	DCC-SC	Pro-DCC-SC	DCC-SC($\gamma = \gamma_t$)	DCC-SC($\gamma = 1$)	Pro-DCC-SC($\gamma = 1$)	DCC-SC($\gamma = 2$)	Pro-DCC-SC($\gamma = 2$)	DCC-NLS	Pro-DCC-NLS	RiskMetrics	RollCov	1/N
DCC-SC	–	−0.57*	−0.01	−0.5	−1.06***	−0.36***	−0.81***	−0.13***	−0.67**	−0.24	7.31***	−0.06
Pro-DCC-SC	0.57*	–	0.56*	0.07	−0.49	0.22	−0.24**	0.44	−0.09**	0.33	7.88***	0.51
DCC-SC($\gamma = \gamma_t$)	0.01	−0.56*	–	−0.49	−1.05***	−0.35***	−0.8***	−0.12***	−0.66**	−0.23	7.32***	−0.05
DCC-SC($\gamma = 1$)	0.5	−0.07	0.49	–	−0.56	0.14	−0.31	0.37	−0.17	0.26	7.81***	0.44
Pro-DCC-SC($\gamma = 1$)	1.06***	0.49	1.05***	0.56	–	0.71***	0.25	0.94***	0.4	0.82	8.37***	1**
DCC-SC($\gamma = 2$)	0.36***	−0.22	0.35***	−0.14	−0.71***	–	−0.46	0.23**	−0.31	0.12	7.67***	0.29
Pro-DCC-SC($\gamma = 2$)	0.81***	0.24**	0.8***	0.31	−0.25	0.46	–	0.69**	0.15**	0.58	8.12***	0.75**
DCC-NLS	0.13***	−0.44	0.12***	−0.37	−0.94***	−0.23**	−0.69**	–	−0.54*	−0.11	7.44***	0.06
Pro-DCC-NLS	0.67**	0.09**	0.66**	0.17	−0.4	0.31	−0.15**	0.54*	–	0.43	7.98***	0.6
RiskMetrics	0.24	−0.33	0.23	−0.26	−0.82	−0.12	−0.58	0.11	−0.43	–	7.55***	0.17
RollCov	−7.31***	−7.88***	−7.32***	−7.81***	−8.37***	−7.67***	−8.12***	−7.44***	−7.98***	−7.55***	–	−7.37***
1/N	0.06	−0.51	0.05	−0.44	−1**	−0.29	−0.75**	−0.06	−0.6	−0.17	7.37***	–

*Indicate that results are significant at the 10%.
**Indicate that results are significant at the 5%.
***Indicate that results are significant at the 1%.

Table 5
Difference in variance test (Ledoit & Wolf, 2011) for different candidate portfolios. Split Date: 2016-12-31. Every cell computes the difference in standard deviation from the column portfolio minus the standard deviation of the row portfolio. Figures are reported in annualized terms i.e. multiplying by $\sqrt{252}$).

	DCC-SC	Pro-DCC-SC	DCC-SC($\gamma = \gamma_t$)	DCC-SC($\gamma = 1$)	Pro-DCC-SC($\gamma = 1$)	DCC-SC($\gamma = 2$)	Pro-DCC-SC($\gamma = 2$)	DCC-NLS	Pro-DCC-NLS	RiskMetrics	RollCov	1/N
DCC-SC	–	−0.33	−0.03**	−0.56*	−0.98***	−0.23***	−0.56***	−0.09***	−0.42**	−0.34	3.51***	−0.31
Pro-DCC-SC	0.33	–	0.29	−0.23	−0.66**	0.09	−0.23***	0.24	−0.09***	−0.01	3.83***	0.01
DCC-SC($\gamma = \gamma_t$)	0.03**	−0.29	–	−0.53*	−0.95***	−0.2***	−0.52***	−0.06**	−0.39*	−0.31	3.54***	−0.28
DCC-SC($\gamma = 1$)	0.56*	0.23	0.53*	–	−0.42***	0.33	0.01	0.47*	0.14	0.22	4.07***	0.25
Pro-DCC-SC($\gamma = 1$)	0.98***	0.66**	0.95***	0.42***	–	0.75***	0.43*	0.89***	0.56**	0.65	4.49***	0.67
DCC-SC($\gamma = 2$)	0.23***	−0.09	0.2***	−0.33	−0.75***	–	−0.32	0.14**	−0.19	−0.11	3.74***	−0.08
Pro-DCC-SC($\gamma = 2$)	0.56***	0.23***	0.52***	−0.01	−0.43*	0.32	–	0.46**	0.13**	0.22	4.06***	0.24
DCC-NLS	0.09***	−0.24	0.06**	−0.47*	−0.89***	−0.14**	−0.46**	–	−0.33*	−0.25	3.6***	−0.22
Pro-DCC-NLS	0.42**	0.09***	0.39*	−0.14	−0.56**	0.19	−0.13**	0.33*	–	0.08	3.93***	0.11
RiskMetrics	0.34	0.01	0.31	−0.22	−0.65	0.11	−0.22	0.25	−0.08	–	3.85***	0.03
RollCov	−3.51***	−3.83***	−3.54***	−4.07***	−4.49***	−3.74***	−4.06***	−3.6***	−3.93***	−3.85***	–	−3.82***
1/N	0.31	−0.01	0.28	−0.25	−0.67	0.08	−0.24	0.22	−0.11	−0.03	3.82***	–

*Indicate that results are significant at the 10%.
**Indicate that results are significant at the 5%.
***Indicate that results are significant at the 1%.

than what is implied by the volatility alone. The main difference is thus explained by the third-to-last column (s), which captures the spillover effects from the asset in question to the rest of the system. In other words, the Pro-DCC model implies that asset 20 is indirectly connected to the remaining assets way more strongly than what is implied by the DCC model, which explains why it has a lower weight.

5. Conclusions

In this paper we contribute to the DCC literature with a novel specification inspired by the literature on Bregman matrix projections and the nearest-correlation matrix problem. We demonstrate the benefits of using our proposed methodology with respect to the standard GARCH-DCC model in a simulated exercise. We also carry out a global minimum variance portfolio exercise using

Table 6
Difference in variance test (Ledoit & Wolf, 2011) for different candidate portfolios. Split Date: 2017-12-31. Every cell computes the difference in standard deviation from the column portfolio minus the standard deviation of the row portfolio. Figures are reported in annualized terms i.e. multiplying by $\sqrt{252}$).

	DCC-SC	Pro-DCC-SC	DCC-SC($\gamma = \gamma_t$)	DCC-SC($\gamma = 1$)	Pro-DCC-SC($\gamma = 1$)	DCC-SC($\gamma = 2$)	Pro-DCC-SC($\gamma = 2$)	DCC-NLS	Pro-DCC-NLS	RiskMetrics	RollCov	1/N
DCC-SC	–	−0.23*	−0.05**	−0.92***	−1.23***	−0.24***	−0.5***	−0.09***	−0.32***	−0.16	4.2***	0.07
Pro-DCC-SC	0.23*	–	0.18	−0.69*	−1***	−0.02	−0.27***	0.14	−0.09***	0.07	4.43***	0.3
DCC-SC($\gamma = \gamma_t$)	0.05**	−0.18	–	−0.87**	−1.18***	−0.2***	−0.45***	−0.04	−0.27**	−0.11	4.25***	0.12
DCC-SC($\gamma = 1$)	0.92***	0.69*	0.87**	–	−0.31***	0.67**	0.42	0.83**	0.59*	0.76	5.12***	0.99*
Pro-DCC-SC($\gamma = 1$)	1.23***	1***	1.18***	0.31***	–	0.98***	0.73**	1.14***	0.91***	1.07*	5.43***	1.3**
DCC-SC($\gamma = 2$)	0.24***	0.02	0.2***	−0.67**	−0.98***	–	−0.25*	0.15**	−0.08	0.09	4.45***	0.32
Pro-DCC-SC($\gamma = 2$)	0.5***	0.27***	0.45***	−0.42	−0.73**	0.25*	–	0.41***	0.18***	0.34	4.7***	0.57
DCC-NLS	0.09***	−0.14	0.04	−0.83**	−1.14***	−0.15**	−0.41***	–	−0.23*	−0.07	4.29***	0.16
Pro-DCC-NLS	0.32***	0.09***	0.27**	−0.59*	−0.91***	0.08	−0.18***	0.23*	–	0.16	4.53***	0.4
RiskMetrics	0.16	−0.07	0.11	−0.76	−1.07*	−0.09	−0.34	0.07	−0.16	–	4.36***	0.23
RollCov	−4.2***	−4.43***	−4.25***	−5.12***	−5.43***	−4.45***	−4.7***	−4.29***	−4.53***	−4.36***	–	−4.13***
1/N	−0.07	−0.3	−0.12	−0.99*	−1.3**	−0.32	−0.57	−0.16	−0.4	−0.23	4.13***	–

*Indicate that results are significant at the 10%.
**Indicate that results are significant at the 5%.
***Indicate that results are significant at the 1%.

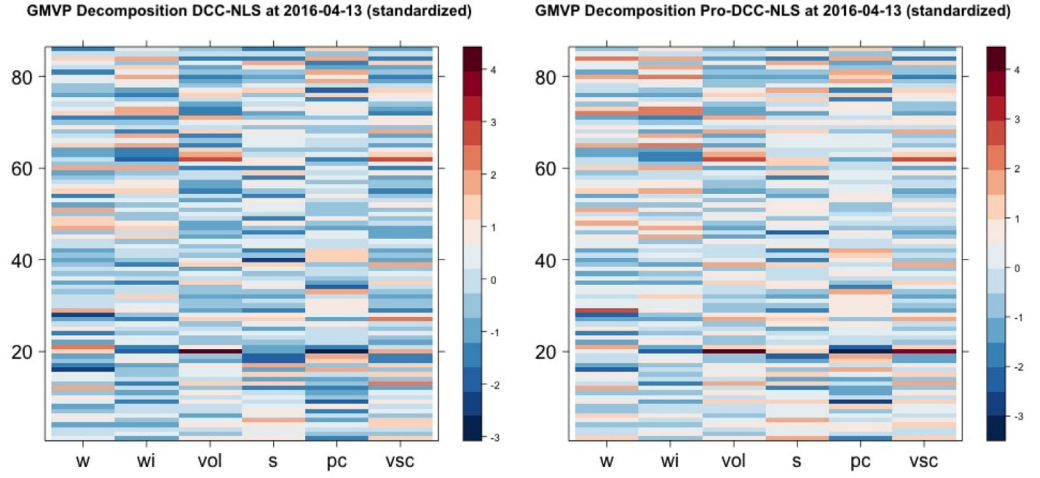

Fig. 7. Heatmap of GMVP at 2016-04-13. Column w shows the weights implied by the model, which are contrasted to the weights under the assumption that $\boldsymbol{\Sigma}_t$ is diagonal (second column). The column vol shows the conditional volatilities for every asset. The column s is the vector that stacks the spillover effects $s_{i,t} := 1/\sqrt{\boldsymbol{\Omega}_{ii,t}}$. The column pc shows the partial correlation effects, and for every asset i they are computed as $\sum_{j \neq i} \boldsymbol{\varrho}_{ij,t}$. Finally, the column vsc contains the reciprocal of the v_t vector, so $vsc_{i,t} = v_{i,t}^{-1}$. For the purposes of visualisation, all columns have been standardized.

a set of constituents of the S&P 100. Results show 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.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

I acknowledge support from the Spanish Ministry of Science and Technology (FPI Grant PRE2019-090839) Christian Brownlees acknowledges support from the Spanish Ministry of Science and Technology (Grant MTM2012-37195) and the Spanish Ministry of Economy and Competitiveness through the Severo Ochoa Programme for Centres of Excellence in R&D (SEV-2011-0075).

Appendix. Proofs

Proof of Lemma 1. The claim follows from Theorem 3.12(*iii*) in Bauschke and Borwein (1997). □

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

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

Therefore,

$$\operatorname{tr}(\mathbf{RK}) = k_1 + 2k_s\rho + k_2, \text{ and}$$
$$\det(\mathbf{RK}) = \det(\mathbf{K})\det(\mathbf{R}) = (k_1k_2 - k_s^2)(1-\rho^2)$$

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

$$\min_{\rho} f(\rho) = \min_{\rho} k_1 + 2k_s\rho + k_2 - \ln(k_1k_2 - k_s^2) - \ln(1-\rho^2) - 2$$

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 the result in (7). □

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

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

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

$$\log \mathbf{R} = \frac{1}{2}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} \ln(1+\rho) & 0 \\ 0 & \ln(1-\rho) \end{bmatrix} \times \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$$

Hence, the problem is equivalent to minimizing f with respect to ρ.[7] 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:

$$\frac{-2\hat{\rho}}{1-\hat{\rho}^2} + \ln\left(\frac{1+\hat{\rho}}{1-\hat{\rho}}\right) + \hat{\rho}\frac{1-\hat{\rho}}{1+\hat{\rho}}\frac{2}{(1-\hat{\rho})^2} - 2k_v = 0$$
$$\implies \hat{\rho} = \frac{e^{2k_v}-1}{e^{2k_v}+1} = \tanh(k_v)$$

To find an analytical expression for k_v, 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 $i = 1, 2$. Their corresponding eigenvectors are $v_i = [v_{i1}, v_{i2}]'$, where

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

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

$$\begin{aligned} k_v &= (\ln\lambda_1)v_{11}v_{12} + (\ln\lambda_2)v_{21}v_{22} \\ &= (\ln\lambda_1)\psi_1 v_{12}^2 + (\ln\lambda_2)\psi_2 v_{22}^2 \\ &= \frac{\psi_1 \ln\lambda_1}{1+\psi_1^2} + \frac{\psi_2\ln\lambda_2}{1+\psi_2^2} \end{aligned}$$ □

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 $\mathbf{R}^{(k)}$:

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

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

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

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

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

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

$$\operatorname{tr}\left(\left[\mathbf{R}^{(k-1)} + \frac{\alpha}{1-\alpha\mathbf{R}_{ii}^{(k-1)}}\mathbf{R}^{(k-1)}e_ie_i'\mathbf{R}^{(k-1)}\right]e_ie_i'\right) = 1$$
$$\alpha = [\mathbf{R}_{ii}^{(k-1)}]^{-1} - 1$$

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

$$\begin{aligned} \frac{\alpha}{1-\alpha\mathbf{R}_{ii}^{(k-1)}} &= \frac{[\mathbf{R}_{ii}^{(k-1)}]^{-1} - 1}{1 - ([\mathbf{R}_{ii}^{(k-1)}]^{-1} - 1)\mathbf{R}_{ii}^{(k-1)}} \\ &= [\mathbf{R}_{ii}^{(k-1)}]^{-2} - [\mathbf{R}_{ii}^{(k-1)}]^{-1} \\ &= [\mathbf{R}_{ii}^{(k-1)}]^{-2}\left(1 - \mathbf{R}_{ii}^{(k-1)}\right) \end{aligned}$$ □

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

References

Aielli, G. P. (2013). Dynamic conditional correlation: On properties and estimation. *Journal of Business & Economic Statistics*, *31*(3), 282–299.

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

Bauschke, H. G., & Borwein, J. M. (1997). Legendre functions and the method of random Bregman projections. *Journal of Convex Analysis*, *4*(1), 27–67.
Bauwens, L., Laurent, S., & 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., & Engle, R. (2017). SRISK: A conditional capital shortfall measure of systemic risk. *The Review of Financial Studies*, *30*(1), 48–79.
Cesa-Bianchi, N., & Lugosi, G. (2006). *Prediction, learning, and games*. New York, NY, USA: Cambridge University Press.
De Nard, G., Ledoit, O., & Wolf, M. (2021). Factor models for portfolio selection in large dimensions: The good, the better and the ugly. *Journal of Financial Econometrics*, Forthcoming.
Dey, D. K., & Srinivasan, C. (1985). Estimation of a covariance matrix under Stein's loss. *The Annals of Statistics*, *13*(4), 1581–1591.
Dhillon, S. I., & 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 & Economic Statistics*, *20*(3), 339–350.
Engle, R. (2009). *Anticipating correlations: A new paradigm for risk management*. Princeton University Press.
Engle, R., & Kelly, B. (2012). Dynamic equicorrelation. *Journal of Business & Economic Statistics*, *30*(2), 212–228.
Engle, R. F., Ledoit, O., & Wolf, M. (2019). Large dynamic covariance matrices. *Journal of Business & Economic Statistics*, *37*(2), 363–375.
Engle, R., & Sheppard, K. (2001). *Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH*: *Technical report*, NBER, Working Paper No. W8554.
Fan, J., Zhang, J., & Yu, K. (2012). Vast portfolio selection with gross-exposure constraints. *Journal of the American Statistical Association*, *107*(498), 592–606.
Hautsch, N., Kyj, L. M., & Malec, P. (2015). Do high-frequency data improve high-dimensional portfolio allocations? *Journal of Applied Econometrics*, *30*(2), 263–290.
Hautsch, N., & Voigt, S. (2019). Large-scale portfolio allocation under transaction costs and model uncertainty. *Journal of Econometrics*, *212*(1), 221–240.
Higham, N. J. (2002). Computing the nearest correlation matrix—a problem from finance. *IMA Journal of Numerical Analysis*, *22*(3), 329–343.
Jagannathan, R., & Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance*, *58*(4), 1651–1683.
Kendall, S., & Stuart, A. (1979). *The advanced theory of statistics, Volume 2* (4th ed.). MacMillan Publishing Co., Inc.
Kulis, B., Sustik, M. A., & Dhillon, I. S. (2009). Low-rank kernel learning with Bregman matrix divergences. *Journal of Machine Learning Research*, *10*, 341–376.
Laurent, S., Rombouts, J., & Violante, F. (2013). On loss functions and ranking forecasting performances of multivariate volatility models. *Journal of Econometrics*, *173*(1), 1–10.
Ledoit, O., & Wolf, M. (2008). Robust performance hypothesis testing with the sharpe ratio. *Journal of Empirical Finance*, *15*(5), 850–859.
Ledoit, O., & Wolf, M. (2011). Robust performances hypothesis testing with the variance. *Wilmott*, *2011*(55), 86–89.
Ledoit, O., & Wolf, M. (2018). Optimal estimation of a large-dimensional covariance matrix under stein's loss. *Bernoulli*, *24*(4B), 3791–3832.
Ledoit, O., & Wolf, M. (2020). Analytical nonlinear shrinkage of large-dimensional covariance matrices. *The Annals of Statistics*, *48*(5), 3043–3065.
Pakel, C., Shephard, N., Sheppard, K. K., & Engle, R. F. (2018). Fitting vast dimensional time-varying covariance models. NYU Working Paper No. FIN-08-009.
Patton, A. (2020). Comparing possibly misspecified forecasts. *Journal of Business & Economic Statistics*, *38*(4), 796–809.
Pourahmadi, M. (2013). *High-dimensional covariance estimation*. John Wiley & Sons, Inc..
Silvennoinen, A., & 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., & Tsui, A. K. C. (2002). Evaluating the hedging performance of the constant-correlation GARCH model. *Applied Financial Economics*, *12*, 791–798.
Van Os, B., & Van Dijk, D. (2021). *Pooling dynamic conditional correlation models*: *Technical report*, Erasmus University.
Zumbach, G. (2007). The riskmetrics 2006 methodology. *SSRN Electronic Journal*.