A Geometric GARCH Framework for Covariance Dynamics

Chulwoo Han[a,*], Frank C. Park[b]

[a]*University of Durham*
[b]*Seoul National University*

Abstract

In this paper, new multivariate GARCH models are developed. These models respect the intrinsic geometric properties of the covariance matrix, and therefore are physically more meaningful. The models can be specified using either asset returns or realized covariances. Tested on three data samples, the empirical results suggest that our models outperform existing models such as BEKK and DCC, and realized covariance based models outperform return based models. In assessing the models, new parameter estimation method and performance evaluation methods are proposed. Limitations of existing evaluation methods are also addressed. Principal geodesic analysis (a version of principal component analysis) shows that time series variation of covariance matrices can be identified by a small number of axes, which suggests potential for a parsimonious specification of covariance dynamics for a large dimensional system.
JEL Classification: C32; C52; C58
Keywords: Geometric GARCH; Multivariate GARCH; Covariance; Realized covariance; Principal geodesic analysis

1. Introduction

Since the seminal work of Bollerslev et al. (1988), there has been significant interest in the development of multivariate GARCH models (see, *e.g.*, Bauwens et al. (2006), Silvennoinen and Teräsvirta (2009)). In generalizing the classical scalar GARCH model to the multidimensional case, two of the most critical issues are how to construct multidimensional models using

*Corresponding author. chulwoo.han@durham.ac.uk

a small number of physically meaningful parameters, and how to preserve positive definiteness in any iterative formula for the covariance matrix.

For multivariate models such as the BEKK model of Engle and Kroner (1995) and its extensions, the number of model parameters scales as at least $O(n^2)$–here n is the dimension of the system–because of the presence of an $n \times n$ constant matrix in the propagation equations, and the computational burden rapidly becomes untenable even for a moderately large n. Covariance targeting, which uses sample analogues to derive the constant matrix instead of estimating it simultaneously with the other parameters, has been proven to be effective in mitigating the effects of dimensional scaling of the model parameters (see, *e.g.*, Cappiello et al. (2006), Noureldin et al. (2012), Noureldin et al. (2014)). However, the estimation error of the sample covariance can be large especially when the length of the available time series is relatively short compared to the dimension of the system, and the power of this method can deteriorate as n increases.

Ensuring positive definiteness of the covariance matrix in the propagation formulas is more subtle. The BEKK model assumes a positive definite quadratic form for the model parameters. The DCC model of Engle (2002) separates variances and correlations, and models the correlation matrix independently via BEKK. GO-GARCH (van der Weide, 2002), as well as other factor models like Vrontos et al. (2003), Lanne and Saikkonen (2007), preserve positive definiteness by transforming the original data to independent factors, and applying GARCH models to these factors. Kawakatsu (2006) extends the exponential GARCH of Nelson (1991) to the multivariate case by exploiting the fact that the matrix exponential of a symmetric matrix is always positive definite.

More recently, covariance time series models that utilize realized covariance have received considerable attention in the literature. Realized covariance not only appears to carry more information about the future covariance, but also allows one to address aforementioned issues more conveniently. For example, Chiriac and Voev (2011) decompose the realized covariance matrices to construct a multivariate vector fractionally integrated ARMA (VARFIMA) process, and obtain positive definite covariance forecasts without imposing any restrictions. The HEAVY model proposed by Noureldin et al. (2012) resembles BEKK but utilizes realized covariances instead of daily returns, and has closed-form formulas for multi-step forecasts. RARCH models proposed by Noureldin et al. (2014) extend the BEKK and DCC models by rotating the returns prior to applying these models. Combined with covariance targeting, RARCH models can be applied to a large dimensional system. One of the limitations of realized covariance based models is that it is applicable only to assets that are traded simultaneously. Also, market microstructure issues can be involved when calculating realized covariances.

While the existing methods manage to preserve positive definiteness of the covariance matrix, they rely on *ad hoc* methods that fail to preserve the intrinsic geometric properties of symmetric positive definite matrices. We explain in further detail in the next section, but the correct way of measuring the distance between a pair of covariance matrices is as the length of the shortest path connecting the pair, and there is in fact a natural notion of length on the space of covariance matrices. These issues are highlighted in Han et al. (2016), whereby connections with known results on the intrinsic geometry of covariance matrices are identified; this allows one to, among other things, formulate coordinate-invariant notions of distances between covariance matrices, and to extend principal component analysis to covariance matrices in a geometrically well-defined, coordinate-invariant manner (via principal geodesic analysis). Distance between covariance matrices, direction from one covariance matrix to another, and other geometric properties defined on the space of covariance matrices can be utilized to develop an intuitive and physically meaningful covariance dynamics model.

The advantages of taking an intrinsic geometric approach in financial applications have been demonstrated in Webber (2000), Björk (2001), Filipović and Teichmann (2004), Park et al. (2011), but covariance dynamics has not been addressed in this fashion. The aim of this paper is to develop new multivariate GARCH models (geometric GARCH) that respect the geometric structure of covariance matrices, provide sufficient flexibility without using a large number of model parameters, and at the same time computationally efficient. Computational efficiency is achieved by making appropriate linear approximations in the propagation equations. Two classes of models, one that utilizes asset returns and the other that utilizes realized covariances are developed. These models can be estimated via the usual quasi-maximum likelihood estimation, but a new model parameter estimation method based on geometrically correct notions of distance between covariance matrices is also developed. Since our focus is on portfolio and risk management, covariance models are assessed based primarily on their out-of-sample performance via various performance metrics. For this, new performance evaluation metrics that are particularly suitable for risk management are developed. Time series variations of covariance matrices are investigated via principal geodesic analysis and potential for a parsimonious specification of the covariance models is discussed. Empirical studies involving three data samples show that our models outperform existing models such as BEKK, DCC, and matrix exponential GARCH.

2. Geometry of Covariance Space

For completeness of the paper, geometric properties of the covariance space that are particularly relevant to the development of the covariance dynamics in this paper are briefly described. More complete description can be found in Fletcher et al. (2003), Fletcher and Joshi (2004), Moakher (2005), Lenglet et al. (2006).

The covariance space $P(n)$ is defined as[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 this space, a Riemannian structure can be constructed via the Riemannian metric given by $\langle X, Y\rangle_P = \mathrm{tr}(P^{-1}XP^{-1}Y)$, $X, Y \in S(n)$. In terms of this metric, the minimal geodesic (shortest path) $\gamma(t) : [0, 1] \to [A, B]$ connecting two points $A, B \in P(n)$ is given by

$$\gamma(t) = G\big(G^{-1}BG^{-\top}\big)^t G^\top, \tag{2}$$

where $GG^\top = A$, $G \in GL^+(n)$. $GL^+(n)$ denotes the identity component of the general linear group $GL(n)$, *i.e.*, a subgroup of $GL(n)$ with positive determinants. The tangent vector of the geodesic at A is defined by the **Riemannian log map**

$$\mathrm{Log}_A(B) = G \log\left(G^{-1}BG^{-\top}\right) G^\top. \tag{3}$$

The inverse of the Riemannian log map, the **Riemannian exponential map** is also defined. Given an element $X \in S(n)$, the minimal geodesic emanating from some $A \in P(n)$ in the direction of the tangent vector X can be computed as follows:

$$\mathrm{Exp}_A(X) = G \exp\left(G^{-1}XG^{-\top}\right) G^\top. \tag{4}$$

Defining the distance between A and B in the usual way by the length of

[1]Strictly speaking, a covariance matrix can be positive semidefinite, but singular cases are ignored here as they are not of interest in financial applications.

the above minimal geodesic, we have

$$d(A, B) = ||\text{Log}_A B|| = \left(\sum_{i=1}^{n} (\log \lambda_i)^2\right)^{1/2} \tag{5}$$

where $||\cdot||$ is the Frobenius norm, and $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of the matrix AB^{-1}. Since AB^{-1} is symmetric positive-definite, the eigenvalues of AB^{-1} are all positive, and $\log \lambda_i$ is well defined for each i. Note also that $d(A, \gamma(t)) = t\, d(A, B)$. Therefore, a point on the geodesic between A and B that divides it in the ratio $\alpha : 1 - \alpha$, $I(A, B, \alpha)$, is given by

$$I(A, B, \alpha) = \text{Exp}_A(\alpha \text{Log}_A B). \tag{6}$$

With the above metric structure on $P(n)$, an intrinsic mean can be defined as follows (Fletcher and Joshi, 2004):

$$\bar{P} = \underset{\bar{P} \in P(n)}{\text{argmin}} \sum_{i=1}^{N} d(\bar{P}, P_i)^2. \tag{7}$$

For the case of two points, the intrinsic mean is simply the midpoint of the minimal geodesic. For an arbitrary N, the above intrinsic mean is unique on $P(n)$. The intrinsic mean has several advantages over the widely used arithmetic mean $\frac{1}{N}\sum_{i=1}^{N} P_i$: *e.g.,* the arithmetic mean of matrices with an equal determinant can have a larger determinant.

2.1. Principal Geodesic Analysis

Principal geodesic analysis (PGA), like the principal component analysis (PCA) which finds a linear subspace in which the variability of the data is best described, seeks a submanifold that best represents the variability of the data in a Riemannian manifold. Fletcher and his colleagues investigate the PGA on a Lie group (Fletcher et al., 2003) and also on the space $P(n)$ (Fletcher and Joshi, 2004). They show that PGA can be essentially carried out by applying PCA to the tangent space of the covariance space at the intrinsic mean. Below is a summary of the algorithm for PGA on $P(n)$. For more details, the reader is referred to the references cited above.

- Given $P_1, \cdots, P_N \in P(n)$,
- Calculate the intrinsic mean of $\{P_i\}$ and denote it by $\bar{P}$.

- Compute the tangent vectors at $\bar{P}$ of the geodesics connecting $\bar{P}$ and P_i:
$$X_i = \mathrm{Log}_{\bar{P}}(P_i). \tag{8}$$
- Apply PCA to $\{X_i\}$, *i.e.*, find eigenvectors and eigenvalues, $\{v_k, \lambda_k\}$, of
$$S = \frac{1}{N}\sum_{i=1}^{N} x_i x_i^\top, \tag{9}$$
where $x_i = vech(X_i)$, half vectorization of X_i. $V_k = vech^{-1}(v_k) \in S(n)$, $k = 1, \cdots, n(n+1)/2$, are the principal directions and λ_k are the corresponding variances.

Each observation P_i can be reproduced by the formula
$$P_i = \mathrm{Exp}_{\bar{P}}\left(\sum_{k=1}^{n(n+1)/2} \alpha_{ki} V_k\right), \tag{10}$$
where $\alpha_{ki} = v_k^\top x_i$. Provided that a time series of a covariance matrix is observable, PGA can identify the principal axes of variation of the covariance matrix. This can provide a valuable information about the evolution of the covariance matrix. Later in this paper, the evolution of a covariance matrix is analyzed via PGA using a realized covariance measure as a proxy for the true covariance matrix.

3. Dynamics of Covariance Matrix

Suppose that an n-variable system, r_t, is governed by the following equation:
$$r_t = \mu + e_t, \qquad e_t \sim N(0, H_t), \tag{11}$$
where $H_t \in P(n)$ is the covariance matrix of e_t. Following Laurent et al. (2012), Becker and Clements (2008), and Hansen and Lunde (2005) among others, r_t is assumed to have a constant mean. It is also found that adding autoregressive terms in the mean equation has little effect on the results of the empirical studies in this paper (not reported). In any event, a generalization to a time-varying mean model should be straightforward.

The covariance matrix of the residuals is assumed to have dynamics of the form
$$dH_t = F_t dt, \tag{12}$$

where F_t is a time-varying $n \times n$ symmetric matrix which depends on the information set at time t. As dH_t is the differential of H_t, it is defined in the tangent space $S(n)$, and it suffices for F_t to be symmetric.

$P(n)$ is a Riemannian symmetric space that is geodesically complete, and as such the minimal geodesics provide a natural way of discretizing general differential equations on $P(n)$. Using the Riemannian exponential map defined in (4), H_t is approximated by the formula

$$H_t = \mathrm{Exp}_{H_{t-1}}(F_{t-1}), \tag{13}$$

where F_{t-1} is the tangent vector at H_{t-1} of the geodesic between H_{t-1} and H_t. The dynamics of the covariance matrix is completely determined by the tangent vector F_{t-1} which is assumed to be a function of the information set at time $t-1$. We call this type of specification **Geometric GARCH** (**GGARCH**) models. In the following, two specifications of F_{t-1}, one that utilizes asset returns and the other that utilizes realized covariance are proposed.

3.1. Geometric GARCH in the Absence of Realized Covariance

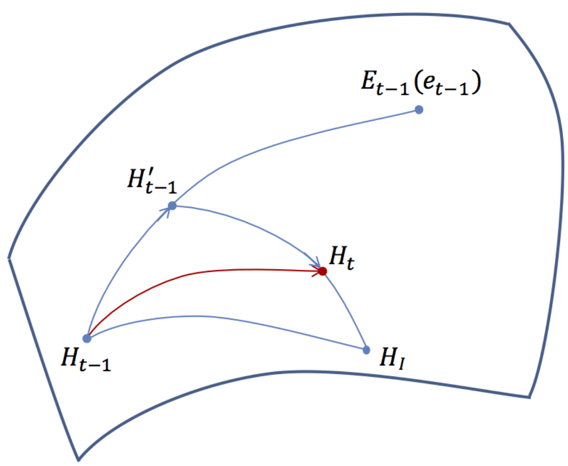


Figure 1: Geometric GARCH in the absence of realized covariance.

The first model is derived from the assumption that the covariance matrix has a mean reverting property such that it has a tendency to move toward a constant covariance matrix H_I: At time $t-1$, the covariance estimate, H_{t-1}, is first updated by the shock e_{t-1} to a new point H'_{t-1}, and then moves toward H_I along the geodesic between H'_{t-1} and H_I. The new covariance estimate H_t is determined on this geodesic. This is illustrated in Figure 1.

More precisely, H'_{t-1} is assumed to be on the geodesic between H_{t-1} and E_{t-1}, a covariance matrix derived from e_{t-1}:

$$H'_{t-1} = I(E_{t-1}, H_{t-1}, a^2), \quad a^2 < 1. \tag{14}$$

E_{t-1} is assumed to have the form

$$E_{t-1} = C \circ \big((1-b^2)e_{t-1}e_{t-1}^\top + b^2 \eta_{t-1}\eta_{t-1}^\top\big), \tag{15}$$

where $\eta_{t-1} = 1/2(|e_{t-1}| - e_{t-1})$ is a variable to reflect the asymmetric effect of the shocks, $b^2 < 1$, $C \in \mathbb{R}^{n \times n}$ is a covariance adjustment term of the form

$$C = \begin{bmatrix} 1 & c & \cdots & c \\ c & \ddots & & \vdots \\ \vdots & & \ddots & c \\ c & \cdots & c & 1 \end{bmatrix}, \quad 0 < c < 1, \tag{16}$$

and $\circ$ is the Hadamard product.

E_{t-1} defined as in (15) is generally in $P(n)$. However, it is possible that E_{t-1} becomes singular if any of the elements of e_{t-1} is zero, in which case, H'_{t-1} in (14) is not well defined. To mitigate this problem, H'_{t-1} is approximated as follows:

$$H'_{t-1} = (1-a^2)H_{t-1} + a^2 E_{t-1}. \tag{17}$$

This approximation turns out to have little effect on the results since the estimate of a^2 tends to be small (H'_{t-1} close to H_t). Also, it reduces the computational burden of the matrix exponential and logarithm involved in Equation (14). A more general form of (17) is also considered:

$$H'_{t-1} = (ii^\top - \vec{a}\vec{a}^\top) \circ H_{t-1} + \vec{a}\vec{a}^\top \circ E_{t-1}, \tag{18}$$

where $\vec{a} \in \mathbb{R}^n$ with i-th element $a_i^2 < 1,\ i = 1, \cdots, n$. This generalized form is to see whether the added degrees of freedom can improve the model.

Once H'_{t-1} is determined, new estimate of the covariance matrix at time t is determined as a point on the geodesic between H'_{t-1} and H_I, and is obtained by the formula:

$$H_t = I(H_I, H'_{t-1}, \alpha), \quad 0 < \alpha < 1. \tag{19}$$

Denoting the tangent vector of the geodesic between H_I and H'_{t-1} by F_{t-1}, *i.e.*,

$$F_{t-1} = \mathrm{Log}_{H_I}(H'_{t-1}), \tag{20}$$

H_t is given by

$$H_t = \mathrm{Exp}_{H_I}(\alpha F_{t-1}). \tag{21}$$

3.1.1. Positive Definiteness and Covariance Stationarity

If H_{t-1} is positive definite, it is trivial to show that H'_{t-1} is also positive definite. Then, by definition, H_t is also positive definite.

To prove covariance stationarity, first note that the sum of the coefficients of H_{t-1} and $e_{t-1}e_{t-1}^\top$ and a half of the coefficient of $\eta_{t-1}\eta_{t-1}^\top$ in (18) has the form

$$(ii^\top - \vec{a}\vec{a}^\top) + \vec{a}\vec{a}^\top(1-b^2) + \frac{1}{2}\vec{a}\vec{a}^\top b^2 = ii^\top - \frac{1}{2}\vec{a}\vec{a}^\top b^2,$$

and its i-th diagonal element, $1 - 1/2a_i^2 b^2 < 0$ for $i = 1, \cdots, n$. Therefore, H'_{t-1} is covariance stationary. Since H_t is a point in $P(n)$ that divides the geodesic between H_I and H'_{t-1} internally, covariance stationarity condition is satisfied.

3.2. Geometric GARCH in the Presence of Realized Covariance

If the assets of interest are traded simultaneously, *e.g.*, stocks traded in the same exchange, realized covariance can be computed and utilized in the covariance dynamics model. Under similar assumptions made for the GGARCH model without realized covariance, the covariance dynamics that incorporates realized covariance is constructed as follows. First, the covariance matrix is assumed to move to a new position H'_{t-1} by the formula[2]

$$H'_{t-1} = (1-a^2)H_{t-1} + a^2 C_{t-1}, \quad a^2 < 1, \tag{22}$$

[2] H'_{t-1} can be defined as a point on the geodesic between H_{t-1} and C_{t-1}. This gave similar results to those from (22), while computationally more demanding. Also, a more general form as in (18) is not considered here as C_{t-1} is a covariance matrix of e_{t-1} and H'_{t-1} as a convex combination of H_{t-1} and C_{t-1} is intuitively appealing.

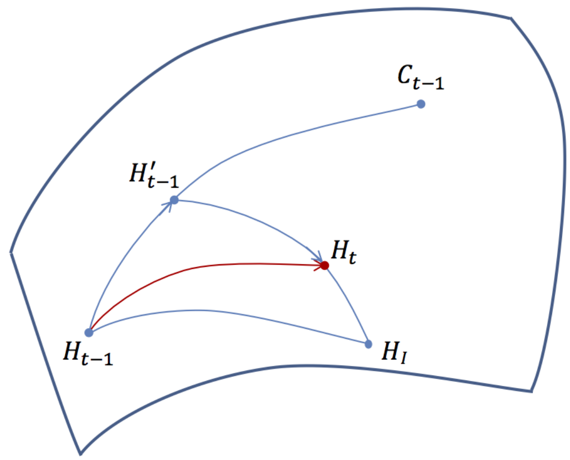


Figure 2: Geometric GARCH in the presence of realized covariance.

where C_{t-1} is a realized covariance matrix at time $t-1$. Then, the new covariance estimate is determined by

$$H_t = I(H_I, H'_{t-1}, \alpha), \quad 0 < \alpha < 1. \tag{23}$$

Denoting the tangent vector of the geodesic between H_I and H'_{t-1} by F_{t-1}, *i.e.*,

$$F_{t-1} = \mathrm{Log}_{H_I}(H'_{t-1}), \tag{24}$$

H_t is given by

$$H_t = \mathrm{Exp}_{H_I}(\alpha F_{t-1}). \tag{25}$$

This specification is conceptually similar to the scalar version of the HEAVY model by Noureldin et al. (2012). Positive definiteness and covariance stationarity can be shown using arguments similar to those in Section 3.1.1.

3.3. Estimation

Like other multivariate GARCH models, the GGARCH models proposed in Section 3.1 and 3.2 can be estimated using the quasi-maximum likelihood estimation (QMLE):

$$\max_{\theta} \sum_{t=1}^{T_i} -\frac{1}{2}\left(n \log(2\pi) + \log|H_t| + e_t^\top H_t^{-1} e_t\right), \tag{26}$$

where θ is the model parameters and T_i is the sample size.

If a realized covariance measure can be obtained, another way to estimate the model parameters is to minimize the distance between the forecast covariance matrix H_t and the realized covariance matrix C_t. One obvious choice of the distance metric between matrices is the Frobenius norm; see Laurent et al. (2012), for example. Another metric that is consistent with our approach is the length of the geodesic between H_t and C_t. The model parameters can be estimated minimizing the error defined by one of these metrics:

$$\min_{\theta} \sum_{t=1}^{T_i} ||H_t - C_t|| \qquad \text{(Frobenius norm)}, \tag{27}$$

$$\min_{\theta} \sum_{t=1}^{T_i} ||\mathrm{Log}_{H_t} C_t|| \qquad \text{(Length of geodesic)}, \tag{28}$$

where $||A|| = \sqrt{tr(AA^\top)}$ is the Frobenius norm. The distance-based estimation criteria assume that the realized covariance is a good proxy for the true covariance.

4. Evaluation

The most important function of the volatility models, either univariate or multivariate, is to forecast future volatilities, and as such the covariance models are evaluated using several out-of-sample performance metrics especially in the context of risk and portfolio management.

4.1. Likelihood

The first performance metric considered is out-of-sample likelihood, which is obtained from one day forecast of the covariance matrix during the out-of-sample period. Given one day forecasts $H_t,\ t = 1, \cdots, T$, the out-of-sample likelihood is defined by the formula

$$\log L = \frac{T_i}{T} \sum_{t=1}^{T} -\frac{1}{2} \left(n \log(2\pi) + \log |H_t| + e_t^\top H_t^{-1} e_t \right), \tag{29}$$

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

4.2. Portfolio Variance

The next performance metric is based on the variance of portfolio return. On each day during the out-of-sample period, the return of a portfolio is computed and normalized by its forecast variance. Then, the standard deviation of the normalized return, S_p, is calculated over the out-of-sample period:

$$S_p = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}(r_{pt} - \bar{r}_p)^2}, \tag{30}$$

where

$$r_{pt} = \frac{w^\top r_t}{\sqrt{w^\top H_t w}}, \quad \bar{r}_p = \frac{1}{T}\sum_{t=1}^{T} r_{pt}, \tag{31}$$

and w is the vector of the portfolio weights. If the forecast H_t is accurate, r_{pt} will follow the standard normal distribution and S_p will converge to 1 as the sample size increases. Accordingly, the performance metric is defined as the difference between S_p and 1:

$$dS_p = |S_p - 1|. \tag{32}$$

For the empirical studies in this paper, an equal-weight portfolio and single-asset portfolios, *i.e.*, portfolios consisting of only one asset, are considered for evaluation. For practical implementation, a financial institution may use their own portfolios.

4.3. Portfolio Conditional Expectation

An accurate estimation of the covariance matrix does not necessarily translate into 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 using a risk measure is still important as most risk measures are based on the covariance matrix. More generally, it would be informative to compare the distribution implied by the model with the actual distribution. In order to measure the overall fitting of the distribution, performance metrics based on conditional expectation are defined as follows. For the normalized return defined in (31), conditional means at different probability levels are calculated on both sides of the distribution. More specifically, for a given 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:

$$CE^p_\alpha = \frac{\sum_{t=1}^T r_{pt} \cdot \delta_{r_{pt} > z_\alpha}}{\sum_{t=1}^T \delta_{r_{pt} > z_\alpha}}, \tag{33}$$

$$CE^n_\alpha = \frac{\sum_{t=1}^T -r_{pt} \cdot \delta_{-r_{pt} > z_\alpha}}{\sum_{t=1}^T \delta_{-r_{pt} > z_\alpha}}, \tag{34}$$

where δ_i is the Kronecker delta, and z_α is the z-score at the probability level α. r_{pt} in (34) is premultiplied by -1 to make CE^n_α positive. CE^n_α with a high value of α, *e.g.,* 0.95 or 0.99, is often called conditional Value-at-Risk or expected shortfall, and is a widely used risk measure. Conditional expectation is preferred as an evaluation metric to a point estimate such as Value-at-Risk since the former contains information about a range of the distribution.

If the forecast of the covariance matrix is accurate and the return is normally distributed, CE^p_α and CE^n_α will converge to the theoretical value

$$CE_\alpha = \frac{1}{1-\alpha} \int_{z_\alpha}^{\infty} z\Phi(z)dz, \tag{35}$$

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

$$dCE^p_\alpha = \frac{|CE^p_\alpha - CE_\alpha|}{CE_\alpha}, \tag{36}$$

$$dCE^n_\alpha = \frac{|CE^n_\alpha - CE_\alpha|}{CE_\alpha}. \tag{37}$$

In the empirical studies, dCE^p_α and dCE^n_α are measured for equal-weight portfolios and single-asset portfolios for $\alpha \in \{0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.99\}$.

4.4. Minimum Variance Portfolio

The next performance metrics are related to the minimum variance portfolio. If the covariance dynamics is correctly specified, a minimum variance portfolio obtained from the forecast covariance matrix is expected to have a minimal variance. In this sense, the variance of the minimum variance portfolio can serve as a performance indicator of the covariance models. Minimum variance portfolios are constructed and their variances are calculated as follows. First, it is assumed that the portfolio is rebalanced whenever the covariance models are recalibrated (every 22 days as illustrated in Section 5.5), and held until the next rebalancing date. No constraints except budget constraint are imposed during the portfolio construction. Then, the minimum variance portfolio is given by the closed form

$$w_t^{\min} = \frac{H_t^{-1} i}{i^\top H_t^{-1} i}, \tag{38}$$

where the superscript "min" denotes a minimum variance portfolio, and i denotes a vector of ones of an appropriate size. The portfolio return is computed every day during the out-of-sample period and its sample standard deviation is calculated:

$$S_p^{\min} = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}(r_{pt}^{\min} - \bar{r}_p^{\min})^2}, \tag{39}$$

where $r_{pt}^{\min} = r_t^\top w_t^{\min}$. A more accurate covariance model will lead to a smaller standard deviation. Only standard deviation is compared for evaluation because mean return or any other performance measures that involve mean return such as Sharpe ratio do not necessarily imply the accuracy of the covariance models.

Another performance metric associated with portfolio rebalancing is the turnover of the portfolio. A covariance model that yields erratic estimates over time will result in a high turnover when the portfolio is rebalanced. An unstable covariance matrix is not only counter-intuitive but also harmful for portfolio management as it will incur a high transaction cost. In this regard, turnover is an appropriate metric to measure the stability of the covariance estimates. When the portfolio is rebalanced at time t, turnover is defined as

$$dw_t = |w_t - w_{t^-}|^\top i, \tag{40}$$

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

$$dw = \frac{1}{K}\sum_{k=1}^{K} dw_{t_k}, \tag{41}$$

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

4.5. Distance Metrics

The final performance metric is applicable when the realized covariance can be measured. If the realized covariances are available, the distance between the forecast covariance and the realized covariance can be used to evaluate the forecast accuracy of the model. The Frobenius norm and the length of the geodesic between the two covariance matrices are chosen as distance measures and the performance metrics are defined as the average distances over the out-of-sample period:

$$D_F = \frac{1}{T}\sum_{t=1}^{T} ||H_t - C_t||, \tag{42}$$

$$D_G = \frac{1}{T}\sum_{t=1}^{T} ||\mathrm{Log}_{H_t} C_t||. \tag{43}$$

D_F is the Frobenius norm-based performance metric and D_R is the geodesic length-based performance metric.

5. The Data and the Models

For the empirical studies, three data samples are selected: Global stock market indexes, currencies, and individual stocks. Testing the models on three data samples of distinct characteristics will enhance the reliability of the conclusions derived from the empirical studies. Stock market indexes are important for international diversification. Three major stock market indexes are chosen for the first sample. Currencies are distinct from the other two data samples in the sense that they feature negative correlation. In the third sample, six stocks from the US stock market are chosen to test the models in a relatively large dimensional environment. The assets in the last two samples are traded simultaneously, and realized covariances can be obtained from high frequency trading data. Details of each sample are described below.

5.1. Stock Market Indexes

Following Kawakatsu (2006), the first data sample consists of three stock market indexes; S&P500, FTSE100, and NIKKEI225. Daily index values are collected from DataStream and daily index returns are generated during the period from 1997-01-02 to 2014-12-31. The return time series are displayed in Figure 3 and their descriptive statistics are reported in Table 1. S&P500 and FTSE100 have a similar level of variance and are highly correlated with

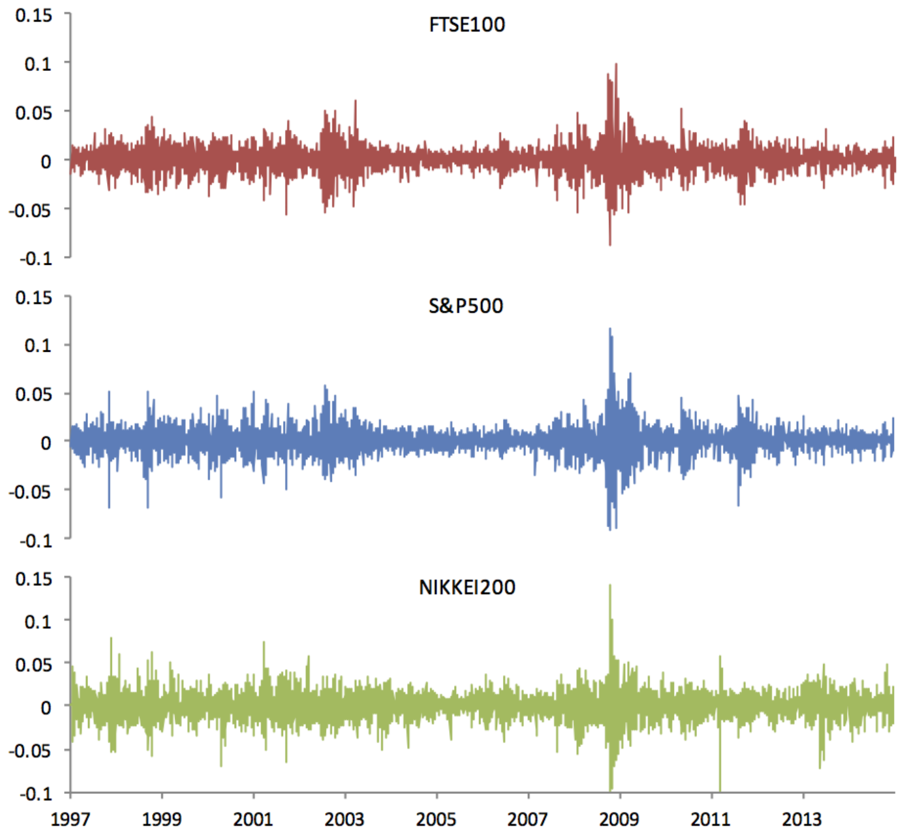


Figure 3: Daily returns of the stock indexes in the first sample from 1997-01-02 to 2014-12-31.

Table 1: Descriptive statistics of the stock index daily returns. The indexes are S&P500 (S&P), FTSE100 (FTSE), and NIKKEI225 (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

each other, whereas NIKKEI225 has a relatively higher variance and a lower return-to-risk ratio. The correlations between NIKKEI225 and the other two markets are also much lower compared to the correlation between S&P500 and FTSE100. In Figure 3, volatility clustering can be observed in all three markets.

5.2. Currencies

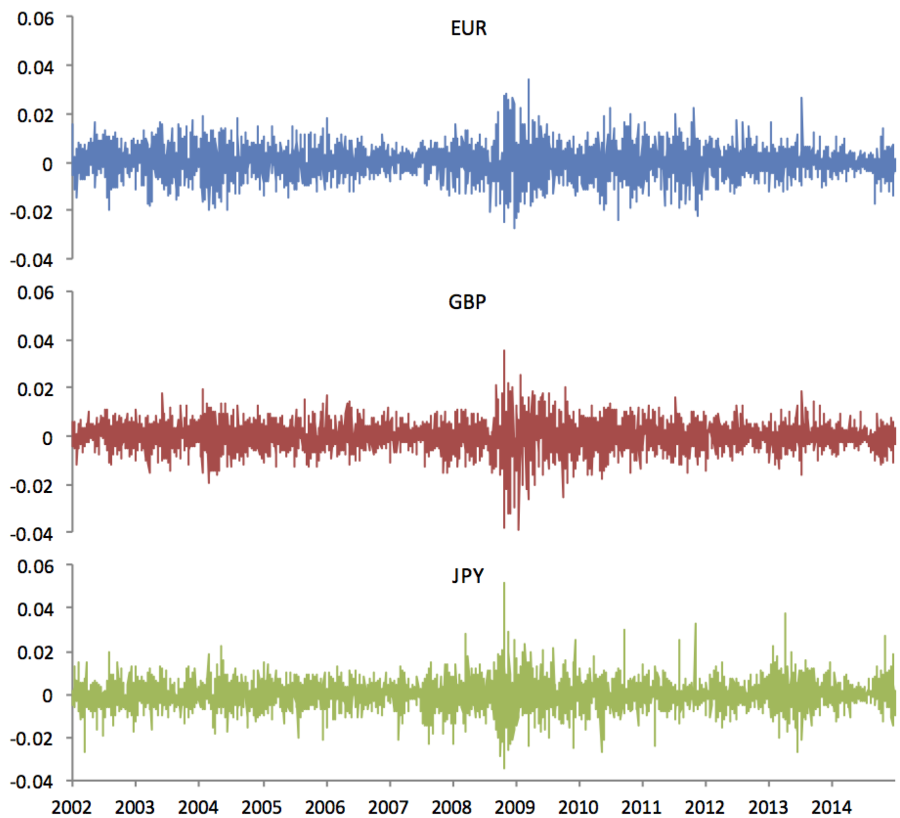


Figure 4: Daily returns of the currencies in the second sample from 2002-01-02 to 2014-12-31.

The second data sample consists of three currencies; Euro (EUR), British Pound (GBP), and Japanese Yen (JPY), which are expressed as US dollar price per unit currency. Laurent et al. (2012) also analyze the covariance of these currencies. The sample period is chosen to be from 2002-01-02 to

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

2014-12-31.[3] These currencies are traded simultaneously and as such both daily returns and realized covariances are calculated. Liu et al. (2014), in their extensive empirical study, find little evidence that a simple 5-minute realized variance is outperformed by other measures. Based on their findings, the realized covariance is computed from 5-minute data as follows. First, spot exchange rates are collected every 5 minutes from 00:00 to 24:00 GMT during the sample period from Thomson Reuters Tick History. Let h denote the time during the trading hours of a day such that $h = 1, 2, \cdots$ represent $5, 10, \cdots$ minutes from the opening of the market. 5-minute log-returns are calculated during the trading hours each day:

$$\rho_{it}^h = \log P_{it}^h - \log P_{it}^{h-1}, \quad h = 1, \cdots, H, \ i = 1, \cdots, n, \tag{44}$$

where P_{it}^h and ρ_{it}^h respectively denote the price and the 5-minute log-return of asset i at time h on day t, and H denotes the closing time of the market. Define ρ_t such that

$$\rho_t = \begin{bmatrix} \rho_{1t}^1 & \cdots & \rho_{nt}^1 \\ \vdots & \ddots & \vdots \\ \rho_{1t}^H & \cdots & \rho_{nt}^H \end{bmatrix}. \tag{45}$$

Then the daily returns and the realized covariance matrix of the assets are given by

$$r_t = \rho_t^\top i, \tag{46}$$

[3]The sample starts from 2002-01-02 considering the fact that euro was introduced in non-physical form on 1999-01-01 and the new euro notes and coins were introduced on 2002-01-01.

$$C_t = \rho_t^\top \rho_t. \tag{47}$$

Only the times when all three exchange rates are available are included for calculation, and the days when the number of available 5-minute returns is less than 80% of the maximum possible number are excluded from the sample. In other words, only those days which have 5-minute returns more than $230(= 0.8\times24\times60/5)$ remain in the sample. This leaves a total number of 3,549 days in the sample. The return time series are described in Figure 4 and their 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.

5.3. Individual 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

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 5 minutes from 09:30 to 16:00 EST from Thomson Reuter Tick History and the daily returns and the realized covariances are computed from these data using the same method described in Section 5.2. This means that the daily return is defined as the open-to-close log price difference and overnight jump is ignored. Applying the same filtering scheme used for the currency data, there are 3,210 days remaining

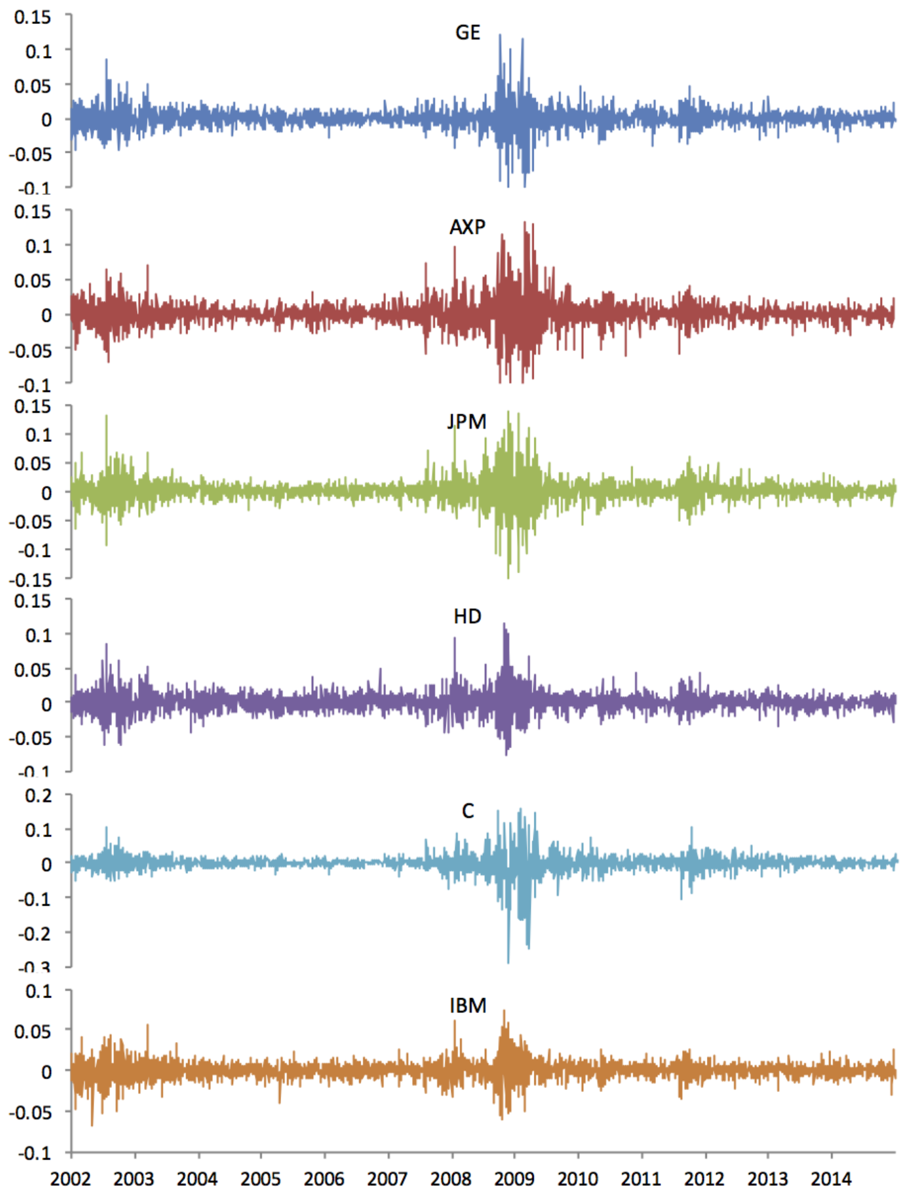


Figure 5: Daily returns of the DJIA stocks in the third sample from 2002-01-02 to 2014-12-31.

in the sample. The return time series are described in Figure 5 and their descriptive statistics are reported in Table 3. All the stocks are highly correlated with each other with the 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), in particular, has the mean return (-0.426) significantly lower than others and the standard deviation (0.376) higher than others.

5.4. Test Models

Four GGARCH-based specifications are considered in the empirical analysis. These models are compared with matrix exponential GARCH by Kawakatsu (2006), two versions of BEKK by Engle and Kroner (1995), and an extended version of DCC (Engle, 2002) by Cappiello et al. (2006). In all models, the covariance matrix is assumed to be a function of only the first lags of the covariance matrix and the shocks, and the asymmetric effect of the shocks is taken into account. The test models are summarized in Table 4 together with the number of model parameters.

Table 4: Test models used in the empirical studies. n in the second column is the number of assets. See Section 5.4 for the details of each model.

Model	No. Parameters	Description
GGS	$1/2n(n+1)+4$	GGARCH with scalar coefficients
GGD	$1/2n(n+3)+3$	GGARCH with diagonal matrix coefficients
EXP	$n(3n+1)$	Matrix exponential GARCH
BEKKS	$1/2n(n+1)+3$	BEKK with scalar coefficients
BEKKD	$1/2n(n+7)$	BEKK with diagonal matrix coefficients
DCC	$1/2n(n+7)+3$	DCC with scalar coefficients
DCCG	$1/2n(n+7)+4$	DCC-GGARCH with scalar coefficients
GGRM	$1/2n(n+1)+2$	GGARCH with realized covariance and QMLE
GGRG	$1/2n(n+1)+2$	GGARCH with realized covariance and geodesic length minimization

For the mean equation

$$r_t = \mu + e_t, \qquad e_t \sim N(0, H_t),$$

covariance matrix H_t is specified in each model as follows.

GGARCH SCALAR (GGS). The first model is GGARCH in Section 3.1 with scalar coefficients:

$$\begin{aligned} &H_t = \mathrm{Exp}_{H_I}(\alpha F_{t-1}), \quad F_{t-1} = \mathrm{Log}_{H_I}(H'_{t-1}), \\ &H'_{t-1} = (1-a^2)H_{t-1} + a^2 C \circ \big((1-b^2)e_{t-1}e_{t-1}^\top + b^2\eta_{t-1}\eta_{t-1}^\top\big). \end{aligned}$$

GGARCH DIAGONAL (GGD). The next model is GGARCH in Section 3.1 with diagonal coefficients:

$$\begin{aligned} &H_t = \mathrm{Exp}_{H_I}(\alpha F_{t-1}), \quad F_{t-1} = \mathrm{Log}_{H_I}(H'_{t-1}), \\ &H'_{t-1} = (ii^\top - \vec{a}\vec{a}^\top) \circ H_{t-1} + \vec{a}\vec{a}^\top \circ C \circ \big((1-b^2)e_{t-1}e_{t-1}^\top + b^2\eta_{t-1}\eta_{t-1}^\top\big). \end{aligned}$$

DCC-GGARCH (DCCG). The third model incorporates GGARCH into the DCC framework. The covariance matrix is decomposed as follows:

$$H_t = D_{H_t} P_t D_{H_t},$$

where D_{H_t} is a diagonal matrix with the square root of the diagonal entries of H_t on its diagonal, and P_t is the correlation matrix of H_t. The variances $h_{it},\ i = 1, \cdots, n$, are assumed to follow GJR-GARCH (Glosten et al., 1993) processes:

$$h_{it} = \alpha_i + \beta_i h_{it-1} + \gamma_i e_{it-1}^2 + \delta_i \eta_{it-1}^2, \quad \alpha_i, \beta_i, \gamma_i, \delta_i > 0,\ \beta_i + \gamma_i + \frac{1}{2}\delta_i < 1,$$

and the correlation matrix is assumed to follow a scalar GGARCH process:

$$\begin{aligned} &P_t = D_{R_t}^{-1} R_t D_{R_t}^{-1}, \\ &R_t = \mathrm{Exp}_{R_I}(F_{t-1}), \quad F_{t-1} = \alpha \mathrm{Log}_{R_I}(R'_{t-1}), \\ &R'_{t-1} = (1-a^2)R_{t-1} + a^2 C \circ \big((1-b^2)\tilde{e}_{t-1}\tilde{e}_{t-1}^\top + b^2\tilde{\eta}_{t-1}\tilde{\eta}_{t-1}^\top)\big), \end{aligned}$$

where D_{R_t} is a diagonal matrix with the square root of the diagonal entries of R_t on its diagonal, and $\tilde{e}_t = D_{H_t}^{-1} e_t$ and $\tilde{\eta}_t = D_{H_t}^{-1}\eta_t$ are standardized shocks and their asymmetric terms.

GGARCH REALISED (GGRM, GGRG). The last model under the GGARCH framework is the GGARCH model in the presence of realized covariances defined in Section 3.2:

$$\begin{aligned} H_t &= \mathrm{Exp}_{H_I}(\alpha F_{t-1}), \quad F_{t-1} = \mathrm{Log}_{H_I}(H'_{t-1}), \\ H'_{t-1} &= (1-a^2)H_{t-1} + a^2 C_{t-1}. \end{aligned}$$

GGRM refers to the model estimated by QMLE, and GGRG refers to the model estimated by geodesic length minimization. Frobenius norm minimization was also considered but is omitted in this paper as the results are similar to those of GGRD whereas the performance is slightly poorer. GGRM and GGRG are applied only to the currency and the DJIA stock samples for which realized covariances are available.

For comparison, the following existing models are also tested. For the details of each model, the reader is referred to the original articles cited below.

BEKK SCALAR (BEKKS). The first comparison model is a scalar version of the BEKK by Engle and Kroner (1995) given by

$$H_t = CC' + d^2 H_{t-1} + a^2 e_{t-1} e_{t-1}^\top + b^2 \eta_{t-1}\eta_{t-1}^\top,$$

where C is a lower triangular matrix, and a, b, and d are scalars.

BEKK DIAGONAL (BEKKD). The next model is a diagonal version of BEKK:

$$H_t = CC' + DH_{t-1}D^\top + Ae_{t-1}e_{t-1}^\top A^\top + B\eta_{t-1}\eta_{t-1}^\top B^\top,$$

where C is a lower triangular matrix, and A, B, and D are diagonal matrices.

DCC (DCC). In DCC, the variances are assumed to follow GJR-GARCH processes as in DCC-GGARCH. For the correlation matrix, a scalar version of the asymmetric DCC (A-DCC) by Cappiello et al. (2006) is used:

$$R_t = ((1-a^2-d^2)\bar{P} - b^2\bar{N}) + d^2 R_{t-1} + a^2 \tilde{e}_{t-1}\tilde{e}_{t-1}^\top + b^2 \tilde{\eta}_{t-1}\tilde{\eta}_{t-1}^\top,$$

where $\bar{P} = E[\tilde{e}_{t-1}\tilde{e}_{t-1}^\top]$ and $\bar{N} = E[\tilde{\eta}_{t-1}\tilde{\eta}_{t-1}^\top]$. $\bar{P}$ is estimated simultaneously with other parameters whilst a sample analogue is used as an estimate of $\bar{N}$.

MATRIX EXPONENTIAL GARCH (EXP). The final model for comparison is the matrix exponential GARCH by Kawakatsu (2006):

$$H_t = \mathrm{Exp}(F_t),$$
$$f_{ij,t} = c_{ij} + d_{ij} f_{ij,t-1} + a_{i,ij} e_{i,t-1} + a_{j,ij} e_{j,t-1} + b_{i,ij} \eta_{i,t-1} + b_{j,ij} \eta_{j,t-1},$$

where $\mathrm{Exp}(\cdot)$ is the matrix exponential function (this is equivalent to the Riemmanian exponential map with an identity matrix as the base), and $f_{ij,t}$ is the (i, j)-th element of F_t. Kawakatsu (2006) assumes that the matrix logarithm of H_t is a linear function of e_{t-1} and η_{t-1} in the manner that its (i, j)-th element is a function of only the i- and j-th elements of e_{t-1} and η_{t-1}.

5.5. Model Estimation and Evaluation

The sample period of each sample is divided into two sub-periods (in-sample and out-of-sample periods), and the model parameters are estimated every 22 days (about one month) during the second period rolling the sample window. Between estimation dates, the covariance matrix is updated daily with the arrival of new returns and/or realized covariance while the model parameters are fixed to the last estimates. Given the covariance estimates, the performance metrics defined in Section 4 are calculated during the out-of-sample period. The size of the sample window is set to 3 years. Therefore, the in-sample period of the stock index sample is from 1997-01-02 to 1999-12-31, and the out-of-sample period is from 2000-01-02 to 2014-12-31. These are respectively from 2002-01-02 to 2004-12-31 and from 2005-01-02 to 2014-12-31 for the currency and DJIA stock samples. Longer sample windows were also tested but the overall conclusions were mostly the same as those derived from the 3-year sample window.

6. Empirical Results

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

6.1. Performance Evaluation

Table 5, 6, and 7 report the out-of-sample performance of the test models evaluated respectively from the stock index, currency, and DJIA stock samples. GGRM and GGRG are applied only to the currency and DJIA samples for which realized covariances are measurable. EXP is applied only to the first two samples as its parameter estimation did not converge well for the DJIA sample. To facilitate the interpretation of the results, the ranking of the models based on each evaluation metric is reported below each metric. Also, two highest rank models are highlighted with dark grey and two lowest rank models are highlighted with light grey to assist visualization.[4] The results from each performance metric are discussed and the overall performance of the models is assessed.

Table 5: Performance evaluation using the stock index sample. The columns represent the test models defined in Section 5.4, and the rows represent the evaluation metrics defined in Section 4 except the first row, $\log L_0$, which is the average in-sample likelihood. dS_p and dCE^n_{99} are for the equally weighted portfolio and before taking absolute values. The second row of each evaluation metric is the ranks of the models based on the associated metric. The two highest rank models are highlighted in dark grey and the two lowest rank models are highlighted in light grey.

	GGS	GGD	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
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
$S^{\min}_p$	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 $\log L$, GGRM and GGRG perform best followed by DCCG. Other models show inconsistent results across samples. It might look surprising that the out-of-sample likelihood, $\log L$, is generally greater than the in-sample likelihood, $\log L_0$. This is because when $\log L$ is calculated, the parameters are updated every 22 days, whereas only one set of parameters are used to calculate the likelihood in each parameter estimation, and $\log L_0$ is the average of them.

The error of the portfolio variance measured by dS_p is very small in

[4] In Table 6 and 7, four highest rank models are highlighted with dark grey. This is because GGRM and GGRG which appear only in these tables are ranked highest with respect to most metrics. Therefore, highlighting four highest rank models make the comparison of the rest models easier.

Table 6: Performance evaluation using the currency sample. The columns represent the test models defined in Section 5.4, and the rows represent the evaluation metrics defined in Section 4 except the first row, $\log L_0$, which is the average in-sample likelihood. dS_p and dCE^n_{99} are for the equally weighted portfolio and before taking absolute values. The second row of each evaluation metric is the ranks of the models based on the associated metric. The four highest rank models are highlighted in dark grey and the two lowest rank models are highlighted in light grey.

	GGS	GGD	EXP	BEKKS	BEKKD	DCC	DCCG	GGRM	GGRG
$\log L_0$	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
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
$S^{\min}_p$	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
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

Table 7: Performance evaluation using the DJIA stock sample. The columns represent the test models defined in Section 5.4, and the rows represent the evaluation metrics defined in Section 4 except the first row, $\log L_0$, which is the average in-sample likelihood. dS_p and dCE^n_{99} are for the equally weighted portfolio and before taking absolute values. The second row of each evaluation metric is the ranks of the models based on the associated metric. The four highest rank models are highlighted in dark grey and the two lowest rank models are highlighted in light grey.

	GGS	GGD	BEKKS	BEKKD	DCC	DCCG	GGRM	GGRG
$\log L_0$	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
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
$S^{\min}_p$	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
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

almost all models. For instance, the largest error in the stock index sample is only -3.5% which comes from BEKKS. The only exception is the large error (19.9%) by GGRG in the DJIA stock sample. This suggests that the realized covariance may not be a good proxy for the true covariance in this particular sample. This, however, does not mean the realized covariance is not informative: GGRM, which also utilizes the realized covariance but is estimated via QMLE, still achieves a good performance. Whether the dimension of the system has an adverse effect on the realized covariance as a proxy is left for future research. None of the models outperforms the others consistently.

Contrary to the results from dS_p, the error in the tail region measured by dCE_{99}^n is much larger often exceeding 10%. GGRM performs best in both the currency and DJIA stock samples, but the error is still substantially larger compared to dS_p. This raises a doubt on the conditional normality assumption of e_t. Possible remedies to reduce the error at the tail is discussed later in Section 6.2.

The rankings based on S_p^{min} and dw are generally similar and consistent across samples favoring GGARCH models, GGRM, GGS, and GGD, in particular. Daily rebalancing and rebalancing subject to no short sale constraint were also tested but the results were not much different from those reported here. However, it should be noted that the variation of $S_p^{\min}$ across models is very small. This makes the role of $S_p^{\min}$ as a performance metric dubious. In fact, the standard deviation of the minimum variance portfolio return turns out to be an inappropriate performance metric in certain circumstances. This is further discussed in Section 6.3.

As expected, GGRG performs best in terms of the distance metrics, D_F and D_G, followed by GGRM. 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 defined in accordance with the model framework.

Comparing the overall performances of the models, the most remarkable finding is that when the realized covariance is available, the models that utilize the realized covariance, *i.e.*, GGRM and GGRG, dominate other models in terms of almost all performance metrics including those metrics that are not directly related to the realized covariance: These models are expected to perform well in terms of D_F and D_R, but the high ranking of GGRG with respect to Log$L$ is surprising considering the fact that the estimation criterion of GGRG is not to maximize the likelihood function but to minimize the distance between the estimated covariance and the realized covariance. Noureldin et al. (2012) also find that their realized covariance-based HEAVY model outperforms conventional GARCH models based on daily returns. Between GGRM and GGRG, GGRM performs more consistently. This is because GGRM utilizes information from both the returns and realized covariances. GGRG, on the other hand, uses only realized covariance and, although geodesic length minimization is easier to perform, the parameter estimation relies heavily on how well the realized covariance reflects the true covariance.

Apart from GGRM and GGRG, it is difficult to pinpoint the best performing model at the first glance. It is rather striking that the performance of the models varies widely not only across data samples but also across performance metrics. Table 8 reports the correlation between the model rankings. Unexpectedly, $\log L$ 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. This emphasizes the necessity of evaluating models using multiple metrics and choosing appropriate metrics for different purposes.

Table 8: Correlation between rankings. Correlations are calculated from the model rankings reported in Table 5, 6, and 7.

	$\log L$	dS_p	dCE^n_{99}	$S_p^{\min}$	dw	D_F	D_G
$\log L$	1.00						
dS_p	0.31	1.00					
dCE^n_{99}	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

Despite the seemingly inconsistent results, a careful inspection reveals several important points addressed below. First, it can be seen that GGARCH models outperform their BEKK counterparts; compare GGS with BEKKS and GGD with BEKKD. Meanwhile, applying the GGARCH model to the evolution of the correlation matrix in the DCC framework (DCCG) does not seem to improve over DCC significantly. EXP, contrary to its high in-sample likelihood values, performs worst out-of-sample. In general, a good in-sample performance in terms of $\log L_0$ does not lead to a good out-of-sample performance. Another important observation is that the diagonal versions of GGARCH and BEKK, despite more parameters, do not outperform their scalar counterparts. This suggests that simply increasing the degrees of freedom under the same model structure does not enhance the performance. Overall, GGARCH models perform best followed by DCC models, which are then followed by BEKK models.

Table 9 provides a summary of the results, in which the numbers are average rankings across data samples. GGRM, GGRG, and EXP are excluded as they are applied to only some of three data samples and their out/under-performances are rather obvious. Without these models, the remaining six models are ranked again from 1 (best) to 6 (worst) in each sample and the ranks are averaged across the samples. The table clearly shows that the best performances (dark grey cells) are usually found in GGARCH models and the worst performances (light grey cells) are usually found in BEKK models. This confirms the previous conclusion.

Table 9: Average ranking of the test models. Without GGRM, GGRG, 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. The two highest rank models are highlighted in dark grey and the two lowest rank models are highlighted in light grey.

	GGS	GGD	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^n_{99}	4.7	3.3	3.0	3.7	3.3	3.0
$S^{\min}_p$	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

Figure 6 and 7 display one day forecasts of the standard deviation of the return on GE and AE, respectively, and Figure 8 displays one day forecasts of their correlation, for some selected models.[5] Comparing GGS with BEKKS, the standard deviation estimated by GGS is more responsive to the shocks, whereas the correlation from GGS is more persistent. The difference between the models is more prominent during volatile market periods. This can be attributed to two facts: With the covariance adjustment term C, GGS produces more stable correlations while allowing the standard deviations move more extremely. The geodesic length between two matrices increases exponentially as one matrix approaches a singular point. This mitigates abrupt correlation changes especially toward singularity.

[5] Only one pair is reported to save the space. Other pairs are available upon request.

DCC, owing to its structure that separates variance and correlation, generates variances that are sensitive to the shocks. However, the correlation from DCC, unexpectedly, turns out to be most persistent among the models tested. Although this is intuitively appealing, the overall results do not support DCC.

Finally, GGRM features the least persistent variance and correlation estimates. This suggests that the realized covariance contains more information about the covariance than the shocks. In fact, the weight on C_{t-1} in GGRM approximated by αa^2 is 0.2437 on average, whilst the weight on $e_{t-1}e_{t-1}^\top$ in GGS approximated by αa^2 is only 0.0235. Even though the size of $e_{t-1}e_{t-1}^\top$ can be much larger than that of C_{t-1}, the weight difference appears to be substantial. It is worth noting that GGRM, despite its sensitive estimates, results in low turnover. Compared to GGRM as a reference, DCC performs well in fitting the variances, whilst BEKKS fits the correlation best. However, GGS seems to find a good compromise between the variance estimation and correlation estimation, and outperforms the other two.

6.2. Risk Measurement

As evidenced in the previous results, the forecast error at the tail measured by dCE^n_{99} is substantially larger than that of variance (dS_p), and there is a lack of consistency in ranking between these performance metrics. This necessitate the need of validating the models with regard to risk measures, especially if the primary purpose of covariance estimation is to measure risk. Figure 9 displays the conditional expectation measures defined in (36) and (37) without the absolute value operation in the numerator. This is to reveal the direction of the estimation error. In an ideal case in which the model is correctly specified, the errors will be close to zero across all probability levels. However, the charts show that all the models underestimate the conditional expectation at both tails, and the underestimation is severer on the negative side. This is in accordance with the well-known fat tailed, left-skewed distribution of asset returns.

If the objective of the covariance estimation is to measure risk, it is worth considering a tailored estimation method for the particular purpose so as to enhance the accuracy of risk measurement. Four methods that aim to enhance risk measurement are considered here. The first method is to replace the normal distribution with a fat-tailed distribution. In particular, student t-distribution is considered. In this case, the QMLE is given by

$$\max_\theta \sum_{t=1}^{T_i} c(n,d) - \frac{1}{2}\log|H_t| - \frac{n+d}{2}\log\left(1 + \frac{e_t^\top H_t^{-1} e_t}{d-2}\right), \qquad (48)$$

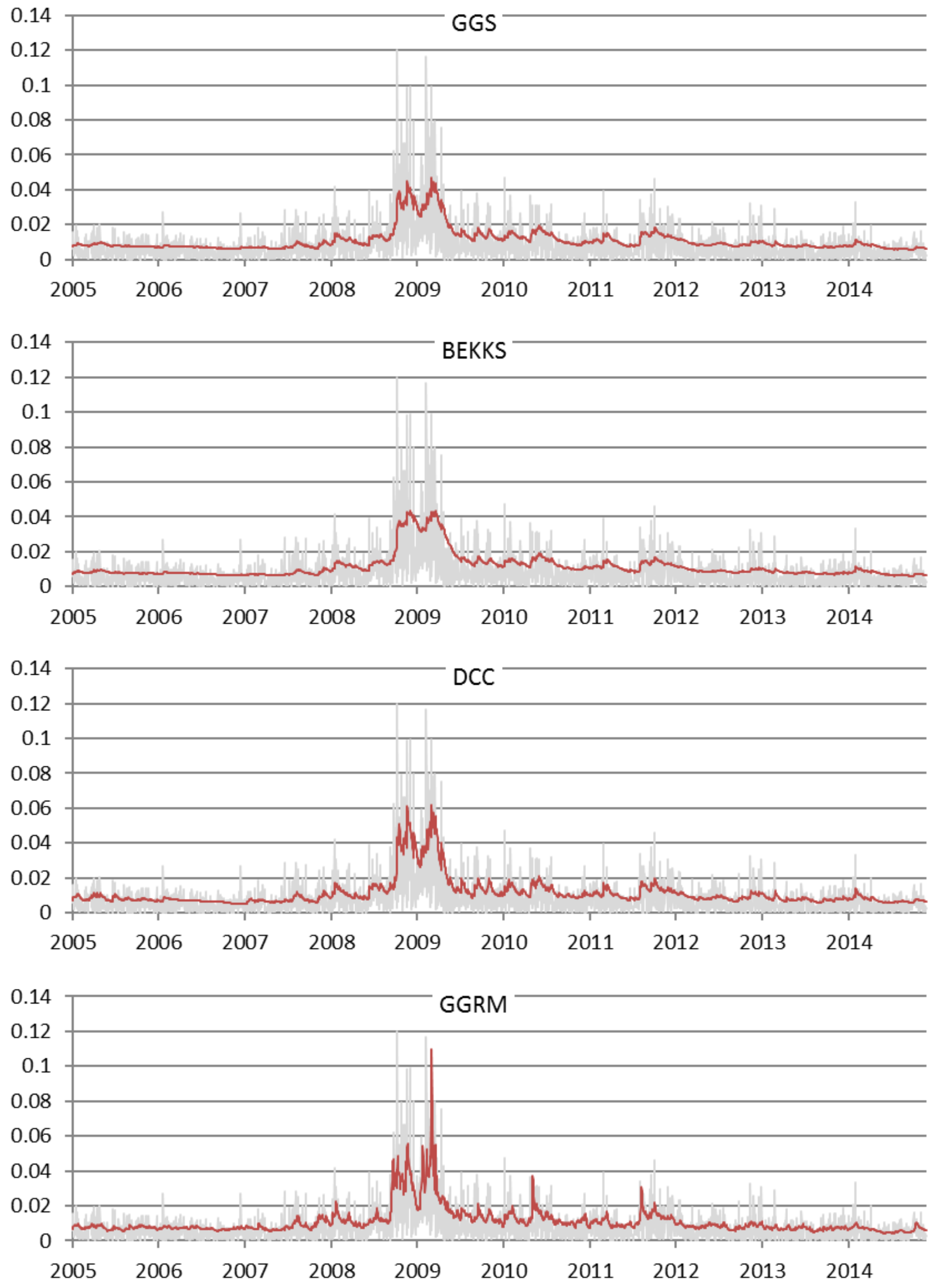


Figure 6: One day forecast of the standard deviation of GE.

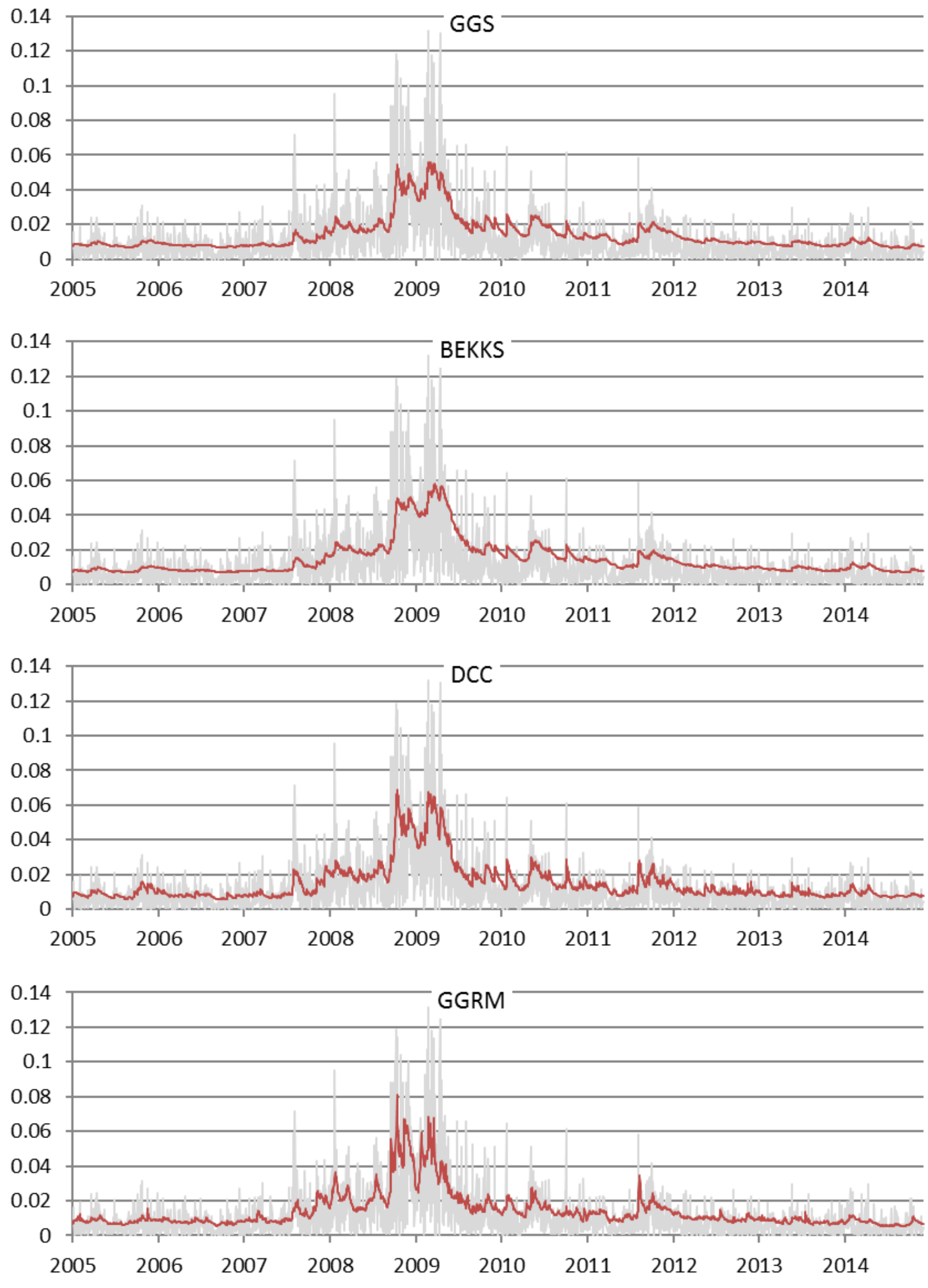


Figure 7: One day forecast of the standard deviation of AE.

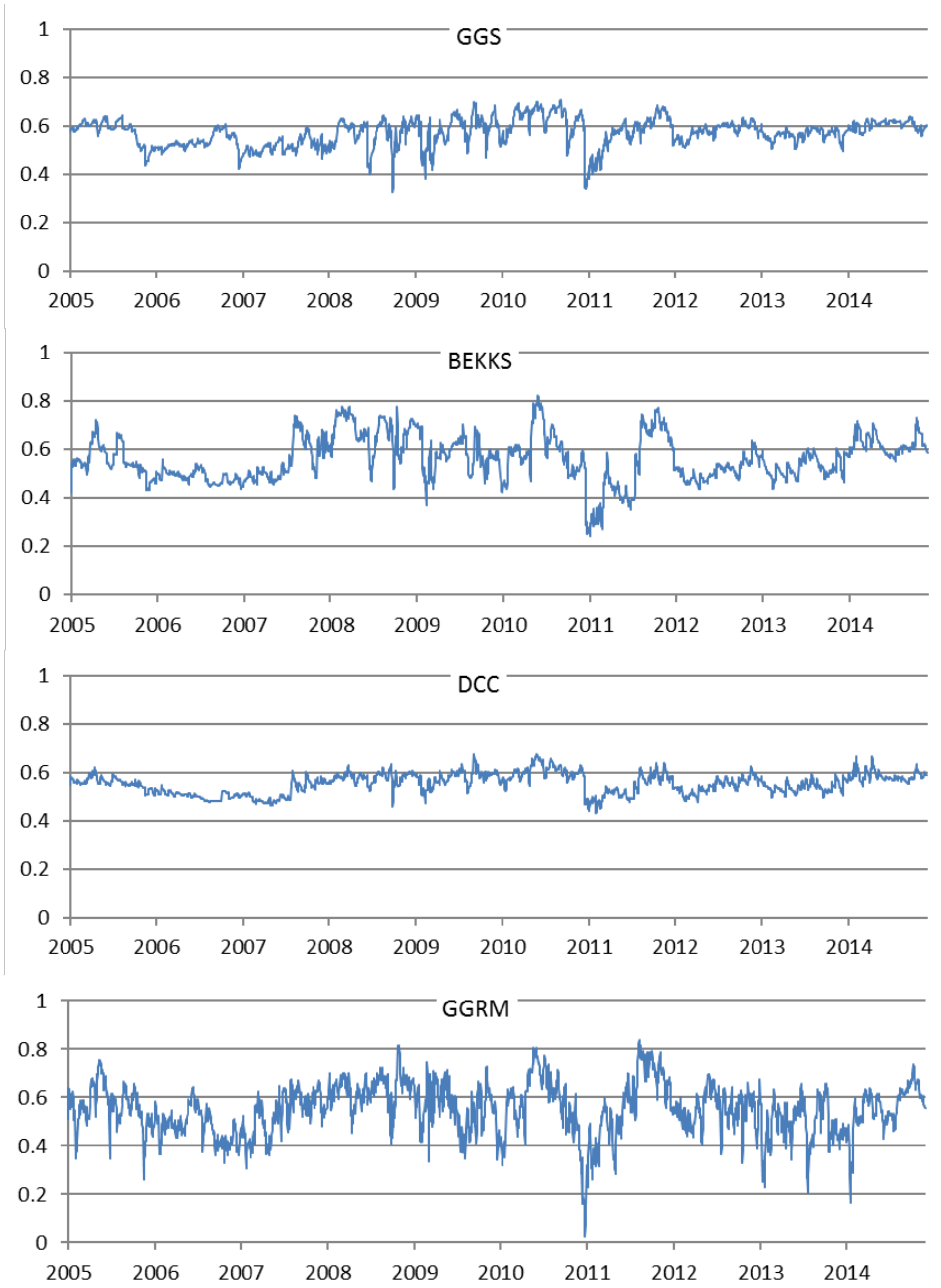
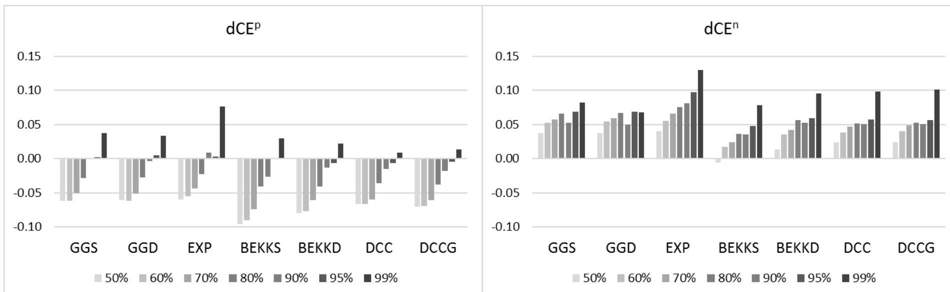
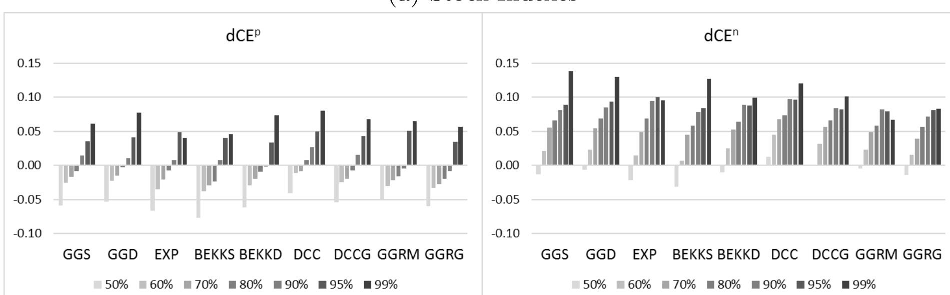


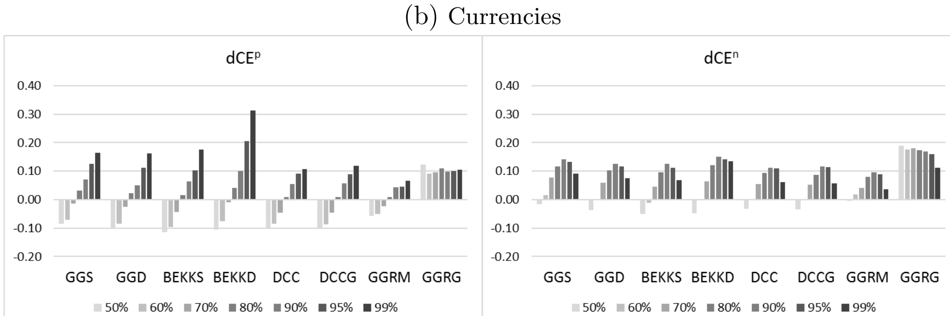
Figure 8: One day forecast of the correlation between GE and AE.



(a) Stock Indexes



(b) Currencies



(c) DJIA stocks

Figure 9: Conditional expectations CE_α^p and CE_α^n.

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 degrees of freedom of the distribution, which is estimated simultaneously with other parameters.[6] When t-distribution is used, the probability density function in (35) also needs to be replaced by the probability density function of the t-distribution.

The second and third methods are rather *ad hoc* in that they selectively choose samples around the tail of the distribution while maintaining the normality assumption. More precisely, the second method uses the historical returns only when the absolute value of the portfolio return exceeds its sample standard deviation:

$$|w^\top e_t| > w^\top \bar{H} w, \tag{49}$$

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

$$e_{it} < -\sqrt{\bar{h}_i}, \tag{50}$$

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

The last method borrows the idea of importance sampling. Importance sampling is a simulation technique that reduces the simulation error by drawing more samples from the region of interest. This is done by replacing the original density function f with a new density function g and multiplying the samples from the new density function by the likelihood ratio f/g:

$$f \to g\frac{f}{g}.$$

For the application of importance sampling in risk measurement, see Glasserman (2003) or Glasserman and Li (2005). The fourth method is to apply importance sampling reversely to give more emphasis on the samples from the tail: Define a new density function that has a higher density around the

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

tail, obtain the likelihood ratio, and multiply the inverse of the likelihood ratio to the data in the sample. 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. Then, the likelihood ratio is 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

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

  where w is the portfolio weight 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 QMLE, multiply the log-likelihood function at time t in (26) by $1/l(e_t)$.

The results of these four alternative estimation methods applied to GGS are reported in Table 10, 11, and 12. All four methods reduce the error at the tail compared to the original method (Normal), but the most prominent enhancement is found in the first (t-dist) and the second (Tail) methods. In particular, t-distribution significantly improves tail risk measurement without any substantial performance loss in terms of other metrics. The third (Negative) and fourth (IS) methods which put more weights on the negative returns are not as effective as the first two methods. Except t-distribution, all three methods incur significant overestimation of the variance (negative dS_p) in order to reduce the error at the tail. This suggests that even after accounting for the heteroscedasticity, the asset returns are still skewed and fat tailed. Figure 10 displays the conditional expectation measures for the alternative methods. The charts show that t-distribution fits the negative side of the distribution remarkably well, but it comes at a cost of slight underestimation on the positive side of the distribution. As far as risk is concerned, t-distribution appears to be a better choice than normal distribution.

Table 10: Performance evaluation of the different estimation methods of GGS using the stock Index sample. Normal is the results of the original QMLE and the other four columns are the results of four alternative methods described in Section 6.2 in the same order.

	Normal	t-dist	Tail	Negative	IS
dS_p	0.676	0.895	-40.106	-24.900	-15.511
	1	2	5	4	3
dCE_{99}^n	8.251	2.716	0.357	4.686	7.594
	5	2	1	3	4
$S_p^{\min}$	15.057	15.056	16.186	15.145	15.157
	2	1	5	3	4
dw	3.481	3.449	4.388	2.848	4.416
	3	2	4	1	5

6.3. Minimum Variance Portfolio

The asset weights of the minimum variance portfolios are summarized in Table 13, 14, and 15 for each sample. As expected from the small variation of $dS_p^{\min}$, the mean weights are similar across models. Only GGRG shows a rather different weight distribution, especially in the DJIA sample. This is probably due to the different estimation method employed. It is worth noting that GGARCH models generally yield more stable weights as evidenced from the standard deviation of the weights. This results in the lower turnover of the portfolios obtained from these models.

The standard deviation of the minimum variance portfolio return is often used to assess covariance estimation models; see Chiriac and Voev (2011), and Han et al. (2016), for example. However, as illustrated below, the results based on this criterion can be misleading. Suppose that the true covariance matrix is H, and the estimated covariance matrix is $\hat{H}$. The minimum variance portfolio is given by

$$w = \frac{\hat{H}^{-1} i}{i^\top \hat{H}^{-1} i},$$

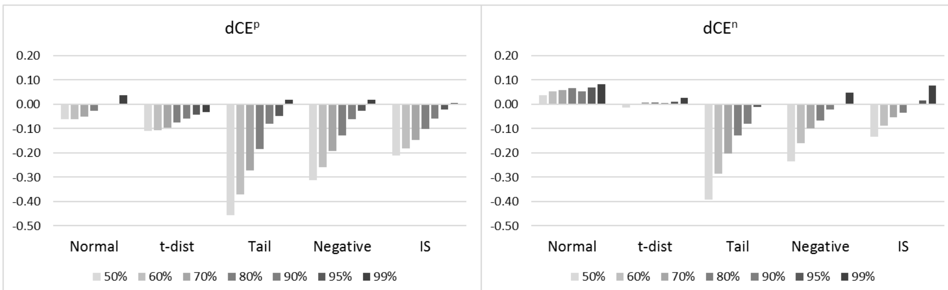
and the variance of the return on this minimum variance portfolio will be $V_p = w^\top H w$. Let dH^{-1} (here -1 is not inverse) denote the difference ($\hat{H}^{-1} -$

Table 11: Performance evaluation of the different estimation methods of GGS using the currency sample. Normal is the results of the original QMLE and the other four columns are the results of four alternative methods described in Section 6.2 in the same order.

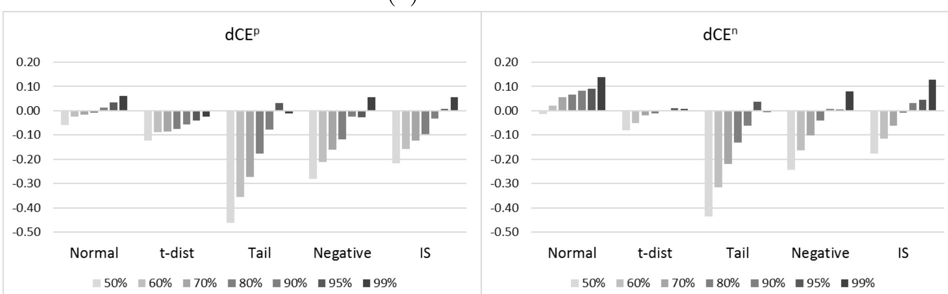
	Normal	t-dist	Tail	Negative	IS
dS_p	0.542	0.766	-41.824	-22.713	-16.136
	1	2	5	4	3
dCE_{99}^n	13.826	0.773	-0.568	8.082	12.832
	5	2	1	3	4
$S_p^{\min}$	6.262	6.258	6.670	6.301	6.384
	2	1	5	3	4
dw	2.781	2.817	3.561	2.348	3.372
	2	3	5	1	4
D_F	0.435	0.434	1.145	0.924	0.518
	2	1	5	4	3
D_G	0.874	0.873	1.501	1.329	0.935
	2	1	5	4	3

Table 12: Performance evaluation of the different estimation methods of GGS using the DJIA stock sample. Normal is the results of the original QMLE and the other four columns are the results of four alternative methods described in Section 6.2 in the same order.

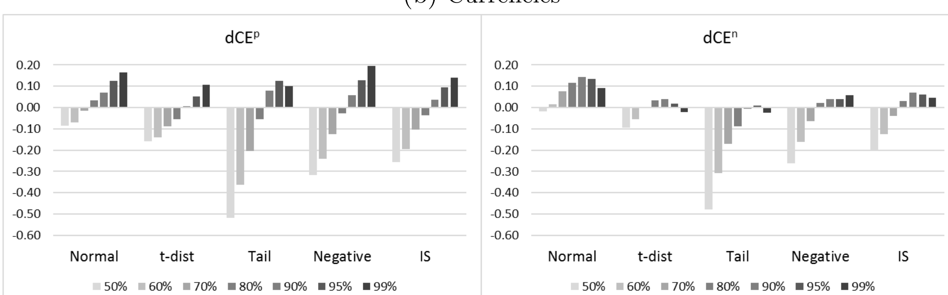
	Normal	t-dist	Tail	Negative	IS
dS_p	1.706	3.072	-42.867	-21.845	-16.784
	1	2	5	4	3
dCE_{99}^n	9.075	-2.251	-2.377	5.819	4.565
	5	1	2	4	3
$S_p^{\min}$	15.012	14.988	16.439	15.341	15.085
	2	1	5	4	3
dw	5.120	4.860	6.588	5.499	6.857
	2	1	4	3	5
D_F	9.320	9.593	27.695	14.058	13.850
	1	2	5	4	3
D_G	1.855	1.870	2.840	2.337	2.105
	1	2	5	4	3



(a) Stock Indexes



(b) Currencies



(c) DJIA stocks

Figure 10: Conditional expectations CE^p_α and CE^n_α from different estimation methods of GGS.

Table 13: Minimum variance portfolio weights: Stock Indexes.

		GGS	GGD	EXP	BEKKS	BEKKD	DCC	DCCG
S&P	Mean	0.403	0.401	0.392	0.410	0.401	0.408	0.410
	Stdev	0.132	0.144	0.193	0.159	0.150	0.174	0.177
	Min	0.103	0.092	-1.072	-0.033	-0.017	0.049	0.062
	Max	0.921	0.860	0.831	1.032	0.900	1.105	1.087
FTSE	Mean	0.338	0.336	0.362	0.339	0.359	0.337	0.337
	Stdev	0.146	0.156	0.205	0.184	0.174	0.188	0.193
	Min	-0.054	-0.063	-0.135	-0.145	-0.058	-0.153	-0.129
	Max	0.640	0.699	1.090	0.948	0.713	0.731	0.772
NIKKEI	Mean	0.259	0.263	0.246	0.251	0.240	0.255	0.253
	Stdev	0.128	0.128	0.165	0.145	0.145	0.142	0.141
	Min	0.012	0.007	-0.097	-0.211	-0.033	0.010	0.008
	Max	0.657	0.659	0.982	0.768	0.768	0.691	0.705

Table 14: Minimum variance portfolio weights: Currencies.

		GGS	GGD	EXP	BEKKS	BEKKD	DCC	DCCG	GGRM	GGRG
EUR	Mean	0.278	0.279	0.276	0.280	0.280	0.281	0.282	0.285	0.315
	Stdev	0.155	0.163	0.181	0.178	0.180	0.178	0.174	0.137	0.113
	Min	0.017	-0.050	-0.064	-0.027	-0.137	-0.056	-0.079	-0.021	0.051
	Max	0.771	0.841	0.746	0.994	0.949	0.925	0.869	0.705	0.674
GBP	Mean	0.286	0.290	0.311	0.292	0.298	0.300	0.301	0.290	0.287
	Stdev	0.151	0.160	0.187	0.169	0.179	0.177	0.170	0.139	0.112
	Min	-0.193	-0.152	-0.075	-0.373	-0.173	-0.186	-0.119	-0.047	0.024
	Max	0.639	0.769	0.826	0.698	0.806	0.932	0.832	0.791	0.662
JPY	Mean	0.436	0.431	0.414	0.428	0.421	0.419	0.418	0.424	0.398
	Stdev	0.077	0.089	0.120	0.082	0.101	0.106	0.112	0.086	0.084
	Min	0.225	0.172	-0.125	0.188	0.084	0.048	0.008	0.121	0.090
	Max	0.618	0.646	0.781	0.713	0.792	0.742	0.756	0.604	0.587

Table 15: Minimum variance portfolio weights: DJIA stocks.

		GGS	GGD	BEKKS	BEKKD	DCC	DCCG	GGRM	GGRG
GE	Mean	0.220	0.222	0.222	0.207	0.219	0.220	0.225	0.168
	Stdev	0.157	0.161	0.205	0.187	0.199	0.198	0.186	0.147
	Min	-0.045	-0.045	-0.103	-0.239	-0.174	-0.183	-0.276	-0.206
	Max	0.823	0.861	1.094	0.976	0.978	0.936	0.830	0.687
AE	Mean	0.034	0.036	0.028	0.049	0.044	0.042	0.027	0.107
	Stdev	0.140	0.153	0.169	0.188	0.168	0.167	0.137	0.120
	Min	-0.268	-0.267	-0.397	-0.402	-0.276	-0.275	-0.405	-0.281
	Max	0.373	0.471	0.428	0.979	0.714	0.717	0.384	0.346
JPM	Mean	0.020	0.012	0.025	0.032	0.023	0.024	0.023	0.049
	Stdev	0.074	0.079	0.135	0.115	0.118	0.121	0.083	0.080
	Min	-0.167	-0.141	-0.432	-0.343	-0.169	-0.183	-0.186	-0.186
	Max	0.266	0.267	0.891	0.371	0.656	0.722	0.396	0.408
HD	Mean	0.152	0.154	0.152	0.154	0.156	0.155	0.152	0.169
	Stdev	0.137	0.134	0.146	0.148	0.170	0.170	0.153	0.116
	Min	-0.058	-0.032	-0.299	-0.301	-0.109	-0.108	-0.147	-0.035
	Max	0.487	0.514	0.546	0.638	0.655	0.663	0.723	0.656
C	Mean	0.011	0.004	0.012	-0.002	0.032	0.032	0.021	0.042
	Stdev	0.142	0.154	0.166	0.158	0.169	0.169	0.164	0.112
	Min	-0.224	-0.224	-0.349	-0.352	-0.238	-0.234	-0.329	-0.268
	Max	0.401	0.438	0.444	0.567	0.630	0.632	0.420	0.344
IBM	Mean	0.563	0.573	0.561	0.560	0.526	0.527	0.552	0.465
	Stdev	0.211	0.216	0.240	0.228	0.236	0.236	0.222	0.227
	Min	0.137	0.151	0.002	0.001	0.039	0.031	0.045	0.022
	Max	1.069	1.054	1.310	1.314	1.032	1.028	1.019	0.932

H^{-1}). The variance of the portfolio return can be approximated as follows.

$$w^\top H w = \frac{i^\top(\hat{H}^{-1} + dH^{-1} + dH^{-1}HdH^{-1})i}{(i^\top \hat{H}^{-1}i)^2} \approx \frac{1}{i^\top \hat{H}^{-1}i} + \frac{i^\top dH^{-1}i}{(i^\top \hat{H}^{-1}i)^2}.$$

While it is straightforward to show that the variance is minimized when $\hat{H} = H$, the effect of the estimation error on the portfolio variance is obscure and will depend on both the size and sign of the error. To see this, consider the following illustrative example.

$$H = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}, \quad \hat{H} = H + dH.$$

Table 16 reports the variance of the minimum variance portfolio return for different values of dH. The results clearly show that the variance of the portfolio is not proportional to the estimation error measured by the Frobenius norm. Although a small variance of the minimum variance portfolio return is likely to indicate a good covariance estimation, the results based on this criterion will always have to validated using other performance metrics.

6.4. Principal Geodesic Analysis

Table 17 reports the results of PGA applied to the DJIA stocks. It also reports the results of the usual PCA applied to the tangent vectors between H_{t-1} and H_t, $\text{Log}_{H_{t-1}}(H_t)$. It is remarkable that only the first 2 principal components out of 21 explain more than 90% of the variation of the covariance matrix. Especially when the PGA is applied, the first component has an explanatory power of remarkable 87%. This result is encouraging as it implies that a parsimonious specification might be obtainable without any significant loss of information. To see the meaning of the first two principal components, a covariance matrix is shifted along the principal axes of these components. For PGA, the intrinsic mean is used as the initial covariance, and for PCA, the sample covariance is used as the initial covariance. The shifted covariances H_{kp} and H_{km} are obtained from

$$H_{kp} = \text{Exp}_{H_I}\left(\overline{|\alpha_k|}V_k\right), \quad H_{km} = \text{Exp}_{H_I}\left(-\overline{|\alpha_k|}V_k\right), \quad k = 1, 2, \qquad (51)$$

where H_I is the initial covariance, and $\overline{|\alpha_k|}$ is the sample mean of the absolute value of α_{ki} in (10). The results are reported in Table 18 for PGA and in Table 19 for PCA. In the tables, the upper triangular part of each matrix contains variances and covariances, and the lower triangular part below the

Table 16: Minimum variance portfolio sensitivity analysis.

dH_1	dH_2	dH_{12}	$\|\|dH\|\|$	V_p
0	0	0	0.000	0.750
0	0	0.1	0.141	0.750
0	0	-0.1	0.141	0.750
0	0.1	0	0.100	0.752
0	0.1	0.1	0.173	0.753
0	0.1	-0.1	0.173	0.751
0	-0.1	0	0.100	0.753
0	-0.1	0.1	0.173	0.755
0	-0.1	-0.1	0.173	0.752
0.1	0	0	0.100	0.752
0.1	0	0.1	0.173	0.753
0.1	0	-0.1	0.173	0.751
0.1	0.1	0	0.141	0.750
0.1	0.1	0.1	0.200	0.750
0.1	0.1	-0.1	0.200	0.750
0.1	-0.1	0	0.141	0.760
0.1	-0.1	0.1	0.200	0.766
0.1	-0.1	-0.1	0.200	0.757
-0.1	0	0	0.100	0.753
-0.1	0	0.1	0.173	0.755
-0.1	0	-0.1	0.173	0.752
-0.1	0.1	0	0.141	0.760
-0.1	0.1	0.1	0.200	0.766
-0.1	0.1	-0.1	0.200	0.757
-0.1	-0.1	0	0.141	0.750
-0.1	-0.1	0.1	0.200	0.750
-0.1	-0.1	-0.1	0.200	0.750

main diagonal contains correlation coefficients. Table 18 shows that the first principal component is related to a co-movement of the variance terms whilst the correlation terms remain still (they move in the same direction but only slightly). The second principal component is, albeit less obvious, related to an independent movement of Citi Bank (C). In fact, Citi Bank has a much larger variance compared to other stocks, as shown in the sample covariance (H_I) in Table 19. The meaning of the principal components obtained form PCA is less clear: Both the first and second principal components seem relevant to the large variance of Citi Bank. The larger contribution and clearer meaning of the first component of PGA makes PGA more attractive than the PCA applied to $\mathrm{Log}_{H_{t-1}}(H_t)$. The results of PGA may serve as a basis for development of a parsimonious covariance model. Proposition of a parsimonious model based on PGA is beyond the scope of this paper and left for future research.

Table 17: PGA (left) and PCA (right) results of the realized covariance matrices of the DJIA stock sample.

Eigenvalue	Percentage	Cumulative	Eigenvalue	Percentage	Cumulative
	$H_t - H_I$			$H_t - H_{t-1}$	
1.23E-07	87.326	87.326	1.57E-05	77.282	77.282
6.75E-09	4.803	92.129	3.00E-06	14.799	92.080
2.31E-09	1.647	93.776	6.89E-07	3.396	95.477
2.12E-09	1.509	95.285	3.07E-07	1.514	96.991
1.91E-09	1.362	96.647	1.84E-07	0.909	97.900

6.5. Parsimonious Specification

In this section, a parsimonious version of the models in which the constant matrix is fixed is investigated. This is often called covariance targeting. The n^2 terms in the number of parameters in Table 4 are from the constant matrix except for EXP, and fixing the constant matrix makes the number of parameters of $O(n)$.

For DCC, covariance targeting is straightforward: a sample analogue can be used as an estimate of $\bar{P}$. But it is less straightforward for GGARCH models as the unconditional expectation of H_t is difficult to obtain due to the exponential form. From Figure 1, it can be seen that H_I can be rewritten as

$$H_I = \mathrm{Exp}_{H'_{t-1}}\left(\frac{1}{1-\alpha}F_{t-1}\right),\ F_{t-1} = \mathrm{Log}_{H'_{t-1}}(H_t). \tag{52}$$

F_{t-1} is the tangent vector of the geodesic emanating from H'_{t-1} to H_t. Let

Table 18: Covariance matrix movement along the first two principal axes obtained by PGA. H_I is the initial covariance, and H_{kp} and H_{km} are the covariances shifted along the k-th principal axis using the formulae in (51). The lower matrix in each shifted covariance is the change defined as $(H_{kp} - H_I)/H_I$ or $(H_{km} - H_I)/H_I$.

	GE	AE	JPM	HD	C	IBM	GE	AE	JPM	HD	C	IBM
			H_I									
GE	0.92	0.40	0.43	0.35	0.45	0.30						
AE	0.41	1.02	0.53	0.37	0.53	0.31						
JPM	0.40	0.47	1.24	0.40	0.69	0.33						
HD	0.36	0.36	0.35	1.02	0.40	0.29						
C	0.38	0.43	0.51	0.32	1.48	0.33						
IBM	0.39	0.38	0.37	0.36	0.34	0.63						
			H_{1p}						H_{1m}			
GE	2.27	1.25	1.35	0.99	1.42	0.82	0.87	0.36	0.40	0.27	0.47	0.23
AE	0.49	2.93	1.79	1.12	1.81	0.89	0.39	0.98	0.53	0.30	0.59	0.24
JPM	0.48	0.56	3.51	1.19	2.21	0.94	0.39	0.49	1.19	0.31	0.80	0.26
HD	0.45	0.45	0.44	2.13	1.23	0.74	0.34	0.35	0.34	0.72	0.36	0.20
C	0.44	0.50	0.55	0.40	4.53	0.96	0.35	0.41	0.51	0.29	2.09	0.29
IBM	0.48	0.46	0.44	0.45	0.40	1.28	0.36	0.36	0.34	0.34	0.29	0.46
GE	1.46	2.16	2.11	1.82	2.17	1.71	-0.49	-0.59	-0.58	-0.55	-0.58	-0.53
AE	0.19	1.86	2.35	2.04	2.41	1.90	-0.15	-0.55	-0.61	-0.57	-0.62	-0.55
JPM	0.18	0.18	1.83	2.01	2.20	1.87	-0.13	-0.15	-0.54	-0.57	-0.59	-0.54
HD	0.25	0.25	0.24	1.08	2.09	1.58	-0.17	-0.17	-0.16	-0.41	-0.57	-0.51
C	0.15	0.15	0.09	0.22	2.07	1.94	-0.11	-0.12	-0.06	-0.15	-0.57	-0.55
IBM	0.22	0.21	0.20	0.26	0.18	1.02	-0.15	-0.13	-0.12	-0.17	-0.10	-0.40
			H_{2p}						H_{2m}			
GE	0.47	0.16	0.18	0.16	0.19	0.14	0.99	0.44	0.49	0.46	0.47	0.39
AE	0.35	0.46	0.21	0.16	0.20	0.14	0.43	1.09	0.55	0.47	0.52	0.39
JPM	0.35	0.40	0.56	0.17	0.28	0.15	0.43	0.46	1.31	0.51	0.65	0.42
HD	0.30	0.30	0.30	0.60	0.17	0.14	0.39	0.38	0.38	1.42	0.48	0.41
C	0.34	0.38	0.48	0.28	0.63	0.15	0.44	0.47	0.53	0.38	1.13	0.40
IBM	0.33	0.33	0.32	0.30	0.30	0.38	0.43	0.40	0.40	0.37	0.41	0.85
GE	-0.06	-0.09	-0.08	-0.23	0.04	-0.24	0.08	0.12	0.12	0.31	0.04	0.31
AE	-0.03	-0.05	0.00	-0.20	0.11	-0.20	0.05	0.06	0.04	0.27	-0.02	0.27
JPM	-0.03	0.04	-0.04	-0.21	0.17	-0.22	0.05	-0.02	0.06	0.30	-0.06	0.29
HD	-0.06	-0.02	-0.04	-0.29	-0.10	-0.31	0.08	0.05	0.07	0.38	0.21	0.42
C	-0.10	-0.04	0.00	-0.10	0.41	-0.13	0.15	0.09	0.05	0.18	-0.24	0.22
IBM	-0.08	-0.04	-0.07	-0.04	-0.14	-0.27	0.09	0.06	0.09	0.04	0.21	0.34

Table 19: Covariance matrix movement along the first two principal axes obtained by PCA. H_I is the initial covariance, and H_{kp} and H_{km} are the covariances shifted along the k-th principal axis using the formulae in (51). The lower matrix in each shifted covariance is the change defined as $(H_{kp} - H_I)/H_I$ or $(H_{km} - H_I)/H_I$.

	GE	AE	JPM	HD	C	IBM	GE	AE	JPM	HD	C	IBM
			H_I									
GE	2.30	1.29	1.38	0.98	1.55	0.81						
AE	0.49	3.01	1.84	1.14	2.00	0.90						
JPM	0.48	0.56	3.59	1.21	2.43	0.95						
HD	0.45	0.46	0.45	2.05	1.32	0.75						
C	0.42	0.48	0.53	0.38	5.85	1.01						
IBM	0.47	0.46	0.45	0.46	0.37	1.28						
			H_{1p}						H_{1m}			
GE	2.55	1.52	1.71	1.12	2.07	0.93	1.78	0.86	0.77	0.71	0.77	0.58
AE	0.51	3.44	2.21	1.32	2.61	1.05	0.45	2.09	1.15	0.78	1.05	0.62
JPM	0.51	0.57	4.34	1.46	3.41	1.18	0.39	0.54	2.18	0.75	1.00	0.54
HD	0.47	0.48	0.47	2.21	1.69	0.85	0.41	0.42	0.39	1.69	0.75	0.53
C	0.44	0.48	0.56	0.39	8.66	1.32	0.41	0.51	0.48	0.41	2.01	0.54
IBM	0.49	0.48	0.48	0.48	0.38	1.40	0.43	0.42	0.36	0.40	0.38	1.02
GE	0.11	0.17	0.24	0.14	0.34	0.15	-0.23	-0.33	-0.44	-0.28	-0.50	-0.28
AE	0.04	0.14	0.20	0.16	0.31	0.16	-0.09	-0.30	-0.38	-0.31	-0.47	-0.31
JPM	0.07	0.02	0.21	0.20	0.40	0.23	-0.18	-0.04	-0.39	-0.38	-0.59	-0.43
HD	0.04	0.04	0.05	0.08	0.28	0.14	-0.09	-0.09	-0.12	-0.18	-0.43	-0.29
C	0.04	0.00	0.05	0.01	0.48	0.30	-0.03	0.08	-0.09	0.07	-0.66	-0.46
IBM	0.04	0.04	0.07	0.05	0.03	0.09	-0.09	-0.08	-0.19	-0.13	0.03	-0.20
			H_{2p}						H_{2m}			
GE	1.73	0.76	0.98	0.65	1.46	0.59	2.84	1.92	1.80	1.32	1.82	1.02
AE	0.46	1.54	1.08	0.61	1.50	0.56	0.52	4.73	2.73	1.73	2.75	1.29
JPM	0.45	0.52	2.75	0.75	2.23	0.69	0.51	0.60	4.40	1.68	2.87	1.22
HD	0.40	0.40	0.37	1.52	0.92	0.50	0.49	0.50	0.50	2.53	1.74	0.98
C	0.38	0.41	0.46	0.25	8.69	0.75	0.48	0.56	0.61	0.49	5.02	1.28
IBM	0.45	0.45	0.42	0.41	0.26	1.00	0.49	0.48	0.47	0.50	0.46	1.52
GE	-0.25	-0.41	-0.29	-0.34	-0.05	-0.27	0.23	0.48	0.31	0.34	0.18	0.26
AE	-0.06	-0.49	-0.41	-0.46	-0.25	-0.38	0.06	0.57	0.48	0.52	0.38	0.42
JPM	-0.06	-0.06	-0.23	-0.38	-0.08	-0.28	0.06	0.07	0.23	0.38	0.18	0.27
HD	-0.12	-0.12	-0.18	-0.26	-0.30	-0.33	0.09	0.09	0.12	0.23	0.32	0.31
C	-0.10	-0.13	-0.14	-0.33	0.48	-0.26	0.15	0.19	0.15	0.28	-0.14	0.26
IBM	-0.05	-0.02	-0.06	-0.12	-0.31	-0.22	0.04	0.04	0.06	0.08	0.25	0.19

$\bar{H}$ and $\bar{M}$ denote $E[e_{t-1}e_{t-1}^\top]$ and $E[\eta_{t-1}\eta_{t-1}^\top]$, respectively. Then,

$$E[H_t] = \bar{H},$$
$$E[H'_{t-1}] = \left(ii^\top - \vec{a}\vec{a}^\top\right) \circ \bar{H} + \vec{a}\vec{a}^\top \circ C \circ ((1-b^2)\bar{H} + b^2\bar{M}).$$

H_I is approximated by substituting H_t and H'_{t-1} in (52) with their expected values:

$$H_I = \mathrm{Exp}_{\bar{H}'}\left(\frac{1}{1-\alpha}F\right), \; F = \mathrm{Log}_{\bar{H}'}(\bar{H}), \tag{53}$$

where $\bar{H}'$ denotes $E[H'_{t-1}]$. Sample analogues are used as estimates of $\bar{H}$ and $\bar{M}$.

The results of covariance targeting for some selected models are reported in Table 20. Surprisingly, fixing the constant matrix does almost no harm to the overall performance regardless of the model. In fact, the fixed constant matrix models outperform their counterparts in terms of some performance metrics. The loss of information seems to be offset by the gain in robustness.

Table 20: Covariance targeting. Est columns are the original results without covariance targeting, and Fix columns are the results from covariance targeting.

	GGS		GGD		DCC		DCCG		GGRM	
	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_{99}^n	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

7. Conclusion

In this paper, new multivariate GARCH models (GGARCH) are developed. These models preserve the geometric structure of the covariance matrix without any arbitrary restrictions by respecting the inherent geometric features of the covariance matrix. One class of models utilizes asset returns whilst the other class utilizes realized covariance. For the latter class of models, a new parameter estimation method based on geometrically correct notions of distance between covariance matrices is developed.

Several versions of the new models are tested on three data samples and compared with existing models: BEKK, DCC, and matrix exponential GARCH. These models are assessed via various out-of-sample performance metrics. Not only the conventional performance evaluation metrics such as out-of-sample likelihood, but also new performance metrics that are particularly suitable for risk management are introduced.

The empirical results suggest that the GGARCH models outperform the existing models, and realized covariance based models outperform return based models. In particular, a realized covariance based GGARCH model estimated via QMLE performs best. This implies 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 that there is a lack of consistency among the performance metrics. This casts a doubt on the previous conclusions in the literature that are drawn from a single or a small number of evaluation methods.

An investigation of time series variation of covariance matrices via principal geodesic analysis shows that the first principal component can explain 87% of the variation of a 6×6 covariance matrix, and the cumulative explanatory power of the first two components is over 92%. This result suggests potential for a parsimonious specification of covariance dynamics for a large dimensional system without sacrificing significant amount of information. We leave this topic as a future research.

References

Bauwens, L., Laurent, S., Rombouts, V., 2006. Multivariate GARCH models: A survey. Journal of Applied Econometrics 21, 79–109.

Becker, R., Clements, A. E., 2008. Are combination forecasts of S&P 500 volatility statistically superior? International Journal of Forecasting 24 (1), 122–133.

Björk, T., 2001. A geometric view of interest rate theory. In: Option Pricing, Interest Rates and Risk Management. Cambridge University Press.

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

Cappiello, L., Engle, R. F., Sheppard, K., 2006. Asymmetric dynamics in the correlations of global equity and bond returns. Journal of Financial Econometrics 4 (4), 537–572.

Chiriac, R., Voev, V., 2011. Modelling and forecasting multivariate realized volatility. Journal of Applied Econometrics 26 (6), 922–947.

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.

Filipović, D., Teichmann, J., 2004. On the geometry of the term structure of interest rates. Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 460 (2041), 129–167.

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). Vol. LNCS 3117. 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. Vol. 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 (2), 117–152.

Glasserman, P., Li, J., 2005. Importance sampling for portfolio credit risk. Management Science 51 (11), 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 (5), 1779–1801.

Han, C., Park, F. C., Kang, J., 2016. A geometric treatment of time-varying volatilities. Working Paper.

Hansen, P. R., Lunde, A., 2005. A forecast comparison of volatility models: does anything beat a GARCH(1,1)? Journal of Applied Econometrics 20 (7), 873–889.

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 (6), 934–955.

Lenglet, C., Rousson, M., Deriche, R., Faugeras, O., 2006. Statistics on the manifold of multivariate normal distributions: Theory and application to diffusion tensor mri processing. Journal of Mathematical Imaging and Vision 25 (3), 423–444.

Liu, L., Patton, A. J., Sheppard, K., 2014. Does Anything Beat 5-Minute RV? A Comparison of Realized Measures Across Multiple Asset Classes. Economics series working paper, University of Oxford, Department of Economics.

Moakher, M., 2005. A differential geometric approach to the geometric mean of symmetric positive-definite matrices. SIAM Journal on Matrix Analysis and Applications 26 (3), 735–747.

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 (6), 907–933.

Noureldin, D., Shephard, N., Sheppard, K., 2014. Multivariate rotated ARCH models. Journal of Econometrics 179 (1), 16–30.

Park, F. C., Chun, C. M., Han, C. W., Webber, N., 2011. Interest rate models on Lie groups. Quantitative Finance 11 (4), 559–572.

Silvennoinen, A., Teräsvirta, T., 2009. Handbook of Financial Time Series. Springer, Ch. Multivariate GARCH models, pp. 201–229.

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

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

Webber, N., 2000. Valuation of financial models with nonlinear state spaces. Working paper.