A brief review of univariate volatility models. Reference: Tsay (2010, chapter 3)

We decompose a financial return series as

\[ r_t = \mu_t + a_t = E(r_t | F_{t-1}) + a_t, \quad a_t = \sigma_t \epsilon_t, \quad \epsilon_t \sim iid(0, 1), \]

where \( F_{t-1} \) denotes the information available at time \( t - 1 \). The conditional mean \( \mu_t \) typically involves an ARMA or linear regression equation. The volatility \( \sigma_t \) is time-varying.

**Two** general approaches are commonly used to model \( \sigma_t \):

1. Fixed function approach: All GARCH-type models, e.g. GARCH(1,1)

\[ \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \]

where \( \alpha_0 > 0, 0 \leq \alpha_1, \beta_1 \leq 1 \) and \( 0 < \alpha_1 + \beta_1 < 1 \).

2. Stochastic function approach: stochastic volatility models

\[ \ln(\sigma_t^2) = \alpha_0 + \alpha_1 \ln(\sigma_{t-1}^2) + v_t, \quad v_t \sim iid N(0, \sigma_v^2), \]

where \( 0 < \alpha_1 < 1 \) and \( \alpha_0 \) is a real number.

The GARCH-type models include (1) ARCH model, (2) GARCH model, (3) TGARCH model, (4) GARCH-M model, (5) IGARCH models, (6) EGARCH models and (7) asymmetric power ARCH (APARCH) models. EGARCH and TGARCH models are proposed to handle the **leverage** effects of asset return. The TGARCH (or GJR) models are special cases of APARCH models. The GARCH-M model is used to take care of risk premium. Also, the model converges to a stochastic diffusion equation when the time interval between observations approaches zero. In general, other types of model do not have this continuous-time limit.

Many distributions are available for \( \epsilon_t \). They are (a) standard Gaussian (norm), (b) standardized Student-\( t \) (std), (c) standardized generalized error distribution (ged), (d) skew normal (snorm) distribution, (e) skew Student-\( t \) distribution (sstd), and (f) skew generalized error distribution (sged). Probability density functions of these distributions can be found in Tsay (2010, chapter 3).

**Estimation**: conditional maximum likelihood method or quasi maximum likelihood method.

Consider a GARCH(1,1) model. The joint density function of the series is

\[
f(r_1, \ldots, r_T | \theta) = \prod_{t=2}^{T} f(r_t | \theta, F_{t-1}) \times f(r_1 | \theta).
\]
The term $f(r_t | \theta)$ is often omitted. It does not have significant impact on parameter estimates asymptotically under the usual stationarity conditions. Under the normality assumption, we have

$$f(r_t, F_{t-1}) = \frac{1}{\sqrt{2\pi\sigma_t}} \exp \left[ -\frac{1}{2} \frac{(r_t - \mu_t)^2}{\sigma_t^2} \right].$$

If $\epsilon_t$ is NOT normally distributed, but the above normal likelihood function is used to obtain estimates, these estimates are called the quasi-maximum likelihood estimates (QMLE). Conditions for consistency and asymptotic normality of QMLE can be found in Francq and Zakoian (2010, Wiley), a book entitled *GARCH models*.

**R packages** available for univariate volatility modeling

1. fGarch of Rmetrics
2. rugarch:
3. stochvol: univariate stochastic volatility models

You may see my lecture notes (Lectures 4 and 5) for BUs41202 about demonstrations.

**Multivariate Volatility Models**

Basic structure:

$$z_t = \mu_t + \alpha_t, \quad \alpha_t = \Sigma_t^{1/2} \epsilon_t, \quad (1)$$

where $\{\epsilon_t\}$ are iid sequence of multivariate random vectors with mean zero and covariance matrix $\text{cov}(\epsilon_t) = I_k$. In many cases, people assume $\epsilon_t$ is either standard multivariate normal random vector or follows a standardized multivariate Student-$t$ distribution with pdf

$$f(\epsilon_t | v) = \frac{\Gamma[(v + k)/2]}{\pi(v - 2)^{k/2} \Gamma(v/2)} [1 + (v - 2)^{-1} \epsilon_t' \epsilon_t]^{-(v+k)/2}, \quad (2)$$

where $\Gamma(.)$ is the Gamma function. In equation (1),

$$\mu_t = E(z_t | \mathcal{F}_{t-1}),$$

is the conditional expectation of $z_t$ given the information available at time $t - 1$. For financial asset returns, $\mu_t$ often assumes a constant vector.

**Two major difficulties**

1. Curse of dimensionality: There are $k(k + 1)/2$ processes in $\Sigma_t$.
2. Positive-definiteness: $\Sigma_t$ must be positive-definitely almost surely for all $t$.

**Testing multivariate conditional heteroscedasticity**

1. Multivariate $Q(m)$ statistics for $\alpha_t^2$ process
2. Univariate $Q(m)$ statistics for transformed series

$$\epsilon_t = a_t'\Sigma^{-1}_a a_t - k.$$ 

3. Rank-based $Q(m)$ statistics: use ranks of the transformed series $\epsilon_t$.

4. Robust multivariate $Q(m)$ statistics: 5% trimming (upper tail) based on $\epsilon_t$.

Simulation results indicate that (1) multivariate $Q(m)$ and univariate $Q(m)$ of $\epsilon_t$ fares poorly in the presence of heavy tails. Rank-based and trimmed $Q(m)$ work reasonably well.

Models available

Some simple models are available in the literature. Here we only discuss selected ones.

- Applications of univariate models
- Exponentially weighted covariance
- Diagonal VEC models [Resulting matrices may not be positive definite.]
- BEKK models
- Dynamic correlation models (a) Tse and Tsui (2002) and (b) Engle (2002)
- Copula-based models
- Recent developments

Applications of univariate models

If the dimension $k$ is small, one can make use of the following identity

$$\text{Cov}(X, Y) = \frac{\text{Var}(X + Y) - \text{Var}(X - Y)}{4}.$$ 

Theoretically, this identity can be used to estimate the time-varying covariance of any pair of asset returns. Therefore, we can estimate the multivariate covariance matrix element-by-element using univariate models and the existing program. However, there is no guarantee that the resulting covariance matrix is positive definite for all $t$.

We demonstrate the method by considering a bivariate example.

Exponentially weighted model

$$\Sigma_t = (1 - \lambda)a_{t-1}a_{t-1}' + \lambda \Sigma_{t-1},$$

where $0 < \lambda < 1$. That is,

$$\Sigma_t = (1 - \lambda) \sum_{i=1}^{\infty} \lambda^{i-1} a_{t-i}a_{t-i}'. $$
Diagonal VEC model
May not be positive definite.
Model elements of $\Sigma_t$ separately
For instance, DVEC(1,1) model

$$
\sigma_{11,t} = c_{11} + \alpha_{11} a_{1,t-1}^2 + \beta_{11} \sigma_{11,t-1}
$$
$$
\sigma_{12,t} = c_{12} + \alpha_{12} a_{1,t-1} a_{2,t-1} + \beta_{11} \sigma_{12,t-1}
$$
$$
\sigma_{22,t} = c_{22} + \alpha_{22} a_{2,t-1}^2 + \beta_{22} \sigma_{22,t-1}
$$

BEKK model Engle and Kroner (1995)
Simple BEKK(1,1) model

$$
\Sigma_t = A_0 A'_0 + A_1 (a_{t-1} a'_{t-1}) A'_1 + B_1 \Sigma_{t-1} B'_1
$$

where $A_0$ is a lower triangular matrix, $A_1$ and $B_1$ are square matrices without restrictions.
Pros: positive definite
Cons: Too many parameters, dynamic relations require further study. Harder to interpret the parameters.

Dynamic correlation models
Write the conditional covariance matrix as

$$
\Sigma_t = D_t R_t D_t
$$

where $D_t = \text{diag}\{\sigma_{1t}, \ldots, \sigma_{kt}\}$ is the diagonal matrix of volatilities of the component series and the diagonal elements of $R_t$ is 1. In other words, $\sigma_{it}^2 = \text{Var}(r_{it} | F_{t-1})$ and $R_t$ is a correlation matrix.
The elements of $D_t$ are often obtained by univariate volatility models.
The dynamic cross-correlation (DCC) models focus on the time-evolution of the off-diagonal elements of $R_t$.

DCC model by Tse and Tsui (2002):

$$
R_t = (1 - \lambda_1 - \lambda_2) R + \lambda_1 \Psi_{t-1} + \lambda_2 R_{t-1},
$$

where $0 \leq \lambda_1, \lambda_2 < 1$ such that $0 \leq \lambda_1 + \lambda_2 < 1$, $R$ is $k \times k$ positive-definite correlation matrix and $\Psi_{t-1}$ is the $k \times k$ sample cross-correlation matrix of some recent asset innovations, e.g., the correlation matrix of $\{u_{t-1}, u_{t-2}, \ldots, u_{t-m}\}$ for some pre-specified positive integer $m$, where $u_t = D_t^{-1} a_t$ and $a_t = r_t - \mu_t$.

DCC model by Engle (2002):

$$
R_t = W_t^{-1} Q_t W_t^{-1},
$$

where $Q_t = [q_{ij,t}]$ is a $k \times k$ positive-definite matrix, $W_t = \text{doag}\{\sqrt{q_{11,t}}, \ldots, \sqrt{q_{kk,t}}\}$ is a normalization matrix, and

$$
Q_t = (1 - \alpha_1 - \alpha_2) \bar{Q} + \alpha_1 u_{t-1} u'_{t-1} + \alpha_2 Q_{t-1},
$$
where $\mathbf{u}_t = D_t^{-1} \mathbf{a}_t$, $Q$ is the sample cross-correlation matrix of $\mathbf{u}_t$, $\alpha_i \geq 0$, and $0 \leq \alpha_1 + \alpha_2 < 1$.

The two DCC models differ in the way by which the cross-correlations are updated. Engle’s model requires normalization because it uses $\mathbf{u}_{t-1}$ in the updating. The Tse and Tsui’s model, on the other hand, uses $m$ innovations in the updating. The choice of $m$ affects the fitted cross-correlations. A larger $m$ provides smoother cross-correlations, but the resulting correlations may not be able to show the impact of a large shock quickly.

**Discussions and some recent research**

1. Dimension reduction
   - Independent component analysis:
   - Dynamic orthogonal components: Matteson and Tsay (2010)
   - Factor models

2. Model simplification: parsimonious models
   - Cholesky decomposition + SV models: Lopes, McCulloch and Tsay
   - Equal dynamic correlation models: Engle and Kelly (2009)
   - Cholesky + penalty: Chang and Tsay (2010, Journal of Statistical Planning and Inference)

Additional references: