Journal of Business & Economic Statistics

ISSN: 0735-0015 (Print) 1537-2707 (Online) Journal homepage: http://amstat.tandfonline.com/loi/ubes20

Large Dynamic Covariance Matrices

Robert F. Engle, Olivier Ledoit & Michael Wolf

To cite this article: Robert F. Engle, Olivier Ledoit & Michael Wolf (2017): Large Dynamic Covariance Matrices, Journal of Business & Economic Statistics, DOI: 10.1080/07350015.2017.1345683

To link to this article: https://doi.org/10.1080/07350015.2017.1345683

Large Dynamic Covariance Matrices

Robert F. ENGLE
Department of Finance, New York University, New York, NY 10012 (rengle@stern.nyu.edu)

Olivier LEDOIT
Department of Economics, University of Zurich, CH-8032 Zurich, Switzerland (olivier.ledoit@econ.uzh.ch);
AlphaCrest Capital Management, New York, NY 10036 (olivier.ledoit@alphacrestcapital.com)

Michael WOLF
Department of Economics, University of Zurich, CH-8032 Zurich, Switzerland (michael.wolf@econ.uzh.ch)

Second moments of asset returns are important for risk management and portfolio selection. The problem of estimating second moments can be approached from two angles: time series and the cross-section. In time series, the key is to account for conditional heteroscedasticity; a favored model is Dynamic Conditional Correlation (DCC), derived from the ARCH/GARCH family started by Engle (1982). In the cross-section, the key is to correct in-sample biases of sample covariance matrix eigenvalues; a favored model is nonlinear shrinkage, derived from Random Matrix Theory (RMT). The present article marries these two strands of literature to deliver improved estimation of large dynamic covariance matrices. Supplementary material for this article is available online.

KEY WORDS: Composite likelihood; Dynamic conditional correlation; GARCH; Markowitz portfolio selection; Nonlinear shrinkage.

1. INTRODUCTION

Multivariate GARCH models derived from the ARCH/GARCH family started by Engle (1982) are popular tools for risk management and portfolio selection. However, the number of assets in the investment universe generally poses a challenge to such models. When this number is large, say on the order of a thousand, many multivariate GARCH models exhibit unsatisfactory performance or cannot even be estimated in the first place due to computational problems. In other words, many multivariate GARCH models suffer from the curse of dimensionality.

The aim of this article is to robustify the dynamic conditional correlation (DCC) model originally proposed by Engle (2002) against large dimensions. To this end we combine two tools. The first tool is the composite likelihood method of Pakel et al. (2014), which makes the estimation of a DCC model in large dimensions computationally feasible: Composite likelihood ensures that DCC can be used in the first place when the number of assets is large. The second tool is the nonlinear shrinkage method of Ledoit and Wolf (2012), which results in improved estimation of the correlation targeting matrix of a DCC model: Nonlinear shrinkage ensures that DCC performs well when the number of assets is large.

Although both methods already exist, the original contribution of the article is that we identify how best to apply nonlinear shrinkage in the DCC model, namely, in the estimation of the intercept matrix—rather than in the "direct" application of nonlinear shrinkage to estimated conditional covariance matrix itself. In addition, we have made substantive improvements to the software designed to estimate the dynamic parameters of the GARCH process in correlation space. Compared to existing toolboxes, these improvements increase the number of assets that can be handled by one order of magnitude.

Related to our proposal is the work of Hafner and Reznikova (2012). The approach that they champion does not use the first tool, which is composite likelihood, and uses linear shrinkage instead of nonlinear shrinkage for the estimation of the intercept matrix. Furthermore, their empirical study only goes to dimension 100, whereas ours can handle at least 1000 assets.

The remainder of the article is organized as follows. Section 2 gives a brief description of the DCC model including the composite likelihood method. Section 3 gives a description of the nonlinear shrinkage method. Section 4 details our loss function, which is custom-tailored to the problem of portfolio selection. Section 5 contains Monte Carlo simulations while Section 6 provides backtesting results based on real-life stock return data. Section 7 gives an extension of our approach to the BEKK model presented in Engle and Kroner (1995). Section 8 concludes. Supplementary material provides additional details and motivation regarding the nonlinear shrinkage method.

Matlab code implementing the DCC model based on nonlinear shrinkage can be downloaded at *http://www.econ.uzh.ch/en/people/faculty/wolf/publications.html* under the link "Programming Code."

2. THE DCC MODEL

Our exposition of the DCC model is primarily based on the works of Engle (2002) and Engle (2009, sec. 11.2).

Journal of Business & Economic Statistics
XXXX 2017, Vol. 0, No. 0
DOI: 10.1080/07350015.2017.1345683

2.1 Notation

In what follows, the subscript i indexes the variables and covers the range of integers from 1 to N, where N denotes the dimension of the covariance matrix. The subscript t indexes the dates and covers the range of integers from 1 to T, where T denotes the sample size. The notation $\mathrm{Diag}(\cdot)$ represents the function that sets to zero all the off-diagonal elements of a matrix.

- $r_{i,t}$: observed data series for variable i at date t, stacked into $\mathbf{r}_t := (r_{1,t}, \ldots, r_{N,t})'$
- $d_{i,t}^2 := \mathrm{var}(r_{i,t}|\mathcal{F}_{t-1})$: conditional variance of the ith variable at date t
- D_t is the N-dimensional diagonal matrix whose ith diagonal element is $d_{i,t}$
- $H_t := \mathrm{cov}(\mathbf{r}_t|\mathcal{F}_{t-1})$: conditional covariance matrix at date t; thus $\mathrm{Diag(H_t) = D_t^2}$
- $s_{i,t} := r_{i,t}/d_{i,t}$: devolatilized series, stacked into $\mathbf{s}_t := (s_{1,t}, \ldots, s_{N,t})'$
- $R_t := \mathrm{Corr}(\mathbf{r}_t|\mathcal{F}_{t-1}) = \mathrm{cov}(\mathbf{s}_t|\mathcal{F}_{t-1})$: conditional correlation matrix at date t
- $\sigma_i^2 := \mathrm{E}(d_{i,t}^2) = \mathrm{var}(r_{i,t})$: unconditional variance of the ith series
- $C := \mathrm{Corr}(\mathbf{r}_t) = \mathrm{cov}(\mathbf{s}_t)$: unconditional correlation matrix

2.2 Model Definition

This exposition is not meant to review the full generality of the DCC family, but to describe a representative member. Certain specific choices have been made for simplicity, tractability, and scalability. For the dynamics of the univariate volatilities, we use a GARCH(1,1) process:

$$d_{i,t}^2 = \omega_i + a_i r_{i,t-1}^2 + b_i d_{i,t-1}^2, \tag{2.1}$$

where (ω_i, a_i, b_i) are the variable-specific GARCH(1,1) parameters.

Remark 1. We use the standard GARCH(1,1) specification here for simplicity. However, it is possible to upgrade to more sophisticated GARCH(-type) models instead; for example, to models incorporating asymmetry effects.

The square roots $d_{i,t}$ of the conditional variances go into the diagonal of the matrix D_t. These N separate univariate models also yield the vector of devolatilized residuals $\mathbf{s}_t := (r_{1,t}/d_{1,t}, \ldots, r_{N,t}/d_{N,t})'$.

We assume that the evolution of the correlation matrix over time is governed by the DCC model with correlation targeting. This is an adaptation of the variance targeting idea of Engle and Mezrich (1996); see eq. (11.7) of Engle (2009). In our notation, it is expressed as

$$Q_t = (1 - \alpha - \beta)C + \alpha\, \mathbf{s}_{t-1}\mathbf{s}_{t-1}' + \beta\, Q_{t-1}, \tag{2.2}$$

where (α, β) are the DCC parameters analogous to (a_i, b_i), but in correlation space instead of variance space. The matrix Q_t can be interpreted as a conditional *pseudo*-correlation matrix, or a conditional covariance matrix of devolatized residuals. It cannot be used directly because its diagonal elements, although close to one, are not exactly equal to one. From this representation, we obtain the conditional correlation matrix and the conditional covariance matrix as

$$R_t := \mathrm{Diag(Q_t)}^{-1/2}\, \mathrm{Q_t}\, \mathrm{Diag(Q_t)}^{-1/2} \tag{2.3}$$

$$H_t := D_t R_t D_t, \tag{2.4}$$

and the data-generating process is driven by the multivariate normal law

$$\mathbf{r}_t|\mathcal{F}_{t-1} \sim \mathcal{N}(0, H_t). \tag{2.5}$$

Remark 2. Aielli (2013) showed that there is a consistency issue related to the estimation of the intercept matrix in the standard DCC model described above. He proposes a correction, labeled cDCC, that apparently solves this problem. However, this correction seems to make very little difference in practice, if any. For the sake of simplicity, we therefore retain the original DCC formulation in the present article.

2.3 Estimation in Large Dimensions

It is well known that estimating the DCC model with a large number of assets is challenging. One of the difficulties is to invert the conditional covariance matrix H_t (for each $t = 1, \ldots, T$) for the computation of the (log-)likelihood. Fortunately, a way to overcome this computational hurdle has been found by Pakel et al. (2014). The basic idea is that, instead of looking at all assets together jointly, it is easier to look at pairs of assets. They call it the *composite likelihood* method: The composite log-likelihood is computed by summing up the log-likelihoods of pairs of assets.

Of the different variants proposed by the authors, the most scalable one is the 2MSCLE method, which is the composite likelihood estimator based on all contiguous pairs. Thus, there are only $N - 1$ bivariate log-likelihoods to compute. Pakel et al. (2014) showed that maximizing the composite (log-)likelihood thus constructed yields consistent, if not efficient, estimators of the two correlation-dynamics parameters α and β.

To summarize, the estimation unfolds as a three-stage process:

1. For each asset, fit a univariate GARCH(1,1) model and use the fitted model to devolatilize the return series.
2. Estimate the unconditional correlation matrix and use it for correlation targeting.
3. Maximize the composite likelihood to estimate correlation dynamics.

The focus of the present article, starting with the next section, is to improve the second step.

Remark 3. The assumption of normality in (2.5) is not a crucial one. In case it does not hold, (composite) likelihood estimation becomes quasi (composite) likelihood estimation, an approach that still enjoys good properties and is commonly accepted by now; for example, see Hamilton (1994, sec. 5.3).

3. LARGE UNCONDITIONAL CORRELATION MATRICES

The critical juncture where the time-series approach of the DCC model meets the cross-sectional results of random matrix theory (RMT) is in the estimation of the unconditional correlation matrix C, which serves as the target or intercept of the ARCH/GARCH dynamics in Equation (2.2).

3.1 Sample Correlation Matrix

It is widely acknowledged that the sample correlation matrix works poorly in large dimensions. This observation goes back at least to Michaud (1989), who famously described portfolio optimization as "error maximization." The sample correlation (and covariance) matrix is a good fit in-sample but it suffers from overfitting, so it underperforms out-of-sample. Portfolio managers invest out-of-sample. Therefore, the sample correlation matrix should be shunned for decision-making.

The reason is that the sample correlation matrix has $N(N-1)/2$ parameters, and the dataset has $N \times T$ noisy observations. When N is of the same order of magnitude as T, these two quantities are similar-sized. It is not possible to estimate accurately $O(N^2)$ parameters from $O(N^2)$ noisy data points: This is the curse of dimensionality in action.

RMT teaches us that the ratio N/T, generally called the *concentration (ratio)*, is the determinant factor. If the concentration is small, say less than $1/100$, then standard (fixed-dimensional) asymptotics represent a good approximation of the true behavior, and the sample correlation matrix can be trusted out of sample. As the concentration gets higher, special methods need to be employed to address the issue of in-sample overfitting due to the excessive number of free parameters.

The most egregious and easily understood example is when the concentration ratio N/T exceeds one, which means that there are more assets than time-series observations, so the sample correlation matrix is singular. In this case, (pseudo) inverting it blows up optimal portfolio weights to plus or minus infinity, which obviously is an economically incorrect solution. But by continuity, as N gets closer to T, problems start to creep in. Indeed, the main lesson of RMT is that, unless N/T is negligible, the sample correlation (and covariance) matrix will systematically misbehave out-of-sample.

3.2 Shrinkage Estimator

Fortunately, recent developments have provided effective solutions to this problem. It is now possible to rectify the in-sample bias of the sample correlation (or covariance) matrix due to overfitting. This rectification takes place at the level of the eigenvalues. The small sample eigenvalues are too small and the large ones too large. So it is just a matter of pushing up the small ones and pulling down the large ones. Since this transformation reduces the spread of the cross-sectional distribution of eigenvalues, it is generally called *shrinkage*.

In this article, we will focus on the nonlinear shrinkage formula of Ledoit and Wolf (2012). The intuition is as follows. Let Σ denote the population covariance matrix, S the sample covariance matrix, and u an eigenvector of S. Then by basic linear algebra, the corresponding sample eigenvalue is equal to $u'Su$. It is the in-sample variance of a portfolio with weights given by the vector u. This is the quantity that needs to be rectified due to overfitting. Nonlinear shrinkage replaces it with (a consistent estimator of) $u'\Sigma u$, the out-of-sample variance of the same portfolio. Clearly, we want to decide our portfolio allocation in the direction of the vector u based on its true out-of-sample risk $u'\Sigma u$, rather than its in-sample counterpart $u'Su$, which is heavily biased due to the curse of dimensionality.

This may seem like magic: given that we do not know Σ, how could we know $u'\Sigma u$? However, Ledoit and Péché (2011) showed that we do not need to know *all* of the true covariance matrix Σ—which would be a hopeless task—but only its eigenvalues. There are just N of them. Given a dataset of dimension $N \times T$, it is clearly impossible to estimate accurately a whole matrix of dimension $N \times N$, but it is theoretically possible to estimate its N eigenvalues. The ratio of parameters to data entries is $1/T$, which is a small number, regardless of how big the matrix dimension N is.

Recovering the population eigenvalues from the sample eigenvalues requires inverting the Marčenko and Pastur (1967) equation, which governs their asymptotic relationship when the dimension is large. El Karoui (2008) and Mestre (2008) were the first to make attempts in this direction. The solutions they proposed suffered from some limitations that made them unsuitable for general use. Subsequently, Ledoit and Wolf (2015) introduced an effective method based on numerical inversion of what they call the QuEST function; this acronym stands for "quantized eigenvalues sampling transform". It is a deterministic N-dimensional function that discretizes the Marčenko–Pastur equation and lends itself to numerical inversion; it is the technology that we will use here.

The basic idea of the article is to use the nonlinear shrinkage estimator of Ledoit and Wolf (2012) to estimate the correlation targeting matrix C in Equation (2.2) instead of the sample correlation matrix. When the dimension N is large and of the same order of magnitude as the sample size T, this approach generates an estimator of the conditional covariance matrix H_t that has better out-of-sample properties.

3.3 Mathematical Formulation

Let $S := [s_{i,t}]$ denote the $N \times T$ matrix of devolatilized returns. Assuming mean zero, its sample covariance matrix is

$$\widehat{C} := \frac{1}{T} SS'. \tag{3.1}$$

Decompose $\widehat{C}$ it into a set of eigenvalues $(\lambda_1, \lambda_2 \ldots, \lambda_N)$, sorted in descending order without loss of generality, and corresponding eigenvectors $(u_1, u_2, \ldots, u_N)$; consequently

$$\widehat{C} = \sum_{i=1}^{N} \lambda_i u_i u_i'. \tag{3.2}$$

Let $Q_{N,T}$ denote the QuEST function defined in Section 2.2 of Ledoit and Wolf (2015). It is a multivariate deterministic function that maps $[0, \infty)^N$ onto itself. Given a set of population eigenvalues $\mathbf{t} := (t_1, \ldots, t_N)$ as input, it returns as output a deterministic equivalent of the sample eigenvalues $Q_{N,T}(\mathbf{t}) = (q^1_{N,T}(\mathbf{t}), q^2_{N,T}(\mathbf{t}), \ldots, q^N_{N,T}(\mathbf{t}))$. Thus, population eigenvalues can be estimated by numerically inverting the QuEST function:

$$\widetilde{\boldsymbol{\tau}} := \underset{\mathbf{t}\in[0,\infty)^N}{\operatorname{argmin}} \frac{1}{N} \sum_{i=1}^{N} \left[q^i_{N,T}(\mathbf{t}) - \lambda_i \right]^2. \tag{3.3}$$

Given estimated population eigenvalues $\widetilde{\boldsymbol{\tau}} = (\widetilde{\tau}_1, \widetilde{\tau}_2, \ldots, \widetilde{\tau}_N)$, we can then use Theorem 4 of Ledoit and Péché (2011) to compute a nonlinear shrinkage formula that is *asymptotically* optimal under large-dimensional asymptotics. Denote the shrunk eigenvalues by $\widetilde{\boldsymbol{\lambda}}(\widetilde{\boldsymbol{\tau}}) := (\widetilde{\lambda}_1(\widetilde{\boldsymbol{\tau}}), \widetilde{\lambda}_2(\widetilde{\boldsymbol{\tau}}), \ldots, \widetilde{\lambda}_N(\widetilde{\boldsymbol{\tau}}))$.

The mathematical expressions for the QuEST function and the nonlinear shrinkage formula are quite involved, as they depend on the Stieltjes transform, an integral transform defined on the upper half of the complex plane, so they will not be repeated here. The way to intuitively understand $\widetilde{\lambda}_i(\widetilde{\boldsymbol{\tau}})$ is that it is a consistent estimator for the out-of-sample variance $u_i'\Sigma u_i$ under large-dimensional asymptotics. The shrinkage estimator of the covariance matrix can then be reconstructed as

$$\widetilde{C} := \sum_{i=1}^{N} \widetilde{\lambda}_i(\widetilde{\boldsymbol{\tau}}) u_i u_i'. \tag{3.4}$$

Two important advantages of this approach are (i) that it does not require the assumption of normality and (ii) that it can handle the challenging case where the number of assets exceeds the sample size.

Supplementary material contains a primer on this technology imported from Probability Theory and from Multivariate Statistics; see Ledoit and Wolf (in press-a) for a more rigorous exposition. In terms of assumptions on the data-generating process, Ledoit and Wolf (in press-a) do not assume normality, but assume that the 12th moment is finite. While this assumption simplifies the mathematical proofs, numerical simulations indicate that a bounded fourth moment would be sufficient in practice.

3.4 Linear Shrinkage

A simpler alternative is to use the linear shrinkage formula of Ledoit and Wolf (2004b). Define the cross-sectional average of sample eigenvalues as

$$\overline{\lambda} := \frac{1}{N} \sum_{i=1}^{N} \lambda_i. \tag{3.5}$$

This method provides a consistent estimator ρ for the optimal *shrinkage intensity*, which is a number between zero and one controlling the amount by which the sample eigenvalues are dragged toward their cross-sectional average $\overline{\lambda}$. The linear shrinkage estimator is expressed as

$$\overline{C} := \sum_{i=1}^{N} \left[\rho\overline{\lambda} + (1-\rho)\lambda_i\right] u_i u_i'. \tag{3.6}$$

This estimator has proven popular with applied researchers but it is optimal only in a class of dimension two, which is nested inside the class of dimension N over which nonlinear shrinkage is optimal. Therefore, asymptotically, when N and T grow large together, nonlinear shrinkage should perform better in the generic case. The main difference is that the shrinkage intensity ρ is constrained to be the same for all eigenvalues under linear shrinkage, whereas it is *individually fitted* to each eigenvalue under nonlinear shrinkage. This approach is obviously more complicated but due to the highly nonlinear nature of the problem, it is also more powerful. Hafner and Reznikova (2012) proposed the application of linear shrinkage to the estimation of DCC models. However, they restricted themselves to smaller dimensions and threw shrinkage into a horse race against composite likelihood, instead of harnessing the combined powers of the two methods. In addition, there is theoretical justification for upgrading from linear to nonlinear shrinkage (Ledoit and Wolf 2012).

Remark 4. As suggested by a referee, one can alternatively consider the linear shrinkage method of Ledoit and Wolf (2004a) for the estimation of the intercept matrix C. This estimator does not belong to the class of rotation-equivariant estimators studied by Ledoit and Wolf (2012, 2015) and also described in supplementary material for completeness.

3.5 Renormalization

In practice, the diagonal elements of $\widehat{C}$, $\widetilde{C}$, and $\overline{C}$ tend to deviate from one slightly, in spite of the fact that devolatilized returns are used as inputs. Therefore, every column and every row has to be divided by the square root of the corresponding diagonal entry, so as to produce a proper correlation matrix.

4. LOSS FUNCTION

This section builds upon Engle and Colacito (2006), hereafter abbreviated by EC. It is couched in terms of the unconditional covariance matrix initially, and the adaptation to conditional covariance matrices is described in Section 4.4. To make the exposition more fluid, the word "return" stands for the raw return on a risky asset minus the risk-free interest rate.

4.1 Out-of-Sample Portfolio Variance

Let Σ denote the N-dimensional covariance matrix of returns, $\widehat{\Sigma}$ a generic estimator of Σ, and m an assumed vector of expected returns. EC's equation (3) gives the out-of-sample variance of the optimal portfolio based on the estimator $\widehat{\Sigma}$ as

$$\mathcal{L}_V(\widehat{\Sigma}, \Sigma, m) := \frac{m'\widehat{\Sigma}^{-1}\Sigma\widehat{\Sigma}^{-1}m}{\left(m'\widehat{\Sigma}^{-1}m\right)^2}. \tag{4.1}$$

This loss function corresponds to the quintessential risk or error minimization objective. It was also adopted by Ledoit and Wolf (in press-a, Definition 2.1), up to rescaling by the squared Euclidian norm of m. In addition, it captures the performance of a covariance matrix estimator for mathematically equivalent problems where variance minimization decisions must be taken, such as Capon (1969) beamforming in signal processing and optimal fingerprinting in climatology (IPCC 2007).

EC's Theorem 1 demonstrates that $\mathcal{L}_V$ is minimized when $\widehat{\Sigma} = \Sigma$, so if we want a loss function that is always nonnegative, and is equal to zero when the covariance matrix estimator happens to be equal to the truth—as is customary—, then we can compute the excess portfolio variance caused by covariance matrix estimation error:

$$\begin{aligned}\mathcal{L}_E(\widehat{\Sigma}, \Sigma, m) &:= \mathcal{L}_V(\widehat{\Sigma}, \Sigma, m) - \mathcal{L}_V(\Sigma, \Sigma, m)\\ &= \frac{m'\widehat{\Sigma}^{-1}\Sigma\widehat{\Sigma}^{-1}m}{(m'\widehat{\Sigma}^{-1}m)^2} - \frac{1}{m'\Sigma^{-1}m} \geq 0.\end{aligned} \tag{4.2}$$

4.2 Expected Returns

The vector of expected returns that we have assumed is not required to be equal to the true one. Different investors will have different models of expected returns. This suggests investigating over a wide range of alternatives for m. The idea is to integrate the excess portfolio variance $\mathcal{L}_E$ over a relevant manifold of expected return vectors.

The relative efficiency of two covariance matrix estimators is unaffected by the norm of the vector m, so we can take $\|m\| = 1$ without loss of generality, where $\| \cdot \|$ denotes the Euclidian norm. For this reason, in dimension 2, EC's sec. 2.1 considers expected returns of the form $m = [\cos(\theta), \sin(\theta)]'$ where θ is some angle. The generalization in dimension N is to consider vectors m that belong to the N-dimensional unit sphere $\mathcal{U}_N := \{x \in \mathbb{R}^N : \|x\| = 1\}$. Averaging out the excess portfolio variance across all possible m's in this manifold leads to a loss function that depends solely on covariance matrices:

$$\begin{aligned}\mathcal{L}_I(\widehat{\Sigma}, \Sigma) &:= \int_{\mathcal{U}_N} \mathcal{L}_E(\widehat{\Sigma}, \Sigma, \mu)\, d\mu \\ &= \int_{\mathcal{U}_N} \left[\frac{\mu' \widehat{\Sigma}^{-1} \Sigma \widehat{\Sigma}^{-1} \mu}{\left(\mu' \widehat{\Sigma}^{-1} \mu\right)^2} - \frac{1}{\mu' \Sigma^{-1} \mu} \right] d\mu, \qquad (4.3)\end{aligned}$$

where $\int_{\mathcal{U}_N}$ denotes the integral over the N-dimensional unit sphere. The intuition is that we want a covariance matrix estimator that is the best "all-rounder" and thus performs well across the board. Given that this article focuses explicitly on the second moments, we are not in a position to take a stance on the orientation of the linear constraint vector.

4.3 Equivalent Formulation Under Large-Dimensional Asymptotics

The integral in $\mathcal{L}_I$ is not easy to evaluate analytically. At this juncture, we can get help from a foundational lemma in random matrix theory (RMT):

$$x'Ax \approx \frac{\mathrm{Tr}(A)\, \|x\|^2}{N}, \qquad (4.4)$$

where $\mathrm{Tr}(\cdot)$ denotes the trace of a matrix, A is some symmetric random matrix of large dimension N, and x is an N-dimensional vector not statistically related to A, that is, x is distributed independently of A and its distribution is rotation-invariant. Rigorous versions of Equation (4.4) appear decisively as Lemma 1 of Marčenko and Pastur (1967), Lemma 3.1 of Silverstein and Bai (1995), and Lemma 1 of Ledoit and Péché (2011). This is basically a cross-sectional law of large numbers. Indeed if, conditional on A, x follows the standard multivariate normal distribution, then even in finite samples the relation

$$\mathrm{E}\left[x'Ax \middle| A \right] = \mathrm{E}\left[\frac{\mathrm{Tr}(A)\, \|x\|^2}{N} \middle| A \right] \qquad (4.5)$$

holds exactly. Injecting (4.4) into the loss function $\mathcal{L}_I$ naturally suggests the alternative loss function $\mathcal{L}$, which is significantly easier to evaluate.

Definition 1. The loss function is defined as

$$\mathcal{L}(\widehat{\Sigma}, \Sigma) := \frac{\mathrm{Tr}(\widehat{\Sigma}^{-1} \Sigma \widehat{\Sigma}^{-1})/N}{[\mathrm{Tr}(\widehat{\Sigma}^{-1})/N]^2} - \frac{1}{\mathrm{Tr}(\Sigma^{-1})/N}. \qquad (4.6)$$

This is the loss function that will be used throughout the rest of the article. The formulations $\mathcal{L}_I$ and $\mathcal{L}$ are interchangeable under large-dimensional asymptotics, as the following proposition attests.

Proposition 1. Under the assumptions of Theorem 3.1 of Ledoit and Wolf (in press-a), both loss functions $\mathcal{L}_I(\widehat{\Sigma}, \Sigma)$ and $\mathcal{L}(\widehat{\Sigma}, \Sigma)$ converge almost surely to the same, nonstochastic limit as $T \to \infty$ with $N/T \to c \in (0, 1)$.

Proof. Let m be a random vector distributed according to the N-dimensional standard multivariate normal distribution, independently from $\widehat{\Sigma}$. Then $m/\|m\|$ is uniformly distributed on $\mathcal{U}_N$. This implies that

$$\mathcal{L}_I(\widehat{\Sigma}, \Sigma) = E\left[\mathcal{L}_E(\widehat{\Sigma}, \Sigma, m/\|m\|) \middle| \widehat{\Sigma} \right]. \qquad (4.7)$$

Proposition 1 then follows directly by injecting Lemma 1 of Ledoit and Péché (2011) into the proof of Theorem 3.1 of Ledoit and Wolf (in press-a). □

A traditional property of a loss function is that if one plugs the population parameters into the loss function, the value of the loss is zero. By convention, there is no loss from estimation if somehow you happen to know the truth; the loss should only be strictly positive if the estimator has error in it. The following theorem, which is equivalent in spirit to EC's Theorem 1, shows that the loss function in Definition 1 possesses the desired property.

Theorem 1. $\mathcal{L}\big(\widehat{\Sigma}, \Sigma\big) \geq 0$; and $\mathcal{L}\big(\widehat{\Sigma}, \Sigma\big) = 0 \iff \exists\, \gamma > 0$ s.t. $\widehat{\Sigma} = \gamma \Sigma$.

Proof. It is a standard result in linear algebra that, for any two symmetric positive-definite matrices $\widehat{\Sigma}$ and Σ, there exists an invertible matrix A such that

$$A' \Sigma A = \mathbb{I}_N \qquad (4.8)$$

$$A' \widehat{\Sigma} A = \Delta, \qquad (4.9)$$

where the superscript $'$ denotes the transpose of a matrix, $\mathbb{I}_N$ denotes the identity matrix of dimension N, and Δ denotes some diagonal matrix with strictly positive diagonal elements. Substituting Equations (4.8)–(4.9) into the definition (4.6) yields

$$\begin{aligned}\mathcal{L}(\widehat{\Sigma}, \Sigma) \geq 0 \iff & \mathrm{Tr}(\Delta^{-2} A'A) \times \mathrm{Tr}(A'A) \\ & \geq \left[\mathrm{Tr}(\Delta^{-1} A'A)\right]^2. \qquad (4.10)\end{aligned}$$

Denote the ith diagonal element of Δ by δ_i and the ith diagonal element of $A'A$ by α_i, for $i = 1, \ldots, N$. Note that $\forall i = 1, \ldots, N$, $\alpha_i > 0$, and $\delta_i > 0$. Then,

$$\mathcal{L}(\widehat{\Sigma}, \Sigma) \geq 0 \iff \left(\sum_{i=1}^{N} \frac{\alpha_i}{\delta_i^2} \right) \left(\sum_{i=1}^{N} \alpha_i \right) \geq \left(\sum_{i=1}^{N} \frac{\alpha_i}{\delta_i} \right)^2. \qquad (4.11)$$

Define $x_i := \sqrt{\alpha_i}/\delta_i$ and $y_i := \sqrt{\alpha_i}$, for $i = 1, \dots, N$. Then

$$\mathcal{L}(\widehat{\Sigma}, \Sigma) \geq 0 \iff \left(\sum_{i=1}^{N} x_i^2\right)\left(\sum_{i=1}^{N} y_i^2\right) \geq \left(\sum_{i=1}^{N} x_i y_i\right)^2. \tag{4.12}$$

On the right-hand side of the equivalency we recognize the Cauchy–Schwarz inequality; therefore, the loss function is always nonnegative. For equality to hold, we need the vectors $(x_1, \dots, x_N)'$ and $(y_1, \dots, y_N)'$ to be collinear, which implies that the diagonal matrix Δ is a scalar multiple of the identity. Inspecting Equations (4.8)–(4.9) reveals that this can only happen if the estimator $\widehat{\Sigma}$ is a scalar multiple of the true covariance matrix Σ. □

4.4 Adaptation for Conditional Covariance Matrices

Given that the loss function from Definition 1 is designed for unconditional covariance matrices, it must be adapted for conditional ones. Let $\widehat{H}_t$ denote a generic estimator of the true conditional covariance matrix H_t (for $t = 1, \dots, T$). Then the average loss is given by

$$\widehat{L} := \frac{1}{T}\sum_{t=1}^{T} \mathcal{L}(\widehat{H}_t, H_t). \tag{4.13}$$

This is the quantity that will be reported in all Monte Carlo simulations below.

5. MONTE CARLO SIMULATIONS

5.1 Base-Case Scenario

To carry out realistic Monte Carlo simulations, we estimate the unconditional population covariance matrix from the $N \in \{100, 500, 1000\}$ most liquid stocks in the CRSP database using 10 years of daily data from 2005 through 2014. This matrix will be taken as the "true" unconditional covariance matrix in all the simulations.

We then simulate a DCC model with parameters $\alpha = 0.05$ and $\beta = 0.93$ as in Pakel et al. (2014, Table 4). According to (2.5), the DCC variates are drawn from a multivariate standard normal distribution, that is,

$$\mathbf{r}_t = H_t^{1/2}\mathbf{z_t} \qquad \text{with} \qquad \mathbf{z}_t \overset{\text{iid}}{\sim} N(0, \mathbb{I}_N). \tag{5.1}$$

The univariate volatility dynamics are governed by GARCH(1,1) models with identical parameters $a_i = 0.05$ and $b_i = 0.90$ across all stocks $i = 1, \dots, N$.

For each simulation, we generate an $N \times T$ matrix of simulated returns, where the sample size is $T = 1250$. This sample size corresponds to approximately 5 years of daily data, an estimation window commonly used in practice. Thus, the concentration ratio for the broadest universe is $c := N/T = 0.8$, which is not negligible, and we can expect substantial shrinkage effects. Even though we could potentially accommodate the case $c > 1$, we choose a value of c less than one so that the sample correlation matrix can be used as a benchmark. In any case, having a universe of 1000 assets in which to invest is sufficiently broad for most portfolio managers.

5.2 Numerical Details

The nonlinear shrinkage estimator of the correlation targeting matrix is computed in Matlab using version 025 of the QuEST package. It contains the QuEST function itself, a routine to invert it numerically by using a nonlinear optimizer, and another routine to compute the optimal nonlinear shrinkage formula. The end user does not have to do anything except input the raw data and collect the shrinkage estimator. The linear shrinkage estimator of the correlation targeting matrix is also computed in Matlab using codes that implement the formulas of Ledoit and Wolf (2004a, 2004b). All codes are available from the University of Zurich faculty website of Michael Wolf.

The program to simulate the DCC model has been adapted from Kevin Sheppard's legacy UCSD GARCH Toolbox. The program to estimate the DCC model has been based on the successor to the UCSD GARCH Toolbox, which is the Oxford MFE Toolbox, also by Kevin Sheppard.

We have made two important modifications to the latter. The first one was to add the flexibility to shrink the eigenvalues of the sample correlation matrix used for correlation targeting as delineated in Section 3. The second one was to rewrite the function that computes the composite likelihood in a way that is improved in terms of speed and memory management; this rewriting required, among other things, translating the Matlab code into the C language, which in this case can generate substantial speed advantage, if used judiciously. The end result is that we can go to $N = 1000$ assets without running out of memory, and it takes less than 3 min to estimate the DCC model with nonlinear shrinkage, a speed gain by a factor of at least 20. (These numbers correspond to using Matlab R2016a on an Apple Mac Pro with a 3.5 GHz Intel Xeon E5 processor and 64GB of memory.)

5.3 List of Candidate Estimators

It is clearly beyond the scope of the present article to compare all the covariance matrix estimators in the literature. Therefore, we focus on the following representative set of six candidates:

- DCC-S: DCC with the correlation targeting matrix C estimated by the sample covariance matrix of the devolatilized returns.
- DCC-L1: DCC with C estimated by the linear shrinkage estimator of Ledoit and Wolf (2004b), whose shrinkage target is (a multiple of) the identity matrix.
- DCC-L2: DCC with C estimated by the linear shrinkage estimator of Ledoit and Wolf (2004a), whose shrinkage target is the equicorrelation matrix.
- DCC-NL: DCC with C estimated by the nonlinear shrinkage estimator; this is the main focus of our article.
- CCC-NL: The constant conditional correlation model using $Q_t \equiv C$, with C estimated by nonlinear shrinkage as in DCC-NL.
- RM-2006: The RiskMetrics 2006 methodology detailed by Zumbach (2007) and implemented in Kevin Sheppard's Matlab routine `riskmetrics2006`.

Table 1. Average loss for six different covariance matrix estimators in dimension $N \in \{100, 500, 1000\}$ with sample size $T = 1250$. The unit is 10^{-3}. The lowest number in each row appears in bold face

N	DCC-S	DCC-L1	DCC-L2	DCC-NL	CCC-NL	RM-2006
100	0.061	0.060	**0.054**	0.055	0.103	0.086
500	0.120	0.098	0.070	**0.066**	0.092	0.194
1000	0.421	0.147	0.087	**0.079**	0.099	0.602

5.4 Results

The results averaged across $10^5/N$ Monte Carlo simulations are presented in Table 1. (Note that in large dimensions, such as $N = 1000$, the results are extremely consistent from one simulation to the next, so there is no need to go to thousands of simulations.)

DCC-NL dominates across the board in large dimensions, but for $N = 100$ linear shrinkage toward the equicorrelation matrix does slightly better. This is probably because we calibrated our population unconditional correlation matrix to actual stock returns data, for which constant correlation happens to be a parsimonious yet accurate model. As the cross-section increases, DCC-NL has more information to pick up on, which enables it to first beat DCC-L2 and to then widen the gap.

An intuitive way to quantify the improvement is to compute the percentage relative improvement in average loss (PRIAL). For example, the PRIAL of DCC-NL with respect to DCC-S is defined as

$$100 \times \left\{ 1 - \frac{\mathrm{E}\left[\widehat{L}^{\text{DCC-NL}}\right]}{\mathrm{E}\left[\widehat{L}^{\text{DCC-S}}\right]} \right\} \%, \tag{5.2}$$

where the loss $\widehat{L}$ for each estimator is calculated as per Equation (4.13) and the expectation is taken across Monte Carlo simulations. By construction, the PRIAL (5.2) of the true conditional covariance matrix with respect to DCC-S is 100%, which is the maximum attainable; on the other hand, 0% means no improvement at all.

The interpretation is that, as long as we do not know the true conditional covariance matrix, estimation error will cause excess out-of-sample portfolio variance, and we want to eliminate as much of it as possible. The PRIAL of DCC-NL with respect to a given estimator (such as DCC-S) says what percentage of the average loss of the given estimator we have managed to eliminate by the use of DCC-NL. Table 2 presents the PRIAL of DCC-NL with respect to each of the other five estimators.

Overall, the magnitude of the improvement from using DCC-NL is substantial. The one exception is DCC-L2 but, once again, this is easily explained by the fact that the equicorrelation shrinkage target is a very parsimonious model that fits closely the population correlation matrix that we happened to choose. The improvement from DCC-NL generally increases in the number of assets N; the one exception is with respect to CCC-NL, but this is likely to be because CCC-NL also benefits from nonlinear shrinkage estimation of the matrix C.

Table 2. PRIAL of DCC-NL with respect to each of the five other candidate estimators. The setup is the same as in Table 1

N	DCC-S	DCC-L1	DCC-L2	CCC-NL	RM-2006
100	10.6%	8.3%	– 0.5%	47.1%	36.2%
500	45.4%	33.1%	6.8%	28.8%	66.2%
1000	81.2%	46.0%	9.3%	20.3%	86.8%

Table 3. Estimated parameter α from correlation matrix dynamics. The panels show the mean and (in parentheses) the standard deviation across Monte Carlo simulations. The setup is the same as in Table 1

N	DCC-S	DCC-L1	DCC-L2	DCC-NL
100	0.0485	0.0485	0.0485	0.0485
	(0.0026)	(0.0026)	(0.0026)	(0.0026)
500	0.0489	0.0490	0.0490	0.0490
	(0.0019)	(0.0019)	(0.0019)	(0.0019)
1000	0.0490	0.0490	0.0491	0.0491
	(0.0023)	(0.0023)	(0.0023)	(0.0023)

5.5 Dynamic Correlation Parameters

There are two possible channels through which the improvement could be happening: either directly because nonlinear shrinkage just gives a better estimate of the correlation targeting matrix or indirectly because it yields better estimates of the dynamic correlation parameters α and β in Equation (2.2). To distinguish between these two hypotheses, we report the mean and standard deviations (across Monte Carlo simulations) of the estimates of the parameters α and β produced by the various methods; remember that the true parameters are $\alpha = 0.05$ and $\beta = 0.93$ as per Section 5.1. Table 3 presents the results for the α estimates and Table 4 for the β estimates. By construction, this exercise cannot be carried out for the non-DCC estimators, namely, CCC-NL and RM-2006, so they are not included.

One can see that there is no pattern. Therefore, the explanation for the outperformance of DCC-NL lies not in the better estimation of the dynamic correlation parameters but in the better estimation of the correlation targeting matrix C.

Remark 5. An alternative way of summarizing the same set of results would be to say that the composite likelihood method manages to obtain accurate estimates of the α and β parameters even when the correlation targeting matrix is the sample correlation matrix, with all the associated difficulties in large

Table 4. Estimated parameter β from correlation matrix dynamics. The panels show the mean and (in parentheses) the standard deviation across Monte Carlo simulations. The setup is the same as in Table 1

N	DCC-S	DCC-L1	DCC-L2	DCC-NL
100	0.9301	0.9301	0.9301	0.9301
	(0.0035)	(0.0035)	(0.0035)	(0.0035)
500	0.9296	0.9296	0.9296	0.9297
	(0.0026)	(0.0026)	(0.0026)	(0.0026)
1000	0.9292	0.9293	0.9293	0.9293
	(0.0030)	(0.0030)	(0.0030)	(0.0030)

Table 5. PRIAL of DCC-NL with respect to each of the five other candidate estimators. DCC variates are generated according to the "Student" t-distribution with 5 degrees of freedom, rescaled to have marginal variance one

N	DCC-S	DCC-L1	DCC-L2	CCC-NL	RM-2006
100	10.0%	7.8%	− 0.2%	44.4%	34.2%
500	44.5%	32.6%	6.9%	28.2%	65.4%
1000	80.8%	45.6%	9.3%	19.5%	86.5%

dimensions. This is an advantage of composite likelihood, as we would expect the maximum likelihood estimators of α and β to be hampered in the presence of a rank-deficient, or at least ill-conditioned, correlation targeting matrix.

5.6 Fat Tails

Given that financial data are generally leptokurtotic, it is worth checking whether the DCC-NL estimator is robust against fat-tailedness. To this end, instead of drawing the variates $\mathbf{z}_t$ in (5.1) from the multivariate normal distribution, we draw them from the Cartesian product of the "Student" t-distribution with 5 degrees of freedom, rescaled to have marginal variance one. (This means that the entries of $\mathbf{z}_t$ are independent and identically distributed according to the t-distribution with 5 degrees of freedom, rescaled to have variance one.) All other configurations are as per the base-case scenario presented in Section 5.1. The results are presented in Table 5.

Overall, the changes with respect to Table 2 are minimal, and there is no discernible pattern. Therefore, the conclusions reached in Section 5.4 are confirmed.

5.7 Misspecified Model

Although we know that multivariate GARCH effects are prevalent in financial data, it is possible that the "true" data-generating process differs from the DCC model. In such a case, it would be helpful to know that DCC-NL can still add value by producing conditional covariance matrix estimators that are robust against model misspecification and continue to be relatively accurate.

To this end, we simulate data from the BEKK model presented in Engle and Kroner (1995) with parameter values $\ddot{\alpha} = 0.05$ and $\ddot{\beta} = 0.9$; see Equation (7.1). In practice, we use the Matlab function `scalar_vt_vech_simulate` from Kevin Sheppard's Oxford MFE Toolbox. All the other configurations are as per the base-case scenario. The results are presented in Table 6.

Table 6. PRIAL of DCC-NL with respect to each of the five other candidate estimators. Data are generated from the BEKK model

N	DCC-S	DCC-L1	DCC-L2	CCC-NL	RM-2006
100	20.3%	17.8%	5.1%	72.5%	72.1%
500	54.3%	47.1%	19.3%	41.0%	79.7%
1000	85.4%	65.2%	24.1%	26.8%	91.3%

Compared with Table 2, the PRIAL of DCC-NL with respect to each of its five competitors always goes up. The differences in PRIALs (i.e., the entry in Table 6 minus the corresponding entry in Table 2) range from a low of +4.5% to a high of +35.9%, with an average additive boost of +13.1%. Thus, an additional advantage of DCC-NL is that it appears more robust against model misspecification than other comparable estimators.

6. EMPIRICAL RESULTS

The goal of this section is to examine the out-of-sample properties of Markowitz portfolios based on our newly suggested covariance matrix estimator. There are a myriad of popular investment strategies by now and it is not our goal to compare to an extensive list of them. The only focus of this section is to compare nonlinear DCC to a set of seven other representative portfolio selection strategies.

For compactness of notation, we do not use the subscript T in denoting the covariance matrix itself, an estimator of the covariance matrix, or a return-predictive signal that proxies for the vector of expected returns.

6.1 Data and General Portfolio-Formation Rules

We download daily data from the Center for Research in Security Prices (CRSP) starting in 01/01/1980 and ending in 12/31/2015. For simplicity, we adopt the common convention that 21 consecutive trading days constitute 1 "month." The out-of-sample period ranges from 01/08/1986 through 12/31/2015, resulting in a total of 360 "months" (or 7560 days). All portfolios are updated "monthly." ("Monthly" updating is common practice to avoid an unreasonable amount of turnover and thus transaction costs. During a "month," from one day to the next, we hold number of shares fixed rather than portfolio weights; in this way, there are no transactions at all during a "month.")

We denote the investment dates by $h = 1, \ldots, 360$. At any investment date h, a covariance matrix is estimated using the most recent $T = 1250$ daily returns, which roughly corresponds to using 5 years of past data.

We consider the following portfolio sizes: $N \in \{100, 500, 1000\}$. For a given combination (h, N), the investment universe is obtained as follows. We find the set of stocks that have a complete return history over the most recent $T = 1250$ days as well as a complete return "future" over the next 21 days. (The latter, forward-looking restriction is not a feasible one in real life but is commonly applied in the related finance literature on the out-of-sample evaluation of portfolios.) We then look for possible pairs of highly correlated stocks, that is, pairs of stocks that return with a sample correlation exceeding 0.95 over the past 1250 days. With such pairs, if they should exist, we remove the stock with the lower volume of the two on investment date h. (The reason is that we do not want to include highly similar stocks; in the early years, there are no such pairs; in the most recent years, there are never more than three such pairs.) Of the remaining set of stocks, we then pick the largest N stocks (as measured by their market capitalization on investment date h) as our investment universe. In this way, the investment universe changes slowly from one investment date to the next.

6.2 Global Minimum Variance Portfolio

We consider the problem of estimating the global minimum variance (GMV) portfolio, in the absence of short-sales constraints. The problem is formulated as

$$\min_{w} w' H_t w \tag{6.1}$$

$$\text{subject to} \quad w'\mathbf{1} = 1, \tag{6.2}$$

where $\mathbf{1}$ denotes a vector of ones of dimension $N \times 1$. It has the analytical solution

$$w = \frac{H_t^{-1}\mathbf{1}}{\mathbf{1}' H_t^{-1}\mathbf{1}}. \tag{6.3}$$

The natural strategy in practice is to replace the unknown H_t by an estimator $\widehat{H}_t$ in formula (6.3), yielding a feasible portfolio

$$\widehat{w} := \frac{\widehat{H}_t^{-1}\mathbf{1}}{\mathbf{1}' \widehat{H}_t^{-1}\mathbf{1}}. \tag{6.4}$$

Estimating the GMV portfolio is a "clean" problem in terms of evaluating the quality of a covariance matrix estimator, since it abstracts from having to estimate the vector of expected returns at the same time. In addition, researchers have established that estimated GMV portfolios have desirable out-of-sample properties not only in terms of risk but also in terms of reward-to-risk, that is, in terms of the information ratio, for example, see Haugen and Baker (1991), Jagannathan and Ma (2003), and Nielsen and Aylursubramanian (2008). As a result, such portfolios have become an addition to the large array of products sold by the mutual-fund industry.

The following eight portfolios are included in the study.

- $1/N$: the equal-weighted portfolio. This portfolio is a standard benchmark and has been promoted by DeMiguel, Garlappi, and Uppal (2009), among others.
- DCC-S: the portfolio (6.4), where the estimator $\widehat{H}_t$ is obtained from DCC based on the sample correlation matrix.
- DCC-L1: the portfolio (6.4), where the estimator $\widehat{H}_t$ is obtained from DCC based on the linear shrinkage of Ledoit and Wolf (2004b) with shrinkage target given by (a multiple of) the identity matrix.
- DCC-L2: the portfolio (6.4), where the estimator $\widehat{H}_t$ is obtained from DCC based on the linear shrinkage of Ledoit and Wolf (2004a) with shrinkage target given by the equicorrelation matrix.
- DCC-NL: the portfolio (6.4), where the estimator $\widehat{H}_t$ is obtained from DCC based on nonlinear shrinkage.
- NL-DCC: the portfolio (6.4), where the estimator $\widehat{H}_t$ is obtained by post-processing the DCC-S estimator of H_t with nonlinear shrinkage.
- CCC-NL: the portfolio (6.4), where the estimator $\widehat{H}_t$ is obtained from CCC based on nonlinear shrinkage. This means that instead of the DCC dynamics of model (2.2), one considers the static constant-conditional-correlation model $Q_t \equiv C$; in practice, the estimator of C is obtained by nonlinear shrinkage just as in DCC-NL.
- RM-2006: the portfolio (6.4), where the estimator $\widehat{H}_t$ is obtained from the RiskMetrics 2006 methodology; see Zumbach (2007). (We use the Matlab routine `riskmetrics2006` provided by Kevin Sheppard.)

We report the following three out-of-sample performance measures for each scenario. (All of them are annualized and in percent for ease of interpretation.)

- AV: We compute the average of the 7560 out-of-sample log returns and then multiply by 252 to annualize.
- SD: We compute the standard deviation of the 7560 out-of-sample log returns and then multiply by $\sqrt{252}$ to annualize.
- IR: We compute the (annualized) information ratio as the ratio AV/SD.

Our stance is that in the context of the GMV portfolio, the most important performance measure is the out-of-sample standard deviation, SD. The true (but unfeasible) GMV portfolio is given by (6.3). It is designed to minimize the variance (and thus the standard deviation) rather than to maximize the expected return or the information ratio. Therefore, any portfolio that implements the GMV portfolio should be primarily evaluated by how successfully it achieves this goal. A high out-of-sample average return, AV, and a high out-of-sample information ratio, IR, are naturally also desirable, but should be considered of secondary importance from the point of view of evaluating the quality of a covariance matrix estimator.

We also consider the question of whether DCC-NL delivers a lower out-of-sample standard deviation than DCC-S at a level that is statistically significant. For a given universe size N, a two-sided p-value for the null hypothesis of equal standard deviations is obtained by the prewhitened HAC_{PW} method described by Ledoit and Wolf (2011, sec. 3.1).

The results are presented in Table 7 and can be summarized as follows; unless stated otherwise, the findings are with respect to the out-of-sample standard deviation as performance measure.

- All other portfolios consistently outperform $1/N$ by a wide margin. In addition, all DCC and CCC portfolios also outperform RM-2006 by substantial margin.
- Among the DCC portfolios, there is a consistent ranking across all portfolio sizes N: DCC-NL, NL-DCC, DCC-L2, DCC-L1, and DCC-S.
- The second-best portfolio consistently is CCC-NL, that is, its performance ranks between DCC-NL and NL-DCC.
- The outperformance of DCC-NL over DCC-S is always statistically significant and it is also economically meaningful for $N = 500$, 1000.
- DeMiguel, Garlappi, and Uppal (2009) claimed that it is difficult to outperform $1/N$ in terms of the out-of-sample Sharpe ratio with "sophisticated" portfolios (i.e., with Markowitz portfolios that estimate input parameters). It can be seen that all other portfolios consistently outperform $1/N$ in terms of the out-of-sample information ratio, which translates into outperformance in terms of the out-of-sample Sharpe ratio. For $N = 100$, CCC-NL is best overall, whereas for $N = 500$, 1000, DCC-NL is best overall.

Table 7. Annualized performance measures (in percent) for various estimators of the GMV portfolio

	1/N	DCC-S	DCC-L1	DCC-L2	DCC-NL	NL-DCC	CCC-NL	RM-2006
Period: 01/08/1986–12/31/2015								
$N = 100$								
AV	12.10	9.92	9.91	9.91	9.95	10.24	11.37	8.41
SD	21.56	13.36	13.33	13.27	**13.17*****	13.22	13.20	14.69
IR	0.56	0.74	0.74	0.74	0.76	0.77	0.86	0.57
$N = 500$								
AV	13.46	13.94	13.88	13.76	13.38	10.87	13.25	11.26
SD	19.53	10.57	10.40	10.16	**9.64*****	9.93	9.74	12.60
IR	0.69	1.32	1.33	1.35	1.39	1.09	1.36	0.89
$N = 1000$								
AV	14.21	11.77	12.15	11.88	12.17	11.12	11.72	11.37
SD	19.04	10.59	9.14	8.55	**8.02*****	8.79	8.13	14.86
IR	0.75	1.11	1.33	1.38	1.52	1.27	1.44	0.77

NOTES: AV stands for average; SD stands for standard deviation; and IR stands for information ratio. All measures are based on 7560 daily out-of-sample returns from 01/08/1986 through 12/31/2015. In the rows labeled SD, the lowest number appears in bold face. In the columns labeled DCC-S and DCC-NL, significant outperformance of one of the two portfolios over the other in terms of SD is denoted by asterisks: ***denotes significance at the 0.01 level; ** denotes significance at the 0.05 level; and * denotes significance at the 0.1 level.

6.3 Markowitz Portfolio with Momentum Signal

We now turn attention to a "full" Markowitz portfolio with a signal.

By now a large number of variables have been documented that can be used to construct a signal in practice. For simplicity and reproducibility, we use the well-known momentum factor (or simply momentum for short) of Jegadeesh and Titman (1993). For a given investment period h and a given stock, the momentum is the geometric average of the previous 252 returns on the stock but excluding the most recent 21 returns; in other words, one uses the geometric average over the previous "year" but excluding the previous "month." Collecting the individual momentums of all the N stocks contained in the portfolio universe yields the return-predictive signal m.

In the absence of short-sales constraints, the investment problem is formulated as

$$\min_{w} w' H_t w \tag{6.5}$$

$$\text{subject to} \quad w'm = b, \text{ and} \tag{6.6}$$

$$w'\mathbf{1} = 1, \tag{6.7}$$

where b is a selected target expected return. The problem has the analytical solution

$$w = c_1 H_t^{-1}\mathbf{1} + c_2 H_t^{-1} m, \tag{6.8}$$

$$\text{where } c_1 := \frac{C - bB}{AC - B^2} \quad \text{and} \quad c_2 := \frac{bA - B}{AC - B^2}, \tag{6.9}$$

$$\text{with } A := \mathbf{1}' H_t^{-1}\mathbf{1}, \quad B := \mathbf{1}' H_t^{-1} b, \quad \text{and}$$

$$C := m' H_t^{-1} m. \tag{6.10}$$

The natural strategy in practice is to replace the unknown H_t by an estimator $\widehat{H}_t$ in formulas (6.8)–(6.10), yielding a feasible portfolio

$$\widehat{w} := c_1 \widehat{H}_t^{-1}\mathbf{1} + c_2 \widehat{H}_t^{-1} m, \tag{6.11}$$

$$\text{where } c_1 := \frac{C - bB}{AC - B^2} \quad \text{and} \quad c_2 := \frac{bA - B}{AC - B^2}, \tag{6.12}$$

$$\text{with } A := \mathbf{1}' \widehat{H}_t^{-1}\mathbf{1}, \quad B := \mathbf{1}' \widehat{H}_t^{-1} b, \quad \text{and}$$

$$C := m' \widehat{H}_t^{-1} m. \tag{6.13}$$

The following eight portfolios are included in the study.

- EW-TQ: The equal-weighted portfolio of the top-quintile stocks according to momentum m. This strategy does not make use of the momentum signal beyond sorting of the stocks in quintiles.
  The value of the target expected return b for the remaining four portfolios below is then given by the arithmetic average of the momentums of the stocks included in this portfolio (i.e., the expected return of EW-TQ according to the signal m).
- DCC-S: the portfolio (6.8)–(6.10), where the estimator $\widehat{H}_t$ is obtained from DCC based on the sample correlation matrix.
- DCC-L1: the portfolio (6.8)–(6.10), where the estimator $\widehat{H}_t$ is obtained from DCC based on the linear shrinkage of Ledoit and Wolf (2004b) with shrinkage target given by (a multiple of) the identity matrix.
- DCC-L2: the portfolio (6.8)–(6.10), where the estimator $\widehat{H}_t$ is obtained from DCC based on the linear shrinkage of Ledoit and Wolf (2004a) with shrinkage target given by the equicorrelation matrix.
- DCC-NL: the portfolio (6.8)–(6.10), where the estimator $\widehat{H}_t$ is obtained from DCC based on nonlinear shrinkage.
- NL-DCC: the portfolio (6.8)–(6.10), where the estimator $\widehat{H}_t$ is obtained by post-processing the DCC-S estimator of H_t with nonlinear shrinkage.
- CCC-NL: the portfolio (6.8)–(6.10), where the estimator $\widehat{H}_t$ is obtained from CCC based on nonlinear shrinkage. This means that instead of the DCC dynamics of

model (2.2), one considers the static constant-conditional-correlation model $Q_t \equiv C$; in practice, the estimator of C is obtained by nonlinear shrinkage just as in DCC-NL. Note that CCC-NL can be interpreted as a robustification of the CCC model of Bollerslev (1990) against large dimensions.

- RM-2006: the portfolio (6.8)–(6.10), where the estimator $\widehat{H}_t$ is obtained from the RiskMetrics 2006 methodology; see Zumbach (2007). (We use the Matlab routine `riskmetrics2006` provided by Kevin Sheppard.)

Our stance is that in the context of a "full" Markowitz portfolio, the most important performance measure is the out-of-sample information ratio, IR. In the "ideal" investment problem (6.8)–(6.10), minimizing the variance (for a fixed target expected return b) is equivalent to maximizing the information ratio (for a fixed target expected return b). In practice, because of estimation error in the signal, the various strategies do not have the same expected return and, thus, focusing on the out-of-sample standard deviation is inappropriate.

We also consider the question whether DCC-NL delivers a higher out-of-sample information ratio than DCC-S at a level that is statistically significant. For a given universe size N, a two-sided p-value for the null hypothesis of equal information ratios is obtained by the prewhitened HAC_{PW} method described in Ledoit and Wolf (2008, sec. 3.1).

The results are presented in Table 8 and can be summarized as follows; unless stated otherwise, the findings are with respect to the out-of-sample information ratio as performance measure.

- With the exception of RM-2006, all other portfolios consistently outperform EW-TQ by a wide margin.
- Among the DCC portfolios, DCC-NL is consistently best followed by DCC-L2. NL-DCC does well for $N = 1000$ but badly for $N = 100, 500$.
- The CCC-NL portfolio is second best for $N = 500, 1000$ but does badly for $N = 100$.
- The outperformance of DCC-NL over DCC-S is statistically significant and also economically meaningful for $N = 500, 1000$.
- The performance of RM-2006 is disappointing. It is also worse than DCC-S and for $N = 1000$ is also worse than EW-TQ.
- DeMiguel, Garlappi, and Uppal (2009) claimed that it is difficult to outperform $1/N$ in terms of the out-of-sample Sharpe ratio with "sophisticated" portfolios (i.e., with Markowitz portfolios that estimate input parameters). Comparing Table 8 with Table 7, it can be seen that all DCC variants of the "full" Markowitz portfolio consistently outperform $1/N$ in terms of the out-of-sample information ratio, which translates into outperformance in terms of the out-of-sample Sharpe ratio.
  Even though momentum is not a very powerful return-predictive signal, the differences can be enormous. For example, for $N = 1000$, the information ratio of $1/N$ is only 0.75 whereas the information ratio of DCC-NL is 1.62, more than twice as large.
- Engle and Colacito (2006) argued for the use of the out-of-sample standard deviation, SD, as a performance measure also in the context of a "full" Markowitz portfolio. Also for this alternative performance measure, all DCC variants consistently outperform EW-TQ by a wide margin. Furthermore, DCC-NL consistently has the smallest out-of-sample standard deviation and its outperformance over DCC-S is always statistically significant.

7. THE BEKK-NL MODEL

Although the present article focuses on the DCC model, which works at the level of correlations and devolatilized

Table 8. Annualized performance measures (in percent) for various estimators of the Markowitz portfolio with momentum signal

	Period: 01/08/1986–12/31/2015							
	EW-TQ	DCC-S	DCC-L1	DCC-L2	DCC-NL	NL-DCC	CCC-NL	RM-2006
				$N = 100$				
AV	17.13	15.79	15.79	15.77	15.77	15.09	15.65	15.96
SD	28.43	17.05	17.03	16.99	16.90***	16.88	17.02	18.87
IR	0.60	0.93	0.93	0.93	**0.93**	0.89	0.92	0.85
				$N = 500$				
AV	17.15	16.60	16.66	16.55	16.78	14.53	16.13	16.50
SD	24.42	12.36	12.16	11.87	11.31***	11.87	11.55	16.14
IR	0.70	1.34	1.37	1.40	**1.48****	1.22	1.40	1.02
				$N = 1000$				
AV	17.35	12.78	13.96	13.96	14.92	14.08	14.03	15.55
SD	22.89	13.07	10.76	9.85	9.20***	10.31	9.45	29.29
IR	0.76	0.98	1.30	1.42	**1.62*****	1.36	1.48	0.53

NOTES: AV stands for average; SD stands for standard deviation; and IR stands for information ratio. All measures are based on 7560 daily out-of-sample returns from 01/08/1986 until 12/31/2015. In the rows labeled IR, the largest number appears in bold face. In the columns labeled DCC-S and DCC-NL, significant outperformance of one of the two portfolios over the other in terms of IR is denoted by asterisks: ***denotes significance at the 0.01 level; **denotes significance at the 0.05 level; and * denotes significance at the 0.1 level.

returns, an alternative approach involves the BEKK model presented by Engle and Kroner (1995), which works in an analogous way at the level of covariances and straight returns. The most scalable version of the BEKK model, and the one most similar to the particular version of DCC presented in Section 2, is the one with scalar dynamics and covariance targeting. Using the notation of Section 2.1, Equations (2.2)–(2.4) are replaced with

$$H_t = (1 - \ddot{\alpha} - \ddot{\beta})\Sigma + \ddot{\alpha}\,\mathbf{r}_{t-1}\mathbf{r}'_{t-1} + \ddot{\beta}\,H_{t-1}, \qquad (7.1)$$

where Σ is the unconditional covariance matrix and $(\ddot{\alpha}, \ddot{\beta})$ are BEKK dynamic parameters analogous to (α, β), but in covariance space instead of correlation space.

BEKK is simpler compared to DCC, but it does not handle well investment universes that include correlated assets with volatilities of different magnitudes. For example, if we have gold and short-term government bonds, both of which can be considered “safe havens” in times of financial crises, volatilities vary by one or two orders of magnitudes, so putting them on the same footing (as BEKK does) may not be the best modeling strategy. To give another example, if we replace one asset, say the S&P 500 index, with a 2-to-1 leveraged version of itself (and such ETFs do exist), then the set of investment opportunities remains the same and DCC adapts automatically, whereas any portfolio allocation based on BEKK will be impacted. In other words, BEKK will be favored by a homogenous, unlevered investment universe.

Estimation of the BEKK model in large dimensions using the composite likelihood method is described by Pakel et al. (2014, Example 2.1 and sec. 3). The BEKK-NL model is obtained by inserting the nonlinear shrinkage estimator of the covariance matrix developed by Ledoit and Wolf (2012, in press-a) in place of the “true” covariance targeting matrix Σ, which is unavailable in practice.

8. CONCLUSION

This article demonstrates that there is a “division of labor” between composite likelihood and nonlinear shrinkage in the estimation of a Dynamic Conditional Correlation (DCC) model: The former takes care of the dynamic correlation parameters (time series) whereas the latter takes care of the correlation targeting matrix (cross-section). Their actions complement each other. Together, they enable DCC to conquer large dimensions on the order of a thousand, which are frequently encountered in modern portfolio theory and risk management. We call the resulting estimator DCC-NL, which stands for DCC based on nonlinear shrinkage. The attractive performance of the DCC-NL estimator has been established both by simulation studies and by backtesting on real-life stock return data; we thus recommend this estimator as the new DCC standard in large dimensions.

SUPPLEMENTARY MATERIAL

The supplementary material provides additional details and motivation regarding the nonlinear shrinkage method.

ACKNOWLEDGMENTS

The authors thank Zhao Zhao (Department of Economics, Huazhong University of Science and Technology, China) for providing research assistance. The authors also thank Kevin Sheppard for having made publicly available the UCSD GARCH Toolbox as well as its successor, the Oxford MFE Toolbox. Last but not least, the authors thank the associate editor and two anonymous referees for helpful comments that have improved the exposition of the article. Any errors are of authors.

[Received July 2016. Revised June 2017.]

REFERENCES

Aielli, G. P. (2013), “Dynamic Conditional Correlation: On Properties and Estimation,” *Journal of Business & Economic Statistics*, 31, 282–299. [2]

Capon, J. (1969), “High-Resolution Frequency-Wavenumber Spectrum Analysis,” *Proceedings of the IEEE*, 57, 1408–1418. [4]

El Karoui, N. (2008), “Spectrum Estimation for Large Dimensional Covariance Matrices Using Random Matrix Theory,” *Annals of Statistics*, 36, 2757–2790. [3]

Engle, R. F. (1982), “Autoregressive Conditional Heteroskedasticity With Estimates of the Variance of United Kingdom Inflation,” *Econometrica*, 50, 987–1007. [1]

——— (2002), “Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models,” *Journal of Business & Economic Statistics*, 20, 339–350. [1]

——— (2009), *Anticipating Correlations: A New Paradigm for Risk Management*, Princeton, NJ: Princeton University Press. [1,2]

Engle, R. F., and Colacito, R. (2006), “Testing and Valuing Dynamic Correlations for Asset Allocation,” *Journal of Business and Economic Statistics*, 24, 238–253. [4]

Engle, R. F., and Kroner, K. (1995), “Multivariate Simultaneous GARCH,” *Econometric Theory*, 11, 122–150. [1,8,12]

Engle, R. F., and Mezrich, J. (1996), “GARCH for Groups,” *Risk*, 9, 36–40. [2]

Hafner, C. M., and Reznikova, O. (2012), “On the Estimation of Dynamic Conditional Correlation Models,” *Computational Statistics & Data Analysis*, 56, 3533–3545. [1,4]

Hamilton, J. D. (1994), *Time Series Analysis*, Princeton, NJ: Princeton University Press. [2]

Haugen, R. A., and Baker, N. L. (1991), “The Efficient Market Inefficiency of Capitalization Weighted Stock Portfolios,” *Journal of Portfolio Management*, 17, 35–40. [9]

IPCC (2007), “Climate Change 2007: The Scientific Basis,” in *Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, eds. S. Solomon, D. Qin, M. Manning, M. Marquis, K. Averyt, M. M. Tignor, H. L. Miller, and Z. Chen, Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press, pp. 996. [4]

Jagannathan, R., and Ma, T. (2003), “Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps,” *Journal of Finance*, 54, 1651–1684. [9]

Jegadeesh, N., and Titman, S. (1993), “Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency,” *Journal of Finance*, 48, 65–91. [10]

Ledoit, O., and Péché, S. (2011), “Eigenvectors of Some Large Sample Covariance Matrix Ensembles,” *Probability Theory and Related Fields*, 150, 233–264. [3,5]

Ledoit, O., and Wolf, M. (2004a), “Honey, I Shrunk the Sample Covariance Matrix,” *Journal of Portfolio Management*, 30, 110–119. [4,6]

——— (2004b), “A Well-Conditioned Estimator for Large-Dimensional Covariance Matrices,” *Journal of Multivariate Analysis*, 88, 365–411. [4,6]

——— (2008), “Robust Performance Hypothesis Testing With the Sharpe Ratio,” *Journal of Empirical Finance*, 15, 850–859. [11]

——— (2011, September), “Robust Performance Hypothesis Testing With the Variance,” *Wilmott Magazine*, pp. 86–89. [9]

——— (2012), “Nonlinear Shrinkage Estimation of Large-Dimensional Covariance Matrices,” *Annals of Statistics*, 40, 1024–1060. [1,3,4,12]

——— (2015), “Spectrum Estimation: A Unified Framework for Covariance Matrix Estimation and PCA in Large Dimensions,” *Journal of Multivariate Analysis*, 139, 360–384. [3,4]

——— (2017), “Nonlinear Shrinkage of the Covariance Matrix for Portfolio Selection: Markowitz Meets Goldilocks.” *Review of Financial Studies*, 30, 4349–4388. [4,5,12]

Marčenko, V. A., and Pastur, L. A. (1967), “Distribution of Eigenvalues for Some Sets of Random Matrices,” *Sbornik: Mathematics*, 1, 457–483. [3,5]

Mestre, X. (2008), "On the Asymptotic Behavior of the Sample Estimates of Eigenvalues and Eigenvectors of Covariance Matrices," *IEEE Transactions on Signal Processing*, 56, 5353–5368. [3]

Michaud, R. (1989), "The Markowitz Optimization Enigma: Is Optimized Optimal?" *Financial Analysts Journal*, 45, 31–42. [3]

Nielsen, F., and Aylursubramanian, R. (2008), "Far From the Madding Crowd — Volatility Efficient Indices," Research Insights, New York: MSCI Barra. [9]

Pakel, C., Shephard, N., Sheppard, K., and Engle, R. F. (2014), "Fitting Vast Dimensional Time-Varying Covariance Models," Technical Report, NYU Working Paper FIN-08-009. Available at SSRN: https://ssrn.com/abstract=1354497. [1,2,6,12]

Silverstein, J. W., and Bai, Z. D. (1995), "On the Empirical Distribution of Eigenvalues of a Class of Large-Dimensional Random Matrices," *Journal of Multivariate Analysis*, 54, 175–192. [5]