

Article

Risk-Based Portfolios with Large Dynamic Covariance Matrices

Kei Nakagawa ^{1,2,*}, Mitsuyoshi Imamura ^{1,3} and Kenichi Yoshida ²¹ Nomura Asset Management Co., Ltd. 1-12-1 Nihonbashi, Chuo-ku, 103-0027 Tokyo, Japan² Graduate School of Business Sciences, University of Tsukuba, 3-29-1 Otsuka, Bunkyo-ku, 112-0012 Tokyo, Japan; yoshida.kenichi.ka@u.tsukuba.ac.jp³ Department of Risk Engineering, University of Tsukuba, 1-1-1 Tennodai, 305-8577 Tsukuba, Japan; ic140tg528@gmail.com

* Correspondence: kei.nak.0315@gmail.com; Tel.: +81-3-4376-6049

Received: 15 February 2018; Accepted: 10 May 2018; Published: 14 May 2018



Abstract: In the field of portfolio management, practitioners are focusing increasingly on risk-based portfolios rather than on mean-variance portfolios. Risk-based portfolios are constructed based solely on covariance matrices, and include methods such as minimum variance (MV), risk parity (RP), and maximum diversification (MD). It is well known that the performance of a mean-variance portfolio depends on the accuracy of the estimations of the inputs. However, no studies have examined the relationship between the performance of risk-based portfolios and the estimated accuracy of covariance matrices. In this research, we compare the performance of risk-based portfolios for several estimation methods of covariance matrices in the Japanese stock market. In addition, we propose a highly accurate estimation method called cDCC-NLS, which incorporates nonlinear shrinkage into the cDCC-GARCH model. The results confirm that (1) the cDCC-NLS method shows the best estimation accuracy, (2) the RP and MD do not depend on the estimation accuracy of the covariance matrix, and (3) the MV does depend on the estimation accuracy of the covariance matrix.

Keywords: (c)DCC-GARCH; nonlinear shrinkage; minimum variance; risk parity; maximum diversification

JEL Classification: C13; C32; G11

1. Introduction

In the field of portfolio management, practitioners are focusing increasingly on risk-based portfolios, which use methods such as the minimum variance (MV), risk parity (RP), and maximum diversification (MD) to construct portfolios based solely on covariance matrices. Risk-based portfolios outperform traditional mean-variance portfolios (Markowitz 1952) and market capitalization-weighted portfolio, as shown in various empirical studies on stock selection (De Carvalho et al. 2012) and asset allocation (Chaves et al. 2011). As a result, the number of investment trusts and pension funds designed based on this concept are increasing.

It is well known that the performance of a mean-variance portfolio depends on the accuracy of the estimations of the inputs (Michaud 1989). In contrast, risk-based portfolios do not need to estimate expected returns, which are notoriously unstable and difficult to predict (Merton 1980). This is one of the main motivations for using a risk-based portfolio rather than a mean-variance portfolio. Recently, Ardia et al. (2017) estimated the accuracy of covariance matrices in terms of the weights of several risk-based portfolios. However, to the best of our knowledge, no studies have examined the relation between the performance of risk-based portfolios and the estimated accuracy of covariance matrices.

In this research, we compare the performances of risk-based portfolios for several estimation methods of covariance matrices. In addition, we propose a highly accurate estimation method that combines nonlinear shrinkage and the cDCC-GARCH model, called the cDCC-NLS method.

To estimate the covariance matrix in financial time series, it is necessary consider two important aspects: the cross section and the time series. With regard to the cross section, we have the difficulty of correcting the biases of the sample covariance matrix eigenvalues in a large number of time series. A popular method used to achieve this is the linear shrinkage method (LS) and its successor, the nonlinear shrinkage method (NLS), proposed by Ledoit and Wolf (2012). With regard to the time series aspect, we have to account for volatility clustering and time-varying correlations. To introduce a time-varying volatility and correlation structure, Engle (2002) proposed using the dynamic conditional correlation GARCH (DCC) model. Aielli (2013) modified the DCC model, proposing the corrected DCC-GARCH (cDCC) model. We combine the cDCC model and the NLS method to deliver an improved estimation of the covariance matrix. We compare various (c)DCC models (DCC, cDCC, DCC-LS, cDCC-LS, DCC-NLS, and cDCC-NLS) to estimate the accuracy of each method. To test whether the estimation accuracy of the covariance matrices affects the performance of the risk-based portfolio, we compare the accuracy of each method, including the proposed method. Then, we check whether the improvement affects the performance of risk-based portfolios in the Japanese stock market.

The proposed method (cDCC-NLS) offers the best estimation accuracy of the six methods. Furthermore, we confirm that, unlike MV, RP and MD do not depend on the estimation accuracy of the covariance matrix.

The remainder of the paper is organized as follows. Section 2 gives a brief description of risk-based portfolios and the estimation method using covariance matrices. Section 3 describes the proposed method, which consists of a cDCC-GARCH model and the nonlinear shrinkage method. Section 4 discusses a simulation of a large number of stocks in Japanese markets. Lastly, Section 5 concludes the paper.

2. Related Works

2.1. Risk-Based Portfolios

In this section, we summarize the well-known risk-based portfolios examined in previous research. Suppose there are N risky assets, with rates of returns given by the random variables $\mathbf{R} = (R_1, \dots, R_N)^T$. Let $\mathbf{w} = (w_1, \dots, w_N)^T$ be a weighting vector for each asset, $\boldsymbol{\mu} = (\mu_1, \dots, \mu_N)^T$ be an expected return vector, and $\boldsymbol{\Sigma} = E[(\mathbf{R} - \boldsymbol{\mu})(\mathbf{R} - \boldsymbol{\mu})^T]$ be a covariance matrix. Thus, a portfolio's variance and risk (standard deviation) can be expressed as $\sigma_p^2 = \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w}$.

2.1.1. The Minimum Variance Portfolio

An MV portfolio has the lowest risk of portfolios based on the mean-variance method developed by Markowitz (1952). Specifically, the optimization problem is to minimize the portfolio variance subject to short-sales constraints and a budget constraint in which the sum of the weights is $\mathbf{1}$, denoting a vector of ones of dimension $N \times 1$:

$$\min_w \sigma_p^2 = \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w} \quad (1)$$

$$s.t., \mathbf{w}^T \mathbf{1} = 1 \quad (2)$$

$$\mathbf{w} > 0. \quad (3)$$

In the absence of the short-sales constraints (Equation (3)), the analytical solution is $\mathbf{w}^* = \frac{\boldsymbol{\Sigma}^{-1} \mathbf{1}}{\mathbf{1}^T \boldsymbol{\Sigma}^{-1} \mathbf{1}}$. The MV portfolio is an efficient portfolio in terms of the mean-variance if the expected returns of all assets are the same.

Haugen and Baker (1991) and Clarke et al. (2006) investigated the MV portfolio and showed that the cumulative excess returns of MV portfolios in the U.S. stock market have been slightly higher than those of the market over the previous 42 years. Scherer (2011) showed that an MV portfolio tends to hold a low beta and low residual-risk stocks. These results are known as low volatility anomalies.

2.1.2. The Risk Parity Portfolio

The RP portfolio assigns the same risk budget to each asset (Qian 2005). This way, no asset class can be dominant in driving the volatility of the portfolio. We present a rigorous mathematical definition for the RP portfolio using the marginal risk contribution (MRC) and the risk contribution (RC), following Maillard et al. (2010). The definition characterizes the RP portfolio in terms of an equal RC for each asset, as follows:

$$MRC = \frac{\partial \sigma_P}{\partial \mathbf{w}} = \frac{\Sigma \mathbf{w}}{\sigma_P}, MRC_i = \frac{(\Sigma \mathbf{w})_i}{\sigma_P} \tag{4}$$

$$\sigma_P = \sum_{i=1}^N w_i \times MRC_i = \mathbf{w}^T MRC \tag{5}$$

$$RC_i = \frac{w_i \times MRC_i}{\sigma_P}. \tag{6}$$

Maillard et al. (2010) propose the following optimization problem (7)–(9) to determine the optimal RP portfolio weights when short selling and when leverage is not allowed. In addition, Maillard et al. (2010) show that the optimization program in (7)–(9) is a convex program and that it has a unique local minimum.

$$\min_{\mathbf{w}} \sum_{i=1}^N \sum_{j=1}^N (RC_i - RC_j)^2 \tag{7}$$

$$s.t., \mathbf{w}^T \mathbf{1} = 1 \tag{8}$$

$$\mathbf{w} > 0 \tag{9}$$

2.1.3. The Maximum Diversification Portfolio

Choueifaty and Coignard (2008) proposed the MD portfolio to benefit most from the diversification effect of the portfolio. They define the diversification ratio of any portfolio as $DR(\mathbf{w})$. Here, $DR(\mathbf{w})$ is the ratio of the weighted average of the volatilities divided by the portfolio volatility. The MD portfolio maximizes the diversification ratio. Intuitively, this portfolio incorporates more low-correlated assets. From the viewpoint of the Sharpe ratio-maximized portfolio, it tries to minimize the volatility and the correlation, rather than the expected return.

$$\max_{\mathbf{w}} DR(\mathbf{w}) = \frac{\sum_{i=1}^N w_i \sigma_i}{\sigma_P}. \tag{10}$$

In order to maximize $DR(\mathbf{w})$, it is necessary to reduce the denominator or to increase the numerator. The numerator and the denominator are linked through the volatility term. Because the denominator has a correlation term, a portfolio with a low correlation is constructed as a result.

2.2. Estimation Method of Covariance Matrices

The difficulty of estimating covariance matrices in financial time series arises from two aspects: the cross section and time series.

In the cross section, the challenge is to correct the biases of the sample covariance matrix eigenvalues. When the number of assets under consideration is large, especially relative to the number of historical return observations, the sample covariance matrix is estimated with a large error.

This is caused by the bias of the eigenvalues of the sample covariance matrix. In order to reduce the estimation error, the linear shrinkage method of [Ledoit and Wolf \(2004\)](#) is often used. This method uniformly corrects all biases of the eigenvalues of the sample covariance matrix.

In general, the eigenvalues of a small sample are estimated as too small, and those of a large sample are estimated as too large. Therefore, the nonlinear shrinkage method of [Ledoit and Wolf \(2012\)](#) individually corrects each bias of the sample eigenvalues using the random matrix theory. In time series, the challenge is to account for volatility clustering and time-varying correlations.

A model that has been shown to be successful in capturing volatility clustering is the autoregressive conditional heteroscedasticity (ARCH) model, introduced by [Engle \(1982\)](#). Then, [Bollerslev \(1986\)](#) generalized the ARCH model (the GARCH model) to make it more realistic.

To introduce a time-varying correlation structure to these conditional variance models, [Engle \(2002\)](#) proposed the dynamic conditional correlation GARCH (DCC) model. However, [Aielli \(2013\)](#) reveals a weak point of the DCC model, in that it possesses a significant asymptotic bias in the estimator of the sample covariance matrix, which is a constituent of the correlation evolution process. [Aielli \(2013\)](#) subsequently modified the DCC model, proposing the corrected DCC-GARCH (cDCC) model.

On the other hand, it is well known that estimating the DCC model with a large number of assets is difficult. One of the reasons for this is that we need to invert the conditional correlation matrix when we calculate the log-likelihood function of the DCC model in order to estimate the parameters. Finding the inverse matrix has a computational load of $O(N^3)$. Fortunately, [Engle et al. \(2008\)](#) discovered a way to overcome this hurdle, called the composite likelihood method. The composite likelihood is computed by summing the log-likelihood functions of all contiguous pairs. This requires a computational load of $O(N)$. We summarize each method in the next section.

2.2.1. Nonlinear Shrinkage

The population covariance matrix Σ is a non-random t by N -dimensional positive definite matrix. Let X_t be a t by N matrix of real independent and identically distributed (i.i.d.) random variables, with a zero mean and unit variance. We only observe $Y_t = X_t \Sigma^{1/2}$. Thus, neither X_t nor Σ are observed on their own.

Let $\hat{\Sigma} = \hat{\Sigma}(Y_t)$ be an estimator of Σ . Then, the estimator is said to be rotation-equivariant if it satisfies $\hat{\Sigma}(AY_t) = A\hat{\Sigma}A^T$ for any orthogonal matrix A . The class of rotation-equivariant estimators of the covariance matrix comprises all estimators that have the same eigenvectors as the sample covariance matrix. Thus, every rotation-equivariant estimator $\hat{\Sigma}$ is of the form $V_t D_t V_t^T$ where D_t is a diagonal matrix with eigenvalues (d_1, \dots, d_t) as its elements, and where V_t is a matrix with the i th column as the sample eigenvector. [Ledoit and Wolf \(2012\)](#) found the best $\hat{\Sigma}$ under the following loss function in the rotation-equivariant estimator.

$$\min_{D_t} \|V_t D_t V_t^T - \Sigma\| \tag{11}$$

Here, $\|x\|$ is the Frobenius norm, defined as $\|M\| = \sqrt{\frac{\text{Tr}(MM^T)}{r}}$ for any $r \times m$ matrix M . Its solution is $D_t^* = \text{diag}(d_1^*, \dots, d_t^*)$, where $d_i^* = v_i^T \Sigma v_i$ and v_i is the i th column of V_t . Therefore, the best rotation-equivariant estimator of the covariance matrix is $S_i^* = V_t D_t^* V_t^T$.

[Ledoit and P ech e \(2011\)](#) showed that d_i^* can be approximated under the following conditions. Let H_T be the population spectral distribution and let F_T be the limiting spectral distribution. We assume that H_T and F_T converge to the non-random limiting spectral distribution H and F , respectively. Here, $\tau_{i,T}$ denotes the population eigenvalues of Σ , and $\lambda_{i,T}$ denotes the sample eigenvalues of $\hat{\Sigma}$.

$$H_T(x) = \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{\{x \leq \tau_{i,T}\}}, \forall x \in \mathbb{R} \tag{12}$$

$$F_T(x) = \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{\{x \leq \lambda_{i,T}\}}, \forall x \in \mathbb{R} \tag{13}$$

$$H_T(x) \longrightarrow H \tag{14}$$

$$F_T(x) \longrightarrow F. \tag{15}$$

From the random matrix theory, m_F is quantitatively related to H , where m_F is the Stieltjes transform of F , and c is p/n (Silverstein 1995).

$$m_F(z) = \int_{-\infty}^{\infty} \frac{1}{\lambda - z} dF(\lambda), \forall z \in \mathbb{C}^+ \tag{16}$$

$$m_F(z) = \int_{-\infty}^{\infty} \frac{1}{\tau(1 - c - czm_F(z)) - z} dH(\tau) \tag{17}$$

$$\hat{m}_F(\lambda) = \lim_{z \in \mathbb{C}^+ \rightarrow \lambda} m_F(z) \tag{18}$$

From the above, d_i^* can be approximated as follows:

$$d_i^* = \frac{\lambda_i}{|1 - c - c\lambda_i \hat{m}_F(\lambda_i)|^2}. \tag{19}$$

2.2.2. DCC-GARCH Model

Suppose there are N risky assets. Let $\mathbf{r}_t = (r_{1t}, \dots, r_{Nt})^T$ be the return vector at time $t (1 \leq t \leq T)$. Then, $E(\mathbf{r}_t | \mathcal{F}_{t-1}) = \mathbf{0}$ is defined as the conditional mean vector and $V(\mathbf{r}_t | \mathcal{F}_{t-1}) = \mathbf{H}_t$ is defined as the conditional covariance matrix. We assume that the joint distribution of returns follows a multivariate Gaussian distribution. Then, \mathcal{F}_{t-1} is the σ -algebra that contains information of what happened up to time $t - 1$.

Engle (2002) defines the DCC model using Equations (21)–(25). Here, \mathbf{H}_t is decomposed into \mathbf{D}_t^2 , the conditional variance of each asset, and \mathbf{R}_t , the conditional correlation matrix.¹

In the DCC model, both the variance and the correlations between the assets are designed to be time-varying. The elements in the diagonal matrix \mathbf{D}_t are standard deviations from univariate GARCH models. Note that the univariate GARCH models can have different orders. Often the simplest model, GARCH(1,1) is adequate. Hereafter, we assume \mathbf{D}_t is driven by GARCH(1,1). Then, \mathbf{R}_t is decomposed into \mathbf{Q}_t to ensure that all elements are less than or equal to one, by definition of the correlation. Several conditions are imposed on the parameters a and b in order to guarantee that \mathbf{H}_t is positive definite. The scalar values a and b must satisfy: $a > 0, b > 0$ and $a + b < 1$. In addition, the starting value of \mathbf{Q}_t has to be positive definite to guarantee that \mathbf{H}_t is positive definite. In addition, \mathbf{S} is the unconditional covariance matrix of the standardized errors, $\boldsymbol{\varepsilon}_t$.

¹ The operator \circ represents the Hadamard product, which is a binary operation that takes two matrices of the same dimensions. For two matrices, \mathbf{A} and \mathbf{B} of the same dimension, the Hadamard product $\mathbf{A} \circ \mathbf{B}$, is a matrix of the same dimension as the operands, with elements given by $A \circ B_{i,j} = (A)_{i,j}(B)_{i,j}$.

$$r_t | \mathcal{F}_{t-1} \sim N(\mathbf{0}, \mathbf{H}_t) \tag{20}$$

$$\mathbf{H}_t = \mathbf{D}_t \mathbf{R}_t \mathbf{D}_t \tag{21}$$

$$\mathbf{D}_t^2 = \text{diag}(\omega_i) + \text{diag}(\alpha_i) \circ r_{t-1}^T r_{t-1} + \text{diag}(\beta_i) \circ \mathbf{D}_{t-1}^2 \tag{22}$$

$$\varepsilon_t = \mathbf{D}_t^{-1} r_t \tag{23}$$

$$\mathbf{R}_t = \mathbf{Q}_t^{-\frac{1}{2}} \mathbf{Q}_t \mathbf{Q}_t^{-\frac{1}{2}} \tag{24}$$

$$\mathbf{Q}_t = S(1 - a - b) + a\varepsilon_{t-1}^T \varepsilon_{t-1} + b\mathbf{Q}_{t-1} \tag{25}$$

Divide the parameters of DCC into two groups: $\theta = (\theta_1, \dots, \theta_N)$, where $\theta_i = (\omega_i, \alpha_i, \beta_i)$ are the parameters of the univariate GARCH model for the i th asset series, and the correlation structure in Equation (24) is driven by Equation (25), which has the parameters $\phi = (a, b)$.

When r_t follows a multivariate Gaussian distribution, the log-likelihood function for $L(\theta, \phi)$ can be written as the sum of the volatility part $L_V(\theta)$ and the correlation part $L_C(\theta, \phi)$.

$$L(\theta, \phi) = L_V(\theta) + L_C(\theta, \phi) \tag{26}$$

$$L_V(\theta) = -\frac{1}{2} \sum_{t=1}^T (n \ln 2\pi + \ln |\mathbf{D}_t|^2 + r_t^T \mathbf{D}_t^{-2} r_t) \tag{27}$$

$$L_C(\theta, \phi) = -\frac{1}{2} \sum_{t=1}^T (\ln |\mathbf{R}_t| + \varepsilon_t^T \mathbf{R}_t^{-1} \varepsilon_t - \varepsilon_t^T \varepsilon_t) \tag{28}$$

However, estimating the correctly specified log-likelihood is difficult. Hence, the DCC model is designed to allow for a two-step estimation.

Compared with the log-likelihood in the univariate case, we find that the log-likelihood in $L_V(\theta)$ is the sum of the log-likelihoods of the univariate GARCH equations of N assets. Thus, the parameters of the different univariate models may be determined separately. In the first step, the parameters of the univariate GARCH models are estimated for each asset series. From this step, the parameter set θ is estimated. When estimating θ , we can also estimate the standardized errors ε_t , which can then be used to estimate S . In the second step, the parameter ϕ is estimated using θ and S . The method used to estimate S with the standardized errors ε_t is called correlation targeting (Engle 2009). Here, S can be estimated using the sample covariance matrix of ε_t

Step 1 Maximize the log-likelihood $L_V(\theta)$ to determine \hat{D}_t .

$$\hat{\theta} = \max_{\theta} L_V(\theta). \tag{29}$$

Step 2 After estimating θ , estimate the standardized errors ε_t . Here, S can be estimated using the moment estimator $\hat{S} = \frac{1}{2} \sum_{t=1}^T \hat{\varepsilon}_t \hat{\varepsilon}_t^T, \hat{\varepsilon}_t = \hat{D}_t^{-1} r_t$.

Then, maximize the log-likelihood $L_C(\theta, \phi)$ to determine ϕ .

$$\hat{\phi} = \max_{\phi} L_C(\hat{\theta}, \phi). \tag{30}$$

2.2.3. cDCC-GARCH Model

Aielli (2013) reveals a weak point in the DCC model. The correlation targeting estimation is based on the assumption that $\bar{Q} = \frac{1}{T} \sum_{t=1}^T \mathbf{Q}_t$ is equal to \hat{S} . However, this assumption cannot be justified because the unconditional expectation of \mathbf{Q}_t is not equal to the unconditional expectation of $\hat{\varepsilon}_t \hat{\varepsilon}_t^T$.

Thus, Aielli (2013) proposes a corrected DCC-GARCH (cDCC) model, which is tractable and easy to implement. In this model, the conditional correlation driving the process \mathbf{Q}_t is defined as follows:

$$Q_t = \Psi(1 - a - b) + a\varepsilon_{t-1}^* \varepsilon_{t-1}^{*T} + bQ_{t-1} \tag{31}$$

$$\varepsilon_t^* = \text{diag}(Q_t)^{\frac{1}{2}} \varepsilon_t \tag{32}$$

The form of the correlation driving the process of the DCC model is modified in such a way that the unconditional expectation of Q_t is equal to the unconditional expectation of $\varepsilon_t^* \varepsilon_t^{*T}$. Then, Ψ can be estimated consistently using the sample covariance matrix of ε_t^* .

The estimation of the cDCC model is based on maximizing the same log-likelihood function as that of the DCC model. In the first step, the parameter vector θ of the univariate GARCH models is estimated. Using θ , estimate the modified errors ε_t^* , the sample covariance matrix of which can be used to estimate Ψ . In the second step, the parameter ϕ is estimated using $\hat{\theta}$.

Aielli (2013) performed a simulation study, showing that there is a considerable bias in the estimator of S for the DCC, even when the cross-section size is small. However, he also shows there is a downward bias in a and b , which is a recognized problem when using DCC with large systems (Engle 2002).

2.2.4. Composite Likelihood

It is well known that estimating the (c)DCC model in the case of a large number of assets is difficult. One of the reasons for this is that we need to invert the conditional correlation matrix when we maximize the log-likelihood $L_C(\theta, \phi)$. Finding the inverse matrix has a computational load of $O(N^3)$. Engle et al. (2008) overcome this problem using what they call the composite likelihood method, which can be traced back to Lindsay (1988). The composite likelihood of the (c)DCC model is based on subsets of observations, for example, a subset of two assets' returns, $Y_{jt} = (r_{1j,t}, r_{2j,t})^T$, $j = 1, \dots, N$. The conditional variance of Y_{jt} is H_{jt} .² Subsets can be selected in various ways: all possible combinations of pairs of assets ($O(N^2)$), contiguous pairs ($O(N)$), or subsets of pairs selected randomly ($O(1)$). The simulation study in Engle et al. (2008) shows that the results for different types of subsets are approximately the same. We use the composite likelihood estimator based on all contiguous pairs because it is the most scalable of those available.

$$Y_{jt} | \mathcal{F}_{t-1} \sim N(\mathbf{0}, H_{jt}) \tag{33}$$

$$H_{jt} = D_{jt} R_{jt} D_{jt} \tag{34}$$

$$D_{jt}^2 = \text{diag}(\omega_i) + \text{diag}(\alpha_i) \circ Y_{jt-1}^T Y_{jt-1} + \text{diag}(\beta_i) \circ D_{jt-1}^2 \tag{35}$$

$$\varepsilon_{jt} = D_{jt}^{-1} Y_{jt} \tag{36}$$

$$Q_{jt} = S_j(1 - a - b) + a\varepsilon_{jt-1}^T \varepsilon_{jt-1} + bQ_{jt-1} \tag{37}$$

$$R_{jt} = Q_{jt}^{-\frac{1}{2}} Q_{jt} Q_{jt}^{-\frac{1}{2}} \tag{38}$$

Here, \hat{S} can be estimated using the moment estimator $\hat{S} = \frac{1}{2} \sum_{t=1}^T \hat{\varepsilon}_{jt} \hat{\varepsilon}_{jt}^T$. The parameter vector $\hat{\phi} = (a, b)$ is estimated by maximizing the average of the log-likelihood function $L_{CL}(\hat{\theta}, \phi)$ among all contiguous pairs, rather than maximizing $\max_{\phi} L_C(\hat{\theta}, \phi)$ in Step 2.

² Note that the parameter vector $\phi = (a, b)$ is the same among all H_{jt} .

$$\hat{\phi} = \max_{\phi} L_{CL}(\hat{\theta}, \phi) \quad (39)$$

$$L_{CL}(\hat{\theta}, \phi) = \sum_{t=1}^T \frac{1}{N} \sum_{j=1}^N \log \mathcal{L}_{j,t}(\phi) \quad (40)$$

$$\mathcal{L}_{j,t}(\phi) = -\frac{1}{2} \sum_{t=1}^T (\ln |R_{jt}| + \varepsilon_{jt}^T R_{jt}^{-1} \varepsilon_{jt} - \varepsilon_{jt}^T \varepsilon_{jt}) \quad (41)$$

3. Combining Nonlinear Shrinkage and the cDCC-GARCH Model

As described above, the estimation of the covariance matrix is problematic in terms of the cross section and the time series. To combine these two aspects, [Hafner and Reznikova \(2012\)](#) proposed an estimation method combining linear shrinkage and the DCC-GARCH model (DCC-LS), and [Engle et al. \(2017\)](#) proposed an estimation method combining nonlinear shrinkage and the DCC-GARCH model (DCC-NLS). We propose a highly accurate estimation method that combines nonlinear shrinkage and the cDCC-GARCH model, called the cDCC-NLS method. To summarize, the cDCC-NLS estimation follows a two-step process:

Step 1 For each asset, fit a univariate GARCH(1,1) model.

Step 2 Estimate the unconditional covariance matrix of the standardized errors using the nonlinear shrinkage method, and then maximize the composite likelihood of the cDCC model.

In the estimation of Step 2, [Hafner and Reznikova \(2012\)](#) used the sample covariance matrix with the linear shrinkage estimator (DCC-LS). [Engle et al. \(2017\)](#) later replaced this with the nonlinear shrinkage estimator (DCC-NLS). We combine nonlinear shrinkage and the cDCC-GARCH model and confirm its advantages in the next section.

4. Simulation Study

In order to check whether the performance of risk-based portfolios depends on the estimation accuracy of the covariance matrices, it is necessary to confirm (1) the estimation accuracy of each method, and (2) the effect of the estimation accuracy on performance.

In this section, we first compare the estimation accuracy of the covariance matrix using various (c)DCC models. We evaluate the covariance matrix estimation accuracy following [Engle et al. \(2017\)](#). Next, we examine the effect of the difference in the estimation accuracy on the performance of the risk-based portfolios.

4.1. Monte Carlo Study

In this section, we perform a Monte Carlo simulation to compare the estimation accuracy of the parameters of various (c)DCC models: the DCC, cDCC, DCC-LS, cDCC-LS, DCC-NLS, and cDCC-NLS models. [Engle et al. \(2017\)](#) defines an error function between the true covariance matrix generated by a (c)DCC model and the estimation. On the other hand, because of the sample estimator, the LS and NLS methods do not handle true covariance matrices at each time point, which means the errors cannot be calculated in the same way as in the (c)DCC model. Therefore, we limit our discussion to the (c)DCC model.

In order to compare our results with those of previous studies ([Hafner and Reznikova 2012](#); [Engle et al. 2017](#)), we run a simulation study using the same setup. Let the covariance matrix at each time point of the DCC model, given a certain unconditional covariance and parameters, be a true covariance matrix. We generate random numbers according to the true DCC model, which we then use to estimate the covariance matrix. An error between the estimated covariance matrix and the true covariance matrix is defined as the estimation accuracy.

As in the previous studies, we estimate the unconditional population covariance matrix from the $N \in \{100, 500, 1000\}$ most liquid stocks in the TOPIX index, using 60 months of daily data from

January 2005 through December 2009. We then simulate a (c)DCC model using the unconditional population covariance matrix and the parameters $a = 0.05$ and $b = 0.93$. The univariate volatility dynamics are GARCH(1,1) models with identical parameters $(\omega_i, \alpha_i, \beta_i) = (0.01, 0.05, 0.9)$ 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 $T = 1250$. We run the simulation study for the same setup, with the exception of the unconditional population covariance matrix. Because we study the Japanese stock market in the next section, we use data on Japanese stocks to define the true unconditional covariance matrix.

First, we check the estimation accuracy of the (c)DCC model parameters $a = 0.05$ and $b = 0.93$. The estimation results for each model are shown in Tables 1 and 2. Tables 1 and 2 show the average values and standard deviations of $10^5/N$ estimates of each parameter (a, b) . There is no downward bias in the parameters described by Engle et al. (2008) or Aielli (2013) in either the estimation method or in the number of stocks. Therefore, it is considered that the difference in each method is the estimation accuracy of the standardized residual sample covariance matrix, which we use for the correlation targeting.

Table 1. The average and standard deviation of the (c)DCC parameter $a = 0.05$ using six correlation methods targeting dimensions $N \in \{100, 500, 1000\}$, with sample size $T = 1250$.

N	DCC	DCC-LS	DCC-NLS	cDCC	cDCC-LS	cDCC-NLS
100	0.0506 (0.0032)	0.0506 (0.0033)	0.051 (0.0034)	0.0498 (0.0028)	0.0501 (0.0029)	0.0503 (0.003)
500	0.0497 (0.0011)	0.0497 (0.0011)	0.0498 (0.0015)	0.0502 (0.0025)	0.0501 (0.0027)	0.0504 (0.003)
1000	0.0496 (0.0016)	0.0498 (0.0013)	0.0501 (0.0015)	0.0476 (0.0017)	0.05 (0.0018)	0.0501 (0.0018)

Table 2. The average and standard deviation of the (c)DCC parameter $b = 0.93$ using six correlation methods targeting dimensions $N \in \{100, 500, 1000\}$, with sample size $T = 1250$.

N	DCC	DCC-LS	DCC-NLS	cDCC	cDCC-LS	cDCC-NLS
100	0.9279 (0.0048)	0.9285 (0.0052)	0.9287 (0.0054)	0.9287 (0.0031)	0.9284 (0.0034)	0.9289 (0.0038)
500	0.9282 (0.0024)	0.9285 (0.0024)	0.9292 (0.0022)	0.927 (0.004)	0.9275 (0.0046)	0.9283 (0.005)
1000	0.9278 (0.0021)	0.9278 (0.0016)	0.9277 (0.0018)	0.9269 (0.0022)	0.9271 (0.0025)	0.9275 (0.0024)

Next, we evaluate the accuracy of the covariance matrix estimation using the loss function of Engle et al. (2017). Let Σ_t denote the N -dimensional “true” covariance matrix at time t , and $\hat{\Sigma}_t$ be an estimator of Σ_t . The loss function $L_t(\hat{\Sigma}_t, \Sigma_t)$ and the average loss \hat{L} are defined below.

$$L_t(\hat{\Sigma}_t, \Sigma_t) = -\frac{\text{Tr}(\hat{\Sigma}_t^{-1}\Sigma_t\hat{\Sigma}_t^{-1})/N}{[\text{Tr}(\hat{\Sigma}_t^{-1}/N)]^2} - \frac{1}{\text{Tr}(\Sigma_t^{-1}/N)} \tag{42}$$

$$\hat{L} = \frac{1}{T} \sum_{t=1}^T L_t \tag{43}$$

Table 3 shows the results for \hat{L} , averaged across $10^5/N$ Monte Carlo simulations.

Table 3. Average loss for (c)DCC estimators using six correlation methods targeting dimensions $N \in \{100, 500, 1000\}$, with sample size $T = 1250$. The unit is 10^{-3} .

N	DCC	DCC-LS	DCC-NLS	cDCC	cDCC-LS	cDCC-NLS
100	8.60257	8.27205	7.47024	7.95906	7.6558	6.8779
500	23.04541	18.36612	14.55736	22.16638	18.07399	13.37496
1000	70.44232	24.46691	18.8293	63.24801	24.4935	15.29444

We find that the average loss of the cDCC-NLS method is the lowest of the various methods for any number of assets. Therefore, the cDCC-NLS method shows the best estimation accuracy.

4.2. Performance of the Risk-Based Portfolios

In the previous section, we compared the accuracy of the covariance matrix estimation of each method using a Monte Carlo study. Thus, we were able to confirm that the cDCC-NLS method shows the best estimation accuracy. Next, we examine the effect of the difference in the estimation accuracy on the performance of risk-based portfolios. Here, we use the minimum variance (MV), minimum variance without short constraint (MVS), risk parity (RP), and maximum diversification (MD). We estimate the covariance matrices using the various (c)DCC models for the $N \in \{100, 500, 1000\}$ most liquid stocks in the TOPIX index, as before. Each portfolio is re-balanced monthly, and the parameter estimation period of each method is 60 months. The out-of-sample period is from January 2000 to December 2015. In other words, we estimate the weights on the previous 60 months of data from January 2000, and then compute the value of the out-of-sample portfolio for the following month. Next, we roll the window forward by one month and repeat the process until December 2015.

We use an annualized Sharpe ratio to evaluate the performance of each portfolio. We also consider whether the cDCC-NLS (the most accurate method) delivers a higher out-of-sample Sharpe ratio than that of the DCC and the cDCC at a level that is statistically significant. For a given sample size N , a two-sided p -value for the null hypothesis of equal Sharpe ratios is obtained using the pre-whitened HACPW method, as described in (Ledoit and Wolf 2008). Tables 4–7 show the annualized return, annualized volatility, and annualized Sharpe ratio for each risk-based portfolio.

Table 4. Annualized return, volatility, and Sharpe ratio for an MV portfolio using six methods for dimensions $N \in \{100, 500, 1000\}$, with sample size $T = 1250$. The out-of-sample period is from January 2000 to December 2015. In the rows labeled Sharpe Ratio, the largest number appears in bold. In the columns labeled DCC and cDCC-NLS, significant out-performance of one of the two portfolios over the other (in terms of the Sharpe ratio) is denoted by asterisks: *** denotes significance at the 0.01 level; ** denotes significance at the 0.05 level; * denotes significance at the 0.1 level. In the columns labeled cDCC and cDCC-NLS, significant out-performance of one of the two portfolios over the other (in terms of the Sharpe ratio) is denoted by asterisks: +++ denotes significance at the 0.01 level; ++ denotes significance at the 0.05 level; + denotes significance at the 0.1 level.

	DCC	DCC-LS	DCC-NLS	cDCC	cDCC-LS	cDCC-NLS
$N = 100$						
Return [%]	8.08	8.23	8.29	7.82	8.09	8.62
Volatility [%]	16.40	16.41	16.36	16.20	16.16	16.18
Sharpe Ratio	0.49	0.50	0.51	0.48	0.50	0.53 +++
$N = 500$						
Return [%]	9.80	9.65	9.71	9.81	9.74	10.09
Volatility [%]	17.15	17.16	17.00	16.92	16.93	16.90
Sharpe Ratio	0.57	0.56	0.57	0.58	0.58	0.60 ++
$N = 1000$						
Return [%]	9.25	9.50	9.97	9.66	9.64	10.05
Volatility [%]	16.61	16.63	16.37	16.23	16.17	16.15
Sharpe Ratio	0.56	0.57	0.61	0.59	0.60	0.62 ** +

Table 5. Annualized return, volatility, and Sharpe ratio for an MV portfolio without short-sales constraints (MVS) using six methods for dimensions $N \in \{100, 500, 1000\}$, with sample size $T = 1250$. The out-of-sample period is from January 2000 to December 2015. In the rows labeled Sharpe Ratio, the largest number appears in bold. In the columns labeled DCC and cDCC-NLS, significant out-performance of one of the two portfolios over the other (in terms of the Sharpe ratio) is denoted by asterisks: *** denotes significance at the 0.01 level; ** denotes significance at the 0.05 level; * denotes significance at the 0.1 level. In the columns labeled cDCC and cDCC-NLS, significant out-performance of one of the two portfolios over the other (in terms of the Sharpe ratio) is denoted by asterisks: +++ denotes significance at the 0.01 level; ++ denotes significance at the 0.05 level; + denotes significance at the 0.1 level.

	DCC	DCC-LS	DCC-NLS	cDCC	cDCC-LS	cDCC-NLS
N = 100						
Return [%]	10.54	10.63	9.43	10.14	10.18	11.57
Volatility [%]	22.65	22.54	20.90	15.19	15.12	15.08
Sharpe Ratio	0.47	0.47	0.45	0.67	0.67	0.77 +++
N = 500						
Return [%]	9.48	9.27	7.41	9.37	9.33	11.09
Volatility [%]	12.64	12.47	13.90	11.63	11.40	11.12
Sharpe Ratio	0.75	0.74	0.53	0.81	0.82	1.00 **
N = 1000						
Return [%]	6.01	6.52	6.98	6.26	7.78	9.30
Volatility [%]	9.96	8.95	9.16	10.56	8.99	8.21
Sharpe Ratio	0.60	0.73	0.76	0.59	0.87	1.13 ***

Table 6. Annualized return, volatility, and Sharpe ratio for an RP portfolio using six methods for dimensions $N \in \{100, 500, 1000\}$, with sample size $T = 1250$. The out-of-sample period is from January 2000 to December 2015. In the rows labeled Sharpe Ratio, the largest number appears in bold. In the columns labeled DCC and cDCC-NLS, significant out-performance of one of the two portfolios over the other (in terms of the Sharpe ratio) is denoted by asterisks: *** denotes significance at the 0.01 level; ** denotes significance at the 0.05 level; * denotes significance at the 0.1 level. In the columns labeled cDCC and cDCC-NLS, significant out-performance of one of the two portfolios over the other (in terms of the Sharpe ratio) is denoted by asterisks: +++ denotes significance at the 0.01 level; ++ denotes significance at the 0.05 level; + denotes significance at the 0.1 level.

	DCC	DCC-LS	DCC-NLS	cDCC	cDCC-LS	cDCC-NLS
N = 100						
Return [%]	6.95	6.97	7.16	7.01	7.01	7.06
Volatility [%]	19.39	19.39	19.45	19.60	19.60	19.56
Sharpe Ratio	0.36	0.36	0.37	0.36	0.36	0.36
N = 500						
Return [%]	9.46	9.35	9.76	9.82	9.81	9.68
Volatility [%]	18.26	18.30	18.41	18.59	18.57	18.48
Sharpe Ratio	0.52	0.51	0.53	0.53	0.53	0.52
N = 1000						
Return [%]	10.41	10.54	10.55	10.54	10.52	10.52
Volatility [%]	16.88	16.93	17.07	17.23	17.19	17.07
Sharpe Ratio	0.62	0.62	0.62	0.61	0.61	0.62

Table 7. Annualized return, volatility, and Sharpe ratio for an MD portfolio using six methods for dimensions $N \in \{100, 500, 1000\}$, with sample size $T = 1250$. The out-of-sample period is from January 2000 to December 2015. In the rows labeled Sharpe Ratio, the largest number appears in bold. In the columns labeled DCC and cDCC-NLS, significant out-performance of one of the two portfolios over the other (in terms of the Sharpe ratio) is denoted by asterisks: *** denotes significance at the 0.01 level; ** denotes significance at the 0.05 level; * denotes significance at the 0.1 level. In the columns labeled cDCC and cDCC-NLS, significant out-performance of one of the two portfolios over the other (in terms of the Sharpe ratio) is denoted by asterisks: +++ denotes significance at the 0.01 level; ++ denotes significance at the 0.05 level; + denotes significance at the 0.1 level.

	DCC	DCC-LS	DCC-NLS	cDCC	cDCC-LS	cDCC-NLS
$N = 100$						
Return [%]	7.06	7.07	7.06	7.00	7.04	7.02
Volatility [%]	20.48	20.49	20.49	20.49	20.49	20.49
Sharpe Ratio	0.35	0.35	0.34	0.34	0.34	0.34
$N = 500$						
Return [%]	10.13	10.13	10.12	10.12	10.12	10.12
Volatility [%]	19.64	19.64	19.64	19.64	19.64	19.64
Sharpe Ratio	0.52	0.52	0.52	0.52	0.52	0.52
$N = 1000$						
Return [%]	10.63	10.63	10.63	10.66	10.66	10.63
Volatility [%]	18.66	18.66	18.66	18.66	18.66	18.66
Sharpe Ratio	0.57	0.57	0.57	0.57	0.57	0.57

All risk-based portfolios show improved performance as the number of stocks N increases. For the RP in Table 6 and the MD in Table 7, we confirm that performance does not depend on the estimation accuracy. There are no statistically significant differences between the DCC and cDCC-NLS or between the cDCC and cDCC-NLS. We further confirm that the weights of the RP and MD are similar across the models until December 2015. We consider that this explains why the performance of the RP and MD are similar.

On the other hand, for the MV in Table 4, the Sharpe ratio of the cDCC-NLS is an improvement over those of the DCC and the cDCC, especially for large numbers of stocks. In addition, the MVS in Table 5, which allows short selling, shows a significant improvement. For the MV and MVS, the cDCC-NLS records the highest Sharpe ratios for any numbers of stocks. According to Scherer (2011), the good performance of the MV portfolio is based on a low-risk anomaly. By being short in high volatility and long in low volatility, low-risk anomalies can be realized. Thus, we think the presence of short sales makes the low-risk anomaly possible. From these results, we conclude the RP and MD do not depend on the estimation accuracy of the covariance matrix, whereas the MV and MVS do depend on the estimation accuracy of the covariance matrix.

5. Conclusions

The contribution of this study is twofold. First, we propose the cDCC-NLS method, a highly accurate estimation method that combines nonlinear shrinkage and the cDCC-GARCH model. This method overcomes the difficulties that arise from the cross-section and time series aspects of estimating covariance matrices. We compared the performance of risk-based portfolios under several estimation methods for covariance matrices: the DCC, cDCC, DCC-LS, DCC-NLS, cDCC-LS, and cDCC-NLS. We performed an empirical analysis using a large number of stocks in Japanese markets and confirm the following results:

- The cDCC-NLS method shows the best estimation accuracy.
- The RP and MD do not depend on the estimation accuracy of the covariance matrix.
- The MV does depend on the estimation accuracy of the covariance matrix.

Our recommendations for investors are as follows. When constructing a minimum variance portfolio, especially one that allows short selling, the covariance matrix should be estimated with good estimation accuracy using the cDCC-NLS method. However, for risk parity and maximum diversification portfolios, the estimation accuracy of the covariance matrix is not so important.

Although not considered in this research, many studies have examined the volatility asymmetry in which volatility tends to show a greater increase after negative returns than it does after positive returns. With regard to the conditional correlation, asymmetry is evident in the correlation between stocks tending to increase after negative returns, as in the case of volatility. In future research, we will investigate the impact on the performance of risk-based portfolios by introducing the asymmetry of volatility and the correlation.

In this study, we limited our scope to the (c)DCC model in order to estimate the error of covariance matrix. However, it is also important to examine the relationship between the performance of a risk-based portfolio and other estimation methods.

Author Contributions: K.N. provided the idea, constructed the hypotheses, analyzed the data and wrote the paper. M.I. performed statistical computations and helped analyzing the data. K.Y. supervised and edited the paper.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Aielli, Gian Piero. 2013. Dynamic conditional correlation: on properties and estimation. *Journal of Business & Economic Statistics* 31: 282–99.
- Ardia, David, Guido Bolliger, Kris Boudt, and Jean-Philippe Gagnon-Fleury. 2017. The impact of covariance misspecification in risk-based portfolios. *Annals of Operations Research* 254: 1–16. [CrossRef]
- Bollerslev, Tim. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31: 307–27. [CrossRef]
- Chaves, Denis, Jason Hsu, Feifei Li, and Omid Shakernia. 2011. Risk parity portfolio vs. other asset allocation heuristic portfolios. *The Journal of Investing* 20: 108–18. [CrossRef]
- Choueifaty, Yves, and Yves Coignard. 2008. Toward maximum diversification. *The Journal of Portfolio Management* 35: 40–51. [CrossRef]
- Clarke, Roger G., Harindra De Silva, and Steven Thorley. 2006. Minimum-variance portfolios in the us equity market. *The Journal of Portfolio Management* 33: 10–24. [CrossRef]
- De Carvalho, Raul Leote, Xiao Lu, and Pierre Moulin. 2012. Demystifying equity risk-based strategies: A simple alpha plus beta description. *The Journal of Portfolio Management* 38: 56–70. [CrossRef]
- Engle, Robert. 2009. *Anticipating Correlations: A New Paradigm for Risk Management*. Princeton: Princeton University Press.
- Engle, Robert F. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society* 50: 987–1007. [CrossRef]
- Engle, Robert F. 2002. Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics* 20: 339–50.
- Engle, Robert F, Olivier Ledoit, and Michael Wolf. 2017. Large dynamic covariance matrices. *Journal of Business & Economic Statistics* 1–13. [CrossRef]
- Engle, Robert F., Neil Shephard, and Kevin Sheppard. 2008. Fitting Vast Dimensional Time-Varying Covariance Models. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1293629 (accessed on 11 May 2018).
- Hafner, Christian M., and Olga Reznikova. 2012. On the estimation of dynamic conditional correlation models. *Computational Statistics & Data Analysis* 56: 3533–45.

- Haugen, Robert A., and Nardin L. Baker. 1991. The efficient market inefficiency of capitalization—Weighted stock portfolios. *The Journal of Portfolio Management* 17: 35–40. [CrossRef]
- Ledoit, Olivier, and Sandrine Péché. 2011. Eigenvectors of some large sample covariance matrix ensembles. *Probability Theory and Related Fields* 151: 233–64. [CrossRef]
- Ledoit, Olivier, and Michael Wolf. 2004. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis* 88: 365–411. [CrossRef]
- Ledoit, Oliver, and Michael Wolf. 2008. Robust performance hypothesis testing with the sharpe ratio. *Journal of Empirical Finance* 15: 850–59. [CrossRef]
- Ledoit, Olivier, and Michael Wolf. 2012. Nonlinear shrinkage estimation of large-dimensional covariance matrices. *The Annals of Statistics* 40: 1024–60. [CrossRef]
- Lindsay, Bruce G. 1988. Composite likelihood methods. *Contemporary Mathematics* 80: 221–39.
- Maillard, Sébastien, Thierry Roncalli, and Jérôme Teiletche. 2010. The properties of equally weighted risk contribution portfolios. *The Journal of Portfolio Management* 36: 60–70. [CrossRef]
- Markowitz, Harry. 1952. Portfolio selection. *The Journal of Finance* 7: 77–91.
- Merton, Robert C. 1980. On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics* 8: 323–61. [CrossRef]
- Michaud, Richard O. 1989. The markowitz optimization enigma: Is 'optimized' optimal? *Financial Analysts Journal* 45: 31–42. [CrossRef]
- Qian, Edward. 2005. *Risk Parity Portfolios: Efficient Portfolios through True Diversification*. Boston: Panagora Asset Management.
- Scherer, Bernd. 2011. A note on the returns from minimum variance investing. *Journal of Empirical Finance* 18: 652–60. [CrossRef]
- Silverstein, Jack W. 1995. Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices. *Journal of Multivariate Analysis* 55: 331–39. [CrossRef]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).