A Geometric Framework for Covariance Dynamics

Chulwoo Han* Frank C. Park†

Abstract

Employing methods of differential geometry, we propose a new framework for covariance dynamics modeling. Our approach respects the intrinsic geometric properties of the space of covariance matrices and allows their natural evolution. We develop covariance models that exploit either asset returns or realized covariances and propose a new estimation method that minimizes the length of the geodesic between the forecast and the realization. The geodesic length is equivalent to the Fisher information metric under the Gaussian assumption and is deemed a proper measure of similarity between two covariance matrices. Empirical studies involving three data samples and various performance metrics suggest that our models outperform existing ones.

Keywords: Geometric covariance dynamics; Realized covariance; Manifold; Riemannian metric; Geodesic

1 Introduction

Multivariate volatility models need to preserve the positive definiteness of the covariance matrix as it propagates. Several methods have been proposed to achieve this. Bollerslev

*Corresponding author. Durham University, Mill Hill Lane Durham, DH1 3LB, UK. E-mail: chulwoo.han@durham.ac.uk.

†Seoul National University, 1 Gwanak-ro Seoul 08826, Korea. E-mail: fcp@snu.ac.kr

et al. (1988) impose constraints on the model parameters to ensure positive definiteness. The BEKK model (Engle and Kroner, 1995) assumes a positive definite quadratic form for the model parameters. The DCC model (Engle, 2002) specifies the time series of the variances and that of the correlation matrix independently using a scalar GARCH model and the BEKK model. The GO-GARCH (van der Weide, 2002), as well as other factor models (*e.g.,* Vrontos et al. (2003), Lanne and Saikkonen (2007)), preserve positive definiteness by transforming the original data to independent factors and applying scalar GARCH models to those factors individually. Kawakatsu (2006) extends the exponential GARCH of Nelson (1991) to a multivariate case by exploiting the fact that the matrix exponential of a symmetric matrix is always positive definite.

Covariance models based on realized covariance can handle the issue more conveniently. Chiriac and Voev (2011) decompose realized covariance matrices using the Cholesky decomposition to construct a multivariate vector fractionally integrated ARMA (VARFIMA) process. The HEAVY model proposed by Noureldin et al. (2012) resembles the BEKK but utilizes realized covariances instead of daily returns. The RARCH model of Noureldin et al. (2014) extends the BEKK and DCC models by rotating the returns prior to applying these models. Modeling the matrix logarithm of realized covariance matrices has also been widely adopted, *e.g.*, Chiriac and Voev (2011), Bauer and Vorkink (2011), Callot et al. (2017).

While these models preserve the positive definiteness of covariance matrices, some of them appear to lack justification. For instance, factor models apply a scalar GARCH model to independent factors, but there is no reason to believe that the transformed data would follow a GARCH process. BEKK-type models update all components of the covariance matrix in a linear fashion. However, this approach is unsatisfactory since it implicitly treats covariance matrices to be elements of a standard Euclidean vector space without attention to the geometric structure of covariance matrices.

Consider a problem where the forecast of a covariance matrix needs to be updated given a current forecast H_t and an observation C_t (*e.g.,* a realized covariance matrix). An obvious

approach must be to combine the two matrices in a linear fashion such that $H_{t+1} = aH_t + (1-a)C_t$ (plus a constant covariance matrix), $0 \leq a \leq 1$. This idea is in the spirit of Engle and Kroner (1995), Noureldin et al. (2014), and many other covariance models. However, with the constraint that covariance matrices should remain positive definite, this is not a geometrically consistent way of combining covariance matrices as H_{t+1} does not extrapolate naturally, *i.e.*, H_{t+1} is no longer positive definite if $a > 1$ or $a < 0$ (geodesically incomplete).

In fact, there exists a notion of straight line in the space of covariance matrices, *i.e.*, *geodesic*. Geodesic can serve as a path connecting H_t and C_t, and its length can be used to measure the distance (similarity) of the pair. Han et al. (2017) adopt this idea and develop a new framework for covariance dynamics using methods of differential geometry. We follow their approach but extend their work in several important aspects as described below.

Firstly, we explain basic concepts of differential geometry and geometric properties of the covariance space in a descriptive, accessible manner so that anyone without a background in differential geometry can quickly grasp the idea behind our model. We also provide further justification for the geometric approach.

Second, we improve the model of Han et al. (2017). Our model is economically more meaningful and numerically more stable. We also develop a new model exploiting realized covariances. Since realized covariance matrices are positive definite and lie in the space of covariance matrices, they are better accommodated in our geometric framework. In contrast, the outer product of asset returns is singular and therefore it is less straightforward to incorporate it in the framework. Associated with the new model is a new estimation method that minimizes the geodesic length between the forecasted and the observed (realized) covariance matrices. As illustrated later, the geodesic length is considered a better measure of error than the Frobenius norm.

Third, we demonstrate the performance of our model via various empirical analyses. For the assessment of covariance matrices, we develop novel performance measures. Empirical studies involving three data samples show that our models outperform existing models such

as BEKK, DCC, and the matrix exponential GARCH.

The rest of the paper is organized as follows. Section 2 describes geometric properties of the covariance space and provides motivating examples. It then develops covariance dynamics models. Section 3 explains the evaluation methods used in the empirical studies, and Section 4 describes the data and models. Section 5 conducts empirical studies and evaluates the models, and Section 6 concludes. Appendix A provides the basics of differential geometry.

2 Geometric Modeling of Covariance Dynamics

In this section, we describe geometric properties of the covariance space that are particularly relevant to the development of our covariance dynamics model. We then motivate the geometric framework via examples and develop covariance models. Readers unfamiliar with differential geometry are encouraged to read the appendix first.

2.1 Space of Covariance Matrices

Covariance space Let $P(n)$ denote the space of $n \times n$ covariance matrices:[1]

$$P(n) = \left\{ P \in \mathbb{R}^{n \times n} \mid P = P^\top, P > 0 \right\}. \tag{1}$$

$P(n)$ is a differentiable manifold whose tangent space at a point $P \in P(n)$ can be identified with $n \times n$ symmetric matrices, $S(n)$. For the space of covariance matrices, the Riemannian metric at $P \in P(n)$ is defined as

$$\langle X, Y \rangle_P = \mathrm{tr}(P^{-1} X P^{-1} Y), \tag{2}$$

[1] We only consider nonsingular covariance matrices with $n > 1$. The geometric properties of the covariance space do not apply to the univariate case ($n = 1$).

where $X, Y \in S(n)$ are tangent vectors at P. For two covariance matrices, $A, B \in P(n)$, the minimal geodesic $\gamma(t) : [0,1] \to [A,B]$ connecting A and B is given by

$$\gamma(t) = G\big(G^{-1}BG^{-\top}\big)^t G^\top, \tag{3}$$

where G is a positive-definite invertible matrix that satisfies $GG^\top = A$, and $G^{-\top}$ denotes $(G^{-1})^\top$. G can be obtained from the eigenvalue decomposition by setting $G = V\sqrt{D}$, where V is the matrix of the eigenvectors and $\sqrt{D}$ is a diagonal matrix whose diagonal entries are the square roots of the eigenvalues.

The tangent vector of the above geodesic at A is given by the Riemannian log map

$$\mathrm{Log}_A(B) = G \log\left(G^{-1}BG^{-\top}\right) G^\top. \tag{4}$$

The log map $\mathrm{Log}_A(B)$ is the direction from A to B under the Riemmanian metric and is equivalent to $(B - A)$ under the flat metric, *i.e.*, when A and B are treated as elements in a vector space. Given a tangent vector $X \in S(n)$ at $A \in P(n)$, the minimal geodesic emanating from A in the direction X is defined by the Riemannian exponential map

$$\mathrm{Exp}_A(X) = G \exp\left(G^{-1}XG^{-\top}\right) G^\top. \tag{5}$$

Note that if $X = \mathrm{Log}_A(B)$, $\mathrm{Exp}_A(X) = B$. Defining the **distance** between A and B in the usual way by the length of the minimal geodesic, we have

$$d(A,B) = \left(\sum_{i=1}^{n} (\log \lambda_i)^2\right)^{1/2}, \tag{6}$$

where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of the matrix AB^{-1}. This distance corresponds to the Fisher information metric for the multivariate normal distribution (Smith, 2005). Note that $d(A, \gamma(t)) = t\, d(A,B)$. Therefore, the internal point on the minimal geodesic that divides A

and B in the ratio $\alpha : 1 - \alpha$ is given by

$$I(A, B, \alpha) = G\big(G^{-1}BG^{-\top}\big)^{\alpha}G^{\top} = \mathrm{Exp}_A(\alpha \mathrm{Log}_A B). \tag{7}$$

If A is an identity matrix, the Riemannian log and exponential maps respectively become the usual matrix logarithm and exponential. In this regard, covariance models that employ the matrix logarithm, *e.g.*, Chiriac and Voev (2011), Bauer and Vorkink (2011), Callot et al. (2017), are essentially modeling the times series of the tangent vectors emanating from the identity matrix towards covariance matrices.

It is important to note that the Riemannian metric in Equation (2) is equivalent to the Fisher information metric (FIM) for Gaussian covariance matrix estimation (Smith, 2005). Let $f(z|\mu, H)$ be a multivariate normal distribution parametrized by $\mu \in \mathbb{R}^n$ and $H \in P(n)$, and $l = \log f$ be the log-likelihood function. For an affine connection ∇ and tangent vectors at $P \in P(n)$, $X, Y \in S(n)$, the Fisher information metric is defined as

$$FIM(X, Y) := E[dl(X)dl(Y)], \quad dl(X) := \frac{d}{dt}\bigg|_{t=0} l(\gamma(t)), \tag{8}$$

where $\gamma(t)$ is the geodesic emanating from P in the direction X. It can be shown that (Smith, 2005)

$$FIM(X, X) = -E[\nabla^2 l] = K \cdot \mathrm{tr}(P^{-1}X)^2 = K \cdot \langle X, X \rangle_P \tag{9}$$

for some constant K. This equivalence implies that the distance measured by the length of the minimal geodesic between a pair of covariance matrices corresponds to the Fisher information metric and measures the similarity of the pair, or the similarity of two multivariate normal distributions with the same mean and different covariance matrices.

2.2 Properties of the Geodesic

As evidenced by the equivalence between the Riemannian metric and the Fisher information metric, respecting geometric properties of covariance matrices is not only mathematically appealing but also econometrically well-defined. Our covariance dynamics framework in Section 2.3 is inspired by this idea and assumes that the covariance matrix evolves along a geodesic and measures the distance (similarity) of two covariance matrices by the geodesic length. To understand the properties of the geodesic, this section computes the length of the geodesic between two arbitrary covariance matrices and compares it with the Frobenius norm (distance with respect to the flat metric). It is revealed that the geodesic length is better aligned with economic intuition.

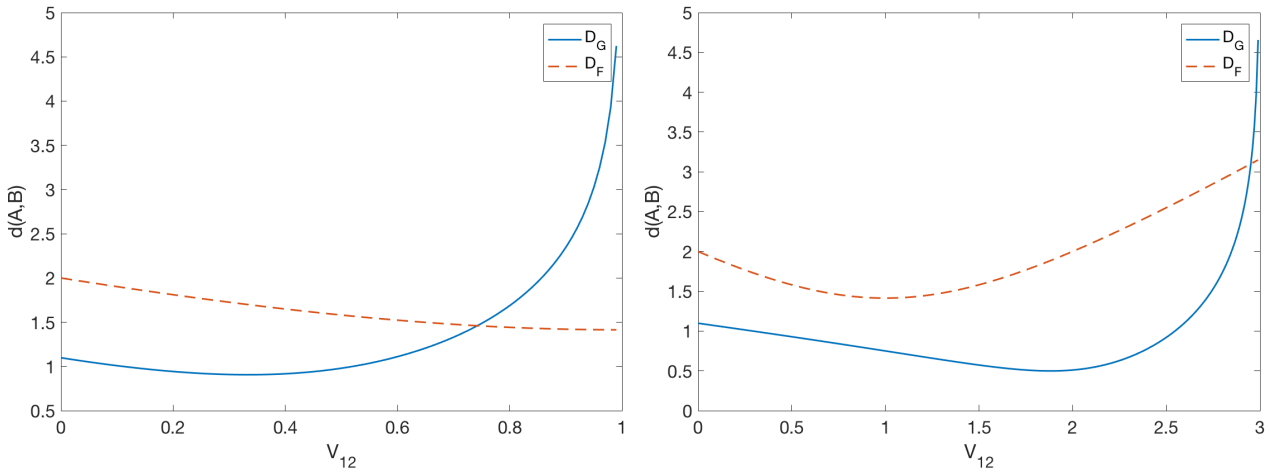
Consider two 2×2 covariance matrices[2]

$$A = \begin{bmatrix} 2.0 & 1.0 \\ 1.0 & 2.0 \end{bmatrix}, \quad B = \begin{bmatrix} V_1 & V_{12} \\ V_{12} & V_2 \end{bmatrix}.$$

We compute the distance between A and B via two methods for various values of B: the length of the shortest path (minimal geodesic) between A and B, D_G, and the Frobenius norm of $(A - B)$, D_F. D_G corresponds to the root-mean-square error (RMSE) with respect to the Riemmanian metric and D_F corresponds to the RMSE with respect to the flat metric. The Frobenius norm is widely used as a measure of error in realized covariance models: *e.g.,* Chiriac and Voev (2011); Bollerslev et al. (2016).

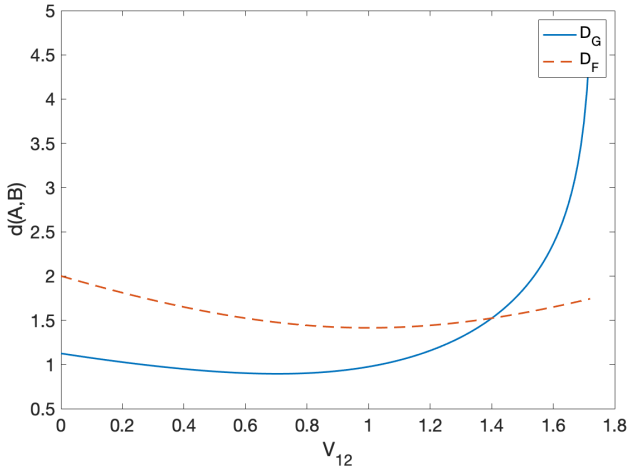
Figure 1(a) shows the distance between A and B for different values of V_{12}, assuming B has smaller variances than A. Figure 1(b) assumes B has larger variances. The Frobenius norm is minimized when the covariance in B is the same as in A, *i.e.,* $V_{12} = A(1,2) = 1.0$, regardless of the variances in B. This causes D_F to have its minimum when the two assets are perfectly correlated if $V_1 = V_2 = 1.0$, and when the correlation is 0.33 if $V_1 = V_2 = 3.0$.

[2]We fix A and the variances of B, while varying the covariance term of B. This is only to facilitate visualization and the findings here can be generalized for any two arbitrary covariance matrices.

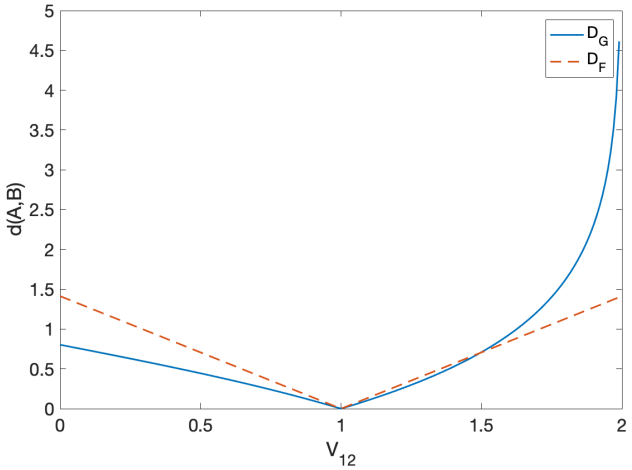


(a) $V_1 = V_2 = 1.0$

(b) $V_1 = V_2 = 3.0$



(c) $V_1 = 1.0, V_2 = 3.0$



(d) $V_1 = V_2 = 2.0$

Figure 1: Comparison of covariance distance measures: Frobenius norm (D_F) vs. geodesic length (D_G)

These correlations are considerably different from the correlation of A, *i.e.,* 0.5. In contrast, the length of the minimal geodesic D_G is minimized at $V_{12} \approx 0.4 < A(1,2)$ if $V_1 = V_2 = 1.0$ and at $V_{12} \approx 2.0 > A(1,2)$ if $V_1 = V_2 = 3.0$. The corresponding correlation coefficients are 0.4 and 0.66, respectively, which are comparably closer to 0.5. From an economic perspective, the covariance matrices minimizing D_G are clearly more similar to A than those minimizing D_F.

Interestingly, B at the minimum D_G has lower correlation than A when its variances are smaller than those of A, while it has higher correlation when its variances are larger. This positive relationship between the correlation and the variances is consistent with what we observe in the market: assets tend to be more highly correlated when the market is turbulent, and less so when the market is calm. Furthermore, D_G goes to infinity when assets are perfectly correlated, which makes sense as it is extremely unlikely to happen. The Frobenius norm D_F, on the other hand, is unable to incorporate correlation in an intuitive manner.

We also calculate the distances when the variance of one asset is smaller while the other is larger, and when both remain the same (Figure 1(c), (d)). In the first case, D_F is again minimized when $V_{12} = 1$, which corresponds to the correlation of 0.58, whereas D_G is minimized when the correlation is around 0.4. When the variances do not change, both measures have their minimum at the correlation of 0.5. The difference is that D_F increases linearly as the correlation drifts away from 0.5, whereas D_G increases exponentially approaching infinity when the two assets are perfectly correlated.[3]

The Frobenius norm is an unsatisfactory measure of error since it treats covariance matrices as elements of a vector space without respecting their geometric structure. On the other hand, the length of the minimal geodesic respects the geometric structure and corresponds to the Fisher information metric, making itself a more appropriate measure of similarity. Likewise, linear combination is not a geometrically consistent way of interpolating two co-

[3] D_G looks linear when $V_{12} < 1$, but it is only because the curvature is small around the minimum.

variance matrices. As one would typically assume that a point in a vector space moves towards another point along the straight line between them, a geometrically consistent way is to connect a pair using the minimal geodesic between them. The following sub-section develops a covariance dynamics model based on this idea.

2.3 Dynamics of Covariance Matrix

Suppose that an n-dimensional vector of asset returns r_t is governed by the following equation:

$$r_t = \mu + e_t, \qquad e_t \sim N(0, H_t), \tag{10}$$

where $H_t \in P(n)$ is the covariance matrix of e_t. We assume

$$H_{t+1} = \mathrm{Exp}_{H_I}(F_t), \tag{11}$$

where $H_I \in P(n)$ is a constant covariance matrix (a long-term mean), and $F_t \in S(n)$ is a tangent vector emanating from H_I, which depends on the information at time t. That is, H_{t+1} is a point on the direction F_t from H_I. We call this framework **Geometric Covariance Dynamics** (**GCD**). In what follows, two specifications of the tangent vector, one that utilizes asset returns and the other that utilizes realized covariances are developed.

2.3.1 Geometric Covariance Dynamics Based on Asset Returns

This model assumes that the tangent vector F_t is determined by the return shocks e_t and the covariance matrix H_t such that (see Figure 2)

$$F_t = \alpha \mathrm{Log}_{H_I}(H'_t), \quad 0 < \alpha < 1, \tag{12}$$

$$H'_t = I(H_t, C_t(e_t), a^2), \quad a^2 < 1, \tag{13}$$

where $C_t(e_t)$ is a covariance matrix derived from the return shocks. In particular, we consider the following form:

$$C_t(e_t) = C \circ \left((1-b^2)e_t e_t^\top + b^2 \eta_t \eta_t^\top\right), \quad b^2 < 1, \tag{14}$$

where $\eta_t = 1/2(|e_t| - e_t)$ is a variable to reflect the asymmetric effect of the shocks, and $C \in \mathbb{R}^{n \times n}$ is an adjustment term required to make C_t positive definite, which is assumed to have 1 on the diagonal and c, $0 < c < 1$, elsewhere. The operator $\circ$ denotes the Hadamard product. Under this specification, the covariance matrix at time $t+1$, H_{t+1}, is a point between H_I and H_t', and H_t' is a point between H_t and C_t. Therefore, ignoring C, GCD can be considered a geometrically consistent version of BEKK, where H_{t+1} is a linear function of H_I, H_t, and C_t.

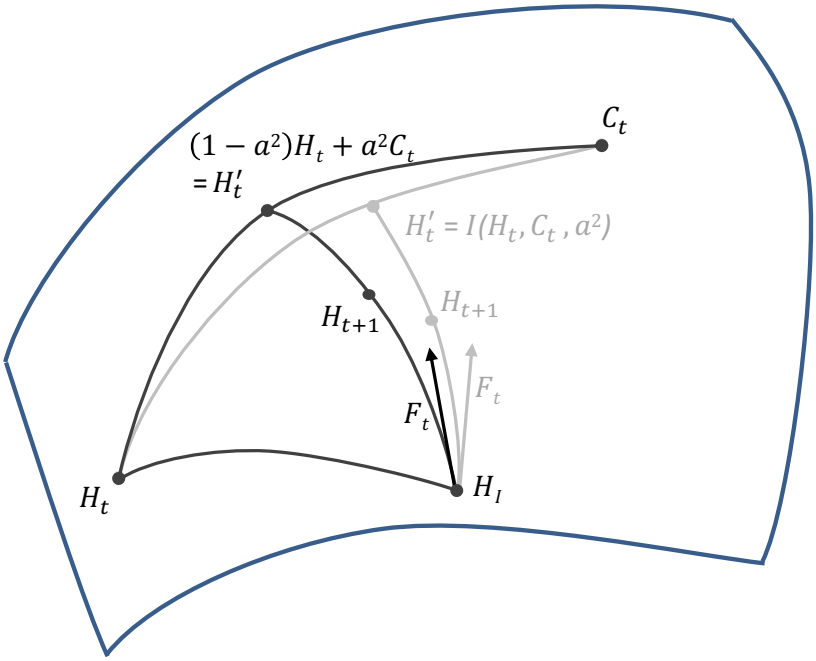


Figure 2: Geometric covariance dynamics

This figure describes how a covariance matrix evolves in the covariance space under the GCD framework. Covariance matrices are treated as points in the covariance space, where a straight line in the Euclidean sense becomes a curved (the black solid line between H_t and C_t).

A pitfall of the above specification is that C_t can become singular in a rare event of some elements of e_t being zero, in which case Equation (13) is not well defined. To avoid this problem, we replace (13) with the following equation

$$H'_t = (1 - a^2)H_t + a^2 C_t(e_t), \quad a^2 < 1. \tag{15}$$

In the empirical studies, we find that a^2 is close to 0, in which case Equation (13) can be approximated by $H_t + a^2 \text{Log}_{H_t}(C_t)$.[4] Hence, by using Equation (15), we are effectively replacing the tangent vector $\text{Log}_{H_t}(C_t)$ with the tangent vector in the Euclidean space, $(C_t - H_t)$. Although $\text{Log}_{H_t}(C_t)$ is different from $(C_t - H_t)$, the effect on H'_t is limited since it is multiplied by a small a^2.

Putting together, H_{t+1}, is given by

$$H_{t+1} = \text{Exp}_{H_I}(\alpha \text{Log}_{H_I}((1 - a^2)H_t + a^2 C_t)). \tag{16}$$

Equation (15) can be generalized by allowing the impact of e_t on H'_t to vary across assets:

$$H'_t = (\iota\iota^\top - dd^\top) \circ H_t + dd^\top \circ C_t, \tag{17}$$

where $d = (d_1, \ldots, d_n)$ is an n-dimensional vector with $d_i^2 < 1$, and ι denotes a vector of ones.

If H_t is positive definite, it is trivial to show that H'_t is also positive definite. Then, by definition, H_{t+1} is also positive definite. Since H_{t+1} is an internal point on the geodesic between H_I and H'_t, and H_I is constant, H_{t+1} is bounded.

[4] By the definition of the log and exponential maps, $H'_t = \text{Exp}_{H_t}(a^2 \text{Log}_{H_t} C_t) = H_t + a^2 \text{Log}_{H_t}(C_t) + \mathcal{O}(a^4)$.

2.3.2 Geometric Covariance Dynamics Based on Realized Covariance

When assets are traded during the same trading hours, *e.g.*, stocks traded on the same exchange, realized covariance can be computed and exploited for the development of covariance dynamics. Replacing C_t in (13) with a realized covariance measure R_t, we construct a realized covariance-based GCD model as follows:

$$F_t = \alpha \mathrm{Log}_{H_I}(H'_t), \quad 0 < \alpha < 1, \tag{18}$$

$$H'_t = I(H_t, R_t, a^2), \quad a^2 < 1. \tag{19}$$

The covariance matrix at $t+1$, H_{t+1}, is then given by

$$\begin{aligned} H_{t+1} &= \mathrm{Exp}_{H_I}(\alpha \mathrm{Log}_{H_I}(I(H_t, R_t, a^2)) \\ &= \mathrm{Exp}_{H_I}(\alpha \mathrm{Log}_{H_I}(\mathrm{Exp}_{H_t}(a^2 \mathrm{Log}_{H_t} R_t)). \end{aligned} \tag{20}$$

This specification is analogous to the scalar version of the HEAVY model of Noureldin et al. (2012) in that both combine the covariance matrix with a realized measure in a "linear" fashion, but the GCD model is defined in the covariance space, whereas the HEAVY model is defined in the vector space.

2.4 Estimation

2.4.1 Maximum Likelihood Estimation

Both versions of GCD can be estimated via the maximum likelihood estimation (MLE):

$$\max_{\theta} \sum_{t=1}^{T_I} -\frac{1}{2} \left(n \log(2\pi) + \log |H_t| + e_t^\top H_t^{-1} e_t \right), \tag{21}$$

where θ is the model parameters, and T_I is the sample size. For the GCD based on asset returns, $\theta = \{H_I, \alpha, a, b, c\}$ (scalar version) or $\theta = \{H_I, \alpha, d, b, c\}$ (diagonal version), and for

the GCD based on realized covariance, $\theta = \{H_I, \alpha, a\}$.

2.4.2 Geodesic Minimization

When realized covariance measures are available, another estimation method is to minimize the distance between the forecast H_t and the realization R_t. Following the discussion in the previous section, we estimate the model by minimizing the length of the minimal geodesic between H_t and R_t:[5]

$$\min_{\theta} \sum_{t=1}^{T_I} d(H_t, R_t). \tag{22}$$

2.4.3 Covariance Targeting

The GCD models have the number of parameters of $\mathcal{O}(n^2)$ due to the inclusion of the constant covariance matrix. This number can be reduced to $\mathcal{O}(n)$ via covariance targeting. Unlike BEKK and DCC, covariance targeting is not straightforward for the GCD models as the unconditional expectation of H_t is difficult to obtain due to the exponential mapping.

As Figure 2 shows, H_{t+1} can also be defined as an internal point on the geodesic emanating from H'_t towards H_I:

$$\begin{aligned} H_{t+1} =& \text{Exp}_{H_I}\left(\alpha \text{Log}_{H_I}(H'_t)\right) \\ =& \text{Exp}_{H'_t}\left((1-\alpha)\text{Log}_{H'_t}(H_I)\right). \end{aligned} \tag{23}$$

Applying $\text{Log}_{Ht'}(\cdot)$ to both sides, we have

$$\text{Log}_{Ht'}(H_{t+1}) = (1-\alpha)\text{Log}_{H'_t} H_I. \tag{24}$$

Diving both sides by $(1-\alpha)$ and applying $\text{Exp}_{H'_t}(\cdot)$, we obtain the equation for H_I as follows.

$$H_I = \text{Exp}_{H'_t}\left(\frac{1}{1-\alpha}\tilde{F}_t\right), \quad \tilde{F}_t = \text{Log}_{H'_t}(H_{t+1}), \tag{25}$$

[5]We test the Frobenius norm in an unreported empirical study and find the geodesic length outperforms the Frobenius norm.

where the matrix $\tilde{F}_t$ is the tangent vector emanating from H'_t towards H_{t+1}. Let $\bar{H}$ and $\bar{M}$ denote $E[e_t e_t^\top]$ and $E[\eta_t \eta_t^\top]$, respectively. Then,

$$\begin{aligned} E[H_{t+1}] &= \bar{H}, \\ E[H'_t] &= \left(\iota\iota^\top - dd^\top\right) \circ \bar{H} + dd^\top \circ C \circ ((1-b^2)\bar{H} + b^2\bar{M}). \end{aligned}$$

The constant covariance matrix H_I is approximated by substituting H_{t+1} and H'_t in (25) with their expected values:

$$H_I = \mathrm{Exp}_{\bar{H}'}\left(\frac{1}{1-\alpha}\tilde{F}\right), \quad \tilde{F} = \mathrm{Log}_{\bar{H}'}(\bar{H}), \tag{26}$$

where $\bar{H}'$ denotes $E[H'_t]$. Sample analogs are used to estimate $\bar{H}$ and $\bar{M}$.

For GCD with realized covariance, we have $E[H_t] = E[R_t] = \bar{H}$, which results in $H'_t = \bar{H}$ as H'_t is an internal point between H_t and R_t. Then $\tilde{F}_t$ becomes a $n \times n$ zero matrix and $H_I = H'_t = \bar{H}$ from Equation (25). This can also be derived graphically from Figure 2. When H_t, H_{t+1} and R_t (C_t in the figure) are equal to $\bar{H}$, both H'_t and H_I also have to be equal to $\bar{H}$ for Equation (20) to hold.

3 Evaluation

As the primary purpose of volatility estimation is to measure risk, we evaluate covariance models via out-of-sample performance metrics, especially in the context of risk and portfolio management.

3.1 Out-of-Sample Log-Likelihood

The first performance metric is the out-of-sample log-likelihood, which is obtained from the one-day forecasts of the covariance matrix during the out-of-sample period. Given one-day

forecasts $H_t,\ t = 1, \cdots, T$, the out-of-sample log-likelihood is given by the formula

$$\log L = \frac{T_I}{T} \sum_{t=1}^{T} -\frac{1}{2} \left(n \log(2\pi) + \log |H_t| + e_t^\top H_t^{-1} e_t \right), \tag{27}$$

where T_I and T are respectively the in-sample and out-of-sample sizes. The scaling factor T_I/T is multiplied to make the out-of-sample log-likelihood comparable to the in-sample value.

3.2 Portfolio Variance

The next performance metric is based on the variance of portfolio returns. During the out-of-sample period, a portfolio return is computed and normalized by the forecast of the portfolio variance every day. The standard deviation of the normalized returns S_p is then calculated as follows:

$$S_p = \sqrt{\frac{1}{T-1} \sum_{t=1}^{T} (r_{pt} - \bar{r}_p)^2}, \tag{28}$$

where

$$r_{pt} = \frac{w^\top r_t}{\sqrt{w^\top H_t w}}, \quad \bar{r}_p = \frac{1}{T} \sum_{t=1}^{T} r_{pt}, \tag{29}$$

and w is the portfolio weight vector. If the forecasts H_t are accurate, S_p will converge to 1 as the sample size increases. The performance metric is defined as the difference between S_p and 1:

$$dS_p = |S_p - 1|. \tag{30}$$

For empirical studies, the equal-weight portfolio is considered.

3.3 Portfolio Conditional Expectation

An accurate estimation of the covariance matrix does not necessarily lead to an accurate estimation of tail distribution unless the actual distribution of the data is equal to the

assumed distribution, which is normal in our case. Nonetheless, evaluating the models based on the tail distribution is important from a risk management perspective as most risk measures are derived from the covariance matrix. More generally, it is informative to compare the distribution implied by the model with the observed distribution.

In order to measure the distribution's overall goodness of fit, conditional expectation-based performance metrics are defined. For the normalized return defined in (29), conditional means at different probability levels are calculated on both sides of the distribution: given a probability level $\alpha > 0.5$, the conditional mean on the positive side, CE^p_α, and the conditional mean on the negative side, CE^n_α, are computed using the formulae

$$CE^p_\alpha = \frac{\sum_{t=1}^T r_{pt} \cdot \delta_{r_{pt}>z_\alpha}}{\sum_{t=1}^T \delta_{r_{pt}>z_\alpha}}, \quad CE^n_\alpha = \frac{\sum_{t=1}^T -r_{pt} \cdot \delta_{-r_{pt}>z_\alpha}}{\sum_{t=1}^T \delta_{-r_{pt}>z_\alpha}}, \tag{31}$$

where δ_i is the Kronecker delta, and z_α is the z-score at the probability level α. The portfolio return r_{pt} in CE^n_α is premultiplied by -1 so that $CE^n_\alpha > 0$. When α is large, *e.g.*, 0.95 or 0.99, CE^n_α becomes the conditional Value-at-Risk, also known as the expected shortfall.

If the covariance dynamics is correctly specified and asset returns normally distributed, CE^p_α and CE^n_α will converge to their theoretical value

$$CE_\alpha = \frac{1}{1-\alpha} \int_{z_\alpha}^{\infty} z\Phi(z)dz, \tag{32}$$

where $\Phi(z)$ is the standard normal probability density function. A natural choice of the performance metric is then the difference between the estimated value and the theoretical value:

$$dCE^p_\alpha = \frac{|CE^p_\alpha - CE_\alpha|}{CE_\alpha}, \quad dCE^n_\alpha = \frac{|CE^n_\alpha - CE_\alpha|}{CE_\alpha}. \tag{33}$$

In the empirical studies, dCE^p_α and dCE^n_α are measured for equal-weight portfolios.

3.4 Minimum Variance Portfolio

The next two performance metrics are related to the minimum variance portfolio. The standard deviation of the minimum variance portfolio return is often employed to assess covariance models, *e.g.*, Chiriac and Voev (2011). When there are no constraints other than the budget constraint, the minimum variance portfolio has the closed form:

$$w_t^{\min} = \frac{H_t^{-1}\iota}{\iota^\top H_t^{-1}\iota}, \tag{34}$$

where the superscript "min" denotes the minimum variance portfolio, and ι denotes a vector of ones of an appropriate size. Note that the minimum variance portfolio is invariant to a scalar multiplication of H_t. Therefore, the standard deviation of the minimum variance portfolio return is invariant to the overall level of the covariance matrix and only evaluates the cross-sectional variation within the covariance matrix.

The minimum variance portfolio is rebalanced when the covariance model is calibrated (every 22 days as illustrated in Section 4.3) and held until the next rebalancing date. The portfolio return is computed every day during the out-of-sample period, and its sample standard deviation is calculated as follows:

$$S_p^{\min} = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}(r_{pt}^{\min} - \bar{r}_p^{\min})^2}, \tag{35}$$

where $r_{pt}^{\min} = r_t^\top w_t^{\min}$. A more accurate covariance model will lead to a smaller standard deviation.

Another performance metric associated with portfolio rebalancing is turnover. A covariance model that yields inconsistent forecasts over time will result in high turnover. Although turnover is not directly related to the accuracy of the model, an unstable covariance forecast is not only counter-intuitive but also harmful to portfolio management as it incurs high transaction costs. In this regard, measuring the stability of the covariance model via turnover

is worthwhile. When a portfolio is rebalanced at time t, the turnover is defined as

$$dw_t = |w_t - w_{t^-}|^\top \iota, \tag{36}$$

where w_{t^-} and w_t are the portfolio weights immediately before and after rebalancing at time t. The performance metric is defined as the average turnover during the out-of-sample period:

$$dw = \frac{1}{K} \sum_{k=1}^{K} dw_{t_k}, \tag{37}$$

where K is the number of rebalancing during the out-of-sample period, and t_k is the k-th rebalancing time.

3.5 Distance Metrics

The last performance metrics are based on distance measures and applicable when a realized covariance matrix is available. Using the Frobenius norm or the geodesic length as the distance measure, the average distance over the out-of-sample period is calculated as follows:

$$D_F = \frac{1}{T} \sum_{t=1}^{T} ||H_t - R_t||_F, \tag{38}$$

$$D_G = \frac{1}{T} \sum_{t=1}^{T} d(H_t, R_t), \tag{39}$$

where $|| \cdot ||_F$ denotes the Frobenius norm.

4 Data and Models

4.1 Data

For the empirical part of the paper, three data samples; global stock market indexes, currencies, and individual stocks are chosen. The assets in the last two samples are traded

during the same trading hours and realized covariances can be obtained from high-frequency trading data.

4.1.1 Stock Market Indexes

The first data sample consists of three stock market indexes; S&P 500, FTSE 100, and NIKKEI 225. These indexes are also used in Kawakatsu (2006). Daily index values are collected from DataStream, and daily index returns are generated during the period from 1997-01-02 to 2014-12-31. Descriptive statistics are reported in Table 1. S&P 500 and FTSE 100 have a similar level of variance and are highly correlated with each other, whereas NIKKEI 225 has a relatively higher variance and a lower return-to-risk ratio. The correlations between NIKKEI 225 and the other two indexes are also much lower compared to the correlation between S&P 500 and FTSE 100.

Table 1: Descriptive statistics of the stock index daily returns

The indexes are S&P 500 (S&P), FTSE 100 (FTSE), and NIKKEI 225 (NIKKEI). The sample period is from 1997-01-02 to 2014-12-31, and the mean (Mean) and the standard deviation (Stdev) values are annualized, assuming 250 business days per year.

	Mean	Stdev	Correlation		
			S&P	FTSE	NIKKEI
S&P	0.074	0.197	1.000	0.512	0.112
FTSE	0.043	0.190	0.512	1.000	0.284
NIKKEI	0.023	0.240	0.112	0.284	1.000

4.1.2 Currencies

The second data sample consists of three currencies; euro (EUR), British pound (GBP), and Japanese yen (JPY) expressed as US dollar price per unit currency. Laurent et al. (2012) also analyze the covariance of these currencies. The currency sample is distinct from the other data samples in that it features a negative correlation. The sample period is from 2002-01-02 to 2014-12-31.[6] These currencies are traded 24 hours, so both daily returns and

[6]The sample starts from 2002-01-02 because euro was introduced in a non-physical form on 1999-01-01, and the new euro notes and coins were introduced on 2002-01-01.

realized covariances are calculated. In an extensive empirical study, Liu et al. (2015) find little evidence that a simple 5-minute realized variance is outperformed by other measures. Based on their findings, we employ a simple 5-minute realized covariance.

Descriptive statistics are reported in Table 2. While EUR and GBP are highly correlated with each other, they are negatively correlated with JPY. All three currencies have a similar level of variance.

Table 2: Descriptive statistics of the currency daily returns

The currencies are euro (EUR), British pound (GBP), and Japanese yen (JPY). The sample period is from 2002-01-02 to 2014-12-31, and the mean (Mean) and the standard deviation (Stdev) values are annualized assuming 250 business days per year.

	Mean	Stdev	Correlation		
			EUR	GBP	JPY
EUR	0.024	0.099	1.000	0.673	-0.237
GBP	0.016	0.091	0.673	1.000	-0.128
JPY	-0.008	0.102	-0.237	-0.128	1.000

4.1.3 Individual Stocks

The last data sample consists of six stocks from the Dow Jones Industrial Average (DJIA) index; General Electric (GE), American Express (AXP), JP Morgan (JPM), Home Depot (HD), Citi Bank (C), and IBM (IBM). These stocks are chosen for their liquidity following Chiriac and Voev (2011). The sample period is from 2002-01-02 to 2014-12-31. During the sample period, stock prices are collected every minute during the trading hours (09:30 to 16:00 EST) from Thomson Reuter Tick History, and the daily returns and the 5-minute realized covariances are computed. The daily return is defined as the open-to-close return and overnight jump is ignored. For the close-to-close return, one could introduce an additional parameter in the model to scale up the covariance matrix: *e.g.,* the covariance matrix of the close-to-close returns could be modeled as λH_t with λ being a scalar or a diagonal matrix.

Descriptive statistics are reported in Table 3. All the stocks are highly correlated with each other, with correlation coefficients being greater than 0.5. Four out of six stocks (GE,

HD, C, and IBM) show negative mean returns during the sample period, and the standard deviations of the returns range between 0.166 and 0.376. Citi (C) has a significantly low mean return (-0.426) and a high standard deviation (0.376) compared to the other stocks.

Table 3: Descriptive statistics of the DJIA stock returns

The stocks are General Electric (GE), American Express (AXP), JP Morgan (JPM), Home Depot (HD), Citi Bank (C), and IBM (IBM). The sample period is from 2002-01-02 to 2014-12-31, and the mean (Mean) and the standard deviation (Stdev) values are annualized assuming 250 business days per year.

	Mean	Stdev	Correlation					
			GE	AE	JPM	HD	C	IBM
GE	-0.055	0.215	1.000	0.595	0.629	0.559	0.567	0.596
AE	0.073	0.276	0.595	1.000	0.722	0.579	0.642	0.576
JPM	0.046	0.301	0.629	0.722	1.000	0.592	0.698	0.582
HD	-0.014	0.218	0.559	0.579	0.592	1.000	0.457	0.588
C	-0.426	0.376	0.567	0.642	0.698	0.457	1.000	0.442
IBM	-0.018	0.166	0.596	0.576	0.582	0.588	0.442	1.000

4.2 Models

Four GCD specifications are considered in the empirical analysis: three from Section 2.3, *i.e.,* Equation (15), (17), and (19), and a hybrid model of the DCC and GCD, where the correlation matrix is assumed to follow the scalar GCD process in (15) and the variances follow the GJR-GARCH (Glosten et al., 1993) process. These models are compared with the matrix exponential GARCH of Kawakatsu (2006), two versions of the BEKK of Engle and Kroner (1995), and the DCC extension of Cappiello et al. (2006). We chose BEKK because its specification is similar to our GCD model except it is defined in the Euclidean space, so we can reveal how defining the dynamics in the covariance space can improve the estimation. The DCC model is chosen because it is one of the most widely used models and performs well. Finally, the matrix exponential GARCH is chosen as it is similar to the GCD framework in that it uses matrix logarithm. A crucial difference is that the matrix exponential GARCH defines the dynamics in the tangent space.

In all models, the covariance matrix is assumed to be a function of only the first lags of the covariance matrix and return shocks, and the asymmetric effect term, η_t, is included.

The test models are summarized in Table 4, together with the number of model parameters.

Table 4: Test models

Model	No. Parameters	Description
GCDS	$1/2n(n+1)+4$	GCD with scalar coefficients (Eq. (15)).
GCDD	$1/2n(n+3)+3$	GCD with diagonal coefficients (Eq. (17)).
EXP	$n(3n+1)$	Matrix exponential GARCH (Kawakatsu, 2006).
BEKKS	$1/2n(n+1)+3$	BEKK with scalar coefficients (Engle and Kroner, 1995).
BEKKD	$1/2n(n+7)$	BEKK with diagonal coefficients (Engle and Kroner, 1995).
DCC	$1/2n(n+7)+3$	DCC with scalar coefficients (Cappiello et al., 2006).
DCCG	$1/2n(n+7)+4$	DCC-GCD with scalar coefficients.
GCRM	$1/2n(n+1)+2$	GCD based on realized covariance estimated via MLE (Eq. (19)).
GCRG	$1/2n(n+1)+2$	GCD based on realized covariance estimated via geodesic length minimization (Eq. (19)).

4.3 Model Estimation and Evaluation

Model parameters are estimated monthly (every 22 days) during the sample period rolling a three-year estimation window.[7] Hence, the out-of-sample period starts from 2000-01-02 for the stock index dataset and 2005-01-02 for the currency and DJIA stock datasets. Between estimation dates, the covariance matrix is updated daily with the arrival of new returns and realized covariances while the model parameters are fixed to the last estimates. The performance metrics defined in Section 3 are calculated during the out-of-sample period.

5 Empirical Results

This section analyzes empirical results. For the sake of space, parameter estimation results are not reported, and the focus is given to out-of-sample evaluation.

5.1 Performance Evaluation

Table 5, 6, and 7 report the out-of-sample performance of the test models evaluated respectively using the stock index, currency, and DJIA stock samples. GCRM and GCRG are

[7] A five-year estimation window was also tested and the results were qualitatively similar.

applied only to the currency and DJIA samples, for which realized covariances are available. EXP is applied only to the first two samples as its parameters do not converge in the DJIA sample. The tables also report the ranks of the models under each evaluation metric and Hansen et al. (2011)'s model confidence set (MCS) p-values when applicable. When the MCS test is conducted on all models, the realized covariance-based models (GCRM, GCRG) dominate the asset return-based models and the p-values of the latter become near zero. Therefore, to compare the asset return-based models, we report p-values obtained from the MCS tests conducted on the asset return-based models only.

Table 5: Performance evaluation using the stock index sample

The rows represent the evaluation metrics defined in Section 3 except the first row, $\log L_0$, which is the average in-sample likelihood. dS_p and dCE^n_{99} are calculated from the equal-weight portfolio and reported without taking absolute value to show the direction of the errors. The second row of each evaluation metric is the ranks of the models based on the associated metric. 'mcs p-val' denotes the MCS p-value. The MCS test is conducted on the asset return-based models only.

	GCDS	GCDD	EXP	BEKKS	BEKKD	DCC	DCCG
$\log L_0$	7386	7391	7425	7345	7375	7395	7397
$\log L$	7449	7453	7420	7407	7427	7459	7460
	4	3	6	7	5	2	1
mcs p-val	0.00	0.06	0.00	0.00	0.00	0.06	1.00
dS_p	0.676	0.658	1.795	-3.468	-1.536	-0.557	-0.668
	4	2	6	7	5	1	3
dCE^n_{99}	8.251	6.805	12.978	7.875	9.520	9.879	10.077
	3	1	7	2	4	5	6
$S_p^{\min}$	15.057	15.025	18.545	15.302	15.146	15.115	15.115
	2	1	7	6	5	3	4
dw	3.481	3.865	4.376	4.242	4.285	4.793	4.734
	1	2	5	3	4	7	6

In terms of the out-of-sample log-likelihood, $\log L$, GCRM and GCRG perform best, followed by DCCG. Other models show inconsistent results across samples. To our surprise, the out-of-sample log-likelihood is usually greater than the mean of the in-sample log-likelihood, $\log L_0$. This can be attributed to the fact that $\log L$ is calculated using the parameters updated every 22 days, whereas only one set of parameters are used to calculate the in-sample likelihood in each parameter estimation.

The error of the portfolio variance measured by dS_p is trivial in most models. For instance, the largest error in the stock index sample is only -3.5% of BEKKS. The only

Table 6: Performance evaluation using the currency sample

The rows represent the evaluation metrics defined in Section 3 except the first row, $\log L_0$, which is the average in-sample likelihood. dS_p and dCE^n_{99} are calculated from the equal-weight portfolio and reported without taking absolute value to show the direction of the errors. The second row of each evaluation metric is the ranks of the models based on the associated metric. 'mcs p-val' denotes the MCS p-value. The MCS test is conducted on the asset return-based models only.

	GCDS	GCDD	EXP	BEKKS	BEKKD	DCC	DCCG	GCRM	GCRG
$\log L_0$	8764	8767	8785	8757	8773	8772	8776	8780	
$\log L$	8777	8778	8697	8776	8777	8775	8786	8811	8797
	6	4	9	7	5	8	3	1	2
mcs p-val	0.71	0.71	0.00	0.71	0.71	0.00	1.00		
dS_p	0.542	1.098	-0.289	-1.330	0.428	2.998	1.306	0.640	-0.494
	4	6	1	8	2	9	7	5	3
dCE^n_{99}	13.826	12.962	9.530	12.645	9.940	12.011	10.095	6.650	8.268
	9	8	3	7	4	6	5	1	2
$S_p^{\min}$	6.262	6.268	6.346	6.360	6.318	6.325	6.300	6.199	6.188
	3	4	8	9	6	7	5	2	1
dw	2.781	3.107	3.600	3.036	3.459	3.395	3.436	2.239	2.081
	3	5	9	4	8	6	7	2	1
D_F	0.435	0.435	0.505	0.435	0.436	0.438	0.432	0.385	0.367
	5	4	9	6	7	8	3	2	1
mcs p-val	0.56	0.60	0.00	0.60	0.10	0.00	1.00		
D_G	0.874	0.887	1.042	0.870	0.906	0.895	0.902	0.806	0.660
	4	5	9	3	8	6	7	2	1
mcs p-val	0.00	0.00	0.00	1.00	0.00	0.00	0.00		

Table 7: Performance evaluation using the DJIA stock sample

The rows represent the evaluation metrics defined in Section 3 except the first row, $\log L_0$, which is the average in-sample likelihood. dS_p and dCE^n_{99} are calculated from the equal-weight portfolio and reported without taking absolute value to show the direction of the errors. The second row of each evaluation metric is the ranks of the models based on the associated metric. 'mcs p-val' denotes the MCS p-value. The MCS test is conducted on the asset return-based models only.

	GCDS	GCDD	BEKKS	BEKKD	DCC	DCCG	GCRM	GCRG
$\log L_0$	14257	14265	14197	14146	14260	14262	14325	
$\log L$	14354	14346	14277	14075	14336	14345	14465	14416
	3	4	7	8	6	5	1	2
mcs p-val	1.00	0.77	0.00	0.00	0.00	0.13		
dS_p	1.706	-0.453	-2.236	2.132	-1.380	-1.565	0.530	19.893
	5	1	7	6	3	4	2	8
dCE^n_{99}	9.075	7.448	6.807	13.514	6.202	5.642	3.534	11.105
	6	5	4	8	3	2	1	7
$S_p^{\min}$	15.012	14.905	14.956	15.395	15.520	15.522	15.091	15.259
	3	1	2	6	7	8	4	5
dw	5.120	5.332	6.457	7.446	8.753	8.791	6.265	5.329
	1	3	5	6	7	8	4	2
D_F	9.320	9.590	10.249	10.363	10.380	10.399	8.151	7.602
	3	4	5	6	7	8	2	1
mcs p-val	1.00	0.01	0.01	0.00	0.01	0.00		
D_G	1.855	1.872	1.943	2.099	1.872	1.868	1.655	1.485
	3	6	7	8	5	4	2	1
mcs p-val	1.00	0.00	0.00	0.00	0.03	0.03		

exception is the significant error (19.9%) from GCRG in the DJIA stock sample. GCRG is the only model that minimizes the distance from the realized covariance and performs best in terms of D_F and D_G. This contradicting performance of GCRG suggests that the realized covariance may not be a good proxy for the true covariance in this particular sample. This result, however, does not mean the realized covariance is uninformative. The GCRM model, which is also based on the realized covariance but estimated via MLE utilizing both the returns and realized covariance, outperforms most other models.

Contrary to the results from dS_p, the error in the tail region measured by dCE^n_{99} is much larger, often exceeding 10%. GCRM performs best in both the currency and DJIA stock samples, but the error is still substantial compared to dS_p. This casts a doubt on the conditional normality assumption of asset returns. Possible remedies to reduce the error at the tail is discussed later in Section 5.2.

The rankings based on S_p^{min} and dw are generally similar and consistent across samples favoring GCD models; GCRM, GCDS, and GCDD, in particular. As a robustness check, daily rebalancing and rebalancing subject to short-sale constraints were also tested, and the results were qualitatively similar to those reported here. Compared to other performance metrics, the variation of $S_p^{\min}$ across the models is trivial. This is because the minimum variance portfolio is invariant to the overall level of the covariance matrix and depends only on the cross-sectional variation, which is comparably stable over time.

As expected, GCRG performs best in terms of the distance metrics, D_F and D_G, followed by GCRM. The rankings based on D_F and D_G are similar but not identical. Although both measures are valid, D_G is preferred in our context as it is consistent with the geometric framework.

When we conduct the MCS test on all models, the p-values of the asset return-based models are almost 0, indicating that the realized covariance models dominate these. Therefore, we report the p-values obtained from the MCS test conducted only on the asset return-based models. For the stock indexes, DCCG is the only model in the 90% MCS. For the currencies,

many models are contained in the 90% MCS implying that these models are not significantly different. Still, all GCD models are included in the MCS based on $\log L$ or D_F. For the DJIA stocks, only the GCD models are in the 90% MCS. In particular, GCDS performs best in terms of all losses.

Comparing overall performances of the models, we find that the models based on realized covariance, *i.e.,* GCRM and GCRG, dominate the other models in terms of most performance metrics. While these models are expected to perform well in terms of D_F and D_R, the high ranking of GCRG with respect to Log$L$ is remarkable as it does not maximize the likelihood function. Noureldin et al. (2012) also find that their realized covariance-based HEAVY model outperforms conventional GARCH models based on daily returns. Between GCRM and GCRG, GCRM performs more consistently, which is perhaps because GCRM exploits information from both returns and realized covariances. GCRG, on the other hand, uses only realized covariance, and its performance depends heavily on the quality of the realized covariance as a proxy for the true covariance.

Apart from GCRM and GCRG, it is difficult to pinpoint the best performing model at first glance. It is rather striking that the performances of the models vary widely not only across data samples but also across performance metrics. Table 8 reports the correlation between the model rankings. To our surprise, $\log L$ is highly correlated with both D_F and D_G, whereas dS_p and dCE^n_{99} are weakly, sometimes even negatively correlated with other metrics. The inconsistent performance across performance metrics suggests that it is crucial to evaluate models using multiple criteria and choose the most appropriate metric for a given purpose.

Nevertheless, the GCD models appear to outperform the benchmark models, especially in terms of the out-of-sample log-likelihood: as models are estimated via MLE except for GCRG, $\log L$ is most relevant to the objective function. Comparison between GCDS and BEKKS and between GCDD and BEKKD reveals that the GCD models outperform their BEKK counterparts. Incorporating the GCD model into the DCC framework (DCCG) also

Table 8: Correlation between rankings

Correlations are calculated from the model rankings reported in Table 5, 6, and 7.

	$\log L$	dS_p	dCE_{99}^n	$S_p^{\min}$	dw	D_F	D_G
$\log L$	1.00						
dS_p	0.31	1.00					
dCE_{99}^n	0.26	0.41	1.00				
$S_p^{\min}$	0.60	0.34	0.21	1.00			
dw	0.40	-0.15	0.08	0.63	1.00		
D_F	0.86	-0.10	0.07	0.72	0.81	1.00	
D_G	0.71	-0.14	0.17	0.31	0.77	0.69	1.00

improves the performance, as evidenced by the increased log-likelihoods in all three samples. EXP, contrary to its high in-sample log-likelihood values, performs worst out-of-sample. In general, good in-sample performance does not imply good out-of-sample performance. Another important observation is that the diagonal versions of GCD and BEKK, despite more parameters, do not outperform their scalar counterparts. Overall, the GCD models perform best followed by the BEKK models and DCC, and EXP performs worst.

Table 9 provides a summary of the results, in which the numbers are average rankings across data samples. GCRM, GCRG, and EXP are excluded as they are applied to only some of the three data samples, and their out/under-performance is apparent. Without these models, the remaining six models are ranked from 1 (best) to 6 (worst) in each sample, and the ranks are averaged across the samples. The table clearly shows that the best performance is usually associated with the GCD models, while the worst performance is associated with the BEKK models, which confirms the previous conclusion.

Table 9: Average ranking of the test models

Without GCRM, GCRG, and EXP, the remaining six models are ranked again from 1 (best) to 6 (worst) in each sample, and the ranks are averaged across the samples.

	GCDS	GCDD	BEKKS	BEKKD	DCC	DCCG
$\log L$	3.0	2.3	5.3	4.7	4.0	1.7
dS_p	3.3	2.0	5.7	3.7	3.0	3.3
dCE_{99}^n	4.7	3.3	3.0	3.7	3.3	3.0
$S_p^{\min}$	2.0	1.3	4.7	4.3	4.3	4.3
dw	1.0	2.3	2.7	4.7	5.0	5.3
D_F	2.0	2.0	3.5	4.5	5.5	3.5
D_G	1.5	3.5	3.0	6.0	3.5	3.5

Covariance Targeting The results of covariance targeting for some selected models are reported in Table 10. It is remarkable that fixing the constant matrix does little harm to the overall performance regardless of the model. The fixed constant matrix models often outperform their counterparts in terms of some performance metrics. The loss in flexibility appears to be offset by the gain in robustness.

Table 10: Covariance targeting

Est columns are the results without covariance targeting, and Fix columns are the results from covariance targeting.

	GCDS		GCDD		DCC		DCCG		GCRM	
	Est	Fix	Est	Fix	Est	Fix	Est	Fix	Est	Fix
$\log L$	14354	14334	14346	14312	14336	14341	14345	14345	14465	14424
	3	9	4	10	8	7	6	5	1	2
dS_p	1.706	-1.651	-0.453	-0.886	-1.380	-1.859	-1.565	-2.020	0.530	-1.198
	8	7	1	3	5	9	6	10	2	4
dCE^n_{99}	9.075	9.354	7.448	9.465	6.202	6.034	5.642	5.469	3.534	3.496
	8	9	7	10	6	5	4	3	2	1
$S_p^{\min}$	15.012	14.928	14.905	14.923	15.520	15.544	15.522	15.538	15.091	14.915
	5	4	1	3	7	10	8	9	6	2
dw	5.120	5.542	5.332	5.593	8.753	8.814	8.791	8.877	6.265	6.011
	1	3	2	4	7	9	8	10	6	5
D_F	9.320	9.733	9.590	9.516	10.380	10.438	10.399	10.461	8.151	8.190
	3	6	5	4	7	9	8	10	1	2
D_G	1.855	1.922	1.872	1.933	1.872	1.871	1.868	1.871	1.655	1.734
	3	9	8	10	7	6	4	5	1	2

Paths of the Forecasts Figure 3 displays one-day forecasts of the standard deviation of the return on GE, and Figure 4 displays one-day forecasts of the correlation between GE and AE obtained from selected models. We report the results from only one pair to save the space, but the findings below can be generalized to other pairs.

When GCDS is compared with BEKKS, the standard deviation forecast from GCDS is more responsive to the shocks, whereas the correlation from GCDS is more stable. The difference between the models is more prominent during volatile market periods. This can be attributed to two facts. Firstly, the geodesic length between two matrices increases exponentially as one matrix approaches a singular point. This prevents a sudden increase in correlation. Secondly, with the covariance adjustment term C, GCDS is able to produce

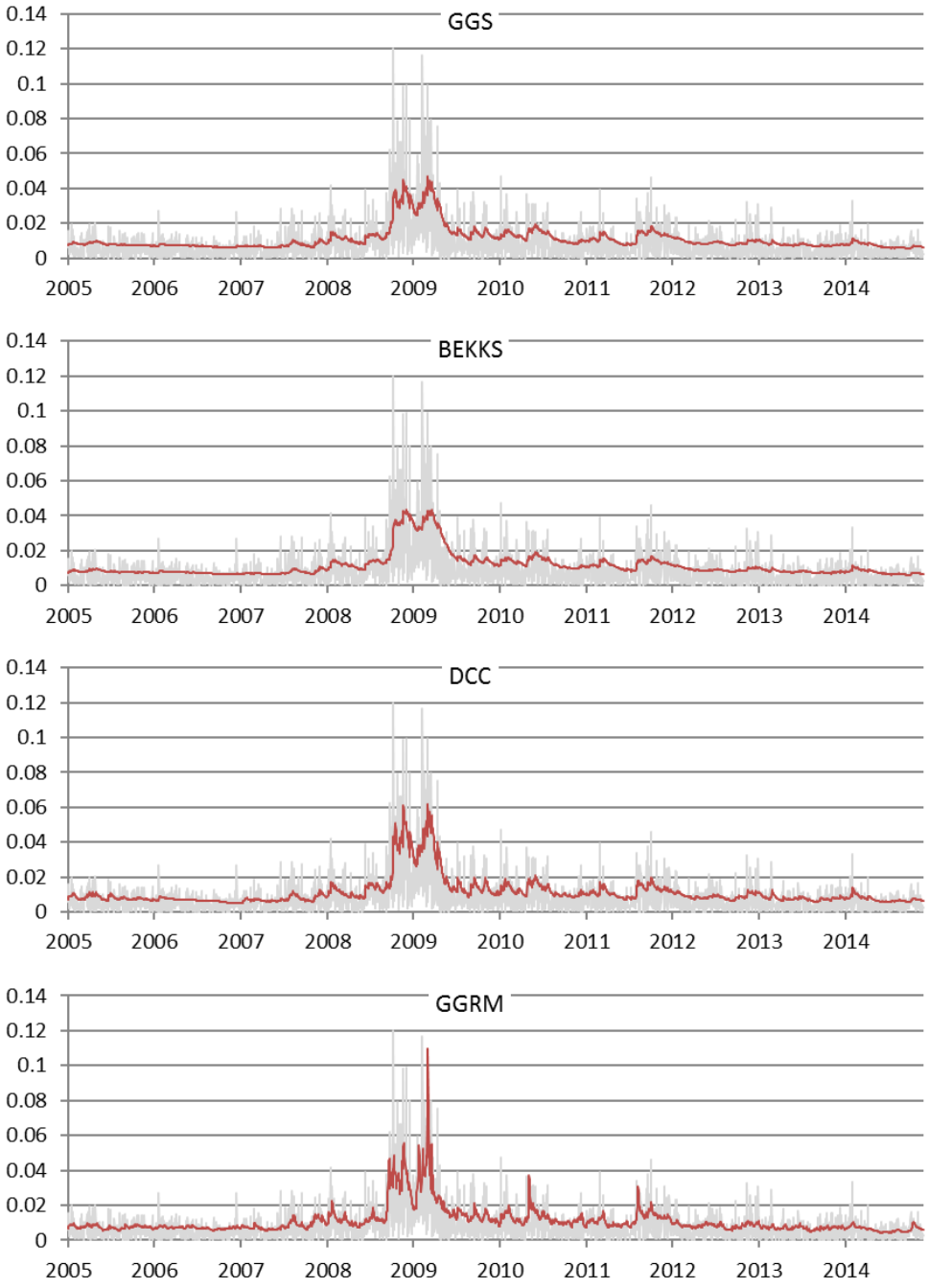


Figure 3: One-day forecast of the standard deviation of GE

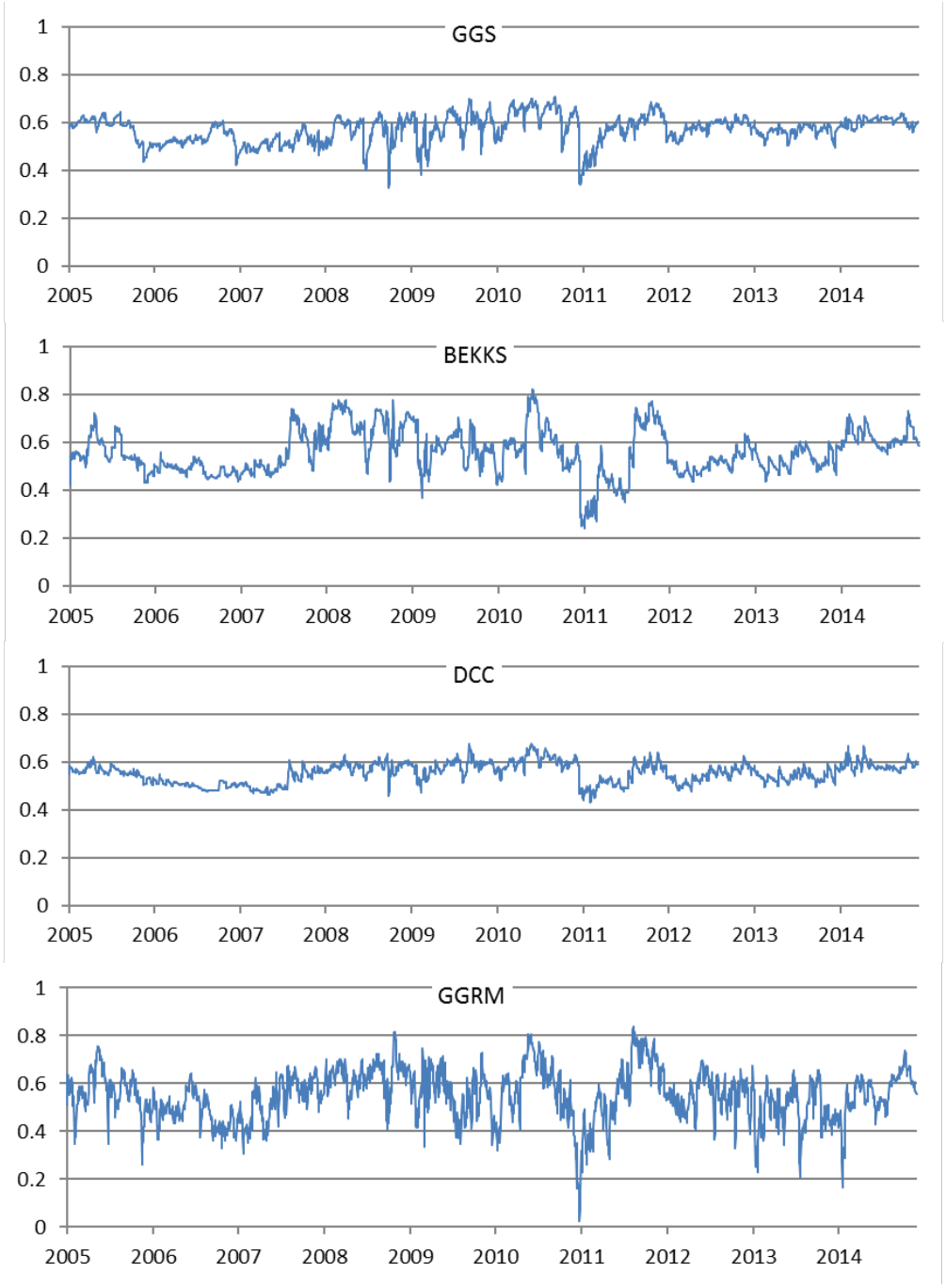


Figure 4: One-day forecast of the correlation between GE and AE

stable correlations while allowing volatile movements of the standard deviations.

Treating variances and correlations independently, DCC allows variances to be responsive to the shocks while maintaining stable correlations. As a result, it generates the most stable correlations among the models tested.

GCRM features the most responsive variance and correlation forecasts. This suggests that the realized covariance contains more information about the future covariance than the return shocks. The loading on R_t in GCRM approximated by αa^2 is 0.244 on average, while the loading on $e_t e_t^\top$ in GCDS is only 0.024. Even though the magnitude of $e_t e_t^\top$ is much larger than that of R_t, the difference in loadings is substantial. It is worth noting that the volatile forecasts from GCRM does not result in high turnover of the minimum variance portfolio.

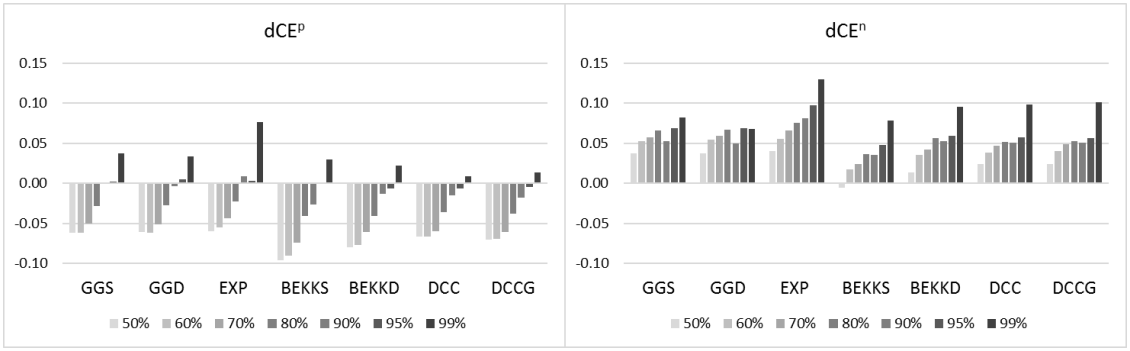
5.2 Risk Measurement

As evidenced by the previous results, the forecast error at the tail measured by dCE_{99}^n is substantially larger than that of variance (dS_p), and there is a lack of consistency in ranking between these performance metrics. The significant tail error necessitates validation of the models in terms of risk measures, especially if the primary purpose of covariance estimation is to measure risk. Figure 5 displays the conditional expectations defined in (33), without the absolute value operation in the numerator.

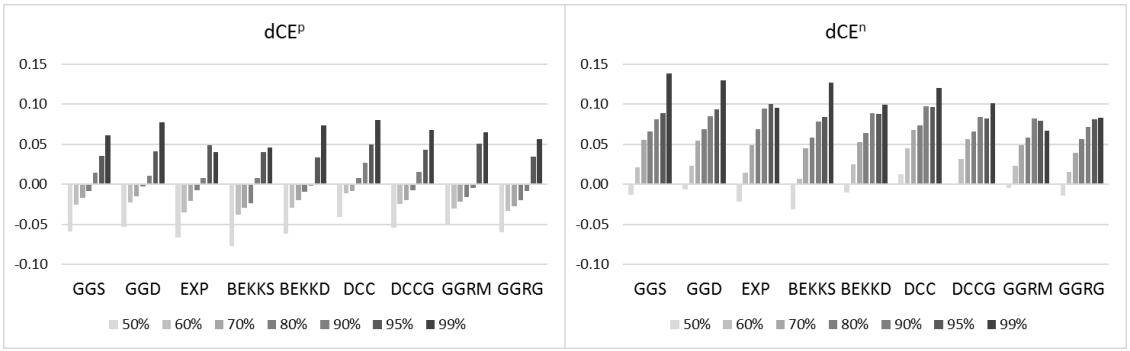
If a model were correctly specified, the errors would be close to zero across all probability levels. However, the graphs show that all models underestimate the conditional expectation at both tails, and the underestimation is severer on the negative side. This finding is consistent with the well-known fat-tailed, left-skewed distribution of asset returns.

For risk management, it is worth considering a tailored estimation method to enhance the accuracy at the tail. Four methods that aim to achieve this are considered.

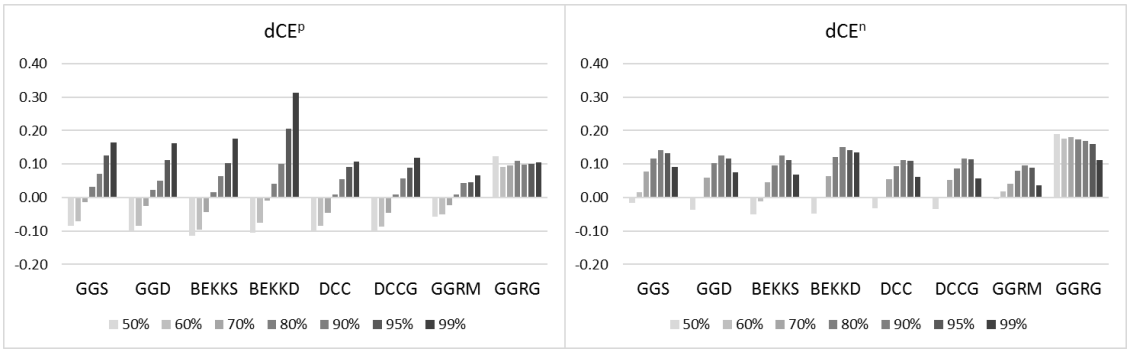
The first method is to replace the normal distribution with a fat-tailed distribution. In



(a) Stock Indexes



(b) Currencies



(c) DJIA stocks

Figure 5: Conditional expectations CE_α^p and CE_α^n

particular, the student t-distribution is considered, in which case, the QMLE is given by

$$\max_{\theta} \sum_{t=1}^{T_I} c(n,d) - \frac{1}{2}\log|H_t| - \frac{n+d}{2}\log\left(1 + \frac{e_t^\top H_t^{-1} e_t}{d-2}\right), \tag{40}$$

where

$$c(n,d) = \log\Gamma\left(\frac{n+d}{2}\right) - \log\Gamma\left(\frac{d}{2}\right) - \frac{n}{2}\log\pi - \frac{n}{2}\log(d-2),$$

$\Gamma(\cdot)$ is the gamma function, and d is the degree of freedom of the distribution, which is estimated simultaneously with other parameters.[8] When the t-distribution is used, the probability density function in (32) also needs to be replaced by the probability density function of the t-distribution.

The second and third methods select samples around the tail of the distribution while retaining the normality assumption. More specifically, the second method uses historical returns only when the absolute value of the portfolio return exceeds its sample standard deviation:

$$|w^\top e_t| > \sqrt{w^\top \bar{H} w}, \tag{41}$$

where $\bar{H}$ is the sample covariance of e_t. The third method chooses samples only when at least one asset return is below the negative of its sample standard deviation:

$$e_{it} < -\sqrt{\bar{h}_i}, \tag{42}$$

where $\bar{h}_i$ is the sample variance of e_{it}. One drawback of these methods is the loss of information from the omitted data.

The last method borrows the idea of importance sampling. Importance sampling is a simulation technique that reduces simulation error by drawing more samples from the region of interest, which is done by replacing the original density function f with a new density

[8] Following Kawakatsu (2006), we use a likelihood function slightly different from the usual form so that the variance of e_t becomes H_t.

function g, and multiplying the samples from the new density function by the likelihood ratio: $f \to g\frac{f}{g}$. For the application of importance sampling in risk measurement, the reader is referred to Glasserman (2003) and Glasserman and Li (2005).

The final method applies importance sampling reversely to assign more weights to the samples from the tail. This is done by i) defining a new density function that has a higher density around the tail; ii) calculating the likelihood ratio; iii) multiplying the data by the inverse of the likelihood ratio. The procedure is summarized below.

- **Likelihood Ratio.** Assume that e_t are random samples from $N(0, \bar{H})$. The new density function is defined by shifting the mean of e_t by λ. The likelihood ratio is then given by

$$l(e_t) = \exp\left(-e_t^\top \bar{H}^{-1}\lambda + \frac{1}{2}\lambda^\top \bar{H}^{-1}\lambda\right).$$

- **New Density Function.** Following Glasserman et al. (1999), λ is determined by solving

$$\begin{aligned} \lambda = \underset{x}{\text{argmax}} \ & \exp\left(-\frac{1}{2}x^\top \bar{H}^{-1}x\right) \\ \text{subject to } & x^\top w \leq L, \end{aligned}$$

  where w is the portfolio weights, and L is a loss level. The above problem has a closed-form solution

$$\lambda = \frac{L\bar{H}w}{w^\top \bar{H}w}.$$

  L is set to $-0.5\sqrt{w^\top \bar{H}w}$.

- **Estimation.** In the ML estimator, multiply the log-likelihood function at time t by $1/l(e_t)$.

The results from these four alternative estimation methods applied to GCDS are reported in Table 11 and Figure 6. All four methods reduce the error at the tail compared to the original method (Normal), and the first (t-dist) method makes the most noticeable improvement.

Using the t-distribution significantly improves tail risk measurement without any substantial performance loss in terms of other metrics. Figure 6 shows that the t-distribution fits the negative side of the distribution remarkably well, while it underestimates the positive side only slightly. Except for the t-distribution, all other methods incur significant overestimation of the variance (negative dS_p). In fact, these methods perform much worse than the original method in terms of other metrics. These results suggest that, even after accounting for heteroscedasticity, the asset returns are skewed and fat-tailed. As far as risk is concerned, the t-distribution appears to be a better choice than the normal distribution.

Table 11: Performance evaluation of the alternative estimation methods applied to GCDS

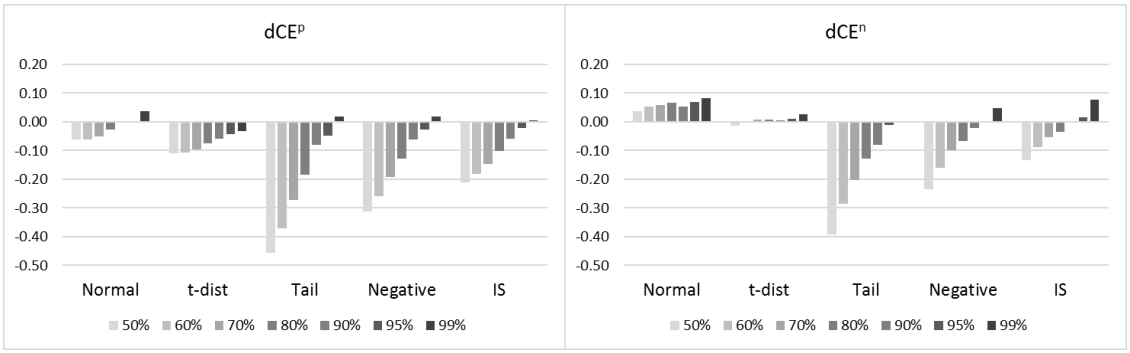
"Normal" refers to the original MLE, and the other four columns refer to the four alternative methods depicted in Section 5.2.

	Normal	t-dist	Tail	Negative	IS
			Stock Indexes		
dS_p	0.676	0.895	-40.106	-24.900	-15.511
dCE^n_{99}	8.251	2.716	0.357	4.686	7.594
			Currencies		
dS_p	0.542	0.766	-41.824	-22.713	-16.136
dCE^n_{99}	13.826	0.773	-0.568	8.082	12.832
			DJIA Stocks		
dS_p	1.706	3.072	-42.867	-21.845	-16.784
dCE^n_{99}	9.075	-2.251	-2.377	5.819	4.565

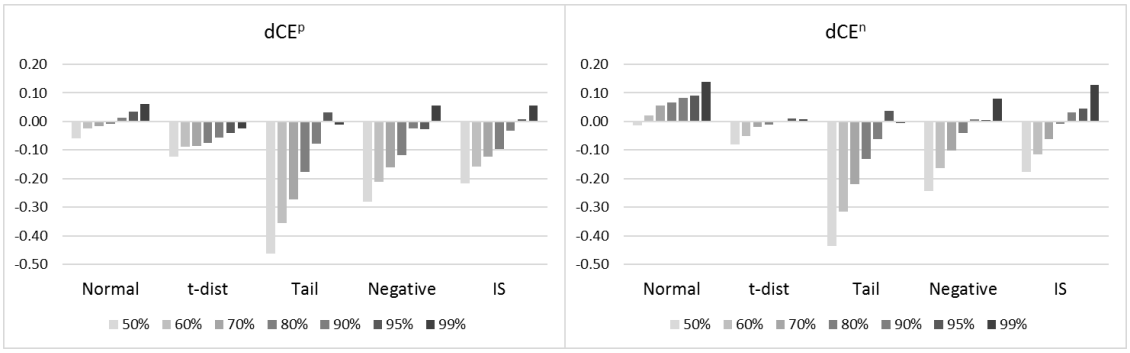
5.3 Portfolio Performance

This section compares the covariance models from a perspective of portfolio management. Two types of portfolios, minimum variance portfolio and tangent portfolio, are considered. For the tangent portfolios, the expected return is assumed known: the sample mean over the out-of-sample period is used as the expected return. We make this assumption so that the performance of the tangent portfolio is affected only by the estimation error of the covariance matrix. Table 12 reports the results.

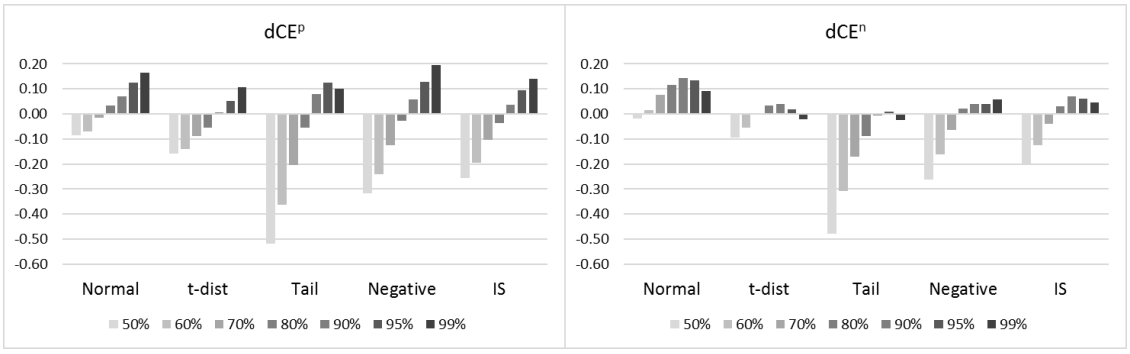
From all three samples, we find that the GGARCH models usually outperform the bench-



(a) Stock Indexes



(b) Currencies



(c) DJIA stocks

Figure 6: Conditional expectations of the alternative estimation methods applied to GCDS

marks in terms of the Sharpe ratio. For the minimum variance portfolio, the highest Sharpe ratio is attained by GCDD, GCRG, and GCRG, respectively under the stock index, currency, and DJIA stock samples. For the tangent portfolio, the highest Sharpe ratio is attained by GCDS, GCRM, and GCDS, respectively for the three samples. Other GGARCH models also perform better or comparably with the benchmarks. The fact that DCCG outperforms DCC further supports the advantages of the GGARCH models.

The tangent portfolios generally outperform the minimum variance portfolios but they are also much more volatile. This result is anticipated since we assume that the expected returns are known. The equally-weighted portfolio (EW) does not perform very well in all the samples. This is partly because the global market did not perform well in the out-of-sample periods.

6 Conclusion

In this paper, we develop new covariance dynamics models (GCD) using methods of differential geometry. These models preserve the geometric structure of the covariance matrix without any arbitrary restrictions by respecting the inherent geometric features of the covariance matrix. The GCD models are tested on three data samples and compared with existing models; BEKK, DCC, and the matrix exponential GARCH, using various out-of-sample performance metrics including new ones that are particularly relevant to risk management.

Empirical results suggest that the GCD models outperform the existing models, and realized covariance based models outperform return based models. In particular, a realized covariance-based GCD model estimated via MLE performs best. These findings imply that realized covariances carry more information on the future covariance, but they do not cover the informational contents of asset returns. Another important finding is the lack of consistency across the performance metrics, which suggests that it is crucial to evaluate models using multiple criteria and choose the most relevant metric for a given purpose.

Table 12: Portfolio performance

This table reports the performance of the minimum variance portfolios and the tangent portfolios, which are constructed using different covariance models. For the tangent portfolios, the expected return is assumed known: the sample mean over the out-of-sample period is used as the expected return. We make this assumption so that the performance of the tangent portfolio is affected only by the estimation error of the covariance matrix. The mean return (Mean) and the standard deviation (Std) are annualized assuming 250 business days per year. EW denotes the equally-weighted portfolio.

	GCDS	GCDD	EXP	BEKKS	BEKKD	DCC	DCCG	EW
Minimum Variance								
Mean	0.012	0.012	0.013	0.008	0.008	0.009	0.010	0.083
Std	0.151	0.150	0.185	0.153	0.151	0.151	0.151	0.460
Sharpe	0.077	0.078	0.072	0.051	0.055	0.057	0.068	0.180
Tangent Portfolio								
Mean	0.046	0.043	0.055	0.022	0.034	0.033	0.038	0.083
Std	0.183	0.182	0.719	0.223	0.187	0.183	0.182	0.460
Sharpe	0.251	0.238	0.076	0.098	0.179	0.179	0.210	0.180

(a) Stock indexes

	GCDS	GCDD	EXP	BEKKS	BEKKD	DCC	DCCG	GCRM	GCRG	EW
Minimum Variance										
Mean	0.002	0.002	-0.011	0.004	-0.003	-0.002	0.004	0.005	0.005	-0.003
Std	0.063	0.063	0.063	0.064	0.063	0.063	0.063	0.062	0.062	0.192
Sharpe	0.024	0.027	-0.174	0.065	-0.048	-0.029	0.062	0.076	0.080	-0.014
Tangent Portfolio										
Mean	2.166	1.164	0.917	-0.954	2.109	0.387	1.265	3.668	1.326	-0.003
Std	9.218	4.712	5.159	2.696	10.621	1.441	5.222	4.419	2.816	0.192
Sharpe	0.235	0.247	0.178	-0.354	0.199	0.269	0.242	0.830	0.471	-0.014

(b) Currencies

	GCDS	GCDD	BEKKS	BEKKD	DCC	DCCG	GCRM	GCRG	EW
Minimum Variance									
Mean	0.026	0.045	0.017	0.035	0.034	0.045	0.030	0.049	-0.536
Std	0.150	0.149	0.150	0.154	0.155	0.155	0.151	0.153	1.293
Sharpe	0.176	0.304	0.113	0.228	0.222	0.287	0.196	0.318	-0.415
Tangent Portfolio									
Mean	6.887	4.968	13.390	6.969	5.428	5.452	8.644	8.379	-0.536
Std	3.580	5.770	20.025	8.140	4.795	3.971	5.678	4.995	1.293
Sharpe	1.924	0.861	0.669	0.856	1.132	1.373	1.522	1.677	-0.415

(c) DJIA stocks

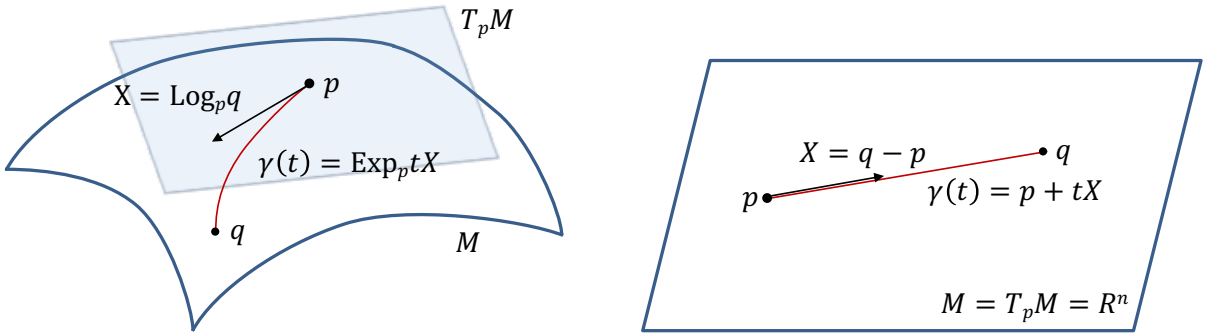
This paper demonstrates the advantages of taking an intrinsic geometric approach in covariance modeling. There are many areas of research that can be developed based on this paper, *e.g.,* an extension of the GCD model for a large dimensional system. We hope to see more papers in this direction.

A Basics of Differential Geometry

This section explains concepts from differential geometry that are essential to understand the model of the paper. The exposition below is by no means rigorous but intended to make the concepts as accessible as possible to the reader with no background in differential geometry. A more complete description can be found in Fletcher et al. (2003), Fletcher and Joshi (2004), Moakher (2005), Smith (2005), Lenglet et al. (2006a), Lenglet et al. (2006b), Pennec et al. (2006), Dryden et al. (2009), Ben-David and Marks (2014) and the references therein.

Manifold A manifold M is a space that looks like $\mathbb{R}^n$ (Euclidean space) locally. To understand the notion of "locally Euclidean," imagine looking at the earth from a rocket. When the rocket is far from the earth, you can see the entire shape of the earth, which is a sphere, but when it lands, the surface of the earth around looks like a flat plane. That is, a sphere (2-dimensional curved space) looks like a 2-dimensional Euclidean space locally. Examples of manifolds are a sphere, a flat plane, a torus, and so forth. A manifold is differentiable if it is everywhere smooth and continuous so that one can define a differentiation operator in a consistent manner. Figure 7 compares manifold with Euclidean space.

Tangent Space The tangent space of a manifold at $p \in M$, denoted by T_pM, is the set of vectors tangent to the point. It is a vector space with the same dimension as M. For a flat plane, the manifold and its tangent space coincide. When the manifold is a curved space, the tangent spaces at different points are different from each other (see Figure 8 (a)). This makes



(a) Manifold (b) Euclidean space

Figure 7: Manifold vs. Euclidean space

comparing two tangent vectors at different points or taking their difference complicated. To illustrate this, consider a tangent vector moving along a closed loop on a sphere (Figure 8 (b)). When the tangent vector moves parallel from $A \to N \to B \to A$, the tangent vector at the end does not coincide with the tangent vector at the beginning even though they are at the same point.

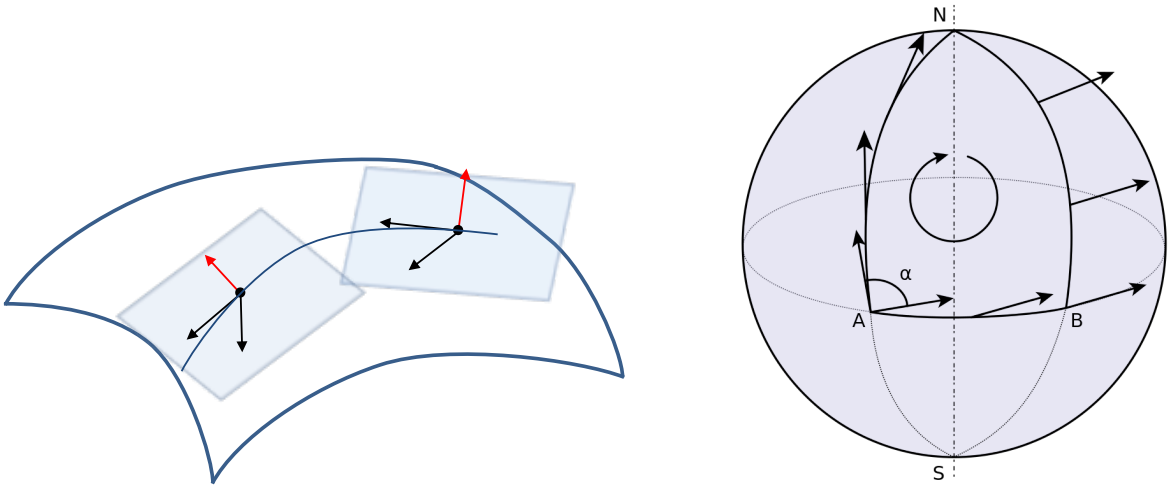
Because a tangent vector at one point is not necessarily tangent at another, the differentiation of tangent vectors is not well defined: subtraction between two different vector spaces is not intrinsically defined. The additional structure required to differentiate tangent vectors is called the affine connection, which is defined later in this section.

Riemannian metric Let M be a differentiable manifold of dimension n. A Riemannian metric (or Riemannian metric tensor) on M is a family of (positive-definite) inner products on the manifold's tangent space:[9]

$$g : T_pM \times T_pM \to \mathbb{R}. \tag{43}$$

If X is a tangent vector, the square of its length is given by $\langle X, X \rangle := g(X, X)$. The Riemannian metric becomes the dot product in a Euclidean space.

[9] A Riemannian metric depends on the location (p) of the tangent vector, and some texts use g_p instead of g to emphasize the dependency.



(a) Transportation on a manifold (b) Transportation along a closed loop

Figure 8: Location dependency of the tangent space

In a system of coordinates $x_1, x_2, \ldots, x_n$, the vector fields $\left\{\frac{\partial}{\partial x^1}, \frac{\partial}{\partial x^2}, \ldots, \frac{\partial}{\partial x^n}\right\}$ serve as a basis of tangent vectors at each point of M, and tangent vectors $X, Y \in T_pM$ can be written as

$$X = \sum_i X^i \frac{\partial}{\partial x^i}, \qquad Y = \sum_j Y^j \frac{\partial}{\partial x^j}. \tag{44}$$

Then, the Riemannian metric can be expressed as a sum of $n \times n$ components:

$$g(X, Y) = \sum_{i,j} X^i Y^j g\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right) = \sum_{i,j} g_{ij}(p) X^i Y^j, \tag{45}$$

where $g_{ij}(p) := g\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right)$. In the Euclidean space, $\frac{\partial}{\partial x^i}$ is identified with the basis vector $e_i = (0, ..., 1, ..., 0)$, and $g_{ij} = \langle e_i, e_j \rangle = \delta_{ij}$.

A Riemannian metric allows us to define several geometric notions on a Riemannian manifold, such as angle at an intersection, length of a curve, area of a surface, and so forth. A real, smooth manifold M equipped with a Riemannian metric g is called a Riemannian manifold (M, g).

Affine Connection An affine connection ∇ is an operator on a smooth manifold that connects nearby tangent spaces:

$$\nabla : TM \times TM \to TM, \tag{46}$$

where TM is the tangent bundle, *i.e.*, the union of all tangent spaces of a manifold M. It identifies a tangent vector at one point with a unique tangent vector at another point, so allowing differentiation of tangent vectors.

If (M, g) is a Riemannian manifold, there exists a unique affine connection, which is called the Levi-Civita connection. Under the Levi-Civita connection, the differentiation (covariant differential) of a tangent vector is defined as follows. Given a coordinate system $x^1, x^2, \ldots, x^n$, the ij-th component of the covariant derivative of $X = (X^1, X^2, \ldots, X^n) \in T_pM$ is given by

$$\nabla_j X^i = \frac{\partial X^i}{\partial x^j} + \sum_k \Gamma^i_{jk} X^k, \tag{47}$$

where

$$\Gamma^k_{ij} = \sum_s \frac{1}{2} g^{ks} \left(\frac{\partial g_{is}}{\partial x^j} + \frac{\partial g_{js}}{\partial x^i} - \frac{\partial g_{ij}}{\partial x^s} \right) \tag{48}$$

are the Christoffel symbols of the second kind, and g^{ks} denote the ks-th entry of the inverse of g_{ij}.

The covariant differential is a generalization of the Hessian: the covariant derivative $\nabla_j X^i$ is the sum of the partial derivative $\partial X^i / \partial x^j$ and a term to account for the change of the tangent space. If the manifold is flat, the second term vanishes and the covariant derivative becomes the partial derivative.

Geodesic A notion of straight line on a manifold is called a geodesic. A geodesic curve $\gamma(t) : [0, 1] \to M$ on a smooth manifold M must satisfy the equation $\nabla_{\dot\gamma} \dot\gamma = 0$, where

$\dot{\gamma} = d\gamma/dt$. It can be shown that $\nabla_{\dot{\gamma}}\dot{\gamma} = 0$ implies the differential equations

$$\ddot{\gamma}^k + \sum_{i,j} \Gamma^k_{ij}\dot{\gamma}^i\dot{\gamma}^j = 0, \quad k = 1, 2, \ldots, n, \tag{49}$$

where $\ddot{\gamma} = d^2\gamma/dt^2$. Equation (49) has a unique solution given an initial position and an initial velocity (tangent vector). The solution of Equation (49) is a curve with its Hessian (curvature) equal to zero, *i.e.,* "straight line."

A geodesic connecting two points on a manifold is not always unique: *e.g.*, a geodesic on a sphere is a great circle, and there are two segments of the great circle that connect two points. The minimal geodesic is a geodesic with the shortest path.

From a classical mechanics perspective, a geodesic can be thought of as a path of a free particle on a manifold, *i.e.*, the trajectory of a particle when there is no external force and the bending of the surface entirely determines its motion.

Distance In a Riemannian manifold (M, g), the length of a continuously differentiable curve $\gamma(t) : [0, 1] \to M$ is defined as

$$L(\gamma) = \int_0^1 \sqrt{g(\dot{\gamma}(t), \dot{\gamma}(t))}dt. \tag{50}$$

The distance between two points on a manifold is defined as the length of the shortest path connecting them, *i.e.,* the length of the minimal geodesic.

Exponential map The (Riemannian) exponential map $\mathrm{Exp}_p : T_pM \to M$ is a map from the tangent space at a point $p \in M$ to points on M. If $X \in T_pM$, the map $\mathrm{Exp}_p(X)$ transports p to a point in the direction X. Consider a geodesic curve $\gamma(t) : [0, 1] \to [p, q]$ emanating from p in a direction X. It follows that $\gamma(1) = q = \mathrm{Exp}_p(X)$. More generally,

$$\gamma(t) = \mathrm{Exp}_p(tX). \tag{51}$$

The log map $\text{Log}_p : M \to T_pM$ is the inverse of the exponential map. For the example above,

$$\text{Log}_p(q) = X. \tag{52}$$

References

Bauer, G.H., Vorkink, K., 2011. Forecasting multivariate realized stock market volatility. Journal of Econometrics 160, 93–101.

Ben-David, A., Marks, J., 2014. Geodesic paths for time-dependent covariance matrices in a Riemannian manifold. IEEE Geoscience and Remote Sensing Letters 11, 1499–1503.

Bollerslev, T., Engle, R., Wooldridge, J., 1988. A capital asset pricing model with time varying covariances. Journal of Political Economy 96, 116–131.

Bollerslev, T., Patton, A.J., Quaedvlieg, R., 2016. Exploiting the errors: A simple approach for improved volatility forecasting. Journal of Econometrics 192, 1–18.

Callot, L.A., Kock, A.B., Medeiros, M.C., 2017. Modeling and forecasting large realized covariance matrices and portfolio choice. Journal of Applied Econometrics 32, 140–158.

Cappiello, L., Engle, R.F., Sheppard, K., 2006. Asymmetric dynamics in the correlations of global equity and bond returns. Journal of Financial Econometrics 4, 537–572. doi:10.1093/jjfinec/nbl005.

Chiriac, R., Voev, V., 2011. Modelling and forecasting multivariate realized volatility. Journal of Applied Econometrics 26, 922–947. doi:10.1002/jae.1152.

Dryden, I.L., Koloydenko, A., Zhou, D., 2009. Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. The Annals of Applied Statistics 3, 1102–1123.

Engle, R., 2002. Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. Journal of Business and Economic Statistics 20, 339–350.

Engle, R., Kroner, F., 1995. Multivariate simultaneous generalized ARCH. Econometric Theory 11, 122–150.

Fletcher, P., Joshi, S., 2004. Principal geodesic analysis on symmetric spaces: Statistics of diffusion tensors, in: ECCV 2004 Workshop on Computer Vision Approaches to Medical Image Analysis (CVAMIA), Springer-Verlag. pp. 87–98.

Fletcher, P., Lu, C., Joshi, S., 2003. Statistics of shape via principal geodesic analysis on Lie groups, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 95–101.

Glasserman, P., 2003. Monte Carlo Methods in Financial Engineering. volume 53 of *Stochastic Modelling and Applied Probability*. Springer New York, New York, NY.

Glasserman, P., Heidelberger, P., Shahabuddin, P., 1999. Asymptotically optimal importance sampling and stratification for pricing path-dependent options. Mathematical Finance 9, 117–152.

Glasserman, P., Li, J., 2005. Importance sampling for portfolio credit risk. Management Science 51, 1643–1656.

Glosten, L.R., Jagannathan, R., Runkle, D.E., 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. The Journal of Finance 48, 1779–1801. doi:10.1111/j.1540-6261.1993.tb05128.x.

Han, C., Park, F.C., Kang, J., 2017. A geometric treatment of time-varying volatilities. Review of Quantitative Finance and Accounting 49, 1121–1141.

Hansen, P.R., Lunde, A., Nason, J.M., 2011. The model confidence set. Econometrica 79, 453–497.

Kawakatsu, H., 2006. Matrix exponential GARCH. Journal of Econometrics 134, 95–128.

Lanne, M., Saikkonen, P., 2007. A multivariate generalized orthogonal factor GARCH model. Journal of Business and Economic Statistics 25, 61–75.

Laurent, S., Rombouts, J.V.K., Violante, F., 2012. On the forecasting accuracy of multivariate GARCH models. Journal of Applied Econometrics 27, 934–955. doi:10.1002/jae.1248.

Lenglet, C., Rousson, M., Deriche, R., 2006a. DTI segmentation by statistical surface evolution. IEEE Transactions on Medical Imaging 25, 685–700.

Lenglet, C., Rousson, M., Deriche, R., Faugeras, O., 2006b. Statistics on the manifold of multivariate normal distributions: Theory and application to diffusion tensor MRI processing. Journal of Mathematical Imaging and Vision 25, 423–444.

Liu, L.Y., Patton, A.J., Sheppard, K., 2015. Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes. Journal of Econometrics 187, 293–311.

Moakher, M., 2005. A differential geometric approach to the geometric mean of symmetric positive-definite matrices. SIAM Journal on Matrix Analysis and Applications 26, 735–747. doi:10.1137/S0895479803436937.

Nelson, D., 1991. Conditional heteroskedasticity in asset returns: A new approach. Econometrica 59, 347–370.

Noureldin, D., Shephard, N., Sheppard, K., 2012. Multivariate high-frequency-based volatility (HEAVY) models. Journal of Applied Econometrics 27, 907–933. doi:10.1002/jae.1260.

Noureldin, D., Shephard, N., Sheppard, K., 2014. Multivariate rotated ARCH models. Journal of Econometrics 179, 16–30. doi:10.1016/j.jeconom.2013.10.003.

Pennec, X., Fillard, P., Ayache, N., 2006. A Riemannian framework for tensor computing. International Journal of computer vision 66, 41–66.

Smith, S., 2005. Covariance, subspace, and intrinsic Cramér-Rao bounds, in: IEEE Transactions on Signal Processing, pp. 1610–1630.

Vrontos, I.D., Dellaportas, P., Politis, D.N., 2003. A full-factor multivariate GARCH model. Econometrics Journal 6, 312–334.

van der Weide, R., 2002. GO-GARCH: a multivariate generalized orthogonal GARCH model. Journal of Applied Econometrics 17, 549–564.