# Lecture 11: Multivariate Volatility Modeling

## Outline

## 11.1 Basic Concepts

How do the correlations between asset returns change/evolve over time?

- Focus on two asset return series for ease of demonstration:

$$\boldsymbol{r}_t = \begin{bmatrix} r_{1,t} \\ r_{2,t} \end{bmatrix}$$

- In general: $\boldsymbol{r}_t$ is of dimension $k$.
- Data: $\boldsymbol{r}_1, \ldots, \boldsymbol{r}_T$.
- $F_{t-1}$: the information set up to time $t-1$.
- Partion $\boldsymbol{r}_t$ as:

$$\boldsymbol{r}_t = \boldsymbol{\mu}_t + \boldsymbol{a}_t, \qquad \boldsymbol{a}_t = \Sigma_t^{1/2}\boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t \sim (\boldsymbol{0}, I_k),$$

  where

$$\begin{aligned} \boldsymbol{\mu}_t &= \mathbb{E}(\boldsymbol{r}_t|F_{t-1}) = \begin{bmatrix} \mu_{1,t} \\ \mu_{2,t} \end{bmatrix} = \text{conditional mean} \\ \Sigma_t &= \text{Cov}(\boldsymbol{a}_t|F_{t-1}) = \begin{bmatrix} \sigma_{11,t} & \sigma_{12,t} \\ \sigma_{21,t} & \sigma_{22,t} \end{bmatrix} = \text{conditional covariance} \end{aligned}$$

- Concisely: $\boldsymbol{r}_t|F_{t-1} \sim (\boldsymbol{\mu}_t, \Sigma_t)$.
- Restrictions: none on $\boldsymbol{\mu}_t$, but $\Sigma_t$ must be **symmetric** and **positive definite** (for all $t$):

$$\sigma_{12,t} = \sigma_{21,t}, \quad \sigma_{11,t} > 0, \quad \sigma_{22,t} > 0, \quad \sigma_{11,t}\sigma_{22,t} - \sigma_{12,t}^2 > 0.$$

- Dynamic (conditional) correlation between $\boldsymbol{r}_{1,t}$ and $\boldsymbol{r}_{2,t}$:

$$\rho_{12,t} = \frac{\sigma_{12,t}}{\sqrt{\sigma_{11,t}\sigma_{22,t}}}.$$

- Complications about $\Sigma_t$:
  (i) positive definite requirement is difficult to enforce;
  (ii) computational burden increases rapidly with $k$ since there are $k(k+1)/2$ independent components (curse of dimensionality).
- Additional decompositions:
  - Let $V_t = \text{Diag}(\sqrt{\sigma_{11,t}}, \ldots, \sqrt{\sigma_{kk,t}})$ be the diagonal matrix consisting of just the diagonal elements of $\Sigma_t$. E.g., when $k=2$ we have:

$$V_t = \begin{bmatrix} \sqrt{\sigma_{11,t}} & 0 \\ 0 & \sqrt{\sigma_{22,t}} \end{bmatrix}.$$

  - Let $R_t = \left[\rho_{ij,t}\right]$ be the conditional correlation matrix, so that:

$$\Sigma_t = V_t R_t V_t, \qquad \text{and} \qquad R_t = V_t^{-1}\Sigma_t V_t^{-1}.$$

- Coverage: Ch. 10 of Tsay's AFTS (2010), and Ch. 7 of Tsay's MTS (2014). These both rely on the R package `MTS`.

## 11.2 Simple Models

Starting from the WN residuals:

$$\hat{\boldsymbol{a}}_t = \begin{bmatrix} \hat{a}_{1,t} \\ \vdots \\ \hat{a}_{k,t} \end{bmatrix} = \boldsymbol{r}_t - \hat{\boldsymbol{\mu}}_t, \tag{1}$$

where the conditional mean can be estimated via, e.g., a VAR model, we can fit the following (simple) models to capture the volatility $\Sigma_t$.

**EWMA.** A crude estimate of $\Sigma_t$ is to use the (unconditional) empirical covariance estimate of the innovations:

$$\widehat{\Sigma}_t = \frac{1}{t-1} \sum_{i=1}^{t-1} \boldsymbol{a}_i \boldsymbol{a}_i', \tag{2}$$

which assigns equal weight of $1/(t-1)$ to all observations. The **exponentially weighted moving average (EWMA)** model, improves on this by using a weighted combination of the latest observations and the previous estimate:

$$\widehat{\Sigma}_t = (1-\lambda)\boldsymbol{a}_i \boldsymbol{a}_i' + \lambda \widehat{\Sigma}_{t-1},$$

where $0 < \lambda < 1$. Iterating this leads to the infinite sum representation:

$$\widehat{\Sigma}_t = (1-\lambda) \sum_{i=1}^{\infty} \lambda^{i-1} \boldsymbol{a}_{t-i} \boldsymbol{a}_{t-i}'.$$

In order to estimate $\lambda$, we assume $\boldsymbol{a} = \boldsymbol{r}_t - \boldsymbol{\mu}_t \sim N(\mathbf{0}, \Sigma_t)$, and maximize this Gaussian likelihood (a function of $\lambda$ and the parameters in $\boldsymbol{\mu}_t$.) The function `EWMAvol` can be used for this.

**BEKK.** Introduced by Engle and Kroner (1995); tries to mimic a GARCH$(m, s)$, but the curse of dimensionality prevents its usage beyond the simple $m = 1 = s$ case, so that the BEKK$(1,1)$ model is:

$$\Sigma_t = A_0 A_0' + A_1(\boldsymbol{a}_{t-1}\boldsymbol{a}_{t-1}')A_1' + B_1 \Sigma_{t-1} B_1',$$

where $A_0$ is a lower triangular matrix, and $A_1$ and $B_1$ are unrestricted square matrices.

**Summary Remarks:**

- Pros: positive definite.
- Cons: Many parameters; dynamic relations require further study.
- Estimation: quasi-MLE (QMLE) similar to GARCH; can use function `BEKK11` ($k = 2$ and $k = 3$ dimensional cases only).

## 11.3 Dynamic Conditional Correlation (DCC) Models

These are the most popular models, and involve a three-step process. We start once again from the WN residuals in (1).

**Step 1.** Fit a univariate GARCH to each component residual series $\{\hat{a}_{i,t}\}$ in order to obtain the volatility estimate $\hat{\sigma}_{ii,t}$, and compute the standardized innovations:

$$\hat{\eta}_{i,t} = \hat{a}_{i,t}/\sqrt{\hat{\sigma}_{ii,t}}.$$

As vectors, we then have:

$$\hat{\boldsymbol{\eta}}_t = \widehat{V}_t^{-1}\hat{\boldsymbol{a}}_t, \qquad \text{and} \qquad \hat{\boldsymbol{a}}_t = \widehat{V}_t\hat{\boldsymbol{\eta}}_t.$$

**Step 2.** Use a joint dependence model on $\hat{\boldsymbol{\eta}}_t$ to capture the time evolution of the conditional correlation $R_t$ of $\hat{\boldsymbol{a}}_t$ (and ultimately the evolution of $\Sigma_t = V_t R_t V_t$). Let $\boldsymbol{\theta}$ denote the parameters of this model.

**Step 3.** With $\ell_t(\boldsymbol{\theta})$ the log of the pdf of $\hat{\boldsymbol{\eta}}_t$, estimate $\boldsymbol{\theta}$ by maximizing the (conditional) log-likelihood:

$$\ell(\boldsymbol{\theta}) = \sum_{t=1}^{T} \ell_t(\boldsymbol{\theta}).$$

A sensible choice for the pdf of $\boldsymbol{\eta}_t$ is a multivariate $t_\nu$. The pdf of $\boldsymbol{x} \sim t_\nu(\boldsymbol{\mu}, \Sigma)$, where $\boldsymbol{\mu}$ and $\Sigma$ are location and scale parameters, is proportional to:

$$f_\nu(\boldsymbol{x};\boldsymbol{\mu},\Sigma) \propto |\Sigma|^{-1/2}\left[1 + \frac{(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}{\nu}\right]^{-(k+\nu)/2}.$$

If the degrees of freedom $\nu > 2$, then:

$$\mathbb{E}(\boldsymbol{x}) = \boldsymbol{\mu}, \qquad \text{and} \qquad \mathbb{V}(\boldsymbol{x}) = \frac{\nu}{\nu-2}\Sigma.$$

However, since $\boldsymbol{\eta}_t$ is standardized,

$$\mathbb{E}(\boldsymbol{\eta}_t) = \mathbf{0}, \qquad \text{and} \qquad \mathbb{V}(\boldsymbol{\eta}_t) = R_t,$$

we should use the $t_\nu^*$, a standard $t_\nu$ whose pdf is proportional to:

$$f_\nu^*(\boldsymbol{\eta}_t;\boldsymbol{\mu}_t = \mathbf{0}, \Sigma_t = R_t) \propto |R_t|^{-1/2}\left[1 + \frac{\boldsymbol{\eta}_t' R_t^{-1}\boldsymbol{\eta}_t}{\nu-2}\right]^{-(k+\nu)/2}.$$

For modeling $R_t$, three types of DCC are available in the literature.

(i) **Engle (2002)**. Let $W_t = \mathrm{Diag}(\sqrt{Q_{11,t}}, \ldots, \sqrt{Q_{kk,t}})$ be a diagonal matrix whose elements are taken from those of $Q_t$, defined recursively as:

$$\begin{aligned} Q_t &= (1-\theta_1-\theta_2)R_0 + \theta_1 Q_{t-1} + \theta_2\boldsymbol{\eta}_{t-1}\boldsymbol{\eta}_{t-1}' \\ R_t &= W_t^{-1}Q_tW_t^{-1}, \end{aligned}$$

where $0 \le \theta_i$ and $\theta_1 + \theta_2 < 1$, and $R_0$ is the sample correlation matrix of the $\{\hat{\boldsymbol{\eta}}_t\}$.

(ii) **Tse and Tsui (2002)**. Define $R_t$ recursively as:

$$R_t = (1 - \theta_1 - \theta_2)R_0 + \theta_1 R_{t-1} + \theta_2 \Psi_{t-1},$$

where the $\theta_i$ and $R_0$ are as above, and $\Psi_{t-1}$ is the (local) sample correlation matrix of $\{\boldsymbol{a}_{t-1}, \ldots, \boldsymbol{a}_{t-m}\}$ for a pre-specifed positive integer $m > k$ (acts as a smoothing parameter, larger $m$ give smoother correlations).

(iii) **Van der Weide (2003)**. Decompose $R_t = X_t' X_t$, where $X_t$ is an upper triangular matrix parametrized in terms of sines and cosines of a vector of angles $\boldsymbol{\theta}_t$ evolving via a DCC-type equation:

$$\boldsymbol{\theta}_t = \boldsymbol{\theta}_0 + \lambda_1 \boldsymbol{\theta}_{t-1} + \lambda_2 \boldsymbol{\theta}^*_{t-1},$$

where $\boldsymbol{\theta}^*_{t-1}$ is a local estimate of the angles using data $\{\hat{\boldsymbol{\eta}}_{t-1}, \ldots, \hat{\boldsymbol{\eta}}_{t-m}\}$ for some $m > 1$, the $\lambda_i$ are non-negative numbers satisfying $0 < \lambda_1 + \lambda_2 < 1$, and $\boldsymbol{\theta}_0$ denotes the initial values of the angles.

**Summary Remarks:**

- For the univariate GARCH fits in Step 1: use function `dccPre`.
- DCC models are extremely simple with only two parameters to estimate. The function `dccFit` allows the fitting of models in (i)–(ii), while the model in (iii) can be fitted via `mtCopula` (in the context of copula models to be discussed next).
- Goodness of fit checks tend to reject DCC models. The function `MCHdiag` carries out 4 portmanteau tests to diagnose the fit.

## 11.4 Copula-Based Models

**Copulas:** Tools for modeling the (joint) dependence structure among $k \geq 2$ random variables; from the Latin *copulare* (to connect or join). Thus the copula connects the marginals into a joint distribution.

**Def:** A $k$-dimensional copula is a function:

$$C(u_1, \ldots, u_k) : [0,1]^k \mapsto [0,1],$$

which is a $k$-dimensional cdf with Unif$[0,1]$ marginals: $U_1, \ldots, U_k$.

**Key Idea ($k = 2$):** For any r.v.s $X_i = F_i^{-1}(U_i)$, where $U_i \sim$ Unif$[0,1]$ and $F_i^{-1}$ is the (generalized) inverse cdf of $X_i$, then

$$\begin{aligned} P(X_1 \leq x_1, X_2 \leq x_2) &= P\left(F_1^{-1}(U_1) \leq x_1, F_2^{-1}(U_2) \leq x_2\right) \\ &= P\left(U_1 \leq F_1(x_1), U_2 \leq F_2(x_2)\right) := C(u_1, u_2), \end{aligned}$$

where $u_1 = F_1(x_1)$ and $u_2 = F_2(x_2)$. Thus $C(u_1, u_2)$ describes the dependence structure between $X_1$ and $X_2$, stripped of the marginals, w.r.t. standard uniforms (which comprise the reference distribution). Since $C(u_1, u_2)$ is invariant to the marginals, it can be used to connect any two marginals.

**Sklar's Theorem ($k = 2$ case):** Let $F(x_1, x_2)$ be a 2-dim cdf with marginals $F_1(x_1)$ and $F_2(x_2)$. Then there exists a copula $C(u_1, u_2)$ such that $F(x_1, x_2) = C(F_1(x_1), F_2(x_2))$, and it's unique if $F(x_1, x_2)$ is continuous. Conversely, if $C(\cdot, \cdot)$ is a copula and $F_1$ and $F_2$ are univariate cdf's, then the function $C(F_1(x_1), F_2(x_2))$ is a joint cdf with marginals $F_1$ and $F_2$.

**Note:** The main consequence of Sklar's Theorem is that given $F(x_1, x_2)$, the underlying (implied) copula cdf (which always exists) is:

$$C(u_1, u_2) = F\left(F_1^{-1}(u_1), F_2^{-1}(u_2)\right).$$

The corresponding copula pdf, which only exists if $C(u_1, u_2)$ is differentiable, is:

$$c(u_1, u_2) = \frac{\partial^2 C(u_1, u_2)}{\partial u_1 \partial u_2}.$$

**Bounds on Copulas:** Recalling that for $U \sim$ Unif$[0,1]$, $F(u) = uI(0 \leq u \leq 1)$, we have the following bounds, which can be visualized as surfaces in the unit hypercube, and equated with degrees of independence:

- <u>Independence ($U_1$ and $U_2$ are unrelated):</u> If $U_1$ and $U_2$ are independent, then

$$C(u_1, u_2) = P(U_1 \leq u_1, U_2 \leq u_2) = P(U_1 \leq u_1)P(U_2 \leq u_2) = u_1 u_2.$$

- <u>Comonotonic ($U_1$ and $U_2$ are the same):</u> If $U_1$ and $U_2$ are the same, $U_1 = U_2 \equiv U$, so that:

$$C(u_1, u_2) = P(U_1 \leq u_1, U_2 \leq u_2) = P(U \leq u_1, U \leq u_2) = P(U \leq \min(u_1, u_2)) = \min(u_1, u_2).$$

- <u>Countermonotonic ($U_1$ and $U_2$ are opposite):</u> If $U_1$ and $U_2$ are the opposite, $U_2 = 1 - U_1$, so that:

$$\begin{aligned} C(u_1, u_2) &= P(U_1 \leq u_1, U_2 \leq u_2) = P(U \leq u_1, 1 - U \leq u_2) = P(U \leq u_1, U \geq 1 - u_2) \\ &= \begin{cases} u_1 + u_2 - 1, & u_1 + u_2 > 1, \\ 0, & u_1 + u_2 \leq 1. \end{cases} \end{aligned}$$

Here we explore a copula-based model for the joint dependence of $\hat{\eta}_{i,t}$ in order to capture the time evolution of $R_t$ in Step 2 of §11.3. Denote by $f^*_{\nu,R}(\boldsymbol{x})$ the standard $t_\nu$ pdf as above, which is proportional to:

$$f^*_{\nu,R}(\boldsymbol{x}) \propto |R|^{-1/2} \left[ 1 + \frac{\boldsymbol{x}' R^{-1} \boldsymbol{x}}{\nu - 2} \right]^{-(k+\nu)/2},$$

and let $f^*_\nu(x)/F^*_\nu(x)$ denote the pdf/cdf of a univariate standard $t_\nu$. If the data are $\boldsymbol{x} = (x_1, \ldots, x_k)'$, then the pdf of the (std.) $t$-copula evaluated at $\boldsymbol{u} \in [0,1]^k$ is:

$$c^t_{\nu,R}(\boldsymbol{u}) = \frac{f^*_{\nu,R}(\boldsymbol{x})}{\prod_{i=1}^k f^*_\nu(x_i)}, \qquad \text{where} \quad x_i = F^{*-1}_\nu(u_i).$$

Now maximize this copula likelihood for the standardized innovations, i.e., set $x_i = \hat{\eta}_{i,t}$, and use a model for the time evolution of $R \equiv R_t$. One of the DCC models could be used here, but the function `mtCopula` fits a $t$-copula by parametrizing $R_t$ via angles, and the angles evolve via a DCC-type equation.

**Note:** This $t$-copula likelihood is not multvariate $t$; it's called a meta-$t_\nu$.

**Ex:** The left panels of Figure 1 display lag 1 scatterplots $(Z_{t-1}, Z_t)$, for 10,000 simulated values from $Z_t \sim$ ARCH(1) with $\alpha_0 = 1$ and $\alpha_1 = 0.5$ (top panel), and from $Z_t \sim$ GARCH$(1,1)$ with $\alpha_0 = 1$, $\alpha_1 = 0.1$, and $\beta_1 = 0.8$ (bottom panel). The right panels are nonparametric estimates of the corresponding implied copula densities.

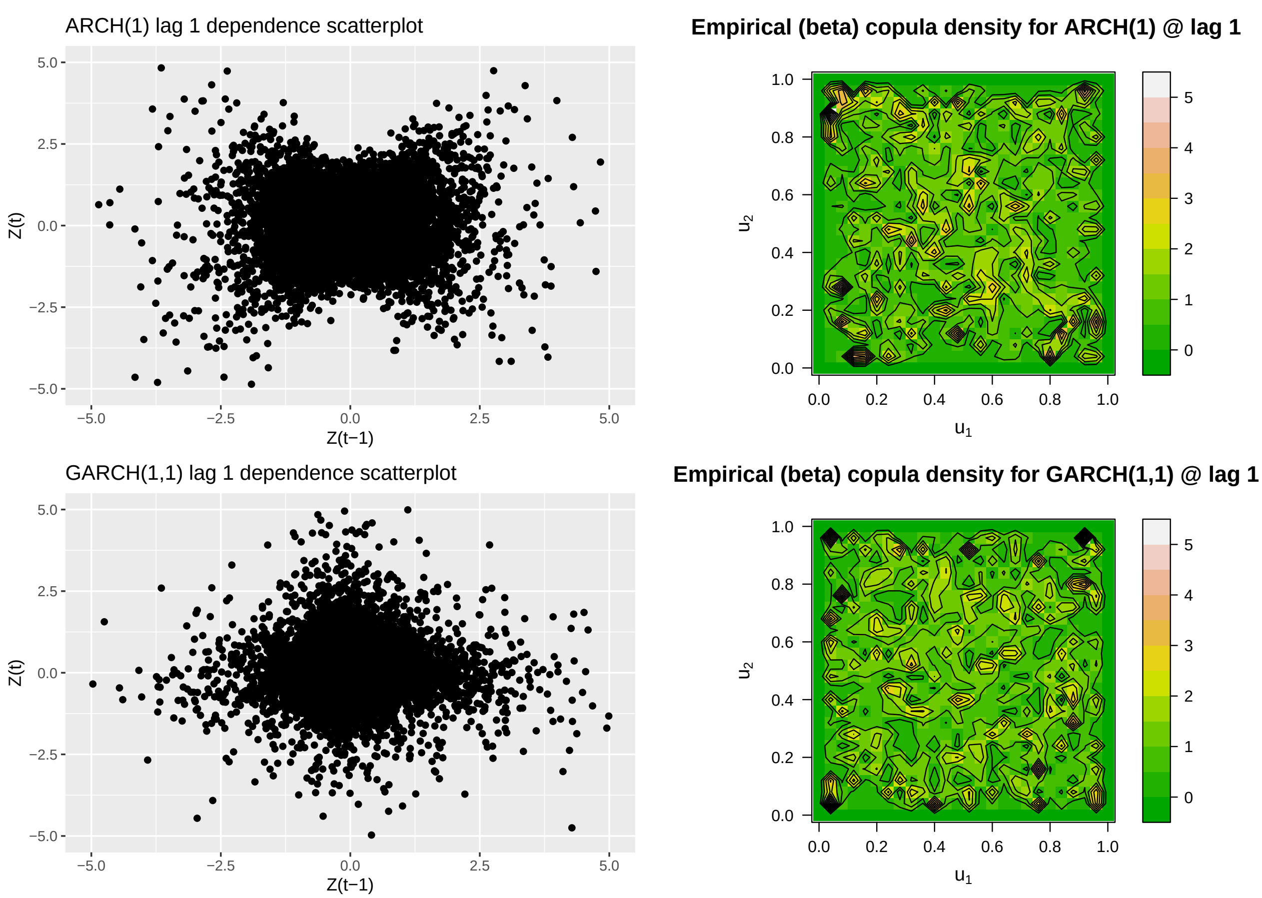


Figure 1: ARCH and GARCH implied copula illustration.

## 11.5 Examples

**Example 1:** (Example 7.7.2 in MTS, Tsay, 2014) Dataset `m-ibmspko-6111.txt` contains the monthly values of IBM stock, S&P composite index, and Coca Cola stock, from January 1961 to December 2011. After converting to log returns, we fit the multivariate $t_\nu$ DCC model of §11.3. (Fig. 2 shows the resulting pairwise $R_t$.)

```
> da=read.table("m-ibmspko-6111.txt",header=T)
> rtn=log(da[,2:4]+1)
> m1=dccPre(rtn,include.mean=T,p=0)
Sample mean of the returns: 0.00772774 0.005023909 0.01059521
Component: 1
Estimates: 0.000419 0.126739  0.788307
se.coef :  0.000162 0.035405  0.055645
t-value :  2.593448 3.57973  14.16662
Component: 2
Estimates: 9e-05    0.127725  0.836053
se.coef :  4.1e-05  0.03084   0.031723
t-value :  2.20126  4.141592 26.35486
Component: 3
Estimates: 0.000256 0.098705  0.830358
se.coef :  8.5e-05  0.022361  0.033441
t-value :  3.015321 4.414112 24.83088

> names(m1)
[1] "marVol" "sresi" "est" "se.coef"

> rtn1=m1$sresi
> Vol=m1$marVol
> m2=dccFit(rtn1)
Estimates: 0.8088086 0.04027318 7.959013
st.errors: 0.1491655 0.02259863 1.135882
t-values:  5.422222  1.782107   7.006898

> names(m2)
[1] "estimates" "Hessian" "rho.t"

> S2.t = m2$rho.t
> m3=dccFit(rtn1,type="Engle")
Estimates: 0.9126634 0.04530917 8.623668
st.errors: 0.0294762 0.01273911 1.332381
t-values: 30.96272   3.556697   6.472376

> S3.t=m3$rho.t
> MCHdiag(rtn1,S2.t)
Test results:
Q(m) of et:
Test and p-value:  20.74262 0.02296152
Rank-based test:
```

```
Test and p-value:  30.20662 0.0007924436
Qk(m) of epsilon_t:
Test and p-value: 132.423   0.002425885
Robust Qk(m):
Test and p-value: 109.9671  0.0750157

> MCHdiag(rtn1,S3.t)
Test results:
Q(m) of et:
Test and p-value:  20.02958 0.02897411
Rank-based test:
Test and p-value:  27.61638 0.002078829
Qk(m) of epsilon_t:
Test and p-value: 131.982   0.002625755
Robust Qk(m):
Test and p-value: 111.353   0.06307334
```

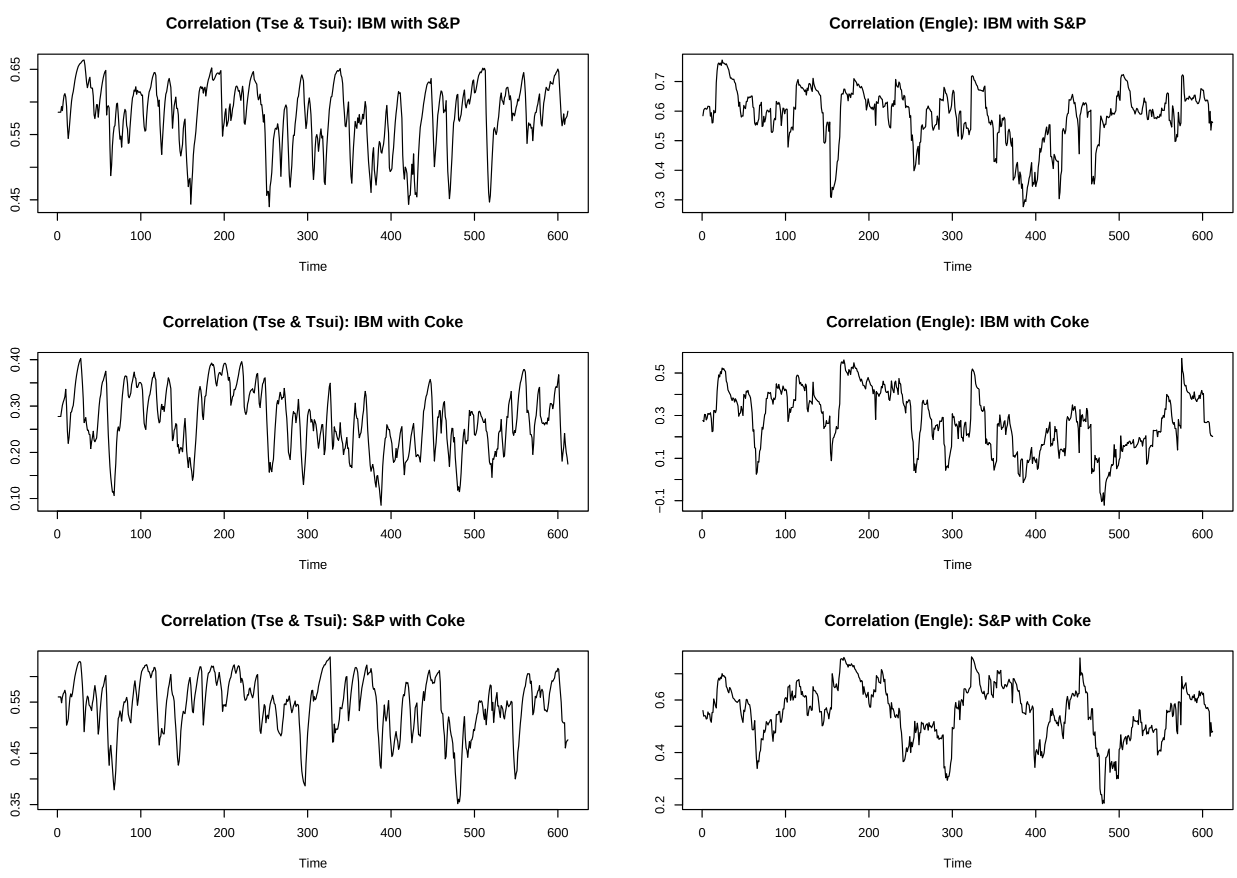


Figure 2: DCC model with $t_\nu$ for the 3 series: IBM, S&P, Coke.

**Example 2:** (Example 7.5 in MTS, Tsay, 2014) Analyzes again `m-ibmspko-6111.txt`: the monthly values of IBM stock, S&P composite index, and Coca Cola stock, from January 1961 to December 2011. Here we fit the meta-$t_\nu$ copula DCC model of §11.4 to the log returns, with the Van der Weide (2003) dynamic correlation structure. (Fig. 3 shows the individual volatility estimates, and pairwise $R_t$.)

```
> da=read.table("m-ibmspko-6111.txt",header=T)
> rtn=log(da[,-1]+1)
> m1=dccPre(rtn,cond.dist="std")
Sample mean of the returns: 0.00772774 0.005023909 0.01059521
Component: 1
Estimates: 0.000388 0.115626  0.805129 9.209269
se.coef :  0.000177 0.036827  0.059471 3.054817
t-value :  2.195398 3.139719 13.5382   3.014671
Component: 2
Estimates: 0.00012  0.130898 0.814531 7.274928
se.coef :  5.7e-05  0.037012 0.046044 1.913331
t-value :  2.102768 3.536655 17.69028 3.802232
Component: 3
Estimates: 0.000216 0.104706  0.837217 7.077138
se.coef :  8.9e-05  0.028107  0.037157 1.847528
t-value :  2.437323 3.725341 22.53208  3.830599

> names(m1)
[1] "marVol" "sresi" "est" "se.coef"

> Vol=m1$marVol; eta=m1$sresi
> m2=mtCopula(eta,0.8,0.04)
Lower limits:  5.1 0.2   1e-04          0.7564334 1.031269 0.8276595
Upper limits: 20   0.95      0.04999999 1.040096  1.417994 1.138032
estimates:        15.38215   0.88189    0.034025  0.919724 1.225322 1.058445
std.errors:        8.222771  0.05117    0.011733  0.041357 0.055476 0.051849
t-values: 1.870677 17.2341 2.899996    22.23883  22.08729 20.41412
Alternative numerical estimates of se:
st.errors:  5.477764  0.051033 0.011714  0.041370 0.055293 0.050793
t-values:   2.808107 17.28091  2.904679 22.23173 22.16072 20.83839

> ### Estimate all parameters (including theta_0)
> names(m2)
[1] "estimates" "Hessian" "rho.t" "theta.t"
> MCHdiag(eta,m2$rho.t)
Test results:
Q(m) of et:
Test and p-value: 19.30177 0.03659304
Rank-based test:
Test and p-value: 27.03262 0.002573576
Qk(m) of epsilon_t:
Test and p-value: 125.9746 0.007387423
Robust Qk(m):
Test and p-value: 107.4675 0.1011374
```

```
> ### Restimate, but fix theta_0 (based on the sample correlations)
> m3=mtCopula(eta,0.8,0.04,include.th0=F)
Value of angles:
[1] 0.9455418 1.2890858 1.0345744
Lower limits:  5.1 0.2  1e-05
Upper limits: 20   0.95 0.0499999
estimates: 14.87427   0.8778   0.03365157
std.errors: 7.959968  0.053013 0.011951
t-values:   1.868635 16.55824  2.815811
Alternative numerical estimates of se:
st.errors: 5.49568  0.0529896 0.01191378
t-values:  2.70654 16.56551   2.824592
```

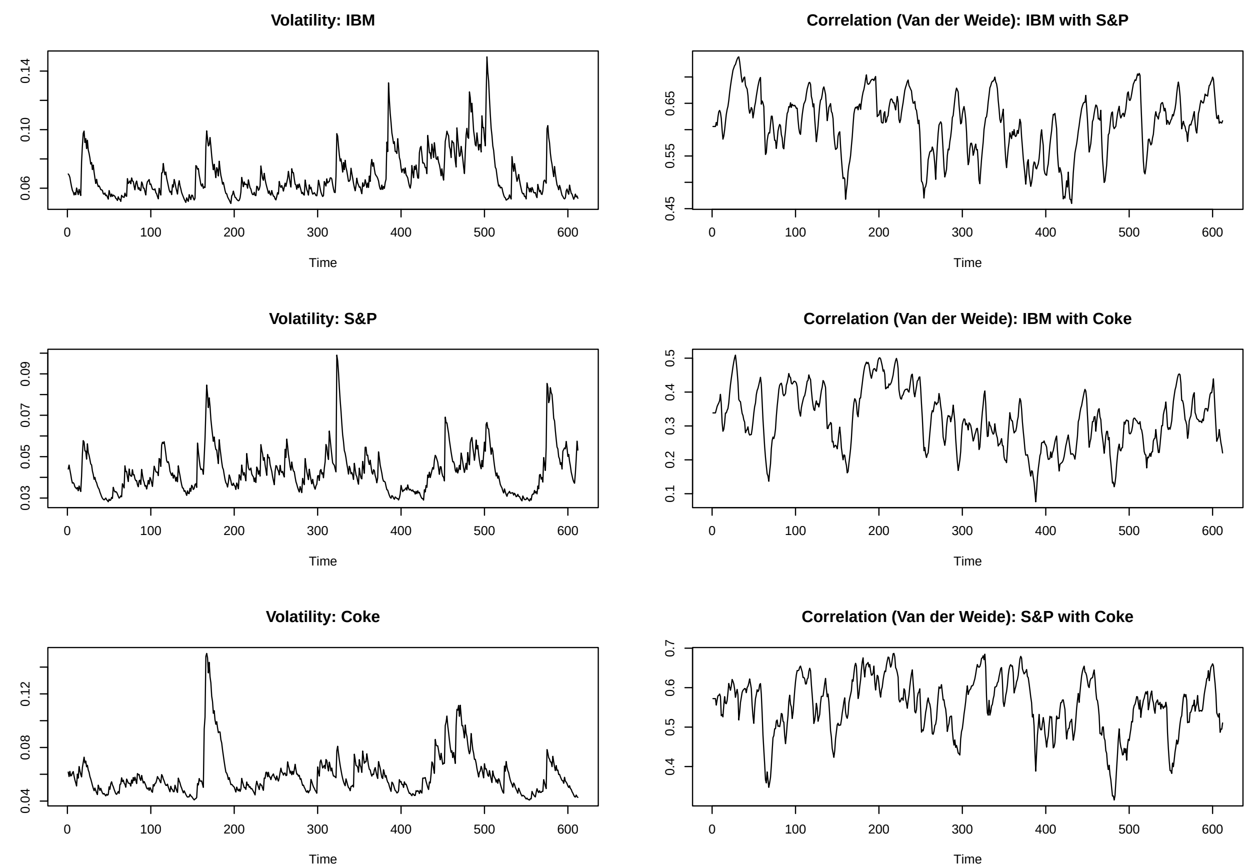


Figure 3: Volatilities and DCC model with meta-$t_\nu$ copula for the 3 series: IBM, S&P, Coke.

## 11.6 VaR and ES via Multivariate Modeling

From our $k$-dim volatility models:

$$\boldsymbol{r}_t = \boldsymbol{\mu}_t + \boldsymbol{a}_t, \qquad \boldsymbol{a}_t = \Sigma_t^{1/2}\boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t \sim (\boldsymbol{0}, I_k),$$

so that conditionally on $F_{t-1}$: $\boldsymbol{r}_t \sim (\boldsymbol{\mu}_t, \Sigma_t)$. Thus, for a portfolio $z_t = \boldsymbol{w}'\boldsymbol{\mu}_t$, where the weights vector $\boldsymbol{w}' = (w_1, \ldots, w_k)$ is composed of deterministic elements, the conditional mean and variance of $z_t$ are:

$$z_t | F_{t-1} \sim (\boldsymbol{w}'\boldsymbol{\mu}_t, \boldsymbol{w}'\Sigma_t\boldsymbol{w}).$$

Most common cases are when $\boldsymbol{r}_t$ is (conditionally) normal or $t$ (both members of the Elliptically Contoured family of distributions).

- Conditionally Normal: If $\boldsymbol{r}_t \sim \mathrm{N}(\boldsymbol{\mu}_t, \Sigma_t)$, then

$$z_t \sim \mathrm{N}(\boldsymbol{w}'\boldsymbol{\mu}_t, \boldsymbol{w}'\Sigma_t\boldsymbol{w}).$$

- Conditionally $t_\nu^*$: If $\boldsymbol{r}_t \sim t_\nu^*(\boldsymbol{\mu}_t, \Sigma_t)$, a standard $t_\nu$ with (conditional) mean and variance $\boldsymbol{\mu}_t$ and $\Sigma_t$ $(\nu \geq 3)$, then

$$z_t \sim t_\nu^*(\boldsymbol{w}'\boldsymbol{\mu}_t, \boldsymbol{w}'\Sigma_t\boldsymbol{w}).$$

**Proofs:** These equations follow from the fact that (conditionally):

$$\boldsymbol{\eta}_t = V_t^{-1}\boldsymbol{a}_t \sim (\boldsymbol{0}, R_t) \quad \Longrightarrow \quad \boldsymbol{a}_t = V_t\boldsymbol{\eta}_t \sim (\boldsymbol{0}, V_tR_tV_t' = \Sigma_t) \quad \Longrightarrow \quad \boldsymbol{r}_t = \boldsymbol{\mu}_t + \boldsymbol{a}_t \sim (\boldsymbol{\mu}_t, \Sigma_t).$$

These results can now be applied using the univariate formulas for VaR and ES from Lecture 9, given forecasts for $\boldsymbol{\mu}_t$ and $\Sigma_t$ (e.g., 1-step ahead). While forecasts of $\boldsymbol{\mu}_t$ are straightforward (e.g., from a fitted VAR), the forecasts for $\Sigma_t$ are more tricky. . . .

For the Tse and Tsui (2002) model, the forecasts of $R_t$ are similar to AR/ARCH:

$$R_t = (1-\theta_1-\theta_2)R_0 + \theta_1 R_{t-1} + \theta_2\Psi_{t-1} \quad \Longrightarrow \quad R_{t+1} = (1-\theta_1-\theta_2)R_0 + \theta_1 R_t + \theta_2\Psi_t.$$

Since $\{R_0, R_t, \Psi_t\} \in F_t$, the best 1-step predictor follows the model:

$$\hat{R}_t(1) = \mathbb{E}(R_{t+1}|F_t) = (1-\theta_1-\theta_2)R_0 + \theta_1 R_t + \theta_2\Psi_t,$$

and similarly for multi-step predictors.