A Geometric Treatment of Time-Varying Volatilities

Chulwoo Han[a],

[a]University of Durham
Mill Hill Lane
Durham DH1 5SZ, UK
chulwoo.han@durham.ac.uk
+44 19 1334 8542

Abstract

In this article, we propose a new framework for addressing multivariate time-varying volatilities. By employing methods of differential geometry, our model respects the geometric structure of the covariance space, *i.e.*, symmetry and positive definiteness, in a way that is independent of any local coordinate parametrization. Its parsimonious specification makes it particularly suitable for large dimensional systems. Simulation studies suggest that our model embraces much of the nonlinear behaviour of the covariance dynamics. Applied to the US and the UK stock markets, the model performs well, especially when applied to risk measurement. In a broad context, our framework presents a new approach treating nonlinear properties observed in the financial market, and numerous areas of application can be further considered.

JEL Classification: C10; C32; C51; C53

Keywords: Multivariate volatility model; Differential geometry; Riemmanian metric; GARCH

1. Introduction

Since the introduction of the ARCH model by Engle (1982), time-varying volatility models have played an important role in finance and have been successfully applied to various financial problems. From a portfolio viewpoint, it is natural to extend the GARCH-type models to a multivariate system, as first done by Bollerslev et al (1988). Since their work, various types of multivariate time-varying volatility models have flourished. Some of the notable examples are the BEKK model of Engle and Kroner (1995), the DCC model of Engle (2002), and the matrix exponential GARCH model of Kawakatsu (2006). For more comprehensive references, we refer the reader to the survey papers by Bauwens et al (2006) and Silvennoinen and Teräsvirta (2009).

The areas of application of time-varying volatility models are extensive; potentially all the areas where covariance dynamics comes into play, such as asset pricing and portfolio optimization, can be considered. It was first applied to asset pricing by Bollerslev et al (1988), and later by De Santis and Gerard (1998) and Hafner and Herwartz (1998). Cross relations in volatilities of several markets were studied by Kearney and Patton (2000) and Karolyi (1995), and Lien and Tse (2002) computed time-varying hedge ratios using a multivariate GARCH model. Wang and Chen (2007) explore the dynamics between the spot market and its derivative markets using a multivariate GARCH-M model.

Despite the wide areas of application, development of multivariate GARCH models has been neither as active nor as rapid as univariate GARCH models. Perhaps the most obvious reason for this is the proverbial "curse of dimensionality": because there are $n(n + 1)/2$ elements in the covariance matrix of an n-variable system and the number of parameters to be estimated, in the

most general case, will be of $O(n^4)$, the computational burden rapidly becomes untenable for even moderately large n.

Another important - and more subtle - difficulty in designing a time series model of a covariance matrix is the need to preserve its geometric structure, *i.e.,* positive definiteness and symmetry. While symmetry can be easily guaranteed, ensuring that a covariance matrix remains positive definite in a mathematically consistent way (we illuminate on this point further below) is not necessarily straightforward. Earlier treatments have often been *ad hoc* and not taken the underlying geometric structure of the positive definite matrices into account. For example, the $vech()$-type model proposed by Bollerslev et al (1988) takes the crude and obvious approach of transforming the covariance matrix into a column vector, and specifying the vector dynamics accordingly; in this case positive definiteness is ensured only after imposing the relevant algebraic constraints on the vector parameters. Some other recent models address the positive definiteness issue by imposing other sets of assumptions: see, for example, Tse and Tsui (2002) and Engle (2002). The well-known BEKK model of Engle and Kroner (1995) and its variants ensure positive definiteness by assuming a quadratic form for each term in the dynamics equation.

More recently, in work that is particularly relevant to ours, Kawakatsu (2006) extends the exponential GARCH model of Nelson (1991) to a multivariate system, and assumes a $vech()$-type time series model for the log of the covariance matrix. Since the matrix exponential of a symmetric matrix is always symmetric and positive definite, the geometric structure of the covariance matrix is preserved in this model without additional constraints. The exponential map, it turns out, plays a particularly important role in the geometric characterization of matrix groups, and the work of Kawakatsu (2006) is a promising first step that implies the important fundamental connections between geometry and the positive definite matrices.

Why is taking geometry into account important? This point can perhaps be best illustrated via an example of a point moving on the surface of a sphere. In principle, one could describe the motion of the point using standard three-dimensional Cartesian coordinates $\boldsymbol{x} \in \Re^3$, with further imposition of the constraint that length of $\boldsymbol{x}$ be one, *i.e.,* $\|\boldsymbol{x}\| = 1$. Imposing the constraint is cumbersome, and it is more convenient to use a set of local two-dimensional coordinates for the sphere; spherical coordinates immediately come to mind. However, there are many alternative choices of two-dimensional coordinates that parametrize the sphere, *e.g.,* stereographic projection. In principle, there should be no reason to prefer one choice of local coordinates over another, as long as essential geometric quantities like length of the curve, and the area of a patch, are preserved. In other words, it is important to distinguish geometric properties from those that are coordinate-dependent. Any meaningful analysis must take into account the geometric properties of the underlying space, and preserve them in a coordinate-invariant way.

Likewise, the space of covariance matrices is not a vector space, but a curved space. More precisely, it has the structure of a differentiable manifold that we denote $P(n)$. $P(n)$ is a differentiable Riemannian manifold and this fact motivates us to specify the dynamics of the covariance matrix using an appropriate set of local coordinates, and also taking care to ensure that the geometric structure of $P(n)$ is always respected. In fact, there is well-developed machinery to specify continuous-time dynamics on differentiable manifolds, and we adopt this as the starting point for developing a general multivariate volatility model.

Beyond the aesthetic appeal of describing the dynamics on a curved space in an intrinsic fashion, however, the most critical reason, that also has a profound effect on computational results, is that ignoring the geometric structure more often than not will lead to arbitrary results without physical meaning. To illustrate this with an example pointed out by Fletcher and Joshi (2004), one can define a linear average of a collection of positive definite symmetric matrices by taking their sum and dividing by the number of elements. This procedure ensures that the resulting average is

also positive-definite. However, such linear averages fail to interpolate natural properties. The linear average of matrices of the same determinant, for example, can result in a matrix with a larger determinant. This issue is even worse for second-order statistics. The standard principal component analysis as formulated on vector spaces is no longer valid, since the "straight lines" corresponding to the eigenmodes of variation do not remain on the space $P(n)$—standard vector space principal component analysis fails to preserve positive-definiteness.

The aforementioned limitations of the traditional approaches can be overcome using methods of differential geometry. With a proper Riemannian metric, a "straight line" (geodesic) connecting two covariance matrices that remains on $P(n)$ can be defined, and the distance can be measured. The notion of geodesic can then be utilized to define statistics on $P(n)$, such as mean and covariance. A version of principal component analysis for covariance matrices can also be developed. By employing these tools from differential geometry, the objective of the paper is to propose a new framework for modelling covariance dynamics from a differential geometry perspective, and develop multivariate volatility models under the framework. Our models preserve symmetry and positive definiteness of the covariance matrix as it evolves, without imposing any *ad hoc* restrictions. We demonstrate benefits of the framework via simulation studies and compare our models with existing ones in empirical studies. Both simulation and empirical studies suggest that our models perform well and possess properties desired for covariance dynamics.

The remainder of the paper is organized as follows. Section 2 reviews the geometry of $P(n)$, including its Riemannian structure, evaluation of geodesics in a given direction and also between two arbitrary points, and its natural extension to the computation of intrinsic sample mean and covariance on $P(n)$. Section 3 constructs differential equations on $P(n)$, with a focus on linear and quadratic equations—these are chosen for their ability to model diverse correlation phenomena observed in various multivariate time series data, without sacrificing analytic and computational tractability. From the differential equations, we develop a geometrically well-defined multivariate volatility model. We also propose a variation of the model by employing the principal component analysis. In Section 4, the characteristics of our model are illustrated via simulation analysis and its performance is assessed through empirical studies. We conclude in Section 5 with a discussion of open problems and future extensions of this geometric approach to time-varying volatility.

2. Geometry of $P(n)$

In this section, geometric properties of the covariance space are briefly introduced. 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

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

$P(n)$ is a differentiable manifold whose tangent space at a point $\boldsymbol{P} \in P(n)$ can be identified with $n \times n$ symmetric matrices, $S(n)$. A basis for $S(n)$ can be constructed in the usual way, *i.e.*, the basis element $\boldsymbol{E}_{ij} \in S(n)$, where $i \geq j$, is a symmetric matrix whose ij and ji elements are one, and the remaining elements zero.

A Riemannian structure can be constructed via the Riemannian metric given by $\langle \boldsymbol{X}, \boldsymbol{Y} \rangle_{\boldsymbol{P}} = \mathrm{tr}(\boldsymbol{P}^{-1}\boldsymbol{X}\boldsymbol{P}^{-1}\boldsymbol{Y})$, $\boldsymbol{X}, \boldsymbol{Y} \in S(n)$. In terms of this metric, the length of a curve $\boldsymbol{P}(t) \in P(n)$, $a \leq t \leq b$, is given by

$$L(\boldsymbol{P}) = \int_a^b \sqrt{\mathrm{tr}\left((\boldsymbol{P}^{-1}(t)\dot{\boldsymbol{P}}(t))^2 \right)} \, dt. \tag{2}$$

This notion of length is invariant not only under reparametrization of $[a, b]$, but also under congruent transformations of the form $\boldsymbol{M} \boldsymbol{P} \boldsymbol{M}^\top$, where $\boldsymbol{M}$ is any fixed element in the general linear group, $GL(n)$. Using the fact that $P(n)$ is a complete space (*i.e.*, the geodesics are well-defined for all t), the minimal geodesic $\boldsymbol{\gamma}(t) : [0, 1] \to [\boldsymbol{A}, \boldsymbol{B}]$ connecting two points $\boldsymbol{A}, \boldsymbol{B} \in P(n)$ is given by

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

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

$$\mathrm{Log}_{\boldsymbol{A}}(\boldsymbol{B}) = \boldsymbol{G} \log \big(\boldsymbol{G}^{-1} \boldsymbol{B} \boldsymbol{G}^{-\top} \big) \boldsymbol{G}^\top. \tag{4}$$

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

$$\mathrm{Exp}_{\boldsymbol{A}}(\boldsymbol{X}) = \boldsymbol{G} \exp \big(\boldsymbol{G}^{-1} \boldsymbol{X} \boldsymbol{G}^{-\top} \big) \boldsymbol{G}^\top. \tag{5}$$

Defining the distance between $\boldsymbol{A}$ and $\boldsymbol{B}$ in the usual way by the length of the above minimal geodesic, we have

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

where $\| \cdot \|$ is the Frobenius norm, and $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of the matrix $\boldsymbol{A} \boldsymbol{B}^{-1}$. Since $\boldsymbol{A} \boldsymbol{B}^{-1}$ is symmetric positive-definite, the eigenvalues of $\boldsymbol{A} \boldsymbol{B}^{-1}$ are all positive, and $\log \lambda_i$ is well defined for each i. Note also that $d(\boldsymbol{A}, \boldsymbol{\gamma}(t)) = t \, d(\boldsymbol{A}, \boldsymbol{B})$.

With the above metric structure on $P(n)$, we now discuss sample means and covariances on $P(n)$. A widely used formula for the sample mean of N symmetric positive-definite matrices $\{\boldsymbol{P}_1, \ldots, \boldsymbol{P}_N\}$ is the arithmetic mean $\frac{1}{N} \sum_{i=1}^N \boldsymbol{P}_i$. While the arithmetic mean clearly lies in $P(n)$, it has a number of undesirable properties: *e.g.*, the arithmetic mean of matrices with an equal determinant can have a larger determinant. We therefore focus on the **intrinsic mean**, defined as

$$\underset{\bar{\boldsymbol{P}} \in P(n)}{\arg \min} \sum_{i=1}^N d(\bar{\boldsymbol{P}}, \boldsymbol{P}_i)^2. \tag{7}$$

For the case of two points, the intrinsic mean is simply the midpoint of the minimal geodesic. For arbitrary N, the above intrinsic mean is unique on $P(n)$. Fletcher and Joshi (2004) provides a simple steepest descent algorithm for numerically obtaining the intrinsic mean, using the gradient of (7) given by $\sum_{i=1}^N \log(\bar{\boldsymbol{P}} \boldsymbol{P}_i^{-1})$.

An intrinsic notion of sample covariance on $P(n)$ can also be defined via the Riemannian structure of $P(n)$. Given N elements, $\boldsymbol{P}_1, \ldots, \boldsymbol{P}_N$, and the intrinsic mean $\bar{\boldsymbol{P}}$, the covariance matrix relative to $\bar{\boldsymbol{P}}$ is defined by

$$\Sigma_{\bar{\boldsymbol{P}}} = \frac{1}{N} \sum_{i=1}^N \boldsymbol{X}_i \boldsymbol{X}_i^\top, \tag{8}$$

where $\boldsymbol{X}_i$ is the tangent vector at $\bar{\boldsymbol{P}}$ of the geodesic connecting $\bar{\boldsymbol{P}}$ and $\boldsymbol{P}_i$, *i.e.*, $\boldsymbol{X}_i = \mathrm{Log}_{\bar{\boldsymbol{P}}}(\boldsymbol{P}_i)$. Based on these mean and covariance formulas, a generalized normal distribution on $P(n)$ can be constructed by taking the curvature into account; see Lenglet et al (2006) for details.

The usual principal component analysis defined on vector spaces is no longer valid on $P(n)$. Instead, **principal geodesic analysis (PGA)**, a version of principal component analysis on $P(n)$, can be defined by utilizing the notions of geodesic and intrinsic mean. PGA allows us to identify the principal axes of variation of covariance matrices, which then can be used to develop a parsimonious covariance dynamics model. Details of PGA is provided in Appendix A.

3. Geometrically Well Defined Volatility Models

Assume that an n-variable system, r_t is governed by the following dynamics equation:

$$r_t = \mu + e_t, \qquad e_t \sim N(\mathbf{0}, H_t), \tag{9}$$

where $H_t \in P(n)$ is the covariance matrix of e_t. We assume a constant mean and focus on the covariance dynamics; generalization to a time-varying mean model should be straightforward. The covariance matrix of the shock is assumed to have dynamics of the form

$$dH_t = F_t dt, \tag{10}$$

where F_t is a time-varying $n \times n$ symmetric matrix which depends on the information set at 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 (5), H_t can be obtained by the formula

$$H_t = \text{Exp}_{H_{t-1}}(F_t). \tag{11}$$

Another class of dynamics we consider for H_t is based on the assumption that H_t is mapped by F_t from a constant covariance matrix H_∞, i.e.,

$$H_t = \text{Exp}_{H_\infty}(F_t). \tag{12}$$

In the event that H_∞ is an identity matrix, this model nests the matrix exponential GARCH model by Kawakatsu (2006). In (11), F_t is the tangent vector of the geodesic connecting H_{t-1} and H_t, and it is the tangent vector of the geodesic connecting H_∞ and H_t in (12). Various specifications of F_t can be considered. The remainder of the section is devoted to developing standard models for F_t.

3.1. Geometric GARCH Models

We assume that F_t depends only on the current and past covariances and shocks. In this case, a natural extension of the univariate GARCH model would yield the equation

$$F_t = \sum_{p=1}^{P} (A_p H_{t-p} + H_{t-p} A_p^\top) + \sum_{q=1}^{Q} (B_q e_{t-q} e_{t-q}^\top + e_{t-q} e_{t-q}^\top B_q^\top). \tag{13}$$

We can also construct a quadratic form

$$F_t = \sum_{p=1}^{P} (A_p H_{t-p} + H_{t-p} A_p^\top + H_{t-p} S_p H_{t-p}^\top) + \sum_{q=1}^{Q} (B_q e_{t-q} e_{t-q}^\top + e_{t-q} e_{t-q}^\top B_q^\top). \tag{14}$$

5

In Equation (13) and (14), $\boldsymbol{A}_p, \boldsymbol{S}_p$ and $\boldsymbol{B}_q$ are $n \times n$ matrices. In the event that $\boldsymbol{B}_q$ are zero, we have a standard matrix Riccati equation in $\boldsymbol{H}$, with well-known methods for obtaining solutions:

$$\dot{\boldsymbol{H}} = \boldsymbol{A}\boldsymbol{H} + \boldsymbol{H}\boldsymbol{A}^\top + \boldsymbol{H}\boldsymbol{S}\boldsymbol{H}^\top. \tag{15}$$

The functional form of $\boldsymbol{F}_t$ can be easily extended to accommodate phenomena observed in the market, *e.g.*, asymmetric volatility effect as reported in Kroner and Ng (1998). Asymmetric effect of shocks on $\boldsymbol{F}_t$ is modeled by augmenting Equation (13) with terms of $\boldsymbol{\eta}_t = |\boldsymbol{e}_t| - \boldsymbol{e}_t$:

$$\begin{aligned} \boldsymbol{F}_t &= \sum_{p=1}^{P}(\boldsymbol{A}_p\boldsymbol{H}_{t-p} + \boldsymbol{H}_{t-p}\boldsymbol{A}_p^\top) + \sum_{q=1}^{Q}(\boldsymbol{B}_q\boldsymbol{e}_{t-q}\boldsymbol{e}_{t-q}^\top + \boldsymbol{e}_{t-q}\boldsymbol{e}_{t-q}^\top\boldsymbol{B}_q^\top) \\ &\quad + \sum_{q=1}^{Q}(\boldsymbol{C}_q\boldsymbol{\eta}_{t-q}\boldsymbol{\eta}_{t-q}^\top + \boldsymbol{\eta}_{t-q}\boldsymbol{\eta}_{t-q}^\top\boldsymbol{C}_r^\top). \end{aligned} \tag{16}$$

$\boldsymbol{F}_t$ could be more generally specified by transforming it to a vector as in Kawakatsu (2006), but the above specification is sufficient to capture the essential features of covariance dynamics.

The proposed model is geometrically well-defined: the covariance matrix remains in $P(n)$ as it evolves without further restrictions on the dynamics, as long as the differential $d\boldsymbol{H}_t$ is an element of $S(n)$. We call this specification of time-varying volatilities **Geometric GARCH**, or simply **GGARCH** model.

3.2. PCA Based Specifications

Another formulation of $\boldsymbol{F}_t$ we consider is based on the principal component analysis. The PCA can be applied in two ways depending on the class of the covariance dynamics. For the covariance dynamics defined in (11), the usual PCA can be applied to the tangent vectors of the geodesics connecting $\boldsymbol{H}_{t-1}$ and $\boldsymbol{H}_t$. In conjunction with the dynamics equation in (12), the PGA can be applied to the geodesics connecting $\boldsymbol{H}_\infty$, and $\boldsymbol{H}_t$. In either case, if the variation of the covariance matrix can be approximated by $K < n(n+1)/2$ principal directions, $\{\boldsymbol{V}_k\}$, $\boldsymbol{F}_t$ can be written in the form

$$\boldsymbol{F}_t \approx \sum_{k=1}^{K}\alpha_{kt}\boldsymbol{V}_k, \tag{17}$$

where α_{kt} are time-varying factor loadings. We specify α_{kt} as a scalar function of the lagged terms of the covariance matrix and residuals, $\{\boldsymbol{H}_{t-p}, \boldsymbol{e}_{t-q}\}$. In particular, the following functional form is considered.

$$\alpha_{kt}(\boldsymbol{H}_{t-p}, \boldsymbol{e}_{t-q}) = \text{tr}\left(\sum_{p=1}^{P}\boldsymbol{A}_{k,p}\boldsymbol{H}_{t-p} + \sum_{q=1}^{Q}\boldsymbol{B}_{k,q}\boldsymbol{e}_{t-q}\boldsymbol{e}_{t-q}^\top + \sum_{q=1}^{Q}\boldsymbol{C}_{k,q}\boldsymbol{\eta}_{t-q}\boldsymbol{\eta}_{t-q}^\top\right). \tag{18}$$

where $\boldsymbol{A}_p$, $\boldsymbol{B}_q$, and $\boldsymbol{C}_q$ are $n \times n$ upper triangular matrices. Another method worth consideration is to use the distance function. For example, the following specification can be considered.

$$\alpha_{kt}(\boldsymbol{H}_{t-p}, \boldsymbol{e}_{t-q}) = d\left(\boldsymbol{H}_{t-1}, \sum_{p=1}^{P}a_{k,p}\boldsymbol{H}_{t-p} + \sum_{q=1}^{Q}b_{k,q}\boldsymbol{e}_{t-q}\boldsymbol{e}_{t-q}^\top + \sum_{q=1}^{Q}c_{k,q}\boldsymbol{\eta}_{t-q}\boldsymbol{\eta}_{t-q}^\top\right). \tag{19}$$

where $a_{k,p}$, $b_{k,q}$, and $c_{k,q}$ are nonnegative scalar coefficients.[1] By defining the factor loadings as the distance between the current covariance matrix and a function of the residuals, this specification has a better geometric interpretation. If K is sufficiently small, the PCA-based specifications ensure a parsimonious representation of the covariance dynamics.

Although the principal directions can, in principle, be obtained by applying PCA or PGA to a series of covariance matrices, covariance matrices are unobservable and need to be estimated. A good proxy would be a realized covariance obtained from high frequency data. If high frequency data is not available, sample covariance matrix calculated from a sub-sample could be used.

3.3. Parsimonious Representations

The GGARCH models involve $O(n^2)$ number of parameters in the most general case and unless some restrictions are imposed, the models will be numerically untenable for a large dimensional system. An obvious and popular practice is to restrict the coefficient matrices to diagonal matrices or scalars. Table 1 compares the number of parameters in each GGARCH specification and in some existing models. Note that, while BEKK and DCC models have the number of parameters of $O(n^2)$ regardless of the degree of restriction, diagonal or more restricted versions of the GGARCH model have parameter numbers of $O(n)$ or a constant. This is because the covariance dynamics is defined in the tangent space and there is no constant term in the dynamics equation. A constant term in the tangent space would imply that the covariance matrix tends to move toward a certain direction regardless of its current position, which is not intuitive. Empirical studies also support this: while a constant matrix plays a dominant role in BEKK and DCC models, it turns out to be insignificant in GGARCH models. This gives our model a huge benefit when applied to large dimensional systems.

Model	Parameters	Description
GGARCH PCA DIST	$3K$	α_{kt} is defined as in (19).
GGARCH PCA DIAG	$3nK$	In (18), coefficient matrices are diagonal.
GGARCH PCA FULL	$(1.5n^2 + n)K$	In (18), coefficient matrices are upper triangular.
GGARCH SCALAR	3	In (16), coefficient matrices are scalar.
GGARCH DIAG	$3n$	In (16), coefficient matrices are diagonal.
GGARCH LINEAR	$1.5n^2 + 1.5n$	In (16), coefficient matrices are symmetric and the matrix product is the element-wise product.
GGARCH FULL	$3n^2$	In (16), coefficient matrices are arbitrary $n \times n$ matrices.
BEKK SCALAR	$0.5n^2 + 0.5n + 3$	BEKK model with scalar coefficients.
BEKK DIAGONAL	$0.5n^2 + 3.5n$	BEKK model with diagonal coefficients.
BEKK FULL	$3.5n^2 + 0.5n$	BEKK model with arbitrary $n \times n$ coefficients.
DCC SCALAR	$n^2 + 4n + 3$	DCC model.

Table 1. Multivariate GARCH models and their number of parameters. All models are assumed to have an asymmetric component and be a function of only the first lags.

4. Empirical Studies

In this section, we illustrate the characteristics of the proposed models via simulations. They are then applied to a set of real market data and compared with widely used models, *i.e.*, BEKK

---

[1] $\boldsymbol{H}_{t-1}$ on the right hand side of (19) needs to be replaced by $\boldsymbol{H}_\infty$ under the covariance dynamics specified in (12).

and DCC. Although the case studies introduced in this section are far from exhaustive, they convey some important lessons about our model and shed light on the future directions of research.

4.1. Properties of the Covariance Space

Riemannian exponential map serves as a building block of our covariance dynamics models and it is important to understand its properties in order to correctly specify the tangent vector $\boldsymbol{F}_t$. For illustration purposes, we consider a simple bivariate system; this facilitates visualization and is still sufficient to understand the basic properties.

4.1.1. The Shortest Path

Consider two covariance matrices

$$\boldsymbol{H}_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \text{ and } \boldsymbol{H}_1 = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix},$$

and let

$$\boldsymbol{H}(t) = \begin{bmatrix} h_{11} & h_{12} \\ h_{12} & h_{22} \end{bmatrix}, \ 0 \leq t \leq 1,$$

be a covariance matrix on the geodesic connecting $\boldsymbol{H}_0$ and $\boldsymbol{H}_1$. The geodesic is visualized in Figure 1. In the figure, the first panel plots the trajectory of h_{11} and h_{12}, and the second panel plots the distance between $\boldsymbol{H}(t)$ and $\boldsymbol{H}(t + 0.1)$ for $t = 0, \ldots, 0.9$, with $\boldsymbol{H}(0) = \boldsymbol{H}_0$ and $\boldsymbol{H}(1) = \boldsymbol{H}_1$. The solid lines represent the geodesic and the dotted lines represent the linear interpolation. Interestingly, the trajectory of the variance term is convex and that of the covariance term is slightly concave, contrary to the naive linear interpolation that yields straight lines for both variance and covariance terms. This is because, under the Riemmanian metric, the distance between two covariance matrices increases exponentially as one matrix approaches singularity, *i.e.*, perfect correlation, whereas $\boldsymbol{H}(t)$ moves at a constant speed as shown in Figure 1(b): Convex h_{11} and concave h_{12} implies that the correlation increases more rapidly at the beginning and more slowly later compared to the case of linear interpolation. This results in the equal distance between two adjacent points along the geodesic, contrary to the increasing distance between two adjacent points along the linear interpolation line. This is a desired property of covariance dynamics since, for example, an increase in correlation from 0.0 to 0.5 must be more likely than an increase from 0.5 to 1.0. In fact, for $\boldsymbol{H}_1 = \begin{bmatrix} h & 0.5 \\ 0.5 & h \end{bmatrix}$, the minimum distance between $\boldsymbol{H}_0$ and $\boldsymbol{H}_1$ is achieved when h is near 1.275, whereas it would be achieved when $h = 1$ under a conventional metric such as Frobenius norm. This suggests that an increase in covariance is likely to involve an increase in variance more under the Riemannian metric compared to an Euclidean metric, which is consistent with the fact that the correlations are much more persistent than the variances. This is also supported by the empirical analysis demonstrated later in this section, where the first principal component of covariance matrix variation turns out to be simultaneous changes of the variance and covariance.

4.1.2. Riemmanian Exponential Map

We examine the movement of a covariance matrix when it is repeatedly subject to a tangent vector. This is to assess the effect of repeated shocks in one direction. Suppose the covariance matrix has the initial value

$$\boldsymbol{H}_0 = \begin{bmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{bmatrix}.$$

(a) Trajectory of $\boldsymbol{H}(t)$ (b) $d(\boldsymbol{H}(t), \boldsymbol{H}(t+0.1))$
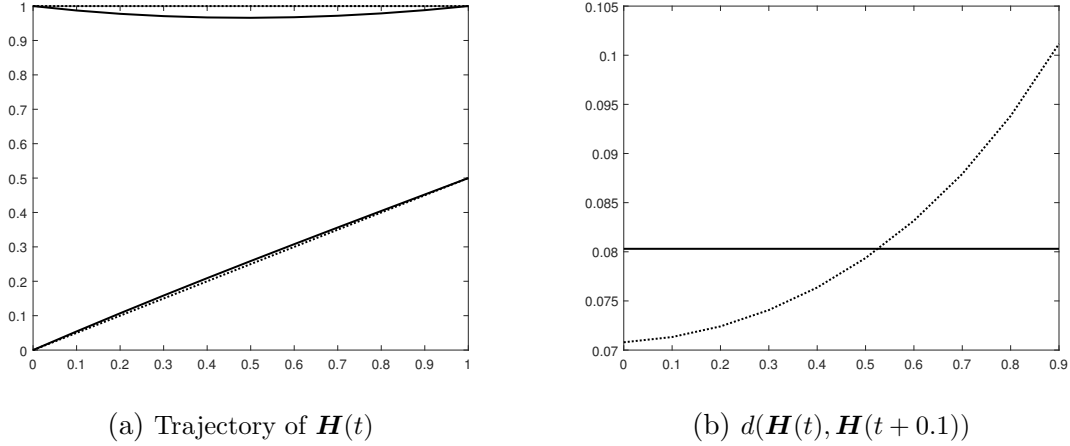
Fig 1. Geodesic between $\boldsymbol{H}_0$ and $\boldsymbol{H}_1$. In (a), the lines at the top are the trajectories of h_{11} and the lines at the bottom are the trajectories of h_{12}. Solid lines represent the geodesic and dotted lines represent the linear interpolation.

The covariance matrix is repeatedly updated by the Riemannian exponential map $\boldsymbol{H}_t = \mathrm{Exp}_{\boldsymbol{H}_{t-1}}(\boldsymbol{F})$ where $\boldsymbol{F}$ has one of three values,

$$\boldsymbol{F}_1 = \begin{bmatrix} 0.05 & 0.00 \\ 0.00 & 0.00 \end{bmatrix}, \ \boldsymbol{F}_2 = \begin{bmatrix} 0.05 & 0.05 \\ 0.05 & 0.05 \end{bmatrix}, \ \text{or } \boldsymbol{F}_3 = \begin{bmatrix} 0.05 & -0.05 \\ -0.05 & 0.05 \end{bmatrix}.$$

$\boldsymbol{F}_1$ implies a shock only on the first variable, while $\boldsymbol{F}_2$ ($\boldsymbol{F}_3$) implies shocks of an equal size in the same (opposite) direction. The results are displayed in Figure 2, 3, and 4. In each figure, σ_1, σ_2, and ρ respectively denote the standard deviations of the two variables and their correlation coefficient. When $\boldsymbol{H}_t$ is subject to $\boldsymbol{F}_1$, σ_1 increases while ρ decreases. This is intuitive as independent shocks would imply reduced correlation. Note that σ_1 increases at a reduced pace when it is high. This is desirable because when the variance is high, the same shock would have a smaller impact. As $\boldsymbol{F}_2$ implies shocks in the same direction, the correlation increases when $\boldsymbol{H}_t$ is subject to $\boldsymbol{F}_2$, and for the same reason, it decreases when $\boldsymbol{H}_t$ is subject to $\boldsymbol{F}_3$. However, the correlation decreases more rapidly, as we would expect from the actual movements of covariance matrices.

While Riemmanian exponential map carries desired properties of covariance dynamics, the behavior of the covariance matrix is not entirely clear due to the nonlinearity of the matrix exponential. This makes specifying the tangent vector and calibrating the model tricky: *e.g.*, one may want to make correlations more persistent than variances. Although the GGARCH model does not allow independent modelling of variances and correlations, one workaround is to incorporate the GGARCH model into the DCC framework, *i.e.*, estimate variances using a univariate GARCH model and estimate correlations using the GGARCH model.

We have investigated various characteristics of the covariance space and related them to some of the features observed in the market. Based on the promising simulation results, we believe that, by correctly specifying the tangent vector, our model can be successfully employed in many financial application domains in which volatilities are best modeled as time-varying.

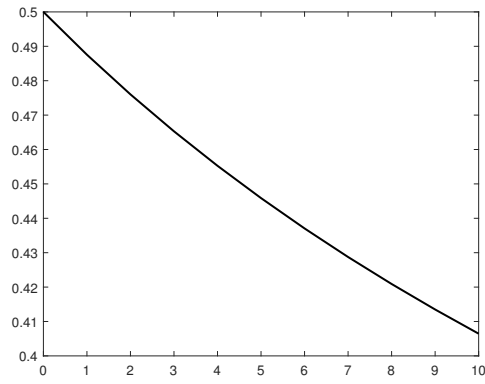### 4.2. Global Market Correlation

In this section, we apply the GGARCH models to a set of real market data and compare them with the BEKK and DCC models. We choose two stock market indexes, S&P500 and FTSE100, both obtained from Thomson Reuters database. Daily index values are collected from October
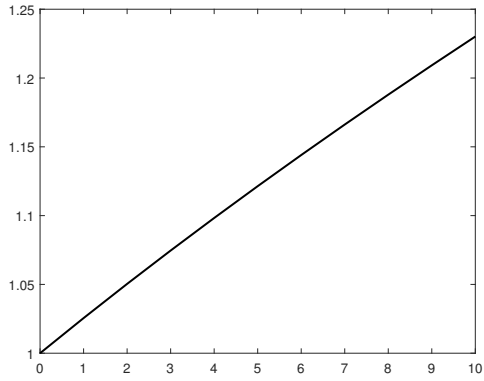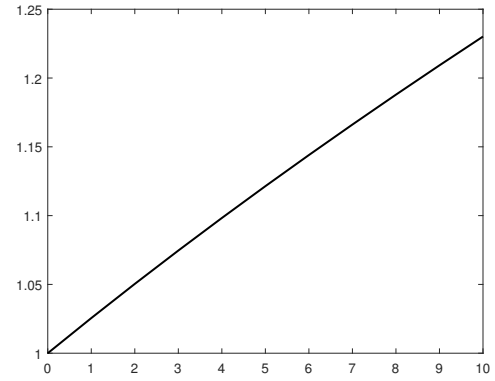
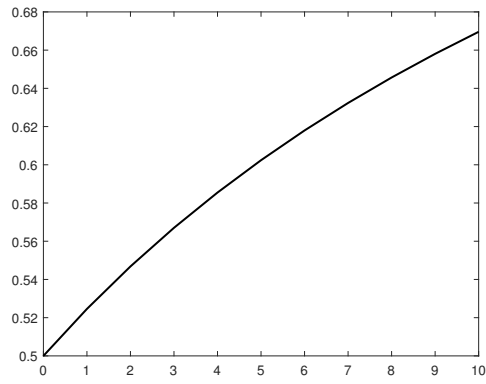(a) σ_1

(b) σ_2

(c) ρ
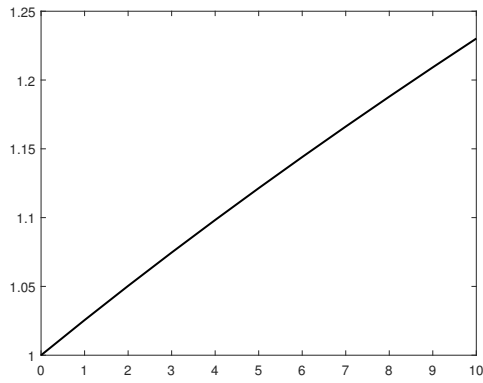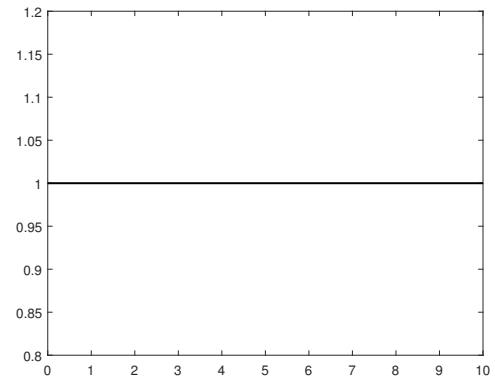
Fig 2. Movement of $\boldsymbol{H}_t$ subject to $\boldsymbol{F}_1$. σ_1, σ_2, and ρ respectively denote the standard deviations of the two variables and their correlation coefficient.

(a) σ_1

(b) σ_2

(c) ρ

Fig 3. Movement of $\boldsymbol{H}_t$ subject to $\boldsymbol{F}_2$. σ_1, σ_2, and ρ respectively denote the standard deviations of the two variables and their correlation coefficient.

(a) σ_1

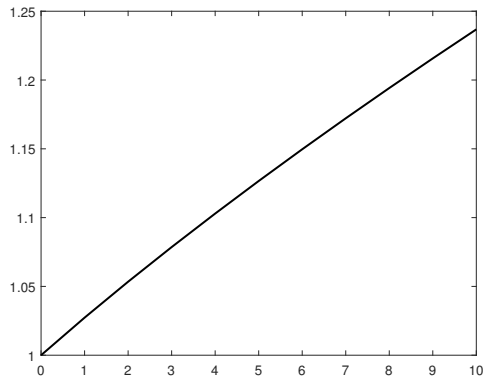

(b) σ_2
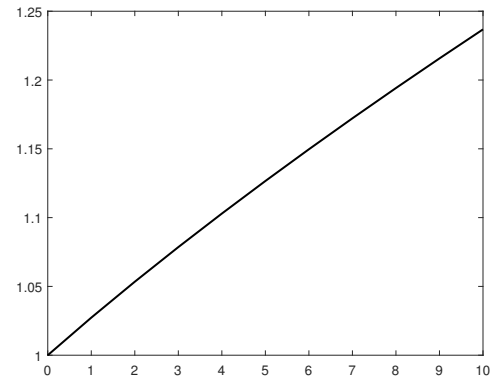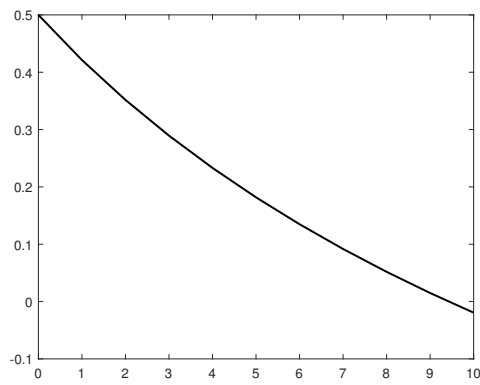

(c) ρ

Fig 4. Movement of $\boldsymbol{H}_t$ subject to $\boldsymbol{F}_3$. σ_1, σ_2, and ρ respectively denote the standard deviations of the two variables and their correlation coefficient.

1, 2003 to September 30, 2013, from which daily log returns are calculated.The sample mean and covariance matrix of the index returns are reported in Table 2. The two markets are highly correlated during the sample period, and have a similar level of mean and variance.

	Mean	Covariance		Correlation	
		S&P500	FTSE100	S&P500	FTSE100
S&P500	1.924E-04	1.600E-04	0.872E-04	1.000	0.578
FTSE100	1.680E-04	0.872E-04	1.425E-04	0.578	1.000

Table 2. Sample Statistics of S&P500 and FTSE100 daily log returns. Sample period is from October 1, 2003 to September 30, 2013.

4.2.1. Test Models

We consider six GGARCH specifications and compare them with two BEKK specifications (scalar and diagonal coefficient matrices) and a DCC specification. The most general form of the BEKK model was also considered but excluded as the estimation results were too sensitive to initial values and therefore unreliable. In all specifications, the covariance matrix is assumed to be a function of only the first lags of the covariance matrix and the residuals. As our primary interest lies in the covariance dynamics, we assume a constant mean return for simplicity.

$$\boldsymbol{r}_t = \boldsymbol{\mu} + \boldsymbol{e}_t, \qquad \boldsymbol{e}_t \sim N(\boldsymbol{0}, \boldsymbol{H}_t), \tag{20}$$

$$\boldsymbol{H}_t = \mathrm{Exp}_{\boldsymbol{H}_{t-1}}(\boldsymbol{F}_t). \tag{21}$$

The six GGARCH specifications are described below. For the details of the BEKK and DCC models, please refer to Engle and Kroner (1995) and Engle (2002), respectively.

- **GGARCH SCALAR**

$$\boldsymbol{F}_t = \boldsymbol{A}\boldsymbol{H}_{t-1} + \boldsymbol{B}\boldsymbol{e}_{t-1}\boldsymbol{e}_{t-1}^{\top} + \boldsymbol{C}\boldsymbol{\eta}_{t-1}\boldsymbol{\eta}_{t-1}^{\top}, \tag{22}$$

where $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$ are scalar.

- **GGARCH DIAG**

$$\boldsymbol{F}_t = \boldsymbol{A}\boldsymbol{H}_{t-1} + \boldsymbol{H}_{t-1}\boldsymbol{A}^{\top} + \boldsymbol{B}\boldsymbol{e}_{t-1}\boldsymbol{e}_{t-1}^{\top} + \boldsymbol{e}_{t-1}\boldsymbol{e}_{t-1}^{\top}\boldsymbol{B}^{\top} + \boldsymbol{C}\boldsymbol{\eta}_{t-1}\boldsymbol{\eta}_{t-1}^{\top} + \boldsymbol{\eta}_{t-1}\boldsymbol{\eta}_{t-1}^{\top}\boldsymbol{C}^{\top}, \tag{23}$$

where $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$ are diagonal.

- **GGARCH LINEAR**

$$\boldsymbol{F}_t = \boldsymbol{A} \otimes \boldsymbol{H}_{t-1} + \boldsymbol{B} \otimes \boldsymbol{e}_{t-1}\boldsymbol{e}_{t-1}^{\top} + \boldsymbol{C} \otimes \boldsymbol{\eta}_{t-1}\boldsymbol{\eta}_{t-1}^{\top}, \tag{24}$$

where $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$ are symmetric, and $\otimes$ is the element-wise matrix product.

- **GGARCH FULL**

$$\boldsymbol{F}_t = \boldsymbol{A}\boldsymbol{H}_{t-1} + \boldsymbol{H}_{t-1}\boldsymbol{A}^{\top} + \boldsymbol{B}\boldsymbol{e}_{t-1}\boldsymbol{e}_{t-1}^{\top} + \boldsymbol{e}_{t-1}\boldsymbol{e}_{t-1}^{\top}\boldsymbol{B}^{\top} + \boldsymbol{C}\boldsymbol{\eta}_{t-1}\boldsymbol{\eta}_{t-1}^{\top} + \boldsymbol{\eta}_{t-1}\boldsymbol{\eta}_{t-1}^{\top}\boldsymbol{C}^{\top}, \tag{25}$$

where $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$ are arbitrary $n \times n$ matrices.

- **GGARCH PCA DIAG**

$$F_t = \sum_{k=1}^{K} \alpha_{kt} V_k \tag{26}$$

$$\alpha_{kt} = \text{tr}(A_k H_{t-1} + B_k e_{t-1} e_{t-1}^\top + C_k \eta_{t-1} \eta_{t-1}^\top),$$

where A_k, B_k, and C_k are diagonal matrices.

- **GGARCH PCA FULL**

$$F_t = \sum_{k=1}^{K} \alpha_{kt} V_k \tag{27}$$

$$\alpha_{kt} = \text{tr}(A_k H_{t-1} + B_k e_{t-1} e_{t-1}^\top + C_k \eta_{t-1} \eta_{t-1}^\top),$$

where A_k, B_k, and C_k are upper triangular matrices.

All the models are estimated via Quasi-Maximum Likelihood Estimation (QMLE).[2]

The PCA-based GGARCH models require a time series of covariance matrices in order to apply PCA prior to model estimation. We generate a covariance matrix time series by calculating sample covariance from the minimum size (two, in our case) subsample at each time t. We tested larger subsamples but the results were qualititively similar. PCA results are reported in Table 3. The first component explains almost 80% of the change and the first two components combined together explain more than 95% of the change. Based on this, we choose $K = 2$. Brief inspection of the eigenvectors reveals an interesting observation: The first component is related to simultaneous changes of the variances and covariance, the second component is related to independent changes of the variances, and finally the third component is related to an independent change of the covariance. The first component implies that the principal variation of the covariance matrix is due to simultaneous changes of the variances while the correlation remaining still.

	1st component	2nd component	3rd component
Eigenvalue	1.703E-06	0.375E-06	0.092E-6
	(78.479%)	(17.281%)	(4.240%)
	0.6470	0.6938	0.3163
Eigenvector	0.4972	-0.0693	-0.8649
	0.5781	-0.7168	0.3898

Table 3. PCA results of the tangent vectors of the geodesics connecting H_{t-1} and H_t.

4.2.2. Estimation and Diagnosis

Estimation and diagnosis results are reported in Table 4 through Table 7. The models are ranked in each table based on their performance: ***, **, and * represent the best three performing models in descending order.

Table 4 compares the log-likelihood value of each model. The gain in the log-likelihood value from increased model flexibility seems trivial. For example, the difference in the log-likelihood values of GGARCH SCALAR and GGARCH LINEAR is mere 0.3%, even though the former has

---

[2]BEKK and DCC models are estimated using MFE Matlab toolbox developed by Kevin Sheppard. The toolbox can be downloaded from `http://www.kevinsheppard.com/wiki/MFE_Toolbox`.

	Log-Likelihood
GGARCH SCALAR	1.7353E+04
GGARCH DIAG	1.7356E+04
GGARCH LINEAR	1.7409E+04
GGARCH FULL	1.7371E+04
GGARCH PCA DIAG	1.7404E+04
GGARCH PCA FULL	1.7478E+04
BEKK SCALAR*	1.7479E+04
BEKK DIAGONAL**	1.7509E+04
DCC SCALAR***	1.7554E+04

Table 4. Log-likelihood as a result of QMLE.

	Q	p-value
GGARCH SCALAR	76.8857	0.0000
GGARCH DIAG	71.4855	0.0004
GGARCH LINEAR	37.5462	0.0646
GGARCH FULL	65.4737	0.0012
GGARCH PCA DIAG	70.6275	0.0000
GGARCH PCA FULL	41.8565	0.0060
BEKK SCALAR*	21.7602	0.1427
BEKK DIAGONAL**	17.2584	0.2642
DCC SCALAR***	14.5448	0.4504

Table 5. Ljung-Box autocorrelation test results. Q denotes Ljung-Box Q statistic.

only 3 parameters, whereas the latter has 9. This means that a model selection criteria such as Bayesian Information Criteria (BIC) would favor the scalar version within each model family. The difference between model families is also small. Still, the DCC and BEKK models achieve higher log-likelihood values than the GGARCH models.

Table 5 reports the results of the autocorrelation test proposed by Ledoit et al (2003). If a model is correctly specified, normalized residuals should be serially uncorrelated. The test results are in favor of the DCC and BEKK models. The null hypothesis of zero autocorrelation cannot be rejected in these models. Among other models, only the GGARCH LINEAR passes the test at 5% level.

A minimum variance portfolio composed of the two market indexes is obtained from the covariance estimate of each model. The portfolio is rebalanced every 22 business days and held until the next rebalancing day. The results are reported in Table 6. It seems difficult to distinguish the models based on the portfolio optimization results. Portfolio composition is similar across the models and the standard deviation of the portfolio return is almost indistinguishable. This is mainly due to the high correlation and the similar sizes of the variances of the two market returns, as reported in Table 2. Nevertheless, subtle differences among the models can still be observed: The GGARCH family allocate consistently lower weights on S&P500 compared to the BEKK and DCC models. Considering the fact that the volatility of S&P500 is higher overall, and the objective is to minimize variance, the allocations by the GGARCH family seem more plausible.

Finally, we calculate Value-at-Risk (VaR) of three portfolios at three different probability levels.

	σ_r	ρ	w_1	w_2
GGARCH SCALAR***	0.0109	0.577	0.5090	0.4910
GGARCH DIAG	0.0110	0.576	0.5076	0.4924
GGARCH LINEAR	0.0110	0.543	0.5046	0.4954
GGARCH FULL	0.0110	0.558	0.5168	0.4832
GGARCH PCA DIAG	0.0110	0.612	0.5055	0.4945
GGARCH PCA FULL	0.0110	0.578	0.4929	0.5071
BEKK SCALAR	0.0110	0.560	0.5215	0.4785
BEKK DIAGONAL	0.0110	0.546	0.5341	0.4659
DCC SCALAR	0.0110	0.520	0.5315	0.4685

Table 6. Minimum variance portfolio. σ_r is the sample standard deviation of the portfolio return, ρ is the average correlation estimate between the two indices, and w_1 and w_2 are average portfolio weights of S&P500 and FTSE100, respectively.

Three portfolios are constructed by mixing the two market indexes at 20:80, 50:50, and 80:20 ratios, and VaR is computed at 95%, 99%, and 99.9% probability levels. Assuming normality of the returns, VaR at time t is given by

$$VaR_{\alpha,t} = z_\alpha \sqrt{\boldsymbol{w}^\top \boldsymbol{H}_t \boldsymbol{w}}, \tag{28}$$

where α is the probability level, z_α is the z-value at α, and $\boldsymbol{w}$ is the portfolio weights. VaR is computed everyday and the probability of loss exceeding VaR is obtained by comparing the VaR with the actual loss of the portfolio over the sample period. The results are reported in Table 7. The values in the table are the probability of loss exceeding VaR. If VaR is correctly estimated, the probabilityt will converge to the probability level α. The models are ranked based on the overall performance across portfolios and probability levels. It turns out that all the models underestimate VaR, except the GGARCH PCA DIAG for 50:50 portfolio at 95% level. Nevertheless, the GGARCH family appear to estimate VaR more accurately, and PCA-based GGARCH models, in particular, perform best.

We have evaluated the proposed models using various diagnostic tools. The results are mixed and do not consistently support any particular model. This could be because the model specifications considered here are not correct, or more fundamentally, our geometric framework has limitations in modelling covariance dynamics. It will remain inconclusive until more exhaustive research is undertaken. Nevertheless, our models demonstrate clear benefits over the existing models: While the BEKK and DCC models perform better in the in-sample tests, GGARCH models are better performers in the out-of-sample tests, *i.e.,* asset allocation and risk measurement, which are more important in practical applications. Remarkably, this is achieved with fewer model parameters.

5. Conclusion

In this paper, we have proposed a new framework for addressing covariance dynamics from a geometric perspective using differential geometry tools developed by Fletcher et al (2003), Fletcher et al (2003), and Moakher (2005) among others. Our framework preserves the geometric structure of the covariance matrix without any arbitrary restrictions by respecting the inherent geometric features of the covariance matrix. It also possesses the desired nonlinear nature of the covariance dynamics observed in the market. Based on our geometric framework, we derived two types of covariance dynamics models, a general geometric GARCH (GGARCH) model, and a PCA-based geometric GARCH (GGARCH PCA) model. The major benefits of our model is 1) it allows us to

	20:80			50:50			80:20		
	0.05	0.01	0.001	0.05	0.01	0.001	0.05	0.01	0.001
GGARCH SCALAR*	0.0518	0.0180	0.0042	0.0541	0.0161	0.0050	0.0525	0.0188	0.0046
GGARCH DIAG	0.0525	0.0165	0.0042	0.0541	0.0165	0.0050	0.0521	0.0196	0.0050
GGARCH LINEAR	0.0548	0.0184	0.0046	0.0567	0.0180	0.0058	0.0579	0.0215	0.0050
GGARCH FULL	0.0529	0.0180	0.0042	0.0533	0.0176	0.0050	0.0560	0.0188	0.0042
GGARCH PCA DIAG***	0.0514	0.0153	0.0035	0.0479	0.0138	0.0038	0.0502	0.0150	0.0023
GGARCH PCA FULL**	0.0548	0.0192	0.0046	0.0521	0.0207	0.0035	0.0506	0.0188	0.0035
BEKK SCALAR	0.0537	0.0176	0.0031	0.0556	0.0192	0.0038	0.0541	0.0199	0.0050
BEKK DIAGONAL	0.0571	0.0207	0.0042	0.0579	0.0215	0.0046	0.0560	0.0219	0.0050
DCC SCALAR	0.0602	0.0211	0.0042	0.0590	0.0226	0.0050	0.0575	0.0226	0.0054

Table 7. Value–at–Risk test. The first row indicates portfolio composition between S&P500 and FTSE100, and the second row indicates probability level. The figures are probability of loss exceeding VaR.

17

specify the dynamics in an intuitive manner; 2) the nonlinear nature of the covariance dynamics is naturally implied; 3) a parsimonious specification can be easily achieved. The last feature is particularly important as it makes the model suitable for large dimensional systems.

Simulation studies reveal potential benefits of our model by showing that it captures many well-known properties of the covariance dynamics. Empirical studies, however, show mixed results and do not fully support our model against the widely recognized BEKK and DCC models. Nevertheless, it is encouraging that our model outperforms these models when forecasting matters, as evidenced in the asset allocation and VaR estimation tests. Different specifications of the dynamics within our framework could improve the overall performance.

We have primarily focused on introducing a new framework and the underlying mathematical concepts from a theoretical perspective, and demonstrated its potential benefits through case studies. More comprehensive empirical studies and comparison analysis are in order: numerous areas of application can be further considered, *e.g.,* credit risk modelling, asset pricing, and portfolio optimization among others. Particularly intriguing is the possibility of a conditional CAPM model within the context of our framework. Econometric theories in the covariance space, *e.g.,* conditions for stationarity, have yet to be established. Extensions of the fundamental models proposed in this paper can also be considered: different specifications of the tangent vector is of particular interest. Finally, a multivariate stochastic volatility model can be specified in the same framework by adding a diffusion term to the continuous-time dynamics. We expect that the geometric framework proposed here can serve as a building block for developing various covariance dynamics models, and more broadly contribute to building a new paradigm for economic modelling.

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

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

De Santis G, Gerard B (1998) How big is the premium for currency risk? Journal of Financial Economics 49:375–412

Engle R (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. Econometrica 50:987–1006

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, vol LNCS 3117, 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

Hafner C, Herwartz H (1998) Time-varying market price of risk in the capm - approaches, empirical evidence and implications. Finance 19:93–112

Karolyi G (1995) A multivariate GARCH model of international transmission of stock returns and volatility:the case of the united states and canada. Journal of Business and Economic Statistics 13:11–25

Kawakatsu H (2006) Matrix exponential GARCH. Journal of Econometrics 134:95–128

Kearney C, Patton A (2000) Multivariate GARCH modelling of exchange rate volatility transmission in the european monetary system. Financial Review 41:29–48

Kroner F, Ng V (1998) Modelling asymmetric comovements of asset returns. The Review of Financial Studies 11:817–844

Ledoit O, Santa-Clara P, Wolf M (2003) Flexible multivariate GARCH modeling with an application to international stock markets. Review of Economics and Statistics 85:735–747

Lenglet C, Rousson M, Deriche R, Faugeras (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

Lien D, Tse Y (2002) Some recent developments in futures hedging. Journal of Economic Surveys 16:357–396

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

Silvennoinen A, Teräsvirta T (2009) Handbook of Financial Time Series, Springer, chap Multivariate GARCH models, pp 201–229

Tse Y, Tsui A (2002) A multivariate GARCH model with time-varying correlations. Journal of Business and Economic Statistics 20:351–362

Wang K, Chen, M (2007) The dynamics in the spot, futures, and call options with basis asymmetries: an intraday analysis in a generalized multivariate GARCH-M MSKST framework. Review of Quantitative Finance and Accounting 29(4):371–394

Appendix A. Principal Geodesic Analysis

Principal component analysis (PCA) finds a linear subspace in which the variability of the data is best described. Similarly, principal geodesic analysis (PGA) 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 first generalize three important concepts in PCA: variance, subspace, and projection, and develop an algorithm for PGA. For completeness, we briefly summarize the process; a more concrete derivation can be found in one of the references cited above.

First, the variance of a set of data on a Riemannian manifold is defined as the expected value of the squared Riemannian distance from the mean:

$$\sigma^2 = E\left[d(\bar{\boldsymbol{P}}, \boldsymbol{P})^2\right]. \tag{A.1}$$

19

Given a sample of N data, the sample variance is defined as

$$s^2 = \frac{1}{N} \sum_{i=1}^{N} d(\bar{\boldsymbol{P}}, \boldsymbol{P}_i)^2. \tag{A.2}$$

Secondly, the manifold counterpart of the linear subspace in the Euclidean space is defined. A submanifold H of a manifold M is said to be *geodesic at* $\boldsymbol{P} \in H$, if all geodesics of H passing through $\boldsymbol{P}$ are also geodesics of M. Since a submanifold geodesic at $\boldsymbol{P}$ preserves the distance from $\boldsymbol{P}$, submanifolds geodesic at the mean can be regarded as the generalization of the linear subspaces.

We now define the projection operator to a manifold. Fletcher and Joshi (2004) define the projection operator, $\pi_H : M \to H$, as

$$\underset{\boldsymbol{Q} \in H}{\operatorname{argmin}} \, d(\boldsymbol{P}, \boldsymbol{Q})^2. \tag{A.3}$$

Projection onto a geodesic submanifold at $\bar{\boldsymbol{P}}$ can be approximated in the tangent space to the mean, $T_{\bar{\boldsymbol{P}}}M$. If $\boldsymbol{V}_1, \ldots, \boldsymbol{V}_K$ form an orthonormal basis for $T_{\bar{\boldsymbol{P}}}H$, then the projection operator can be approximated by the formula

$$\operatorname{Log}_{\bar{\boldsymbol{P}}}(\pi_H(\boldsymbol{P})) \approx \sum_{k=1}^{K} \langle \boldsymbol{V}_k, \operatorname{Log}_{\bar{\boldsymbol{P}}}(\boldsymbol{P}) \rangle_{\bar{\boldsymbol{P}}} \, \boldsymbol{V}_k. \tag{A.4}$$

With the concepts defined above, it can now be shown that PGA can be performed by applying PCA to the tangent space of the manifold, $T_\mu M$. The algorithm for PGA on $P(n)$ is summarized below.

- Given $\boldsymbol{P}_1, \ldots, \boldsymbol{P}_N \in P(n)$,

- Calculate the intrinsic mean of $\{\boldsymbol{P}_i\}$ and denote it $\bar{\boldsymbol{P}}$.

- Compute the tangent vectors at $\bar{\boldsymbol{P}}$ of the geodesics connecting $\bar{\boldsymbol{P}}$ and $\boldsymbol{P}_i$:

$$\boldsymbol{X}_i = \operatorname{Log}_{\bar{\boldsymbol{P}}}(\boldsymbol{P}_i).$$

- Apply PCA to $\{\boldsymbol{X}_i\}$, *i.e.*, find eigenvectors and eigenvalues, $\{\boldsymbol{v}_k, \lambda_k\}$ of

$$\boldsymbol{S} = \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{x}_i \boldsymbol{x}_i^\top$$

where $\boldsymbol{x}_i = vech(\boldsymbol{X}_i)$. $\boldsymbol{V}_k = vech^{-1}(\boldsymbol{v}_k) \in S(n)$ are the principal directions and λ_k are the corresponding variances.

Each observation $\boldsymbol{P}_i$ can be reproduced by the formula

$$\boldsymbol{P}_i = \operatorname{Exp}_{\bar{\boldsymbol{P}}} \left(\sum_{k=1}^{n(n+1)/2} \alpha_k \boldsymbol{V}_k \right),$$

where $\alpha_k = \boldsymbol{v}_k^\top \boldsymbol{x}_i$.