# MULTIVARIATE GARCH MODELS

**Eduardo Rossi**

University of Pavia
ITALY

September 2012

# Outline

Economics and financial economics present problems whose solutions need the specification and estimation of a multivariate distribution.

- the standard portfolio allocation problem
- the risk management of a portfolio of assets
- pricing of derivative contracts based on a more than one underlying asset (e.g., Quanto options)
- Financial contagion (shocks transmission volatility and returns)

Stylized facts:

- Volatility clustering
- Time-varying dynamic covariances and dynamic correlations

Financial variables have time-dependent second order moments.
Parametric models:

1. Multivariate GARCH models
2. Multivariate Stochastic volatility models
3. Multifactor models
4. Multifactor realized volatility models

Vector of returns:

$$\mathbf{y}_t = (y_{1t}, \ldots, y_{Nt})' \quad (N \times 1)$$

$$\mathbf{y}_t - \boldsymbol{\mu}_t = \boldsymbol{\epsilon}_t = \mathbf{H}_t^{-1/2}\mathbf{z}_t$$

Let $\{\mathbf{z}_t\}$ be a sequence of $(N \times 1)$ i.i.d. random vector with the following characteristics:

$$E[\mathbf{z}_t] = \mathbf{0}$$

$$E[\mathbf{z}_t\mathbf{z}_t'] = \mathbf{I}_N$$

$$\mathbf{z}_t \sim G(\mathbf{0}, \mathbf{I}_N)$$

with $G$ continuous density function.

$$E_{t-1}(\boldsymbol{\epsilon}_t) = 0$$

$$\begin{aligned} E_{t-1}(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t') &= \mathbf{H}_t \\ E(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t') &= \boldsymbol{\Sigma} \end{aligned}$$

$$E_{t-1}[\cdot] = E[\cdot|\Phi_{t-1}]$$

$\Phi_{t-1}$ is the $\sigma$-field generated by past values of observable variables. where $\mathbf{H}_t$ is a matrix ($N \times N$) positive definite and measurable with respect to the information set $\Phi_{t-1}$, that is the $\sigma$-field generated by the past observations: $\{\boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, \ldots\}$.
The correlation matrix:

$$\mathrm{Corr}_{t-1}(\boldsymbol{\epsilon}_t) = \mathbf{R}_t = \mathbf{D}_t^{-1/2}\mathbf{H}_t\mathbf{D}_t^{-1/2}$$

$$\mathbf{D}_t = \mathrm{diag}(h_{11,t}, \ldots, h_{NN,t})$$

MVMs provide a parametric structure for the dynamic evolution of $\mathbf{H}_t$. MVMs must satisfy:

1. Diagonal elements of $\mathbf{H}_t$ must be strictly positive;
2. Positive definiteness of $\mathbf{H}_t$;
3. Stationarity: $E[\mathbf{H}_t]$ exists, finite and constant w.r.t. $t$.

Ideal characteristics of a MVM:

1. Estimation should be flexible for increasing $N$
2. It should allow for covariance spillovers and feedbacks;
3. Coefficients should have an economic or financial interpretation

Three approaches for constructing multivariate GARCH models:

1. direct generalizations of the univariate GARCH model of Bollerslev (1986); (VEC, BEKK and factor models)
2. linear combinations of univariate GARCH models; ((generalized) orthogonal models and latent factor models.)
3. nonlinear combinations of univariate GARCH models; (constant and dynamic conditional correlation models, copula-GARCH models)

Caporin and McAleer (2009). Covariance targeting if the conditions are met:

- The model intercept is an explicit function of the model long-run covariance (or correlation) The long-run covariance (or correlation) solution is given by the $E[\mathbf{H}_t]$ or $E[\mathbf{R}_t]$.
- The long-run solution is replaced by a consistent estimator.

GARCH(1,1):

$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2$$

Long-run variance (if $(\alpha + \beta) < 1$:

$$\sigma^2 = E[\sigma_t^2] = \omega(1 - \alpha - \beta)^{-1}$$

*Variance targeting*:

$$\sigma_t^2 = \widehat{\sigma}^2(1 - \alpha - \beta) + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2$$

$$\widehat{\sigma}^2 = T^{-1}\sum_t \widehat{\epsilon}_t^2$$

Introduction of targeting transforms the model estimation into a two-step estimation approach:

1. $\widehat{\sigma}^2$
2. $\alpha, \beta$

$N$ assets: $N$ variances $+\ \frac{1}{2}N(N-1)$ covariances $= \frac{N}{2}(N+1)$.
Two alternative approaches:

- Models of $\mathbf{H}_t$
- Models of $\mathbf{D}_t$ and $\mathbf{R}_t$

The parametrization of $\mathbf{H}_t$ as a multivariate GARCH, which means as a function of the information set $\Phi_{t-1}$, allows each element of $\mathbf{H}_t$ to depend on $q$ lagged of the squares and cross-products of $\epsilon_t$, as well as $p$ lagged values of the elements of $\mathbf{H}_t$. So the elements of the covariance matrix follow a vector of ARMA process in squares and cross-products of the disturbances.

Let **vech** denote the vector-half operator, which stacks the lower triangular elements of an $N \times N$ matrix as an $[N(N+1)/2] \times 1$ vector.
Let $\mathbf{A}$ be $(2 \times 2)$, then vech($\mathbf{A}$)

$$\text{vech}(\mathbf{A}) = \begin{bmatrix} a_{11} \\ a_{21} \\ a_{22} \end{bmatrix}$$

Since the conditional covariance matrix $\mathbf{H}_t$ is symmetric,
$\text{vech}(\mathbf{H}_t), (N(N+1)/2 \times 1)$ contains all the unique elements in $\mathbf{H}_t$.

A natural **multivariate extension** of the univariate GARCH(p,q) model is

$$\begin{aligned} \mathrm{vech}\left(\mathbf{H}_t\right) &= \mathbf{W} + \sum_{i=1}^{q} \mathbf{A}_i^* \mathrm{vech}\left(\boldsymbol{\epsilon}_{t-i}\boldsymbol{\epsilon}_{t-i}'\right) + \sum_{j=1}^{p} \mathbf{B}_j^* \mathrm{vech}\left(\mathbf{H}_{t-j}\right) \\ &= \mathbf{W} + \mathbf{A}^*\left(L\right) \mathrm{vech}\left(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t'\right) + \mathbf{B}^*\left(L\right) \mathrm{vech}\left(\mathbf{H}_t\right) \end{aligned}$$

$$\begin{aligned} \mathbf{A}^*\left(L\right) &= \mathbf{A}_1^* L + \ldots + \mathbf{A}_q^* L^q \\ \mathbf{B}^*\left(L\right) &= \mathbf{B}_1^* L + \ldots + \mathbf{B}_q^* L^p \end{aligned}$$

$$N^* \equiv \frac{N(N+1)}{2}$$

$$\begin{array}{rcl} \mathbf{W} & : & [N(N+1)/2] \times 1 \\ \mathbf{A}_i^*,\ \mathbf{B}_j^* & : & [N^* \times N^*] \end{array}$$

$N = 2$, Vech-GARCH(1,1):

$$\begin{bmatrix} h_{11,t} \\ h_{21,t} \\ h_{22,t} \end{bmatrix} = \begin{bmatrix} w_1^* \\ w_2^* \\ w_3^* \end{bmatrix} + \begin{bmatrix} a_{11}^* & a_{12}^* & a_{13}^* \\ a_{21}^* & a_{22}^* & a_{23}^* \\ a_{31}^* & a_{32}^* & a_{33}^* \end{bmatrix} \begin{bmatrix} \epsilon_{1,t-1}^2 \\ \epsilon_{1,t-1}\epsilon_{2,t-1} \\ \epsilon_{2,t-1}^2 \end{bmatrix} + \begin{bmatrix} b_{11}^* & b_{12}^* & b_{13}^* \\ b_{21}^* & b_{22}^* & b_{23}^* \\ b_{31}^* & b_{32}^* & b_{33}^* \end{bmatrix} \begin{bmatrix} h_{11,t-1} \\ h_{21,t-1} \\ h_{2,t-1} \end{bmatrix}$$

- This general formulation is termed *vec representation* by Engle and Kroner (1995).
- The number of parameters is $\left[1+(p+q)\left[N\left(N+1\right)/2\right]^2\right]$.
- Even for low dimensions of $N$ and small values of $p$ and $q$ the number of parameters is very large; for $N=5$ and $p=q=1$ the unrestricted version of (1) contains 465 parameters.
- The number of parameters is of order $O(N^4)$: the *curse of dimensionality*.

For any parametrization to be sensible, we require that $\mathbf{H}_t$ be positive definite for all values of $\epsilon_t$ in the sample space in the *vech* representation this restriction can be difficult to check, let alone impose during estimation.

A natural restriction is the *diagonal representation*, in which each element of the covariance matrix depends only on past values of itself and past values of $\epsilon_{jt}\varepsilon_{kt}$.
In the diagonal model the $\mathbf{A}_i^*$ and $\mathbf{B}_j^*$ matrices are all taken to be diagonal.
For $N = 2$ and $p = q = 1$, the diagonal model is written as:

$$\begin{bmatrix} h_{11,t} \\ h_{21,t} \\ h_{22,t} \end{bmatrix} = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} + \begin{bmatrix} a_{11}^* & 0 & 0 \\ 0 & a_{22}^* & 0 \\ 0 & 0 & a_{33}^* \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t-1}^2 \\ \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon_{2,t-1}^2 \end{bmatrix} + \begin{bmatrix} b_{11}^* & 0 & 0 \\ 0 & b_{22}^* & 0 \\ 0 & 0 & b_{33}^* \end{bmatrix} \begin{bmatrix} h_{11,t-1} \\ h_{21,t-1} \\ h_{22,t-1} \end{bmatrix}$$

$$h_{ij,t} = w_i^* + a_{ii}^* \epsilon_{i,t-1}\epsilon_{j,t-1} + b_{ii}^* h_{ij,t-1}$$

Thus the $(i,j)$ *th* element in $\mathbf{H}_t$ depends on the corresponding $(i,j)$ *th* element in $\varepsilon_{t-1}\varepsilon'_{t-1}$ and $\mathbf{H}_{t-1}$. This restriction reduces the number of parameters to $[N(N+1)/2](1+p+q)$. This model does not allow for causality in variance, co-persistence in variance and asymmetries.
The number of parameters is of order $O(N^2)$.
The *diagonal vech* is equivalent to:

$$\mathbf{H}_t = \mathbf{W} + \mathbf{A} \odot \epsilon_{t-1}\epsilon'_{t-1} + \mathbf{B} \odot \mathbf{H}_{t-1}$$

where $\mathbf{A}$ and $\mathbf{B}$ are symmetric matrices. $\odot$ is the Hadamard product.

Given that

$$\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t' = \mathbf{H}_t + \mathbf{V}_t$$

with $E_{t-1}(\mathbf{V}_t) = 0$.

$$\text{vech}(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t') = \text{vech}(\mathbf{H}_t) + \text{vech}(\mathbf{V}_t)$$

$$\text{vech}(\mathbf{V}_t) \quad \text{vector m.d.s.}$$

with $E(\text{vech}(\mathbf{V}_t)) = \text{vech}(E(\mathbf{V}_t)) = 0$.

For a GARCH(1,1), the unconditional covariance matrix, when it exists, is given by

$$
\begin{aligned}
\mathrm{vech}\left(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t'\right) &= \mathbf{W} + \mathbf{A}_1^*\mathrm{vech}\left(\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1}'\right) \\
&\quad + \mathbf{B}_1^*\left[\mathrm{vech}\left(\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1}'\right) - \mathrm{vech}\left(V_{t-1}\right)\right] + \mathrm{vech}\left(\mathbf{V}_t\right)
\end{aligned}
$$

$$E\left(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t'\right) = \boldsymbol{\Sigma}$$

$$\mathrm{vech}\left(E\left(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t'\right)\right) = \mathbf{W} + \left(\mathbf{A}_1^* + \mathbf{B}_1^*\right)\left[\mathrm{vech}\left(E\left(\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1}'\right)\right)\right]$$

$$\mathrm{vech}\left(\boldsymbol{\Sigma}\right) = \left[\mathbf{I}_{N^*} - \mathbf{A}_1^* - \mathbf{B}_1^*\right]^{-1}\mathbf{W}.$$

For a GARCH(p,q) model

$$\mathrm{vech}\left(\boldsymbol{\Sigma}\right) = \left[\mathbf{I}_{N^*} - \mathbf{A}^*\left(1\right) - \mathbf{B}^*\left(1\right)\right]^{-1}\mathbf{W}$$

Targeting cannot easily introduced in the model.
MGARCH(1,1)-vech, unconditional var-cov matrix

$$[\mathbf{I}_{N^*} - \mathbf{A}_1^* - \mathbf{B}_1^*]\text{vech}(\boldsymbol{\Sigma})$$

$$\widehat{\boldsymbol{\Sigma}} = T^{-1} \sum_t \widehat{\boldsymbol{\epsilon}}_t \widehat{\boldsymbol{\epsilon}}_t'$$

Targeting allows reducing by $N^*$ the parameters to be estimated. The total number of parameters is still $O(N^4)$.

$$\begin{aligned} \text{vech}\left(\mathbf{H}_t\right) &= \text{vech}\left(\widehat{\boldsymbol{\Sigma}}\right) + \mathbf{A}_1^* \text{vech}\left[\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1}' - \text{vech}\left(\widehat{\boldsymbol{\Sigma}}\right)\right] \\ &\quad + \mathbf{B}_1^* \left[\text{vech}\left(\mathbf{H}_{t-1}\right) - \text{vech}\left(\widehat{\boldsymbol{\Sigma}}\right)\right] \end{aligned}$$

Engle and Kroner (1995) propose a parametrization that imposes positive definiteness restrictions.

Consider the following model

$$\mathbf{H}_t = \mathbf{C}\mathbf{C}' + \sum_{k=1}^{K}\sum_{i=1}^{q} \mathbf{A}_{ik}\epsilon_{t-i}\epsilon_{t-i}'\mathbf{A}_{ik}' + \sum_{k=1}^{K}\sum_{i=1}^{p} \mathbf{B}_{ik}\mathbf{H}_{t-i}\mathbf{B}_{ik}' \tag{1}$$

where $\mathbf{C}$, $\mathbf{A}_{ik}$ and $\mathbf{B}_{ik}$ are $(N \times N)$.

- The intercept matrix is decomposed into $\mathbf{C}\mathbf{C}'$, where $\mathbf{C}$ is a lower triangular matrix.
- Without any further assumption $\mathbf{C}\mathbf{C}'$ is positive semidefinite.
- This representation is general, it includes all positive definite diagonal representations and nearly all positive definite *vech* representations.

For exposition simplicity we will assume that $K = 1$:

$$\mathbf{H}_t = \mathbf{C}\mathbf{C}' + \sum_{i=1}^{q} \mathbf{A}_i \epsilon_{t-i} \epsilon'_{t-i} \mathbf{A}'_i + \sum_{i=1}^{p} \mathbf{B}_i \mathbf{H}_{t-i} \mathbf{B}'_i$$

Consider the simple GARCH(1,1) model:

$$\mathbf{H}_t = \mathbf{C}\mathbf{C}' + \mathbf{A}_1 \epsilon_{t-1} \epsilon'_{t-1} \mathbf{A}'_1 + \mathbf{B}_1 \mathbf{H}_{t-1} \mathbf{B}'_1 \tag{2}$$

## *BEKK* (Engle and Kroner (1995))

Suppose that the diagonal elements in $\mathbf{C}$ are restricted to be positive and that $a_{11}$ and $b_{11}$ are also restricted to be positive. Then if $K = 1$ there exists no other $\mathbf{C}$, $\mathbf{A}_1$, $\mathbf{B}_1$ in the model (2) that will give an equivalent representation.

The purpose of the restrictions is to eliminate all other observationally equivalent structures.

For example, as relates to the term $\mathbf{A}_1\epsilon_{t-1}\epsilon'_{t-1}\mathbf{A}'_1$ the only other observationally equivalent structure is obtained by replacing $\mathbf{A}_1$ by $-\mathbf{A}_1$. The restriction that $a_{11}$ ($b_{11}$) be positive could be replaced with the condition that $a_{ij}$ ($b_{ij}$) be positive for a given $i$ and $j$, as this condition is also sufficient to eliminate $-\mathbf{A}_1$ from the set of admissible structures.

MGARCH(1,1)-BEKK, $N = 2$:

$$\begin{aligned}
\mathbf{H}_t &= \mathbf{CC}' + \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} \varepsilon_{1t-1}^2 & \varepsilon_{1t-1}\varepsilon_{2t-1} \\ \varepsilon_{2t-1}\varepsilon_{1t-1} & \varepsilon_{2t-1}^2 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}' \\
&\quad + \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} h_{11t-1} & h_{12t-1} \\ h_{21t-1} & h_{22t-1} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}'
\end{aligned}$$

BEKK-GARCH(p,q) model (Engle and Kroner (1995)):

## Sufficient condition for positive definiteness of $\mathbf{H}_t$

If $\mathbf{H}_0$, $\mathbf{H}_{-1}, \ldots, \mathbf{H}_{-p+1}$ are all positive definite, then the BEKK parametrization (with $K = 1$) yields a positive definite $\mathbf{H}_t$ for all possible values of $\varepsilon_t$ if $\mathbf{C}$ is a full rank matrix or if any $\mathbf{B}_i$ $i = 1, \ldots, p$ is a full rank matrix.

For simplicity consider the GARCH(1,1) model. The BEKK parametrization is

$$\mathbf{H}_t = \mathbf{C}\mathbf{C}' + +\mathbf{A}_1\epsilon_{t-1}\epsilon_{t-1}'\mathbf{A}_1' + \mathbf{B}_1\mathbf{H}_{t-1}\mathbf{B}_1'$$

The proof proceeds by induction.
First $\mathbf{H}_t$ is p.d. for $t = 1$: The term $\mathbf{A}_1\epsilon_0\epsilon_0'\mathbf{A}_1'$ is positive semidefinite because $\epsilon_0\epsilon_0'$ is positive semidefinite. Also if the null spaces of the matrices of $\mathbf{C}$ and $\mathbf{B}_1$ intersect only at the origin, that is at least one of two is full rank then

$$\mathbf{C}\mathbf{C}' + \mathbf{B}_1\mathbf{H}_0\mathbf{B}_1'$$

is positive definite. This is true if $\mathbf{C}$ or $\mathbf{B}_1$ has full rank.

To show that the null space condition is sufficient $\mathbf{CC}' + \mathbf{B}_1\mathbf{H}_0\mathbf{B}_1'$ is p.d. if and only if

$$x'\left(\mathbf{CC}' + \mathbf{B}_1\mathbf{H}_0\mathbf{B}_1'\right)x > 0 \qquad \forall x \neq 0$$

or

$$\left(\mathbf{C}'x\right)'\left(\mathbf{C}'x\right) + \left(\mathbf{H}_0^{1/2}\mathbf{B}_1'x\right)'\left(\mathbf{H}_0^{1/2}\mathbf{B}_1'x\right) > 0 \qquad \forall x \neq 0$$

where $\mathbf{H}_0 = \mathbf{H}_0^{1/2\prime}\mathbf{H}_0^{1/2}$ and $\mathbf{H}_0^{1/2}$ is full rank.

Defining $N(P)$ to be the null space of the matrix $P$, (28) is true if and only if

$$N(\mathbf{C}') \cap N\left(\mathbf{H}_0^{1/2}\mathbf{B}_1'\right) = \emptyset.$$

$N\left(\mathbf{H}_0^{1/2}\mathbf{B}_1'\right) = N(\mathbf{B}_1')$ because $\mathbf{H}_0^{1/2}$ is full rank. This implies that $\mathbf{CC}' + \mathbf{B}_1\mathbf{H}_0\mathbf{B}_1'$ is positive definite if and only if $N(\mathbf{C}') \cap N\left(\mathbf{H}_0^{1/2}\mathbf{B}_1'\right) = \emptyset$. Now suppose that $\mathbf{H}_t$ is positive definite for $t = \tau$.
Then,

$$\mathbf{H}_{\tau+1} = \mathbf{CC}' + \mathbf{A}_1\epsilon_\tau\epsilon_\tau'\mathbf{A}_1' + \mathbf{B}_1\mathbf{H}_\tau\mathbf{B}_1'$$

is positive definite if and only if, given that $\mathbf{A}_1\epsilon_\tau\epsilon_\tau'\mathbf{A}_1'$ is positive semidefinite, the null space condition holds, because $\mathbf{H}_\tau$ is positive definite by the induction assumption.

Consider MGARCH(1,1)-BEKK, $N = 2$ with

$$\mathbf{A}_1 = \mathrm{diag}(a_{11}, a_{22}) \quad \mathbf{B}_1 = \mathrm{diag}(b_{11}, b_{22})$$

the model reduces to

$$\begin{aligned}
\mathbf{H}_t &= \mathbf{CC}' + \begin{bmatrix} a_{11} & 0 \\ 0 & a_{22} \end{bmatrix} \begin{bmatrix} \varepsilon^2_{1t-1} & \varepsilon_{1t-1}\varepsilon_{2t-1} \\ \varepsilon_{2t-1}\varepsilon_{1t-1} & \varepsilon^2_{2t-1} \end{bmatrix} \begin{bmatrix} a_{11} & 0 \\ 0 & a_{22} \end{bmatrix}' \\
&+ \begin{bmatrix} b_{11} & 0 \\ 0 & b_{22} \end{bmatrix} \begin{bmatrix} h_{11t-1} & h_{12t-1} \\ h_{21t-1} & h_{22t-1} \end{bmatrix} \begin{bmatrix} b_{11} & 0 \\ 0 & b_{22} \end{bmatrix}'
\end{aligned}$$

$$\begin{aligned}
h_{11,t} &= c^2_{11} + a^2_{11}\epsilon^2_{1t-1} + b^2_{11}h_{11t-1} \\
h_{12,t} &= c_{21}c_{11} + a_{11}a_{22}\epsilon_{1t-1}\epsilon_{2t-1} + b_{11}b_{22}h_{12t-1} \\
h_{22,t} &= c_{21}c_{11} + c^2_{22} + a^2_{22}\epsilon^2_{1t-1} + b^2_{22}h_{11t-1}
\end{aligned}$$

This model is equivalent to the Hadamard BEKK:

$$\mathbf{H}_t = \mathbf{CC}' + \mathbf{aa}' \odot \boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}'_{t-1} + \mathbf{bb}' \odot \mathbf{H}_{t-1}$$

positive definiteness is not guaranteed. Positive semidefiniteness is obtained by imposing p.s.d. of all terms.

MGARCH(1,1) - Scalar BEKK

$$\mathbf{A}_1 = \alpha \mathbf{I}_N, \quad \mathbf{B}_1 = \beta \mathbf{I}_N$$

$$\mathbf{H}_t = \mathbf{C}\mathbf{C}' + \alpha^2(\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}'_{t-1}) + \beta^2 \mathbf{H}_{t-1}$$

1. BEKK
$$\mathbf{H}_t = \mathbf{\Sigma} + \mathbf{A}_1 \left(\epsilon_{t-1}\epsilon'_{t-1} - \mathbf{\Sigma}\right) \mathbf{A}'_1 + \mathbf{B}_1 \left(\mathbf{H}_{t-1} - \mathbf{\Sigma}\right) \mathbf{B}_1$$
or
$$\mathbf{H}_t = \left(\mathbf{\Sigma} - \mathbf{A}_1\mathbf{\Sigma}\mathbf{A}_1 - \mathbf{B}_1\mathbf{\Sigma}\mathbf{B}'_1\right) + \mathbf{A}_1 \left(\epsilon_{t-1}\epsilon'_{t-1}\right) \mathbf{A}'_1 + \mathbf{B}_1\mathbf{H}_{t-1}\mathbf{B}'_1$$
To have p.d-ness of $\mathbf{H}_t$, $(\mathbf{\Sigma} - \mathbf{A}_1\mathbf{\Sigma}\mathbf{A}_1 - \mathbf{B}_1\mathbf{\Sigma}\mathbf{B}'_1)$ must be p.d..
2. Hadamard BEKK
$$\mathbf{H}_t = \mathbf{\Sigma} + \mathbf{A}_1 \odot \left(\epsilon_{t-1}\epsilon'_{t-1} - \mathbf{\Sigma}\right) + \mathbf{B}_1 \odot \left(\mathbf{H}_{t-1} - \mathbf{\Sigma}\right)$$
$\left(\epsilon_{t-1}\epsilon'_{t-1} - \mathbf{\Sigma}\right)$ must be p.s.d., while $(\mathbf{H}_{t-1} - \mathbf{\Sigma})$ must be p.s.d.
3. Scalar BEKK
$$\mathbf{H}_t = \mathbf{\Sigma} + \alpha^2 \left(\epsilon_{t-1}\epsilon'_{t-1} - \mathbf{\Sigma}\right) + \beta^2 \left(\mathbf{H}_{t-1} - \mathbf{\Sigma}\right)$$
for $\alpha + \beta < 1$.

We now examine the relationship between the BEKK and *vech* parameterizations. The mathematical relationship between the parameters of the two models can be found simply vectorizing the BEKK equation:

$$\mathrm{vec}(\mathbf{H}_t) = \mathrm{vec}(\mathbf{CC}') + \sum_{i=1}^{q} \mathrm{vec}(\mathbf{A}_i \epsilon_{t-i} \epsilon'_{t-i} \mathbf{A}'_i) + \sum_{i=1}^{p} vec(\mathbf{B}_i \mathbf{H}_{t-i} \mathbf{B}'_i)$$

where $\mathrm{vec}\,()$ is an operator such that given a matrix $\mathbf{A}$ $(n \times n)$, $\mathrm{vec}(\mathbf{A})$ is a $\left(n^2 \times 1\right)$ vector. The $\mathrm{vec}\,()$ satisfies

$$\mathrm{vec}\,(\mathbf{ABC}) = (\mathbf{C}' \otimes \mathbf{A})\,\mathrm{vec}\,(\mathbf{B})$$

For a symmetric $\mathbf{A}$, $(n \times n)$:

- $\text{vech}(\mathbf{A})$ contains precisely the $n(n+1)/2$ distinct elements of $\mathbf{A}$;
- the elements of $\text{vec}(\mathbf{A})$ are those of $\text{vech}(\mathbf{A})$ with some repetitions;
- There exists a unique $n^2 \times n(n+1)/2$ which transforms, for symmetric $\mathbf{A}$, $\text{vech}(\mathbf{A})$ into $\text{vec}(\mathbf{A})$. This matrix is called the *duplication matrix* and is denoted $\mathbf{D}_n$:

$$\text{vec}(\mathbf{A}) = \mathbf{D}_n \text{vech}(\mathbf{A})$$

  where $\mathbf{D}_n$ is the duplication matrix.

Then

$$\begin{aligned} \mathrm{vec}(\mathbf{H}_t) &= \mathrm{vec}(\mathbf{C}\mathbf{C}') + \sum_{i=1}^{q} (\mathbf{A}_i \otimes \mathbf{A}_i)\, \mathrm{vec}(\varepsilon_{t-i}\boldsymbol{\epsilon}'_{t-i}) \\ &\quad + \sum_{i=1}^{p} (\mathbf{B}_i \otimes \mathbf{B}_i)\, \mathrm{vec}(\mathbf{H}_{t-i}) \end{aligned}$$

$$\begin{aligned} \mathbf{D}_N \mathrm{vech}\,(\mathbf{H}_t) &= \mathbf{D}_N \mathrm{vech}(\mathbf{C}\mathbf{C}') + \sum_{i=1}^{q} (\mathbf{A}_i \otimes \mathbf{A}_i)\, \mathbf{D}_N \mathrm{vech}(\boldsymbol{\epsilon}_{t-i}\boldsymbol{\epsilon}'_{t-i}) \\ &\quad + \sum_{i=1}^{p} (\mathbf{B}_i \otimes \mathbf{B}_i)\, \mathbf{D}_N \mathrm{vech}(\mathbf{H}_{t-i}) \end{aligned}$$

If $\mathbf{D}_N$ is a full column rank matrix we can define the generalized inverse of $\mathbf{D}_N$ as:

$$\mathbf{D}_N^+ = \left(\mathbf{D}_N'\mathbf{D}_N\right)^{-1}\mathbf{D}_N'$$

that is a $\left(N\left(N+1\right)/2\right) \times \left(N^2\right)$ matrix, where

$$\mathbf{D}_N^+\mathbf{D}_N = \mathbf{I}_N$$

This implies that premultiplying by $\mathbf{D}_N^+$

$$\begin{aligned}\text{vech}\left(\mathbf{H}_t\right) &= \text{vech}(\mathbf{C}\mathbf{C}') + \mathbf{D}_N^+ \left(\sum_{i=1}^{q} (\mathbf{A}_i \otimes \mathbf{A}_i)\right) \mathbf{D}_N \text{vech}(\epsilon_{t-i}\epsilon_{t-i}') \\ &\quad + \mathbf{D}_N^+ \left(\sum_{i=1}^{p} (\mathbf{B}_i \otimes \mathbf{B}_i)\right) \mathbf{D}_N \text{vech}(\mathbf{H}_{t-i})\end{aligned}$$

- The *vech* model implied by any given BEKK model is unique, while the converse is not true.
- The transformation from a *vech* model to a BEKK model (when it exists) is not unique, because for a given $\mathbf{A}_1^*$ the choice of $\mathbf{A}_1$ is not unique.

- This can be seen recognizing that $(\mathbf{A}_i \otimes \mathbf{A}_i) = (-\mathbf{A}_i \otimes -\mathbf{A}_i)$ so while $\mathbf{A}_i^* = \mathbf{D}_N^+ (\mathbf{A}_i \otimes \mathbf{A}_i) \mathbf{D}_N$ is unique, the choice of $\mathbf{A}_i$ is not unique. It can also be shown that all positive definite diagonal *vech* models can be written in the BEKK framework.

Given $\mathbf{A}_i$ diagonal matrix, then $\mathbf{D}_N^+ (\mathbf{A}_i \otimes \mathbf{A}_i) \mathbf{D}_N$ is also diagonal, with diagonal elements given by $a_{ii} a_{jj}$ $(1 \leq j \leq i \leq N)$ (See Magnus).

Given the *vech* model

$$\text{vech}\left(\mathbf{H}_t\right) = \mathbf{W} + \mathbf{A}^*\left(L\right)\text{vech}\left(\varepsilon_t\varepsilon_t'\right) + \mathbf{B}^*\left(L\right)\text{vech}\left(\mathbf{H}_t\right)$$

the necessary and sufficient condition for covariance stationary of $\{\varepsilon_t\}$ is that all the eigenvalues of $\mathbf{A}^*\left(1\right) + \mathbf{B}^*\left(1\right)$ are less than one in modulus. But defining

$$\begin{aligned}
\mathbf{A}^*\left(1\right) &= \mathbf{D}_N^+\left(\sum_{i=1}^{q}\left(\mathbf{A}_i \otimes \mathbf{A}_i\right)\right)\mathbf{D}_N \\
\mathbf{B}^*\left(1\right) &= \mathbf{D}_N^+\left(\sum_{i=1}^{q}\left(\mathbf{B}_i \otimes \mathbf{B}_i\right)\right)\mathbf{D}_N
\end{aligned}$$

This implies also that in the BEKK model, $\{\epsilon_t\}$ is covariance stationary if and only if all the eigenvalues of

$$\mathbf{D}_N^+ \left( \sum_{i=1}^{q} (\mathbf{A}_i \otimes \mathbf{A}_i) \right) \mathbf{D}_N + \mathbf{D}_N^+ \left( \sum_{i=1}^{p} (\mathbf{B}_i \otimes \mathbf{B}_i) \right) \mathbf{D}_N$$

are less than one in modulus.

Let $\lambda_1, \ldots, \lambda_N$ be the eigenvalues of $\mathbf{A}_i$, the eigenvalues of

$$\mathbf{D}_N^+ \left( \sum_{i=1}^{q} (\mathbf{A}_i \otimes \mathbf{A}_i) \right) \mathbf{D}_N$$

are $\lambda_i \lambda_j \, (1 \leq j \leq i \leq N)$ (Magnus).

BEKK model

$$
\begin{aligned}
\text{vec}\left(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t'\right) &= \text{vec}\left(\mathbf{CC}'\right) + \left(\mathbf{A}_1 \otimes \mathbf{A}_1\right) \text{vec}\left(\boldsymbol{\epsilon}_{t-1} \boldsymbol{\epsilon}_{t-1}'\right) \\
&\quad + \left(\mathbf{B}_1 \otimes \mathbf{B}_1\right) \left[\text{vec}\left(\boldsymbol{\epsilon}_{t-1} \boldsymbol{\epsilon}_{t-1}'\right) - \text{vec}\left(V_{t-1}\right)\right] + \text{vec}\left(\mathbf{V}_t\right)
\end{aligned}
$$

$$
E\left[\text{vec}\left(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t'\right)\right] = \text{vec}\left(\mathbf{CC}'\right) + \left[\left(\mathbf{A}_1 \otimes \mathbf{A}_1\right) + \left(\mathbf{B}_1 \otimes \mathbf{B}_1\right)\right] E\left[\text{vec}\left(\boldsymbol{\epsilon}_{t-1} \boldsymbol{\epsilon}_{t-1}'\right)\right]
$$

$$
\text{vec}\left(\boldsymbol{\Sigma}\right) = \left[\mathbf{I}_{N^2} - \left(\mathbf{A}_1 \otimes \mathbf{A}_1\right) - \left(\mathbf{B}_1 \otimes \mathbf{B}_1\right)\right]^{-1} \text{vec}\left(\mathbf{CC}'\right)
$$

or in *vech* representation as

$$
\begin{aligned}
\mathbf{D}_N \text{vech}\left(E\left(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t'\right)\right) &= \mathbf{D}_N \text{vech}\left(\mathbf{CC}'\right) + \left(\mathbf{A}_1 \otimes \mathbf{A}_1\right) \mathbf{D}_N \text{vech}\left(E\left(\boldsymbol{\epsilon}_{t-1} \boldsymbol{\epsilon}_{t-1}'\right)\right) \\
&\quad + \left(\mathbf{B}_1 \otimes \mathbf{B}_1\right) \mathbf{D}_N \text{vech}\left(E\left(\boldsymbol{\epsilon}_{t-1} \boldsymbol{\epsilon}_{t-1}'\right)\right)
\end{aligned}
$$

$$
\text{vech}\left(\boldsymbol{\Sigma}\right) = \left[\mathbf{I}_{N^*} - \mathbf{D}_N^+ \left(\mathbf{A}_1 \otimes \mathbf{A}_1\right) \mathbf{D}_N - \mathbf{D}_N^+ \left(\mathbf{B}_1 \otimes \mathbf{B}_1\right) \mathbf{D}_N\right]^{-1} \text{vech}\left(\mathbf{CC}'\right)
$$

$N^* = N\left(N+1\right)/2.$

- The diagonal *vech* model is stationary if and only if the sum $a_{ii}^* + b_{ii}^* < 1$ for all $i$.
- In the diagonal BEKK model the covariance stationary condition is that $a_{ii}^2 + b_{ii}^2 < 1$.

Only in the case of diagonal models the stationarity properties are determined solely by the diagonal elements of the $\mathbf{A}_i$ and $\mathbf{B}_i$ matrices.

A general multivariate model can be written as:

$$\mathbf{y}_t = \mu + \mathbf{\Pi}(L)\,\mathbf{y}_{t-1} + \mathbf{\Psi}\mathbf{x}_{t-1} + \mathbf{\Lambda}\mathrm{vech}(\mathbf{H}_t) + \boldsymbol{\epsilon}_t \tag{3}$$

$$\mathbf{y}_t : (N \times 1)$$

$$\mathbf{\Pi}(L) = \mathbf{\Pi}_1 + \mathbf{\Pi}_2 L + \cdots + \mathbf{\Pi}_k L^{k-1} \quad (N \times N)$$

$$\mathbf{\Psi} : (N \times L)$$

$$\mathbf{\Lambda} : (N \times N(N+1)/2)$$

$$\mathbf{x}_t : (L \times 1)$$

$\mathbf{x}_{t-1}$ contains predetermined variables. $\epsilon_t$ is the vector of innovation with respect to the information set formed exclusively of past realizations of $\mathbf{y}_t$.

$$\mathbf{H}_t = E_{t-1}\left(\epsilon_t \epsilon_t'\right)$$

$$\mathbf{H}_t = \mathbf{C}\mathbf{C}' + \sum_{i=1}^{q} \mathbf{A}_i \left(\epsilon_{t-i} + \gamma\right)\left(\epsilon_{t-i} + \gamma\right)' \mathbf{A}_i' + \sum_{j=1}^{p} \mathbf{B}_j \mathbf{H}_{t-j} \mathbf{B}_j'$$

We can consider a multivariate generalization of the size effect and sign effect:

$$\mathbf{H}_t = \mathbf{C}\mathbf{C}' + \mathbf{A}_1\epsilon_{t-1}\epsilon_{t-1}'\mathbf{A}_1' + \mathbf{B}_1\mathbf{H}_{t-1}\mathbf{B}_1' + \mathbf{D}\mathbf{v}_{t-1}\mathbf{v}_{t-1}'\mathbf{D}' + \mathbf{G}\epsilon_{t-1}\epsilon_{t-1}'\mathbf{G}'$$

where $\mathbf{v}_t = |\mathbf{z}_t| - E\,|\mathbf{z}_t|$, with $z_{it} = \varepsilon_{it}/\sqrt{h_{ii,t}}$ and

$$\mathbf{G} = \begin{bmatrix} I\left(\varepsilon_{1t-1} < 0\right) g_{11} & 0 & \dots & 0 & \\ 0 & & \ddots & \vdots & \\ \vdots & & & \ddots & 0 \\ 0 & & \dots & 0 & I\left(\varepsilon_{Nt-1} < 0\right) g_{NN} \end{bmatrix}$$

When $N = 2$

$$
\begin{aligned}
\mathbf{v}_{t-1}\mathbf{v}_{t-1}' &= \left[ \begin{array}{c} \left|\varepsilon_{1t-1}/\sqrt{h_{11,t-1}}\right| - E\left|\varepsilon_{1t-1}/\sqrt{h_{11,t-1}}\right| \\ \left|\varepsilon_{2t-1}/\sqrt{h_{22,t-1}}\right| - E\left|\varepsilon_{2t-1}/\sqrt{h_{22,t-1}}\right| \end{array} \right] \times \\
& \left[ \begin{array}{c} \left|\varepsilon_{1t-1}/\sqrt{h_{11,t-1}}\right| - E\left|\varepsilon_{1t-1}/\sqrt{h_{11,t-1}}\right| \\ \left|\varepsilon_{2t-1}/\sqrt{h_{22,t-1}}\right| - E\left|\varepsilon_{2t-1}/\sqrt{h_{22,t-1}}\right| \end{array} \right]' \\
&= \left[ \begin{array}{ll} \left(|z_{1t}| - E\,|z_{1t}|\right)^2 & \left(|z_{1t}| - E\,|z_{1t}|\right)\left(|z_{2t}| - E\,|z_{2t}|\right) \\ \left(|z_{2t}| - E\,|z_{2t}|\right)\left(|z_{1t}| - E\,|z_{1t}|\right) & \left(|z_{2t}| - E\,|z_{2t}|\right)^2 \end{array} \right]
\end{aligned}
$$

$$\mathbf{G}\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1}'\mathbf{G}' = \begin{bmatrix} g_{11}^{*2}\varepsilon_{1t-1}^{2} & g_{11}^{*}g_{22}^{*}\varepsilon_{1t-1}\varepsilon_{2t-1} \\ g_{11}^{*}g_{22}^{*}\varepsilon_{1t-1}\varepsilon_{2t-1} & g_{22}^{*2}\varepsilon_{2t-1}^{2} \end{bmatrix}$$

$$= \begin{bmatrix} I\left(\varepsilon_{1t-1} < 0\right) g_{11}^{2}\varepsilon_{1t-1}^{2} & \delta_{12} g_{11} g_{22} \varepsilon_{1t-1}\varepsilon_{2t-1} \\ \delta_{12} g_{11} g_{22} \varepsilon_{1t-1}\varepsilon_{2t-1} & I\left(\varepsilon_{2t-1} < 0\right) g_{22}^{2}\varepsilon_{2t-1}^{2} \end{bmatrix}$$

$$\delta_{12} = I\left(\varepsilon_{1t-1} < 0\right) I\left(\varepsilon_{2t-1} < 0\right)$$

Given the model (3)-(44), the log-likelihood function for $\{\varepsilon_T, \ldots, \varepsilon_1\}$ obtained under the assumption of conditional multivariate normality is:

$$\log L_T(\boldsymbol{\epsilon}_T, \ldots, \boldsymbol{\epsilon}_1; \boldsymbol{\theta}) = -\frac{1}{2}\left[TN\log(2\pi) + \sum_{t=1}^{T}\left(\log|\mathbf{H}_t| + \boldsymbol{\epsilon}_t'\mathbf{H}_t^{-1}\boldsymbol{\epsilon}_t\right)\right]$$

- The assumption of conditional normality can be quite restrictive.
- The symmetry imposed under normality is difficult to justify, and the tails of even conditional distributions often seem fatter than that of normal distribution.

Let $\{(\mathbf{y}_t, \mathbf{x}_t) : t = 1, 2, \ldots\}$ be a sequence of observable random vectors with $\mathbf{y}_t$ $(N \times 1)$ and $\mathbf{x}_t$ $(L \times 1)$.
The vector $y_t$ contains the "endogenous" variables and $x_t$ contains contemporaneous "exogenous" variables.

$$w_t = (\mathbf{x}_t, \mathbf{y}_{t-1}, \mathbf{x}_{t-1}, \ldots, y_1, \mathbf{x}_1).$$

The conditional mean and variance functions are jointly parameterized by a finite dimensional vector $\boldsymbol{\theta}$:

$$\{\boldsymbol{\mu}_t (w_t, \boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}$$

$$\{\mathbf{H}_t (w_t, \boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}$$

where $\Theta \subset \mathbb{R}^P$ and $\boldsymbol{\mu}_t$ and $\mathbf{H}_t$ are known functions of $w_t$ and $\boldsymbol{\theta}$.

The validity of most of the inference procedures is proven under the null hypothesis that the first two conditional moments are correctly specified, for some $\boldsymbol{\theta}_0 \in \Theta$,

$$E\left(\mathbf{y}_t \left| w_t \right.\right) = \mu_t\left(w_t, \boldsymbol{\theta}_0\right)$$
$$Var\left(\mathbf{y}_t \left| w_t \right.\right) = \mathbf{H}_t\left(w_t, \boldsymbol{\theta}_0\right) \quad t = 1, 2, \ldots$$

The procedure most often used to estimate $\boldsymbol{\theta}_0$ is the maximization of a likelihood function that is constructed under the assumption that

$$\mathbf{y}_t \left| w_t \right. \sim N\left(\boldsymbol{\mu}_t, \mathbf{H}_t\right).$$

The approach taken here is the same, but the subsequent analysis does not assume that $y_t$ has a conditional normal distribution.

For observation $t$ the quasi-conditional log-likelihood is

$$\begin{aligned} l_t(\boldsymbol{\theta}; \mathbf{y}_t, w_t) &= -\frac{N}{2}\ln(2\pi) - \frac{1}{2}\ln|\mathbf{H}_t(w_t, \boldsymbol{\theta})| \\ &\quad -\frac{1}{2}(\mathbf{y}_t - \boldsymbol{\mu}_t(w_t, \boldsymbol{\theta}))' \mathbf{H}_t^{-1}(w_t, \boldsymbol{\theta})(\mathbf{y}_t - \boldsymbol{\mu}_t(w_t, \boldsymbol{\theta})) \end{aligned}$$

Letting

$$\boldsymbol{\epsilon}_t(\mathbf{y}_t, w_t, \boldsymbol{\theta}_0) \equiv \mathbf{y}_t - \boldsymbol{\mu}_t(w_t, \boldsymbol{\theta}) : (N \times 1)$$

denote the residual function

$$l_t(\boldsymbol{\theta}) = -\frac{N}{2}\log(2\pi) - \frac{1}{2}\log|\mathbf{H}_t(\boldsymbol{\theta})| - \frac{1}{2}\boldsymbol{\epsilon}_t'(\boldsymbol{\theta})\mathbf{H}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\epsilon}_t(\boldsymbol{\theta})$$

$$\log L_T(\boldsymbol{\theta}) = \sum_{t=1}^{T} l_t(\boldsymbol{\theta})$$

If $\boldsymbol{\mu}_t\left(w_t, \boldsymbol{\theta}\right)$ and $\mathbf{H}_t\left(w_t, \boldsymbol{\theta}\right)$ are differentiable on $\Theta$ for all relevant $w_t$, and if $\mathbf{H}_t\left(w_t, \boldsymbol{\theta}\right)$ is nonsingular with probability one for all $\boldsymbol{\theta} \in \Theta$, then the differentiation of the loglik yields the $(1 \times P)$ score function $s_t\left(\boldsymbol{\theta}\right)$:

$$
\begin{aligned}
s_t\left(\boldsymbol{\theta}\right)' &= \nabla_{\boldsymbol{\theta}} l_t\left(\boldsymbol{\theta}\right)' - \nabla_{\boldsymbol{\theta}} \mu_t\left(\boldsymbol{\theta}\right)' \mathbf{H}_t^{-1}\left(\boldsymbol{\theta}\right) \boldsymbol{\epsilon}_t\left(\boldsymbol{\theta}\right) + \\
& \frac{1}{2} \nabla_{\boldsymbol{\theta}} \mathbf{H}_t\left(\boldsymbol{\theta}\right)' \left[\mathbf{H}_t^{-1}\left(\boldsymbol{\theta}\right) \otimes \mathbf{H}_t^{-1}\left(\boldsymbol{\theta}\right)\right] vec\left[\boldsymbol{\epsilon}_t\left(\boldsymbol{\theta}\right) \boldsymbol{\epsilon}_t\left(\boldsymbol{\theta}\right)' - \mathbf{H}_t\left(\boldsymbol{\theta}\right)\right]
\end{aligned}
$$

where

$$
\nabla_{\boldsymbol{\theta}} \boldsymbol{\mu}_t\left(\boldsymbol{\theta}\right) : (N \times P)
$$

$$
\nabla_{\boldsymbol{\theta}} \mathbf{H}_t\left(\boldsymbol{\theta}\right) : \left(N^2 \times P\right)
$$

If the first conditional two moments are correctly specified, the true error vector is defined as

$$\boldsymbol{\epsilon}_t^0 \equiv \boldsymbol{\epsilon}_t(\boldsymbol{\theta}_0) = \mathbf{y}_t - \boldsymbol{\mu}_t(w_t, \boldsymbol{\theta}_0)$$

and $E\left(\boldsymbol{\epsilon}_t^0 \mid w_t\right) = 0$,

$$E\left(\boldsymbol{\epsilon}_t^0 \boldsymbol{\epsilon}_t^{0\prime} \mid w_t\right) = \mathbf{H}_t(w_t, \boldsymbol{\theta}_0)$$

It follows that under correct specification of the first two conditional moments of $y_t$ given $w_t$:

$$E\left[s_t(\boldsymbol{\theta}_0) \mid w_t\right] = 0$$

The score evaluated at the true parameter is a **vector of martingale difference** with respect to the $\sigma - \textit{fields}$ $\{\sigma(\mathbf{y}_t, w_t) : t = 1, 2, \ldots\}$. This result can be used to establish **weak consistency of the quasi-maximum likelihood estimator** (QMLE).

For **robust inference** we also need an expression for the hessian $h_t(\boldsymbol{\theta})$ of $l_t(\boldsymbol{\theta})$. Define the positive semidefinite matrix

$$a_t(\boldsymbol{\theta}_0) = -E\left[\nabla_{\boldsymbol{\theta}} s_t(\boldsymbol{\theta}_0)\,|w_t\right] = E\left[-h_t(\boldsymbol{\theta}_0)\,|w_t\right] : (P \times P)$$

$$\begin{aligned} a_t(\boldsymbol{\theta}_0) \quad = \quad & \nabla_{\boldsymbol{\theta}}\mu_t(\boldsymbol{\theta}_0)' \, \mathbf{H}_t^{-1}(\boldsymbol{\theta}_0) \nabla_{\boldsymbol{\theta}}\mu_t(\boldsymbol{\theta}_0) \\ & + \frac{1}{2}\nabla_{\boldsymbol{\theta}}\mathbf{H}_t(\boldsymbol{\theta})' \left[\mathbf{H}_t^{-1}(\boldsymbol{\theta}) \otimes \mathbf{H}_t^{-1}(\boldsymbol{\theta})\right] \nabla_{\boldsymbol{\theta}}\mathbf{H}_t(\boldsymbol{\theta}) \end{aligned}$$

When the normality assumption holds the matrix $a_t(\boldsymbol{\theta}_0)$ is the conditional information matrix. However, if $y_t$ does not have a conditional normal distribution then

$$Var\left[s_t(\boldsymbol{\theta}_0)\,|w_t\right] \neq a_t(\boldsymbol{\theta}_0)$$

and the information matrix equality is violated.

The QMLE has the following properties:

$$\left[\mathbf{A}_T^{0-1}\mathbf{B}_T^0\mathbf{A}_T^{0-1}\right]^{-1/2}\sqrt{T}\left(\widehat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0\right)\xrightarrow{d} N\left(\mathbf{0},\mathbf{I}_P\right)$$

where

$$\mathbf{A}_T^0 \equiv -\frac{1}{T}\sum_{t=1}^{T}E\left[h_t\left(\boldsymbol{\theta}_0\right)\right]=\frac{1}{T}\sum_{t=1}^{T}E\left[a_t\left(\boldsymbol{\theta}_0\right)\right]$$

and

$$\mathbf{B}_T^0 \equiv Var\left[T^{-1/2}S_T\left(\boldsymbol{\theta}_0\right)\right]=\frac{1}{T}\sum_{t=1}^{T}E\left[s_t\left(\boldsymbol{\theta}_0\right)' s_t\left(\boldsymbol{\theta}_0\right)\right]$$

in addition

$$\widehat{\mathbf{A}}_T-\mathbf{A}_T^0\xrightarrow{p}\mathbf{0}$$

$$\widehat{\mathbf{B}}_T-\mathbf{B}_T^0\xrightarrow{p}\mathbf{0}$$

The matrix $\widehat{\mathbf{A}}_T^{-1}\widehat{\mathbf{B}}_T\widehat{\mathbf{A}}_T^{-1}$ is a consistent estimator od the robust asymptotic covariance matrix of $\sqrt{T}\left(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0\right)$.

In practice,

$$\widehat{\boldsymbol{\theta}}_T \approx N\left(\boldsymbol{\theta}, \widehat{\mathbf{A}}_T^{-1}\widehat{\mathbf{B}}_T\widehat{\mathbf{A}}_T^{-1}/T\right)$$

Under normality, the variance estimator can be replaced by $\widehat{\mathbf{A}}_T^{-1}/T$ (Hessian form) or $\widehat{\mathbf{B}}_T^{-1}/T$ (outer product of the gradient form).

The null hypothesis is

$$H_0 : r\left(\boldsymbol{\theta}_0\right) = 0$$

where $r : \Theta \to \mathbb{R}^Q$ is continuously differentiable on $int\left(\Theta\right)$ and $Q < P$. Let

$$\mathbf{R}\left(\boldsymbol{\theta}\right) = \nabla_{\boldsymbol{\theta}} r\left(\boldsymbol{\theta}\right) : \left(Q \times P\right)$$

be the gradient of $r$ on $int\left(\Theta\right)$. If $\boldsymbol{\theta}_0 \in int\left(\Theta\right)$ and $rank\left(\mathbf{R}\left(\boldsymbol{\theta}_0\right)\right) = Q$ then the Wald statistic

$$\xi_W = Tr\left(\widehat{\boldsymbol{\theta}}_T\right)' \left[\mathbf{R}\left(\widehat{\boldsymbol{\theta}}_T\right)\widehat{\mathbf{A}}_T^{-1}\widehat{\mathbf{B}}_T\widehat{\mathbf{A}}_T^{-1}\mathbf{R}\left(\widehat{\boldsymbol{\theta}}_T\right)'\right]^{-1} r\left(\widehat{\boldsymbol{\theta}}_T\right) \underset{H_0}{\overset{d}{\to}} \chi_Q^2.$$

The Factor GARCH model, introduced by Engle et al. (1990), can be thought of as an alternative simple parametrization of the BEKK model.

## Factor model

Suppose that the $(N \times 1)$ $\mathbf{y}_t$ has a factor structure with $K$ factors given by the $K \times 1$ vector $\mathbf{f}_t$ and a time invariant factor loadings given by the $N \times K$ matrix $B$:

$$\mathbf{y}_t = \mathbf{B}\mathbf{f}_t + \boldsymbol{\epsilon}_t$$

Assume that the idiosyncratic shocks $\boldsymbol{\epsilon}_t$ have conditional covariance matrix $\Psi$ which is constant in time and positive semidefinite, and that the common factors are characterized by

$$E_{t-1}(\mathbf{f}_t) = 0$$

$$E_{t-1}(\mathbf{f}_t \mathbf{f}_t') = \boldsymbol{\Lambda}_t$$

$\boldsymbol{\Lambda}_t = \text{diag}(\lambda_1, \ldots, \lambda_K)$ and positive definite. The conditioning set is $\{\mathbf{y}_{t-1}, \mathbf{f}_{t-1}, \ldots, \mathbf{y}_1, \mathbf{f}_1\}$. Also suppose that $E(\mathbf{f}_t \boldsymbol{\epsilon}_t') = 0$. The conditional covariance matrix of $\mathbf{y}_t$ equals

$$E_{t-1}(\mathbf{y}_t \mathbf{y}_t') = \mathbf{H}_t = \Psi + \mathbf{B}\boldsymbol{\Lambda}_t\mathbf{B}' = \Psi + \sum_{k=1}^{K} \boldsymbol{\beta}_k \boldsymbol{\beta}_k' \lambda_{kt}$$

where $\boldsymbol{\beta}_k$ denotes the $k$th column in $\mathbf{B}$. Thus, there are $K$ statistics which determine the full covariance matrix.

Forecasts of the variances and covariances or of any portfolio of assets, will be based only on the forecasts of these $K$ statistics.

## *Factor-representing portfolios*

Portfolio weights are orthogonal to all but one set of factor loadings:

$$r_{kt} = \phi_k' \mathbf{y}_t$$

$$\phi_k' \boldsymbol{\beta}_j = \begin{cases} 1 & k = j \\ 0 & \textit{otherwise} \end{cases}$$

the vector of factor-representing portfolios is

$$\mathbf{r}_t = \Phi' \mathbf{y}_t$$

where the columns of matrix $\Phi$ are the $\phi_k$ vectors.

The conditional variance of $r_{kt}$ is given by

$$\begin{aligned} \mathrm{Var}_{t-1}(r_{kt}) &= \boldsymbol{\phi}_k' E_{t-1}(\mathbf{y}_t\mathbf{y}_t')\,\boldsymbol{\phi}_k = \boldsymbol{\phi}_k'\mathbf{H}_t\boldsymbol{\phi}_k \\ &= \boldsymbol{\phi}_k'(\Psi + \mathbf{B}\boldsymbol{\Lambda}_t\mathbf{B}')\,\boldsymbol{\phi}_k \\ &= \psi_k + \lambda_{kt} \end{aligned}$$

where $\psi_k = \boldsymbol{\phi}_k'\Psi\boldsymbol{\phi}_k$. The portfolio has the exact time variation as the factors, which is why they are called factor-representing portfolios. In order to estimate this model, the dependence of the $\lambda_{kt}$'s upon the past information set must also be parameterized:

$$\boldsymbol{\theta}_{kt} \equiv \boldsymbol{\phi}_k'\mathbf{H}_t\boldsymbol{\phi}_k = Var_{t-1}(r_{kt}) = \psi_k + \lambda_{kt}$$

So we get that

$$\sum_{k=1}^{K}\boldsymbol{\beta}_k\boldsymbol{\beta}_k'\boldsymbol{\theta}_{kt} = \sum_{k=1}^{K}\boldsymbol{\beta}_k\boldsymbol{\beta}_k'\psi_k + \sum_{k=1}^{K}\boldsymbol{\beta}_k\boldsymbol{\beta}_k'\lambda_{kt}$$

$$\sum_{k=1}^{K}\boldsymbol{\beta}_k\boldsymbol{\beta}_k'\lambda_{kt} = \sum_{k=1}^{K}\boldsymbol{\beta}_k\boldsymbol{\beta}_k'\boldsymbol{\theta}_{kt} - \sum_{k=1}^{K}\boldsymbol{\beta}_k\boldsymbol{\beta}_k'\psi_k$$

$$H_t = \Psi + \sum_{k=1}^{K}\boldsymbol{\beta}_k\boldsymbol{\beta}_k'\lambda_{kt} = \Psi + \sum_{k=1}^{K}\boldsymbol{\beta}_k\boldsymbol{\beta}_k'\boldsymbol{\theta}_{kt} - \sum_{k=1}^{K}\boldsymbol{\beta}_k\boldsymbol{\beta}_k'\psi_k$$

The simplest assumption is that there is a set of factor-representing portfolios with univariate GARCH(1,1) representations. The conditional variance $\boldsymbol{\theta}_{kt}$ follows a GARCH(1,1) process

$$\begin{aligned}
\boldsymbol{\theta}_{kt} &= \omega_k + \alpha_k \left(\boldsymbol{\phi}_k' \boldsymbol{\epsilon}_{t-1}\right)^2 + \gamma_k E_{t-2}\left(r_{kt-1}^2\right) \\
&= \omega_k + \alpha_k \boldsymbol{\phi}_k' \left(\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1}'\right) \boldsymbol{\phi}_k + \gamma_k E_{t-2}\left[\left(\boldsymbol{\phi}_k' y_t\right)\left(\boldsymbol{\phi}_k' y_t\right)\right] \\
&= \omega_k + \alpha_k \boldsymbol{\phi}_k' \left(\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1}'\right) \boldsymbol{\phi}_k + \gamma_k \left[\boldsymbol{\phi}_k' E_{t-2}\left(\mathbf{y}_t \mathbf{y}_t'\right) \boldsymbol{\phi}_k\right] \\
&= \omega_k + \alpha_k \boldsymbol{\phi}_k' \left(\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1}'\right) \boldsymbol{\phi}_k + \gamma_k \left[\boldsymbol{\phi}_k' \mathbf{H}_{t-1} \boldsymbol{\phi}_k\right]
\end{aligned}$$

The conditional variance-covariance matrix of $\mathbf{y}_t$ can be written as

$$
\begin{aligned}
\mathbf{H}_t &= \Psi^* + \sum_{k=1}^{K} \boldsymbol{\beta}_k \boldsymbol{\beta}_k' \boldsymbol{\theta}_{kt} \\
&= \Psi^* + \sum_{k=1}^{K} \boldsymbol{\beta}_k \boldsymbol{\beta}_k' \left\{ \omega_k + \alpha_k \left[ \boldsymbol{\phi}_k' \left( \boldsymbol{\epsilon}_{t-1} \boldsymbol{\epsilon}_{t-1}' \right) \boldsymbol{\phi}_k \right] + \gamma_k \left[ \boldsymbol{\phi}_k' \mathbf{H}_{t-1} \boldsymbol{\phi}_k \right] \right\} \\
&= \left( \Psi^* + \sum_{k=1}^{K} \boldsymbol{\beta}_k \boldsymbol{\beta}_k' \omega_k \right) \\
&\quad + \sum_{k=1}^{K} \boldsymbol{\beta}_k \boldsymbol{\beta}_k' \left\{ \alpha_k \left[ \boldsymbol{\phi}_k' \left( \boldsymbol{\epsilon}_{t-1} \boldsymbol{\epsilon}_{t-1}' \right) \boldsymbol{\phi}_k \right] + \gamma_k \left[ \boldsymbol{\phi}_k' \mathbf{H}_{t-1} \boldsymbol{\phi}_k \right] \right\}
\end{aligned}
$$

$$
\mathbf{H}_t = \Gamma + \sum_{k=1}^{K} \boldsymbol{\beta}_k \boldsymbol{\beta}_k' \boldsymbol{\theta}_{kt}
$$

where $\Gamma = \Psi^* + \sum_{k=1}^{K} \boldsymbol{\beta}_k \boldsymbol{\beta}_k' \omega_k$.

Therefore

$$\mathbf{H}_t = \Gamma + \sum_{k=1}^{K} \alpha_k \left[ \boldsymbol{\beta}_k \boldsymbol{\phi}_k' \left( \boldsymbol{\epsilon}_{t-1} \boldsymbol{\epsilon}_{t-1}' \right) \boldsymbol{\phi}_k \boldsymbol{\beta}_k' \right] + \sum_{k=1}^{K} \gamma_k \left[ \boldsymbol{\beta}_k \boldsymbol{\phi}_k' \mathbf{H}_{t-1} \boldsymbol{\phi}_k \boldsymbol{\beta}_k' \right]$$

so that the factor GARCH model is a special case of the BEKK parametrization. Estimation of the factor GARCH model is carried out by maximum likelihood estimation. It is often convenient to assume that the factor-representing portfolios are known a priori.

- The orthogonal models are particular factor models (Kariya (1988) and Alexander and Chibumba (1997)).
- They are based on the assumption that the observed data can be obtained by a *linear transformation* of a set of uncorrelated components by means of an *orthogonal matrix*.

The $(N \times N)$ time-varying variance matrix $H_t$ is generated by $(m \times N)$ univariate GARCH models.
The components are the principal components of the data, or a subset of them.

The diagonal matrix $\mathbf{V}$ contains the population variances of $\mathbf{y}_t$:

$$\mathbf{V} = \text{diag}\{v_1^2, \dots, v_N^2\}$$

the standardized returns are

$$\mathbf{u}_t = \mathbf{V}^{-1/2}\mathbf{y}_t$$

where

$$\mathbf{u}_t = \mathbf{L}\mathbf{f}_t$$

$$E[\mathbf{u}_t] = \mathbf{0} \qquad E[\mathbf{u}_t\mathbf{u}_t'] = \mathbf{R}$$

The population correlation matrix can be decomposed as:

$$\mathbf{R} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}'$$

$\mathbf{P}$ is the orthogonal eigenvectors matrix, $\mathbf{\Lambda}$ is the diagonal matrix of the eigenvalues:

$$\mathbf{\Lambda} = \mathrm{diag}\{\lambda_1, \ldots, \lambda_N\}$$

ranked in descending order.

$\mathbf{P}$ satisfies:

$$\mathbf{P}' = \mathbf{P}^{-1} \qquad \mathbf{P}'\mathbf{P} = \mathbf{I}_N \qquad \mathbf{P}\mathbf{P}' = \mathbf{I}_N$$

It follows that

$$\mathbf{R} = \mathbf{P}\boldsymbol{\Lambda}^{1/2}\boldsymbol{\Lambda}^{1/2}\mathbf{P}' = \mathbf{L}\mathbf{L}'$$

The factor loading matrix is obtained as

$$\mathbf{L} = \mathbf{P}\boldsymbol{\Lambda}^{1/2}$$

such that

$$\mathbf{f}_t = \mathbf{L}^{-1}\mathbf{u}_t$$

with

$$\begin{aligned} E[\mathbf{f}_t\mathbf{f}_t'] &= \mathbf{L}^{-1}E[\mathbf{u}_t\mathbf{u}_t']\mathbf{L}^{-1\prime} = \mathbf{L}^{-1}\mathbf{R}\mathbf{L}^{-1\prime} \\ &= \mathbf{L}^{-1}\mathbf{L}\mathbf{L}'\mathbf{L}'^{-1} = \mathbf{I}_N \end{aligned}$$

Assuming

$$E_{t-1}[\mathbf{f}_t\mathbf{f}_t'] = \mathbf{Q}_t = \mathrm{diag}(\sigma^2_{f_{1,t}}, \dots, \sigma^2_{f_{N,t}})$$

$\mathbf{Q}_t$ is a diagonal matrix.

$$\sigma^2_{f_{i,t}} = (1 - \alpha_{i,1} - \beta_{i,1}) + \alpha_{i,1} f^2_{i,t-1} + \beta_{i,1}\sigma^2_{f_{i,t-1}} \quad i = 1, 2, \dots, N$$

$$E_{t-1}[\mathbf{u}_t\mathbf{u}_t'] = E_{t-1}[\mathbf{L}\mathbf{f}_t\mathbf{f}_t'\mathbf{L}'] = \mathbf{L}\mathbf{Q}_t\mathbf{L}'$$

$$E_{t-1}[\mathbf{y}_t\mathbf{y}_t'] = E_{t-1}[\mathbf{V}^{1/2}\mathbf{u}_t\mathbf{u}_t'\mathbf{V}^{1/2}] = \mathbf{V}^{1/2}\mathbf{L}\mathbf{Q}_t\mathbf{L}'\mathbf{V}^{1/2}$$

The number of parameters is $N(N+5)/2$.

In practice, $\mathbf{V}$ and $\mathbf{L}$ are replaced by their sample counterparts, and $m$ is chosen by principal component analysis applied to the standardized residuals, $\widehat{\mathbf{u}}_t$.
We can work with a reduced number $m < N$ of principal components (eigenvalues), those which explain most of the variation in the data. $\mathbf{L}^{-1}$ is replaced by a matrix $(m \times N)$:

$$\mathbf{\Lambda}_m^{-1/2}\mathbf{P}'_m$$

$\mathbf{P}_m$ is a matrix $(N \times m)$ containing the $m$ eigenvectors of $\mathbf{P}$ corresponding to the $m$ largest eigenvalues.

$$\mathbf{f}_t^m = \mathbf{\Lambda}_m^{-1/2}\mathbf{P}'_m\mathbf{u}_t$$

where $\mathbf{f}_t^m = [f_{1t}, \ldots, f_{mt}]$

$$E_{t-1}[\mathbf{f}_t^m] = \mathbf{0}$$
$$E_{t-1}[\mathbf{f}_t^m \mathbf{f}_t^{m'}] = \mathbf{Q}_{m,t} = \text{diag}(\sigma^2_{f_{1,t}}, \ldots, \sigma^2_{f_{m,t}})$$

Alexander (2001, section 7.4.3) emphasizes that using a small number of principal components compared to the number of assets is the strength of the approach. However, note that the conditional variance matrix has reduced rank (if $m < N$), which may be a problem for applications and for diagnostic tests which depend on the inverse of $\mathbf{H}_t$.

- These models are based on a decomposition of the $\mathbf{H}_t$.
- The conditional var-cov matrix is expressed as

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

  where $\mathbf{R}_t$ is possibly time-varying.
- Conditional correlations and variances are separately modeled.

Bollerslev (1990) Constant Conditional Correlations model:
The time-varying conditional covariances are parameterized to be proportional to the product of the corresponding conditional standard deviations. The model assumptions are:

$$E_{t-1}\left[\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t'\right] = \mathbf{H}_t$$

$$\left\{\mathbf{H}_t\right\}_{ii} = h_{it} \quad i = 1, \ldots, N$$

$$\left\{\mathbf{H}_t\right\}_{ij} = h_{ijt} = \rho_{ij} h_{it}^{1/2} h_{jt}^{1/2} \qquad i \neq j \quad i, j = 1, \ldots, N$$

$$\mathbf{D}_t = \operatorname{diag}\left\{h_{1t}, \ldots, h_{Nt}\right\}$$

The conditional covariance matrix can be written as:

$$\mathbf{H}_t = \mathbf{D}_t^{1/2}\mathbf{R}\mathbf{D}_t^{1/2}$$

$$\mathbf{H}_t = \begin{bmatrix} h_{1t}^{1/2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & h_{Nt}^{1/2} \end{bmatrix} \begin{bmatrix} 1 & \rho_{12} & \ldots & \rho_{1N} \\ \rho_{21} & 1 & \ldots & \vdots \\ \vdots & \vdots & \ldots & \rho_{N-1N} \\ \rho_{N1} & \ldots & \rho_{NN-1} & 1 \end{bmatrix} \begin{bmatrix} h_{1t}^{1/2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & h_{Nt}^{1/2} \end{bmatrix}.$$

When $N = 2$

$$\begin{aligned} \mathbf{H}_t &= \begin{bmatrix} h_{1t}^{1/2} & 0 \\ 0 & h_{2t}^{1/2} \end{bmatrix} \begin{bmatrix} 1 & \rho_{12} \\ \rho_{21} & 1 \end{bmatrix} \begin{bmatrix} h_{1t}^{1/2} & 0 \\ 0 & h_{2t}^{1/2} \end{bmatrix} \\ &= \begin{bmatrix} h_{1t} & \rho_{12} h_{1t}^{1/2} h_{2t}^{1/2} \\ \rho_{12} h_{1t}^{1/2} h_{2t}^{1/2} & h_{2t} \end{bmatrix}. \end{aligned}$$

- The sequence of conditional covariance matrices $\{\mathbf{H}_t\}$ is guaranteed to be positive definite a.s. for all $t$, If the conditional variances along the diagonal in the $\mathbf{D}_t$ matrices are all positive, and the conditional correlation matrix $\mathbf{R}$ is positive definite
- Furthermore the inverse of $\mathbf{H}_t$ is given by

$$\mathbf{H}_t^{-1} = \mathbf{D}_t^{-1/2}\mathbf{R}^{-1}\mathbf{D}_t^{-1/2}.$$

  When calculating the log-likelihood function only one matrix inversion is required for each evaluation.
- CCC is generally estimated in two steps:
  1. conditional variances are estimated employing the marginal likelihoods
  2. $\mathbf{R}$ is estimated using the sample estimator of standardized residuals $\widehat{\mathbf{D}}_t^{-1}\mathbf{y}_t$ (assuming $\boldsymbol{\mu}_t = \mathbf{0}$).

- The CCC solves the *curse of dimensionality* problem of MGARCH models
- The number of parameters is $O(N^2)$ but these are not jointly estimated. The two-step estimation procedure impacts on the computational issues.
- Asymptotic properties of QMLE estimators verified in McAleer and Ling (2003).

The CCC has two main limitations:

1. No spillover neither feedback effects across conditional variances
2. Correlations are static

The evolution of CCC is the Dynamic Conditional Correlation (DCC) Model of Engle (2002). The DCC is an extension of the Bollerslev's CCC Model.

The conditional correlation between two random variables, $X_t$ and $Y_t$ is defined as:

$$\rho_{YX,t} = \frac{Cov_{t-1}(X_t Y_t)}{\sqrt{E_{t-1}(X_t - \mu_{x,t})^2 E_{t-1}(Y_t - \mu_{Y,t})^2}}$$

Assets returns conditional distribution:

$$\mathbf{y}_t | \Phi_{t-1} \sim N(\mathbf{0}, \mathbf{H}_t)$$

$$\mathbf{H}_t = \mathbf{D}_t^{1/2} \mathbf{R}_t \mathbf{D}_t^{1/2}.$$

$$\mathbf{D}_t = \text{diag}(\text{Var}_{t-1}(y_{1t}), \ldots, \text{Var}_{t-1}(y_{Nt}))$$

where the $\text{Var}_{t-1}(y_{it}), i = 1, \ldots, N$ are modeled as univariate GARCH processes.

The standardized returns are:

$$\boldsymbol{\eta}_t = \mathbf{D}_t^{-1/2}\mathbf{y}_t$$

$$E_{t-1}(\boldsymbol{\eta}_t\boldsymbol{\eta}_t') = \mathbf{D}_t^{-1/2}\mathbf{H}_t\mathbf{D}_t^{-1/2} = \mathbf{R}_t = \{\rho_{ij,t}\}$$

we can use the conditional variance of $\boldsymbol{\eta}_t$ to describe the conditional correlation of $\mathbf{y}_t$.

The conditional correlation estimator is

$$\rho_{ij,t} = \frac{q_{ij,t}}{\sqrt{q_{ii,t}q_{jj,t}}}.$$

Where $q_{ij,t}$ are assumed to follow a GARCH(1,1) model

$$q_{ij,t} = \overline{\rho}_{ij} + \alpha(\eta_{i,t-1}\eta_{j,t-1} - \overline{\rho}_{ij}) + \beta(q_{ij,t-1} - \overline{\rho}_{ij}) \qquad (4)$$

The term $\overline{\rho}_{ij}$ is not the unconditional correlation between $\eta_{it}$ and $\eta_{jt}$; the unconditional correlation between $\eta_{it}$ and $\eta_{jt}$ has no closed form.

- Engle (2002) assumes that $\overline{\rho}_{ij} \simeq \overline{q}_{ij}$.
- Aielli (2006) and Engle et al. (2008) suggest to modify the standard DCC in order to correct the asymptotic bias which is due to the fact that $\frac{1}{T}\sum_t \epsilon_t \epsilon_t'$ does not converge to $\overline{\mathbf{Q}}$.
- It is known though that the impact of this is very small (see Engle and Sheppard (2001)).

The conditional covariance matrix is positive definite, $\mathbf{Q}_t$, as long as it is a weighted average of definite matrices and semidefinite matrices.
To ensure p.-d-ness of $\mathbf{Q}_t$ we must impose $\alpha + \beta < 1$ In matrix from:

$$\mathbf{Q}_t = \overline{\mathbf{Q}}(1 - \alpha - \beta) + \alpha(\boldsymbol{\eta}_{t-1}\boldsymbol{\eta}'_{t-1}) + \beta(\mathbf{Q}_{t-1})$$

where $\overline{\mathbf{Q}}$ is the unconditional covariance matrix of $\boldsymbol{\eta}_t$.

DCC model has correlation targeting, when $\alpha + \beta < 1$

$$\begin{aligned} E[\mathbf{Q}_t] &= \mathbf{R} \\ E[\mathbf{Q}_t] &= E[\mathbf{Q}_{t-1}] \\ E[\boldsymbol{\eta}_t \boldsymbol{\eta}_t'] &= E[\mathbf{Q}_t] \end{aligned}$$

$$\begin{aligned} E[\mathbf{Q}_t] &= \overline{\mathbf{Q}}(1-\alpha-\beta) + \alpha E[\boldsymbol{\eta}_{t-1}\boldsymbol{\eta}_{t-1}'] + \beta E[\mathbf{Q}_{t-1}] \\ E[\mathbf{Q}_t] &= \mathbf{R}(1-\alpha-\beta) + \alpha E[\mathbf{Q}_t] + \beta E[\mathbf{Q}_t] \end{aligned}$$

Clearly more complex positive definite multivariate GARCH models could be used for the correlation parametrization as long as the unconditional moments are set to the sample correlation matrix.
For example, the MARCH family of Ding and Engle (2001) can be expressed in first order form as:

$$\mathbf{Q}_t = \overline{\mathbf{Q}} \odot (\boldsymbol{\iota}\boldsymbol{\iota}' - \mathbf{A} - \mathbf{B}) + \mathbf{A} \odot \boldsymbol{\eta}_{t-1}\boldsymbol{\eta}'_{t-1} + \mathbf{B} \odot \mathbf{Q}_{t-1} \tag{5}$$

where $\odot$ denotes the Hadamard product ($\{\mathbf{A} \odot \mathbf{B}\}_{ij} = a_{ij} b_{ij}$).

The Generalized-DCC model specification:

$$
\begin{aligned}
\mathbf{D}_t &= \mathrm{diag}\{\omega_i\} + \mathrm{diag}\{\kappa_i\} \odot \mathbf{y}_{t-1}\mathbf{y}'_{t-1} + \mathrm{diag}\{\lambda_i\} \odot \mathbf{D}_{t-1} \\
\boldsymbol{\eta}_t &= \mathbf{D}_t^{-1/2}\mathbf{y}_t \\
\mathbf{Q}_t &= \mathbf{S} \odot (\boldsymbol{\iota}\boldsymbol{\iota}' - \mathbf{A} - \mathbf{B}) + \mathbf{A} \odot \boldsymbol{\eta}_{t-1}\boldsymbol{\eta}'_{t-1} + \mathbf{B} \odot \mathbf{Q}_{t-1} \\
\mathbf{R}_t &= \mathrm{diag}\{\mathbf{Q}_t\}^{-1/2}\, \mathbf{Q}_t\, \mathrm{diag}\{\mathbf{Q}_t\}^{-1/2}.
\end{aligned}
\tag{6}
$$

**A** and **B** are symmetric matrices.

- The assumption of normality gives rise to a likelihood function.
- Without this assumption, the estimator will still have the QML interpretation.
- The second equation simply expresses the assumption that each of the assets follows a univariate GARCH process.

- A real square matrix $\mathbf{A}$, is positive definite if and only if $\mathbf{B} = \mathbf{A}^{*-1}\mathbf{A}\mathbf{A}^{*-1}$ is positive definite, with $\mathbf{A}^* = \mathrm{diag}\{\mathbf{A}\}$.
- In order to ensure that $\mathbf{H}_t$ is positive definite we must have that $\mathbf{D}_t^{-1/2}\mathbf{H}_t\mathbf{D}_t^{-1/2}$ is positive definite.

$\mathbf{H}_t$ is positive definite $\forall t \in T$, if the following restrictions on the univariate GARCH parameters are satisfied for all series $i \in [1, \ldots, N]$ :

1. $\omega_i > 0$
2. $\kappa_i$ and $\lambda_i$ such that $D_{ii,t} > 0$ with probability 1
3. $D^2_{ii,0} > 0$
4. The roots of $1 - \kappa_i Z - \lambda_i Z$ are outside the unit circle.

and the parameters in the DCC satisfy:

1. $\alpha \geq 0$
2. $\beta \geq 0$
3. $\alpha + \beta \leq 1$
4. The minimum eigenvalue of $Q_0 > \delta > 0$ (where $Q_0$ must be positive definite)

The log-likelihood function can be written as:

$$\begin{aligned}
\log L_T &= -\frac{1}{2}\sum_{t=1}^{T}(N\log(2\pi)+\log|\mathbf{H}_t|+\mathbf{y}_t'\mathbf{H}_t^{-1}\mathbf{y}_t) \\
&= -\frac{1}{2}\sum_{t=1}^{T}(N\log(2\pi)+\log|\mathbf{D}_t^{1/2}\mathbf{R}_t\mathbf{D}_t^{1/2}|+\mathbf{y}_t'\mathbf{D}_t^{-1/2}\mathbf{R}_t^{-1}\mathbf{D}_t^{-1/2}\mathbf{y}_t) \\
&= -\frac{1}{2}\sum_{t=1}^{T}(N\log(2\pi)+\log|\mathbf{D}_t|+\log|\mathbf{R}_t|+\boldsymbol{\eta}_t'\mathbf{R}_t^{-1}\boldsymbol{\eta}_t)
\end{aligned}$$

Adding and subtracting $\mathbf{y}_t'\mathbf{D}_t^{-1/2}\mathbf{D}_t^{-1/2}\mathbf{y}_t = \boldsymbol{\eta}_t'\boldsymbol{\eta}_t$

$$\begin{aligned}
\log L_T &= -\frac{1}{2}\sum_{t=1}^{T}(N\log(2\pi)+\log|\mathbf{D}_t|+\mathbf{y}_t'\mathbf{D}_t^{-1/2}\mathbf{D}_t^{-1/2}\mathbf{y}_t \\
&\quad -\boldsymbol{\eta}_t'\boldsymbol{\eta}_t+\log|\mathbf{R}_t|+\boldsymbol{\eta}_t'\mathbf{R}_t^{-1}\boldsymbol{\eta}_t) \\
&= -\frac{1}{2}\sum_{t=1}^{T}(N\log(2\pi)+\log|\mathbf{D}_t|+\mathbf{r}_t'\mathbf{D}_t^{-1}\mathbf{r}_t) \\
&\quad -\frac{1}{2}\sum_{t=1}^{T}(\boldsymbol{\eta}_t'\mathbf{R}_t^{-1}\boldsymbol{\eta}_t-\boldsymbol{\eta}_t'\boldsymbol{\eta}_t+\log|\mathbf{R}_t|)
\end{aligned}$$

Volatility component:

$$\mathcal{L}_V(\boldsymbol{\theta}) \equiv \log L_{V,T}(\boldsymbol{\theta}) = -\frac{1}{2}\sum_{t=1}^{T}(N\log(2\pi) + \log|\mathbf{D}_t| + \mathbf{y}_t'\mathbf{D}_t^{-1}\mathbf{y}_t)$$

Correlation component:

$$\mathcal{L}_C(\boldsymbol{\theta},\phi) \equiv \log L_{C,T}(\boldsymbol{\theta},\phi) = -\frac{1}{2}\sum_{t=1}^{T}(\boldsymbol{\eta}_t'\mathbf{R}_t^{-1}\boldsymbol{\eta}_t - \boldsymbol{\eta}_t'\boldsymbol{\eta}_t + \log|\mathbf{R}_t|)$$

$\boldsymbol{\theta}$ denotes the parameters in $\mathbf{D}_t$ and $\phi$ the parameters in $\mathbf{R}_t$.

$$\mathcal{L}(\boldsymbol{\theta},\phi) = \mathcal{L}_V(\boldsymbol{\theta}) + \mathcal{L}_C(\boldsymbol{\theta},\phi)$$

$$\mathcal{L}_V(\boldsymbol{\theta}) = -\frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{N}\left(\log(2\pi) + \log(h_{i,t}) + \frac{r_{i,t}^2}{h_{i,t}}\right).$$

The likelihood is apparently the sum of individual GARCH likelihoods, which will be jointly maximized by separately maximizing each term.

Two-step procedure:

1. $$\hat{\boldsymbol{\theta}} = \arg\max\{\mathcal{L}_V(\boldsymbol{\theta})\}$$
2. $$\max_{\boldsymbol{\phi}}\{\mathcal{L}_C(\hat{\boldsymbol{\theta}}, \phi)\}.$$

Under regularity conditions, consistency of the first step will ensure consistency of the second step. The maximum of the second step will a function of the first step parameter estimates. If the first step is consistent then the second step will be too as long as the function is continuous in a neighborhood of the true parameters.

Two step GMM problem (Newey and McFadden, 1994). Consider the moment condition corresponding to the first step

$$\nabla_{\boldsymbol{\theta}} \mathcal{L}_V(\boldsymbol{\theta}) = 0$$

The moment corresponding to the second step is

$$\nabla_{\boldsymbol{\phi}} \mathcal{L}(\boldsymbol{\theta}, \widehat{\boldsymbol{\phi}}) = 0$$

Under regularity conditions the parameter estimates will be consistent, and asymptotically normal, with asymptotic covariance matrix

$$V(\phi) = \left[ E(\nabla_{\boldsymbol{\phi\phi}} \mathcal{L}_C) \right]^{-1} \left[ E\Big( \{ \nabla_{\boldsymbol{\phi}} \mathcal{L}_C - E(\nabla_{\boldsymbol{\phi\theta}} \mathcal{L}_C) [E(\nabla_{\boldsymbol{\theta\theta}} \mathcal{L}_V]^{-1} \nabla_{\boldsymbol{\theta}} \mathcal{L}_V \} \right.$$
$$\left. \{ \nabla_{\boldsymbol{\phi}} \mathcal{L}_C - E(\nabla_{\boldsymbol{\phi\theta}} \mathcal{L}_C) [E(\nabla_{\boldsymbol{\theta\theta}} \mathcal{L}_V]^{-1} \nabla_{\boldsymbol{\theta}} \mathcal{L}_V \}' \Big) \right] \left[ E(\nabla_{\boldsymbol{\phi\phi}} \mathcal{L}_C) \right]^{-1}$$