Multivariate Volatility Models

Feng Li
feng.li@cufe.edu.cn

School of Statistics and Mathematics
Central University of Finance and Economics

Revised on May 19, 2015

Today we are going to learn...

1. **Introduction to Multivariate Volatility Models**
2. **Multivariate GARCH models**
3. **Decomposition of volatility matrix**

Multivariate Volatility Models I

- Multivariate volatilities have many important financial applications. They play an important role in portfolio selection and asset allocation, and they can be used to compute the **value at risk** of a financial position consisting of multiple assets.
- We discuss some simple methods for modeling the dynamic relationships between volatility processes of multiple asset returns.
- By **multivariate volatility**, we mean the conditional covariance matrix of multiple asset returns.
- Consider a multivariate return series $\{\mathbf{r}_t\}$. We adopt the same approach as the univariate case by rewriting the series as

$$\mathbf{r}_t = \boldsymbol{\mu}_t + \mathbf{a}_t$$

where $\boldsymbol{\mu}_t = E(\mathbf{r}_t|F_{t-1})$ is the conditional expectation of $\mathbf{r}_t$ given the past information and $\mathbf{a}_t$ is the **shock**, or **innovation**, of the series at time t. We refer the equation as the **mean equation** of $\mathbf{r}_t$.

Multivariate Volatility Models II

- For most return series, it suffices to employ a simple vector ARMA structure with exogenous variables for $\boldsymbol{\mu}_t$
- The conditional covariance matrix of a_t given F_{t-1} is a positive-definite matrix $\boldsymbol{\Sigma}_t$ defined by $\boldsymbol{\Sigma}_t = Cov(\boldsymbol{a}_t|F_{t-1})$.
- Multivariate volatility modeling is concerned with the time evolution of $\boldsymbol{\Sigma}_t$. We refer to a model for the $\boldsymbol{\Sigma}_t$ **process** as a volatility model for the return series r_t.
- There are many ways to generalize univariate volatility models to the multivariate case, but the **curse of dimensionality** quickly becomes a major obstacle in applications because there are $k(k+1)/2$ quantities in $\boldsymbol{\Sigma}_t$ for a k-dimensional return series.
- **Time-varying correlations** are useful in finance. For example, they can be used to estimate the time-varying beta of the market model for a return series.

Diagonal Vectorization (DVEC) Model I

- Bollerslev, Engle, and Wooldridge (1988) generalize the exponentially weighted moving-average approach to propose the model, also known as the diagonal VEC(m, s) model or DVEC(m, s) model

$$\boldsymbol{\Sigma}_t = \boldsymbol{A}_0 + \sum_{i=1}^{m} \boldsymbol{A}_i \bigodot (\boldsymbol{a}_{t-i}\boldsymbol{a}'_{t-i}) + \sum_{j=1}^{s} \mathbf{B}_j \bigodot \boldsymbol{\Sigma}_{t-i}$$

  where m and s are non-negative integers, $\boldsymbol{A}_i$ and $\mathbf{B}_j$ are symmetric matrices.
- And $\bigodot$ denotes the **Hadamard product**, that is, element-by-element multiplication.
- Aach element of Σ_t depends only on its own past value and the corresponding product term in $\boldsymbol{a}_{t-i}\boldsymbol{a}'_{t-i}$.
- Because Σ_t is symmetric, in practice, only the lower triangular part of the model is given.
- Each element of a DVEC model follows a GARCH(1,1)-type model.

Diagonal Vectorization (DVEC) Model II

- The model is, therefore, simple. However, it may not produce a positive-definite covariance matrix.
- Furthermore, the model **does not** allow for dynamic dependence between volatility series.

Baba-Engle-Kraft-Kroner (BEKK) model I

- To guarantee the positive-definite constraint, Engle and Kroner (1995) propose the Baba-Engle-Kraft-Kroner (BEKK) model

$$\mathbf{\Sigma}_t = \mathbf{A}\mathbf{A}' + \sum_{i=1}^{m} \mathbf{A}_i(\mathbf{a}_{t-i}\mathbf{a}'_{t-i})\mathbf{A}'_i + \sum_{j=1}^{s} \mathbf{B}_i \mathbf{\Sigma}_{t-j} \mathbf{B}'_i$$

  where A is a lower triangular matrix and A_i and B_j are $k \times k$ matrices.
- Based on the symmetric parameterization of the model, $\mathbf{\Sigma}_t$ is almost surely positive definite provided that $\mathbf{A}\mathbf{A}'$ is positive definite.
- This model also allows for dynamic dependence between the volatility series.

Baba-Engle-Kraft-Kroner (BEKK) model II

- On the other hand, the model has several disadvantages.
  1. First, the parameters in $\mathbf{A}_i$ and $\mathbf{B}_j$ do not have direct interpretations concerning lagged values of volatilities or shocks.
  2. Second, the number of parameters increases rapidly with m and s.
  3. Many of the estimated parameters are statistically insignificant, introducing additional complications in modeling.
  4. Furthermore, no methods are currently available to search for simplifying structures embedded in a BEKK model. It seems that an unrestricted BEKK(1,1) model is only applicable in practice when k is small.

Cholesky decomposition I

- For ease in obtaining positive-definite volatility matrices, **Cholesky decomposition** has been used in the literature to model multivariate volatility.
- This approach has some advantages in estimation as it requires no parameter constraints for the positive definiteness of $\mathbf{\Sigma}_t$
- In addition, the reparameterization is an orthogonal transformation so that the resulting likelihood function is extremely simple.
- Because $\mathbf{\Sigma}_t$ is positive definite, there exist a lower triangular matrix $\mathbf{L}_t$ with unit diagonal elements and a diagonal matrix $\mathbf{G}_t$ with positive diagonal elements such that

$$\mathbf{\Sigma}_t = \mathbf{L}_t \mathbf{G}_t \mathbf{L}_t'$$

- A feature of the decomposition is that the lower off-diagonal elements of $\mathbf{L}_t$ and the diagonal elements of G_t have nice interpretations.

Cholesky decomposition II

- For the bivariate case we have

$$\Sigma_t = \begin{bmatrix} \sigma_{11} & \sigma_{21} \\ \sigma_{12} = \sigma_{21} & \sigma_{22} \end{bmatrix}_t, \mathbf{L}_t = \begin{bmatrix} 1 & 0 \\ q_{21} & 1 \end{bmatrix}_t, \mathbf{G}_t = \begin{bmatrix} g_{11} & 0 \\ 0 & g_{22} \end{bmatrix}_t,$$

and thus

$$\Sigma_t = \begin{bmatrix} \sigma_{11} & \sigma_{21} \\ \sigma_{12} = \sigma_{21} & \sigma_{22} \end{bmatrix}_t = \begin{bmatrix} g_{11} & q_{21} g_{11} \\ q_{21} g_{11} & g_{22} + q_{21}^2 g_{11} \end{bmatrix}_t$$

- Solving the prior equations, we have

$$\begin{aligned} g_{11} &= \sigma_{11} \\ q_{21} &= \frac{\sigma_{21}}{\sigma_{11}} = \frac{Cov(a_{1t}, a_{2t})}{Var(a_{1t})} \\ g_{22} &= \sigma_{22} - \frac{\sigma_{21}^2}{\sigma_{11}} \end{aligned}$$

Cholesky decomposition III

- In summary
  - The first diagonal element of $\mathbf{G}_t$ is simply the variance of a_{1t}.
  - The second diagonal element of G_t is the residual variance of the simple linear regression $a_{2t} = \beta a_{1t} + b_{2t}$ where b_{2t} denotes the error term.
  - The element $q21, t$ of the lower triangular matrix L_t is the coefficient β of the regression.
  - The prior properties continue to hold for the higher dimensional case.
  - Based on the prior discussion, using Cholesky decomposition amounts to doing an orthogonal transformation from a_t to b_t where $b_{1t} = a_{1t}$ and b_{it} is defined recursively by OLS regression

$$a_{it} = q_{i1,t}b_{1t} + q_{i2,t}b_{2t} + ... + q_{i(i-1),t}b_{(i-1)t} + b_{it}$$

    where $q_{ij,t}$ is the (i,j)-th element of the lower triangular matrix L_t.
- In matrix form we rewrite this transformation as

$$\mathbf{b}_t = \mathbf{L}_t^{-1}\mathbf{a}_t$$

where $\mathbf{L}_t^{-1}$ is also a lower triangular matrix with unit diagonal elements.

Cholesky decomposition IV

- The covariance matrix of $\mathbf{b}_t$ is the diagonal matrix $\mathbf{G}_t$ of the Cholesky decomposition because

$$\mathrm{Cov}(b_t) = \mathbf{L}_t^{-1}\mathbf{\Sigma}_t(\mathbf{L}_t^{-1})' = \mathbf{G}_t$$

- The previous orthogonal transformation also dramatically simplifies the likelihood function of the data. Using the fact that $|L_t| = 1$, we have

$$|\Sigma_t| = |\mathbf{L}_t\mathbf{G}_t\mathbf{L}_t'| = |\mathbf{G}_t| = \prod g_{ii,t}$$

- The parameter vector relevant to volatility modeling under such a transformation becomes $\mathbf{G}_t$ and $\mathbf{L}_t$ which is also a $k(k+1)/2$-dimensional vector

Cholesky decomposition V

- If the conditional distribution of a_t given the past information is multivariate normal $N(\mathbf{0}, \mathbf{\Sigma}_t)$, then the conditional distribution of the transformed series b_t is multivariate normal $N(\mathbf{0}, \mathbf{G}_t)$, and the log-likelihood function of the data becomes extremely simple.

$$-\frac{1}{2}\sum\left\{\log(g_{ii}) + \frac{b_{it}^2}{g_{ii,t}}\right\}$$

Decomposition by Using of Correlations

- To model the volatility of a_t , it suffices to consider the conditional variances and correlation coefficients of a_{it}.
- This allow us to reparameterize of $\boldsymbol{\Sigma}_t$ is to use the conditional correlation coefficients and variances of $\boldsymbol{a}_t$. Specifically, we write $\boldsymbol{\Sigma}_t$ as

$$\boldsymbol{\Sigma}_t = \mathbf{D_t} \boldsymbol{\rho_t} \mathbf{D_t}$$

  where where $\boldsymbol{\rho}_t$ is the conditional correlation matrix (symmetric) of $\boldsymbol{a}_t$, and D_t is a $k \times k$ diagonal matrix consisting of the conditional standard deviations of elements of $\boldsymbol{a}_t$.
- This reparameterization is useful because it models covariances and correlations directly. Yet the approach has several weaknesses.
  1. The likelihood function becomes complicated when $k \geqslant 3$.
  2. The approach requires a constrained maximization in estimation to ensure the positive definiteness of $\boldsymbol{\Sigma}_t$. The constraint becomes complicated when k is large.

Dynamic Correlation Models I

- Using the correlation parameterization, several authors have proposed parsimonious models for $\boldsymbol{\rho}_t$ to describe the time-varying correlations. We refer to those models as the **dynamic conditional correlation** (DCC) models.
- For k-dimensional returns, Tse and Tsui (2002) assume that the conditional correlation matrix ρ follows the model

$$\rho_t = (1 - \theta_1 - \theta_2)\rho + \theta_1 \boldsymbol{\rho}_{t-1} + \theta_2 \psi_{t-1}$$

  where θ_1 and θ_2 are scalar parameters, ρ is a $k \times k$ positive-definite matrix with unit diagonal elements, and ψ_{t-1} is the $k \times k$ sample correlation matrix using shocks from $t - m, \ldots, t - 1$ for a prespecified m.
- Typically, one assumes that $0 \leqslant \theta_i < 1$ and $\theta_1 + \theta_2 < 1$ so that the resulting correlation matrix ρ_t is positive definite for all t.
- In applications, the choice of ρ and m deserves a careful investigation. One possibility is to let ρ be the sample correlation matrix of the returns. The correlation equation then only employs two parameters.

Dynamic Correlation Models II

- An obvious drawback of the prior model is that θ_1 and θ_2 are scalar so that all the conditional correlations have the same dynamics. This might be hard to justify in real applications, especially when the dimension k is large.
- One may extend the previous DCC models to allow the marginal volatility models have **leverage effects**

$$D_t^2 = \Lambda_0 + \Lambda_1 D_{t-1}^2 + \Lambda_2 A_{t-1}^2 + \Lambda_3 L_{t-1}^2$$

  where where D_t is the diagonal matrix of volatilities.
- The leverage effects are from $L_{i,t-1}$

$$L_{i,t-1} = \begin{cases} a_{i,t-1} & a_{i,t-1} < 0 \\ 0 & \text{elsewhere} \end{cases}$$

- To relax the Gaussian assumption will also bring flexibility in the volitility model.
- **R example**

Suggested Reading

- Tsay (2010) **Chapter 10**
- Tsay (2014) **Chapter 7**