3. Multivariate Volatility models

Consider a $k$ component multivariate return series $r_t = (r_{1t}, \ldots, r_{kt})'$, where the prime denotes transpose.

As in the univariate case, let

$$r_t = \mu_t + u_t,$$

where $\mu_t = \mathbb{E}[r_t|\mathcal{F}_{t-1}]$ is the conditional expectation of $r_t$ given the past information $\mathcal{F}_{t-1}$.

We assume that $\mu_t$ has a vector AR representation.

The conditional covariance matrix of $u_t$

$$\text{Cov}[u_t|\mathcal{F}_{t-1}] = \Sigma_t$$

is a $k \times k$ matrix, assumed positive definite.

There are many ways to generalize the univariate models to multivariate.

The dimensionality rapidly grows to unmanageable magnitudes because there are $k(k + 1)/2$ variance and covariance (or correlation) parameters in $\Sigma_t$ to model.
3.1 Simple bivariate GARCH(1,1) model

Consider the bivariate covariance matrix

\[ \Sigma_t = \begin{pmatrix} \sigma_{11,t} & \sigma_{12,t} \\ \sigma_{21,t} & \sigma_{22,t} \end{pmatrix} \]

The vech operator stacks all columns of a matrix into a column vector.

In the covariance matrix we account for only the distinct elements, such that

\[ \text{vech}(\Sigma_t) = \begin{pmatrix} \sigma_{11,t}^2 \\ \sigma_{21,t}^2 \\ \sigma_{22,t}^2 \end{pmatrix} \]

where \( \sigma_{ii,t}^2 = \sigma_{ii,t}, \ i = 1, 2 \).

The simplest GARCH(1,1) parameterization in a bivariate GARCH is

\[
\begin{align*}
\sigma_{11,t}^2 &= \omega_1 + \alpha_{11} u_{1,t-1}^2 + \beta_{11} \sigma_{11,t-1}^2 \\
\sigma_{22,t}^2 &= \omega_2 + \alpha_{22} u_{2,t-1}^2 + \beta_{22} \sigma_{22,t-1}^2 \\
\sigma_{12,t} &= \omega_1 + \alpha_{12} u_{1,t-1} u_{2,t-1} + \beta_{12} \sigma_{12,t-1}
\end{align*}
\]

Estimation:

EViews (www.eviews.com),
RATS (www.estima.com),
SAS (www.sas.com),
R (www.r-project.org).
Assuming conditional normality of the $u_t$, the estimation can be accomplished with the ML method.

Let $\theta$ denote the vector of all the estimated parameters. Then the likelihood function is of the form as in the univariate case with

$$
\ell_t(\theta) = -\frac{k}{2} \log(2\pi) - \frac{1}{2} \log(|\Sigma_t|) - \frac{1}{2} u_t' \Sigma_t^{-1} u_t,
$$

(6)

where $u_t = y_t - \mu_t$ with $\mu_t$ modeled (usually) with suitable vector ARMA model.

Example 3.1: Consider the forest and metal industry with the covariance matrix specification (5) and

$$
\begin{align*}
  y_{1,t} &= \mu + u_{1,t} \\
  y_{2,t} &= \mu + u_{2,t}
\end{align*}
$$

(7)
The results are:

GARCH Model - Estimation by BFGS
Convergence in 77 Iterations. Final criterion was 0.0000004 <= 0.0
Usable Observations 2206
Log Likelihood -7844.72136260

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coeff</th>
<th>Std Error</th>
<th>T-Stat</th>
<th>Signif</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean(1)</td>
<td>0.0315998874</td>
<td>0.0325636425</td>
<td>0.97040</td>
<td>0.33184518</td>
</tr>
<tr>
<td>Mean(2)</td>
<td>0.1435593579</td>
<td>0.0230536415</td>
<td>6.22719</td>
<td>0.000000003</td>
</tr>
<tr>
<td>C(1,1)</td>
<td>0.0209424100</td>
<td>0.0073302967</td>
<td>2.85697</td>
<td>0.00427711</td>
</tr>
<tr>
<td>C(2,1)</td>
<td>0.0186451969</td>
<td>0.0053642639</td>
<td>3.47582</td>
<td>0.00050930</td>
</tr>
<tr>
<td>C(2,2)</td>
<td>0.0468864922</td>
<td>0.0101689909</td>
<td>4.61073</td>
<td>0.00000401</td>
</tr>
<tr>
<td>A(1,1)</td>
<td>0.0496336132</td>
<td>0.0078433444</td>
<td>6.32812</td>
<td>0.00000000</td>
</tr>
<tr>
<td>A(2,1)</td>
<td>0.0510392579</td>
<td>0.0078265231</td>
<td>6.52132</td>
<td>0.00000000</td>
</tr>
<tr>
<td>A(2,2)</td>
<td>0.0818086550</td>
<td>0.0120582734</td>
<td>6.78444</td>
<td>0.00000000</td>
</tr>
<tr>
<td>B(1,1)</td>
<td>0.9450330948</td>
<td>0.0085771093</td>
<td>110.18084</td>
<td>0.00000000</td>
</tr>
<tr>
<td>B(2,1)</td>
<td>0.9286998103</td>
<td>0.0110740983</td>
<td>83.86234</td>
<td>0.00000000</td>
</tr>
<tr>
<td>B(2,2)</td>
<td>0.8966851334</td>
<td>0.0149509667</td>
<td>59.90818</td>
<td>0.00000000</td>
</tr>
</tbody>
</table>

The estimated coefficient seem reasonable.

Nevertheless, the model has inherently severe shortcomings.

First, covariance does not include conditional variances, and thus does not (explicitly) model the observed fact that correlation tends to increase as variability increases.

Figure 3.1: Conditional correlation and variance processes of the simple bivariate GARCH given in (5)

The (estimated) conditional correlation process depicted in the Figure is defined as

\[ \hat{\rho}_t = \frac{\hat{\sigma}_{12,t}}{\hat{\sigma}_{1,t}\hat{\sigma}_{2,t}}. \]

As seen from the Figure above, the observed relationship between volatility and correlation is, however, captured in some places by the simple model, although the variances are not explicitly included into the covariance equation.
Second, additional constraints should be imposed on the coefficients to ensure positive definiteness of the covariance matrix.

Third, and perhaps the most serious drawback is that the structure is not invariant with respect to linear combinations.

This is not a problem with returns, but for example in exchange rate markets it makes difference in which currency the variables are denominated (e.g. Euro, Dollar, UK pound, Yen, if one is interested in the relationship of these currencies).

Fourth, the model is not invariant under portfolio aggregation (this a drawback for most ARCH models).

This means that aggregating individual stocks to portfolios does not preserve the volatility structure (as an exercise consider a portfolio of the two stocks with weights $w_1$ and $w_2$).

Before going to the more advanced models, let us briefly look at an even simpler model, Exponentially Weighted Moving Average (EWMA), which sometimes preferred by practitioners [see e.g., RiskMetrics™ (www.riskmetrics.com)]

The EWMA is a special case of the simple GARCH in (5)

$$
\begin{align*}
\sigma^2_{1,t} &= (1 - \lambda)u^2_{1,t-1} + \lambda \sigma^2_{1,t-1} \\
\sigma^2_{2,t} &= (1 - \lambda)u^2_{2,t-1} + \lambda \sigma^2_{2,t-1} \\
\sigma_{12,t} &= (1 - \lambda)u_{1,t-1}u_{2,t-1} + \lambda \sigma_{12,t-1},
\end{align*}
$$

(9)

where the $\lambda$-parameter, called persistence is defined by the user.
A smaller $\lambda$ implies a higher reaction of the volatility to the market information in yesterday’s return.

The range of $\lambda$ is usually between 0.75 (highly reactive) and 0.98 (very persistent but not highly reactive).

An $n$-period moving average of a time series $x_t$ is

\[
\hat{x}_t(n) = \frac{x_{t-1} + \lambda x_{t-2} + \lambda^2 x_{t-3} + \cdots + \lambda^{n-1} x_{t-n}}{1 + \lambda + \lambda^2 + \cdots + \lambda^{n-1}}.
\]

Because $0 < \lambda < 1$, $n \to \infty$ the denominator converges to $1 - \lambda$.

Thus

\[
\lim_{n \to \infty} \hat{x}_t(n) = (1 - \lambda) \sum_{i=1}^{\infty} \lambda^{i-1} x_{t-i}.
\]
For volatility and correlation one first calculates the exponentially weighted variance and covariance estimates

\[
\hat{\sigma}_t^2 = (1 - \lambda) \sum_{i=1}^{\infty} \lambda^{i-1} r_{t-i}^2,
\]

and

\[
\hat{\sigma}_{12,t} = (1 - \lambda) \sum_{i=1}^{\infty} \lambda^{i-1} r_{1,t-i} r_{2,t-i}.
\]

Remark 3.1: It is standard to use in above formulas daily returns, \( r_t \), not the residuals, and not deviations from the mean.

Rewriting (10) and (12) in a recursive form gives expressions (9)

Example 3.2: In the graphs below are volatility estimate series, \( \hat{\sigma}_t = \sqrt{(1 - \lambda)r_{t-1}^2 + \lambda \hat{\sigma}_{t-1}^2} \), and correlation estimate series, \( \hat{\rho}_t = \frac{\hat{\sigma}_{12,t}}{\hat{\sigma}_1 \hat{\sigma}_2} \) with \( \lambda = 0.85 \) and \( \lambda = 0.95 \).

![EWMA series for forest return volatility and correlation between forest and metal returns with \( \lambda = 0.85 \) (blue/green) and \( \lambda = 0.96 \) (pink/red) (lower panel average in green).](image)

As indicated by the figure, the effect of \( \lambda \) on EWMA volatility and correlation can be quite substantial.
Remark 3.2: EWMA is a special case of the IGARCH, Integrated GARCH, which is a GARCH(1,1) model with $\alpha + \beta = 1$

$$\sigma_t^2 = \omega + (1 - \lambda)u_{t-1}^2 + \lambda\sigma_{t-1}^2.$$  

(14)

Note that in these models the unconditional variance is not defined (grows to infinity).

3.2 Multivariate GARCH

BEKK model

Generally an $n$-dimensional vec-model can be given as

$$\text{vech}(\Sigma_t) = W + A \text{vech}(u_{t-1}u_{t-1}') + B \text{vech}(\Sigma_{t-1}).$$  

(15)

which becomes (5) if the coefficient matrices are defined as diagonal matrices.
There are altogether
\[ \frac{n(n + 1)}{2} + 2 \times \left( \frac{n(n + 1)}{2} \right)^2 \]
plus the parameters of the mean equation to be estimated.

Thus the model can be used in the general form only in the case of, say, a two or three variables.

Even in the diagonal case it is hard to ensure the positive definiteness.

Different restrictions may lead substantial differences in the estimated model.

This is why the vech model should be employed with caution.

In order to reduce the number of parameters and guarantee positive definiteness several alternatives have been suggested.
A general parameterization that involves the minimum number of parameters, while imposing no gross equation restrictions, and ensuring positive definiteness is the BEKK model (Baba, Engle, Kraft and Kroner, see Engle and Kroner 1995, *Econometric Theory* 11, 122–150).

The model for \( n \) series is of the form
\[
\Sigma_t = C'C + A'u_{t-1}u_{t-1}'A + B'S_{t-1}B,
\]
where \( C \) an \( n \times n \) triangular matrix, \( B \) and \( A \) are \( n \times n \) matrices. It is clear the (50) is positive definite under fairly general assumptions.

\[
\Sigma_t = C'C + \left( \begin{array}{cc} a_{11} & a_{12} \\ a_{21} & a_{22} \end{array} \right)' \left( \begin{array}{cc} u_{1,t-1}^2 & u_{1,t-1}u_{2,t-1} \\ u_{2,t-1}u_{1,t-1} & u_{2,t-1}^2 \end{array} \right) \left( \begin{array}{cc} a_{11} & a_{12} \\ a_{21} & a_{22} \end{array} \right) + \left( \begin{array}{cc} b_{11} & b_{12} \\ b_{21} & b_{22} \end{array} \right)' \Sigma_{t-1} \left( \begin{array}{cc} b_{11} & b_{12} \\ b_{21} & b_{22} \end{array} \right).
\]

(17)

Example 3.2: A bivariate case

\[
\Sigma_t = C'C + \left( \begin{array}{cc} a_{11} & a_{12} \\ a_{21} & a_{22} \end{array} \right)' \left( \begin{array}{cc} u_{1,t-1}^2 & u_{1,t-1}u_{2,t-1} \\ u_{2,t-1}u_{1,t-1} & u_{2,t-1}^2 \end{array} \right) \left( \begin{array}{cc} a_{11} & a_{12} \\ a_{21} & a_{22} \end{array} \right) + \left( \begin{array}{cc} b_{11} & b_{12} \\ b_{21} & b_{22} \end{array} \right)' \Sigma_{t-1} \left( \begin{array}{cc} b_{11} & b_{12} \\ b_{21} & b_{22} \end{array} \right).
\]

(17)

or, suppressing the time subscripts and the GARCH terms
\[
\sigma_1^2 = c_{11}^* + a_{11}^2u_1^2 + 2a_{11}a_{21}u_1u_2 + a_{21}^2u_2^2 \\
\sigma_{12} = c_{12}^* + a_{11}b_{12}u_1^2 + (a_{21}b_{12} + a_{11}b_{22})u_1u_2 + a_{21}a_{22}u_2^2 \\
\sigma_2^2 = c_{22}^* + a_{12}^2u_1^2 + 2a_{12}a_{22}u_1u_2 + a_{22}^2u_2^2,
\]
where \( c_{ij}^* \) are the relevant elements of the \( C^* = C'C \) matrix.

In the general vecm model, excluding the constant terms, there would be 18 parameters to be estimated, compared to the BEKK which has "only" 8 parameters to be estimated. Generally the number of parameters are of order \( n^2 \) in a system of \( n \) variables.
As indicated by the example the formulas become highly non-linear, and there are usually convergence problems in the estimation procedures even in this bivariate case.

This was also the case with our forest-metal data. Even SAS procedure VARMAX which includes a multivariate GARCH option did not find the solution.

Below is an example from RATS manual.

Several simplifications of the general BEKK model have been suggested to alleviate the computational problems.

(a) The Scalar BEKK

In the scalar BEKK the parameters in $A$ are the same (all series react similarly the market information).

Similarly if the persistence in the correlations and variances are assumed to be identical, the parameters in $B$ are the same.

(b) The diagonal BEKK

Matrices $A$ and $B$ are assumed to be diagonal.
Example 3.3: The BEKK model for monthly SP500 and SP Mid Capital stock index (monthly observation from Jan, 1986 to Dec 1996). A version of RATS code can be found at www.estima.com > Proc/Examples > ARCH/GARCH models > garchmv.prg

Modifying the program and running different $\Sigma$, specifications yield the following criterion function values

<table>
<thead>
<tr>
<th>GARCH Model</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple vec</td>
<td>639.97</td>
<td>682.60</td>
</tr>
<tr>
<td>Full vec</td>
<td>641.49</td>
<td>737.85</td>
</tr>
<tr>
<td>Constant corr</td>
<td>636.18</td>
<td>671.05</td>
</tr>
<tr>
<td>Scalar BEKK</td>
<td>766.67</td>
<td>797.67</td>
</tr>
<tr>
<td>Diagonal BEKK</td>
<td>641.67</td>
<td>688.17</td>
</tr>
<tr>
<td>BEKK</td>
<td>641.49</td>
<td>695.74</td>
</tr>
</tbody>
</table>

The constant correlation model (considered in more detail below) seems to fit best according to both criterion functions.

3.3 Covariance matrices based on univariate GARCH

Even in the diagonal vech there are $3n(n + 1)/2$ parameters to be estimated.

For example in a moderate portfolio problem with 10 securities there would be 165 parameters to be estimated! In the diagonal BEKK there are $n(n + 1)/2 + 2n$ parameters to be estimated.

Even in this case with 10 variables there would be 75 parameters to be estimated.
Thus computational problems become soon overwhelming even in relatively low dimensional systems.

This has cast severe doubt to the practical usefulness of full multivariate GARCH modeling.

However, there are some approximations that allow generating multivariate GARCH matrices by univariate GARCH.

(a) Constant correlation model

Assuming the correlation matrix $R$ time invariant, then we can write

$$\Sigma_t = D_t R D_t,$$

where $D_t$ is a diagonal matrix of GARCH volatilities (standard deviations).

Thus the time varying covariances in $\Sigma_t$ are of the form

$$\sigma_{ij,t} = \rho_{ij} \sigma_i \sigma_j,$$

where $\rho_{ij}$ is the correlation of the series $i$ and $j$. 
Example 3.4: A bivariate constant correlation model is defined as

\[
\begin{align*}
\sigma^2_{1,t} &= \alpha_{10} + \alpha_{11} u^2_{1,t-1} + \beta_{11} \sigma^2_{1,t-1} \\
\sigma^2_{2,t} &= \alpha_{20} + \alpha_{22} u^2_{2,t-1} + \beta_{22} \sigma^2_{2,t-1} \\
\sigma_{12,t} &= \rho \sigma_{1,t} \sigma_{2,t},
\end{align*}
\]

(20)

where \( \rho \) is the contemporaneous correlation of the return series.

With the above specification and constant mean, i.e., \( r_{i,t} = \mu_i + u_{i,t}, \quad i = 1, 2 \), the following estimation results are obtained for the data in the previous example.

---

Results show strong evidence of GARCH in the volatility.
(b) Factor ARCH

In order to reduce the dimensionality factor analysis approach may prove useful.

Consider for simplicity a single factor model,
\begin{equation}
    r_{i,t} = \mu + \gamma_i f_t + u_{i,t},
\end{equation}
where $r_{i,t}$ is the return of an asset $i$, and $f_t$ is the common factor for all assets (may be observable or unobservable).

The underlying assumptions are that
\begin{align*}
    \text{Cov}[f_t, u_{i,t}] & = 0, \forall i, \\
    E[u_{i,t}|f_{t-1}, u_{t-1}] & = 0, \forall i, \\
    \text{Cov}[u_t|f_{t-1}, u_{t-1}] & = \Omega, \\
    E[f_t|f_{t-1}, u_{t-1}] & = 0,
\end{align*}
where $\sigma^2_t = \text{Var}[f_t|f_{t-1}, u_{t-1}]$, $\Omega$ is the time invariant contemporaneous covariance matrix of $u_t = (u_{1,t}, \ldots, u_{n,t})'$, the vector of the residual terms, and $f_{t-1}$ and $u_{t-1}$ denote the historical values of the factor and residuals up to time point $t - 1$. 
The common factor may be observable or latent.

All the time varying volatility in the returns are governed by the volatility process of the single common factor.

The conditional covariance matrix of the asset returns becomes then

$$\Sigma_t = \text{Cov}_{t-1}[r_t] = gg'\sigma_t^2 + \Omega,$$

where $g = (\gamma_1, \ldots, \gamma_n)'$.

Engle, et al.* consider a more general factor ARCH and give an application to model bond yield volatilities.

(c) Orthogonal models

In the orthogonal methods Principal Component Analysis (PCA) is utilized to orthogonalize the original returns.

Let

$$r_i = \mu_i + w_i1p_1 + \cdots + w_inp_n,$$

where $r_i$ is the return of the share $i = 1, \ldots, n$, $p_j$ is the $j$th zero mean principal component, $j = 1, \ldots, n$. If only, say, $m < n$ first principal components are utilized, then

$$r_i = \mu_i + w_i1p_1 + \cdots + w_m1p_m + \epsilon_i,$$

where the residual $\epsilon_i$ includes the discarded components.

Remark 3.3: PCA solution is usually calculated from the correlation matrix (standardized solution), in which case the weights must be scaled, such that in place of \(w_{ij}\) must be used \(w^*_ij\sigma_i\), where \(w^*_ij\) is the weight of the standardized solution. Note that in the general case \(w_{ij} \neq w^*_ij\sigma_i\)!

In terms of (25) the covariance matrix of the returns becomes

\[
\Sigma = \mathbf{WDW}' + \mathbf{V}_\epsilon, \tag{26}
\]

where \(\mathbf{W} = (w_{ij})\) is the matrix of weights, and \(\mathbf{D} = \text{diag}(\text{Var}(p_1), \ldots, \text{Var}(p_m))\) is the diagonal matrix of variances of principal components, and \(\mathbf{V}_\epsilon\) is the covariance matrix of residuals. In PCA \(\mathbf{V}_\epsilon\) is ignored and the approximation

\[
\Sigma = \mathbf{WDW}', \tag{27}
\]

is used.
Remark 3.4: Formula (25) defines essentially a Factor Analysis (FA) model with the difference that the residual terms in (26) are correlated.

Orthogonal EWMA

In the basic EWMA model given by equations (12) and (13) the $\lambda$ parameter is assumed the same in all equations. This guarantees that the covariance matrix is positive semidefinite (Exercise: prove it).
Utilizing PCA allows use of different weights in the PCA variances, and the matrix $\Sigma$ defined via (27) will still be p.s.d. (Note that it is not p.d. if $m < n$.)

Remark 3.5: The advantage of PCA to generate risk covariance matrix is the reduction in equations. Instead of generating $n(n+1)/2$ variance and covariance equations, only $m$ variance equations are needed.

Usually the criterion of the successfulness of the PCA approach is how well it can reproduce the direct EWMA series.

Example 3.5: Consider daily returns from Nordic stock exchanges of Copenhagen (Den), Helsinki (Fin), Oslo (Nor) and Stockholm (Swe). The sample period is January 2, 1990 to February 4, 2000.

Below are contemporaneous correlations of the daily returns
Contemporaneous daily return correlations, means and standard deviations of Nordic stock exchanges [Jan 2, 1990 to Feb 4, 2000].

<table>
<thead>
<tr>
<th></th>
<th>DEN</th>
<th>FIN</th>
<th>NOR</th>
<th>SWE</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEN</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FIN</td>
<td>0.420</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NOR</td>
<td>0.486</td>
<td>0.452</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>SWE</td>
<td>0.486</td>
<td>0.550</td>
<td>0.580</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Mean 0.0363 0.0923 0.0379 0.0619
Std. Dev. 0.9625 1.4275 1.1457 1.1866
Obs. 2546 2546 2546 2546

Eigen values and eigen vectors computed from the contemporaneous covariance matrix

<table>
<thead>
<tr>
<th></th>
<th>Comp 1</th>
<th>Comp 2</th>
<th>Comp 3</th>
<th>Comp 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eigenvalue</td>
<td>3.619747</td>
<td>0.956054</td>
<td>0.562894</td>
<td>0.543779</td>
</tr>
<tr>
<td>Variance Prop.</td>
<td>0.637002</td>
<td>0.168246</td>
<td>0.099058</td>
<td>0.095694</td>
</tr>
<tr>
<td>Cumulative Prop.</td>
<td>0.637002</td>
<td>0.805248</td>
<td>0.904306</td>
<td>1.000000</td>
</tr>
</tbody>
</table>

Eigenvectors:

<table>
<thead>
<tr>
<th></th>
<th>Vector 1</th>
<th>Vector 2</th>
<th>Vector 3</th>
<th>Vector 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEN</td>
<td>0.334915</td>
<td>0.270409</td>
<td>0.606230</td>
<td>0.668727</td>
</tr>
<tr>
<td>FIN</td>
<td>0.640295</td>
<td>-0.749719</td>
<td>0.114759</td>
<td>-0.121550</td>
</tr>
<tr>
<td>NOR</td>
<td>0.459146</td>
<td>0.538906</td>
<td>0.240941</td>
<td>-0.665480</td>
</tr>
<tr>
<td>SWE</td>
<td>0.516758</td>
<td>0.276646</td>
<td>-0.749175</td>
<td>0.308488</td>
</tr>
</tbody>
</table>

Defining the EWMA variances \( \hat{\text{Var}}(p_i) \) for the principal components \( p_i \) as

\[
(28) \quad \hat{\text{Var}}(p_i) = (1 - \lambda) \hat{\sigma}_{i,t-1}^2 + \lambda \hat{\text{Var}}_{t-1}(p_i)
\]

we get the conditional variances

\[
(29) \quad \hat{\sigma}_{i,t}^2 = \sum_{k=1}^{4} \hat{w}_{ik}^2 \hat{\text{Var}}(p_k),
\]

covariances

\[
(30) \quad \hat{\sigma}_{ij,t} = \sum_{k=1}^{4} \hat{w}_{ik} \hat{w}_{jk} \hat{\text{Var}}(p_k),
\]

and correlations

\[
(31) \quad \hat{\rho}_{ij,t} = \frac{\hat{\sigma}_{ij,t}}{\hat{\sigma}_{i,t} \hat{\sigma}_{j,t}},
\]

\( i, j = 1, \ldots, 4 \).

Below are graphs of EWMA and PCA-EWMA (annualized) volatilities:

Denmark.
EWMA and PCA-EWMA correlations:

Denmark–Finland

Denmark–Norway

Denmark–Sweden
From the above graphs we observe that the four EWMA-PCA volatilities pretty well can reproduce the 10 EWMA volatilities and correlations.
The real computational strength of the PCA-EWMA comes with highly correlated series, where only few principal components are needed.

For example bond yields of different maturities are this kind of data.

Orthogonal GARCH

As was found earlier it is extremely difficult to use multivariate GARCH to generate co-variance matrices even in the case of low dimension.

PCA-GARCH is obtained by replacing above EWMA principal component variances by GARCH variances.

Note further that PCA is just a special case of multi-factor model with orthogonal factors.
Let $\mathbf{D}_t$ denote the diagonal matrix of $m \leq n$ time varying principal component variances, then the time varying covariance matrix $\mathbf{V}_t$ of the original variables is approximated by

$$ \mathbf{V}_t = \mathbf{W} \mathbf{D}_t \mathbf{W}' $$

(32)

where $\mathbf{W}$ is, as earlier, the $n \times m$ matrix of eigenvectors (principal component weights).

Model (32) is called orthogonal GARCH.

Remark 3.6: The representation (32) is always p.s.d.

Remark 3.7: The principal components are only unconditionally uncorrelated (orthogonal), so the assumed conditional orthogonality is just an approximation.

Example 3.6: Below are GARCH(1,1) models for the principal components of the Nordic indices.

<table>
<thead>
<tr>
<th>Dependent Variable: P1</th>
<th>Method: ML - ARCH (Marquardt)</th>
<th>Date: 02/25/03</th>
<th>Sample(adjusted): 2 2547</th>
</tr>
</thead>
<tbody>
<tr>
<td>Included observations: 2546 after adjusting endpoints</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Convergence achieved after 21 iterations</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variance backcast: ON</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variance Equation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coefficient</td>
<td>Std. Error</td>
<td>z-Statistic</td>
<td>Prob.</td>
</tr>
<tr>
<td>C</td>
<td>0.191986</td>
<td>0.019683</td>
<td>9.75395</td>
</tr>
<tr>
<td>ARCH(1)</td>
<td>0.113826</td>
<td>0.011293</td>
<td>10.07962</td>
</tr>
<tr>
<td>GARCH(1)</td>
<td>0.828716</td>
<td>0.015554</td>
<td>53.28138</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.000000</td>
<td>Mean dependent var</td>
<td>1.98E-16</td>
</tr>
<tr>
<td>Adjusted R-squared</td>
<td>-0.000786</td>
<td>S.D. dependent var</td>
<td>1.902937</td>
</tr>
<tr>
<td>Sum squared resid</td>
<td>9215.875</td>
<td>S.E. of regression</td>
<td>1.903685</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>-4952.133</td>
<td>Akaike info criterion</td>
<td>3.892484</td>
</tr>
<tr>
<td>Durbin-Watson stat</td>
<td>1.699886</td>
<td>Schwarz criterion</td>
<td>3.899368</td>
</tr>
</tbody>
</table>
The conditional variances, covariances and correlations again estimated as in (29)–(31).

Below are graphs of the resulted conditional correlations.
It would be important here to compare the estimation results with BEKK and vech model to see how the relatively simple orthogonal GARCH compares with them.

We however skip the comparison and look at some other approaches to analyze large scale systems.

Before that let us summarize the orthogonal approach.

The idea is simply that:

(i) Find a transformation $W$ such that $y_t = W r_t$, and $\text{Cov}[y_t] = D$ is diagonal.

(ii) Estimate GARCH for the components of the $y_t$ vector and compile the conditional variances to the diagonal matrix $D_t$.

(iii) Under the (strong) assumption $\text{Cov}_{t-1}[y_t] = D_t$, we get

$$\Sigma_t = \text{Cov}_{t-1}[r_t] = W^{-1} D_t W^{-1}.$$
3.4. Dynamic Conditional Correlation

In the Dynamic Conditional Correlation (DCC) proposed by Engle (2002) the conditional correlation is parameterized directly.

Let $r_i, i = 1, \ldots, n$ be random variables with zero mean. The conditional correlations are defined as

$$
\rho_{ij,t} = \frac{E_t[r_{ij,t}]}{\sqrt{E_t[r_{i,t}^2]E_t[r_{j,t}^2]}}.
$$

Let $\sigma_{i,t}^2 = E_t[r_{i,t}^2]$, then $z_{i,t} = r_{i,t}/\sigma_{i,t} \sim \text{WN}(0, 1)$, and the correlation can be written as

$$
\rho_{ij,t} = E_t[z_{i,t}z_{j,t}].
$$

Note: Although $z_{i,t}$ are individually i.i.d it does not follow that they are jointly independent! That is why the conditional correlation may depend on the past. However, joint independence implies independence of individual $z_{i,t}$s.

Engle (2002) suggest to estimate the following GARCH process

$$
q_{ij,t} = \bar{\rho}_{ij} + \alpha (z_{i,t-1}z_{j,t-1} - \bar{\rho}_{ij}) + \beta (q_{ij,t-1} - \bar{\rho}_{ij}),
$$

and

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

where $\bar{\rho}_{ij}$ is the unconditional expectation of the cross product $z_{i,t}z_{j,t}$, i.e., unconditional correlations.

Thus the main difference here compared to the above approaches is that the correlations are modeled individually as GARCH processes with common GARCH parameters $\alpha$ and $\beta$ and separate unconditional expectations $\bar{\rho}_{ij}$ of the cross products.

Thus there are altogether $n(n+1)/2 + n + 2$ parameters to be estimated in the model.

Using these and the separately modeled GARCH variance processes the resulting covariance matrix is obtained as

$$
\Sigma_t = D_tR_tD_t
$$

where $D_t$ is the diagonal matrix of conditional standard deviations [c.f. formula (18)], and $R_t$ is the conditional correlation matrix with elements given by (37).

Remark 3.9: The correlations in (37) will give a p.d. correlation matrix, because $Q_t = (q_{ij,t})$ is a weighted average of p.s.d and p.d matrix.

Remark 3.9: $\omega_{ij}$ is the unconditional correlation, and each term in the denominator of (37) has expected value one.

Remark 3.10: Model (36) is mean reverting as long as $\alpha + \beta < 1$ ($\alpha, \beta \geq 0$)

Estimation

A formulation of the DCC model is

$r_t|\mathcal{F}_{t-1} \sim N(0, D_tR_tD_t)$

$D_t = \text{diag}(\sigma_{1,t}, \ldots, \sigma_{n,t})$

(39)$z_t = D_t^{-1}r_t$

$Q_t = (q_{ij,t})$

$R_t = (\text{diag}(Q_t))^{-\frac{1}{2}}Q_t(\text{diag}(Q_t))^{-\frac{1}{2}},$

where $\sigma_{i,t} = \sqrt{\alpha_0,i + \alpha_i,1r_{i,t-1}^2 + \beta_i\sigma_{i,t-1}^2}$, $i = 1, \ldots, n$ are the GARCH(1,1) standard deviations, and

$(\text{diag}(Q_t))^{-\frac{1}{2}} = \text{diag}(1/\sqrt{q_{11,t}}, \ldots, 1/\sqrt{q_{nn,t}}).$

Let $\theta$ denote the parameters in $D_t$, and $\phi$ the parameters in $R_t$, then the log likelihood function is

(40) $\ell(\theta, \phi) = \sum_{t=1}^{T} \ell_t(\theta, \phi),$

where (ignoring the constant term $n \log 2\pi$)

(41)$\ell_t(\theta, \phi) = -\frac{1}{2} (2\log |D_t| + r_tD_t^{-1}R_t^{-1}D_t^{-1}r_t)$.  

In order to simplify the maximization procedure it is accomplished in two steps. For the purpose, rearrange terms in (41) such that

(42) $\ell_t(\theta, \phi) = -\frac{1}{2} (2\log |D_t| + r_tD_t^{-1}R_t^{-1}r_t - z_t'z_t + \log |R_t| + z_t'R_t^{-1}z_t)$
Thus we can decompose the likelihood function into volatility part

\[ \ell_{V,t}(\theta) = -\frac{1}{2} \left( 2 \log |D_t| + r_t'D_t^{-1}D_t^{-1}r_t \right), \]

and correlation part

\[ \ell_{C,t}(\theta,\phi) = -\frac{1}{2} \left( \log |R_t| + z_t'R_t^{-1}z_t - z'_tz_t \right), \]

so that

\[ \ell_t(\theta,\phi) = \ell_{V,t}(\theta) + \ell_{C,t}(\theta,\phi). \]

Engle (2002) suggest a two step procedure, where in the first step the variance part

\[ \ell_V(\theta) = \sum_{t=1}^{T} \ell_{V,t}(\theta) \]

is maximized, and then given the maximizing value \( \hat{\theta} \), the correlation part

\[ \ell_C(\hat{\theta},\phi) = \sum_{t=1}^{T} \ell_{C,t}(\hat{\theta},\phi) \]

is maximized with respect to \( \phi \).

Remark 3.11: Because \( D_t \) is a diagonal matrix, the variance part (46) is the sum of individual GARCH likelihoods

\[ \ell_V(\theta) = -\frac{1}{2} \sum_{t=1}^{T} \sum_{i=1}^{n} \left( \log(2\pi) + \log(\sigma_{i,t}^2) + \frac{r_{i,t}^2}{\sigma_{i,t}^2} \right), \]

which implies that in the first step the GARCH models can be estimated separately to each series.

Remark 3.12: Because \( z'_tz_t \) terms remain unchanged in (44) you can ignore it and simplify

\[ \ell_{C,t}(\theta,\phi) = -\frac{1}{2} \left( \log |R_t| + z'_tR_t^{-1}z_t \right), \]

Example 3.7: Consider the Nordic stock indices. As noted above in the first step we simply estimate separate GARCH(1,1) models for each index return series. Because there is some autocorrelation in the return series, we adopt the following specifications

\[ y_{i,t} = \phi_{i,0} + \phi_{i,1}y_{i,t-1} + u_{i,t} \]

\[ h_{i,t} = \omega_i + \alpha_iu_{i,t-1}^2 + \beta_ih_{i,t-1}, \]

where in this case \( h_{i,t} = \text{Var}_{t-1}[u_{i,t}] \) denotes the conditional variance, \( i = 1,\ldots,4 \) with 1 = Denmark, 2 = Finland, 3 = Norway, and 4 = Sweden.
RATS estimates for the AR and GARCH parameters are as follows

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coeff</th>
<th>Std Error</th>
<th>T-Stat</th>
<th>Signif</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHI0(1)</td>
<td>0.0258</td>
<td>0.0208</td>
<td>1.24</td>
<td>0.21</td>
</tr>
<tr>
<td>PHI0(2)</td>
<td>0.0463</td>
<td>0.0366</td>
<td>1.26</td>
<td>0.21</td>
</tr>
<tr>
<td>PHI0(3)</td>
<td>0.0572</td>
<td>0.0264</td>
<td>2.16</td>
<td>0.03</td>
</tr>
<tr>
<td>PHI0(4)</td>
<td>0.0844</td>
<td>0.0209</td>
<td>4.02</td>
<td>0.00</td>
</tr>
<tr>
<td>PHI1(1)</td>
<td>0.1653</td>
<td>0.0219</td>
<td>7.55</td>
<td>0.00</td>
</tr>
<tr>
<td>PHI1(2)</td>
<td>0.1995</td>
<td>0.0195</td>
<td>10.24</td>
<td>0.00</td>
</tr>
<tr>
<td>PHI1(3)</td>
<td>0.1846</td>
<td>0.0277</td>
<td>6.66</td>
<td>0.00</td>
</tr>
<tr>
<td>PHI1(4)</td>
<td>0.1463</td>
<td>0.0177</td>
<td>8.63</td>
<td>0.00</td>
</tr>
<tr>
<td>OMEGA(1)</td>
<td>0.0813</td>
<td>0.0361</td>
<td>2.25</td>
<td>0.02</td>
</tr>
<tr>
<td>OMEGA(2)</td>
<td>0.0716</td>
<td>0.0280</td>
<td>2.55</td>
<td>0.01</td>
</tr>
<tr>
<td>OMEGA(3)</td>
<td>0.0513</td>
<td>0.0177</td>
<td>2.98</td>
<td>0.01</td>
</tr>
<tr>
<td>OMEGA(4)</td>
<td>0.0574</td>
<td>0.0177</td>
<td>3.25</td>
<td>0.01</td>
</tr>
<tr>
<td>ALPHA(1)</td>
<td>0.1358</td>
<td>0.0355</td>
<td>3.85</td>
<td>0.00</td>
</tr>
<tr>
<td>ALPHA(2)</td>
<td>0.1330</td>
<td>0.0231</td>
<td>5.75</td>
<td>0.00</td>
</tr>
<tr>
<td>ALPHA(3)</td>
<td>0.1640</td>
<td>0.0436</td>
<td>3.78</td>
<td>0.00</td>
</tr>
<tr>
<td>ALPHA(4)</td>
<td>0.1366</td>
<td>0.0234</td>
<td>5.82</td>
<td>0.00</td>
</tr>
<tr>
<td>BETA(1)</td>
<td>0.7759</td>
<td>0.0611</td>
<td>12.68</td>
<td>0.00</td>
</tr>
<tr>
<td>BETA(2)</td>
<td>0.8309</td>
<td>0.0316</td>
<td>26.22</td>
<td>0.00</td>
</tr>
<tr>
<td>BETA(3)</td>
<td>0.8054</td>
<td>0.0505</td>
<td>15.93</td>
<td>0.00</td>
</tr>
<tr>
<td>BETA(4)</td>
<td>0.8245</td>
<td>0.0251</td>
<td>32.84</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Thus e.g., for Finland

\[ y_{2,t} = 0.0463 + 0.1995 y_{2,t-1} + \hat{u}_{2,t} \]

\[ h_{2,t} = 0.0716 + 0.1330 u_{2,t-1}^2 + 0.8310 h_{2,t-1} \]

with standard errors in parentheses.

For the sake of simplicity we use the simplified GARCH(1,1) specification for the correlation part as

\[ q_{ij,t} = \bar{\rho}_{ij} + \alpha (z_{i,t-1} z_{j,t-1} - \bar{\rho}_{ij}) + \beta (q_{ij,t-1} - \bar{\rho}_{ij}) \]

where \( \bar{\rho}_{ij} \) are estimated by the sample contemporaneous correlations. That is, only the GARCH(1,1)-parameters \( \alpha \) and \( \beta \) are estimated, and \( \bar{\rho}_{ij} \) are replaced by the sample contemporaneous correlations of the standardized series \( z_{i,t} \). The correlations are as follows

<table>
<thead>
<tr>
<th></th>
<th>Den</th>
<th>Fin</th>
<th>Nor</th>
<th>Swe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Den</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fin</td>
<td>0.371</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nor</td>
<td>0.461</td>
<td>0.416</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Swe</td>
<td>0.464</td>
<td>0.504</td>
<td>0.542</td>
<td>1</td>
</tr>
</tbody>
</table>

These are generally a little bit lower than the return contemporaneous correlations considered earlier.

The (quasi) ML estimation of the GARCH(1,1) parameters yields \( \hat{\alpha} = 0.0336(0.001703) \) and \( \hat{\beta} = 0.9247(0.004116) \) with standard errors in parentheses. The respective \( t \)-values are 19.7 and 224.7, and thus highly significant.
The estimate of $\beta$ is close to one, which implies that the correlations should be highly persistent. Nevertheless, as indicated by the graphs below there is considerable variation in the correlations. The general tendency is increasing. These can be compared with the earlier unconditional contemporaneous correlations.

**Conditional Correlations**

Denmark, Finland, Norway and Sweden (Jan 1990 - Jan 2000)

Figure. Estimated dynamic conditional correlations.

Bellow is RATS code. If experimented with another data sets only those parts indicated by ‘<<<<<<’ should be changed.

```r
** Estimated Dynamic Conditional Correlation DCC Example**
* Cal(Irregular)
* COMPUTE GSTART=4, GEND = 2547 ;* <<<<<<
all gend
Open Data ind012000pg.xls ;* <<<<<<
data(format=xls,org=cols)
comp nSeries = 4 ;* number of series in the analysis <<<<<<
* Define return series
dec vect[series] y(nSeries) ;* Return series
set y(1) = 100*log(deClose/deClose{1}) ;* <<<<<<
set y(2) = 100*log(fiClose/fiClose{1}) ;* <<<<<<
set y(3) = 100*log(noClose/noClose{1}) ;* <<<<<<
set y(4) = 100*log(swClose/swClose{1}) ;* <<<<<<
* YOU NEED NOT CHANGE THE PROGRAM BELOW THIS
* * Parameters for the regression function*
* *
dec vect phi0(nSeries) phi1(nSeries) ;* AR-parameters
dec vect omega(nSeries) alpha(nSeries) beta(nSeries) ;* GARCH params
dec vect yt(nSeries)
dec vect yt1(nSeries)
dec vect[series] z(nSeries) ;* standardized series
dec vect[series] h(nSeries) ;* GARCH processes
dec vect ht(nSeries)
dec vect ht1(nSeries)
dec frm1[vect] resid ;* AR residual vector functions
dec vect residt1(nSeries)
dec frm1[vect] hf ;* variance vector function
dec symm sigma(nSeries,nSeries)
dec vec zVec(nSeries)
```

*I want to thank Martin Richter from Danskebank for careful examination of an earlier version of the code and helpful suggestions.*
NONLIN(parmset=arParms) phi0 phi1
NONLIN(parmset=garchParms) omega alpha beta
FRML RESID = yt=%xt(y,t),yt1=%xt(y,t-1),yt - phi0 - %diag(phi1)*yt1
frml hf = residt1=%xt(z,t-1),ht1=%xt(h,t-1),$
  residt1=%diag(%outerxx(residt1)),$
  omega + %diag(alpha)*residt1 + %diag(beta)*ht1
frml glogl = %pt(z,t,resid(t)),$
  %pt(h,t,hf(t)),$
  sigma = %diag(%xt(h,t)), zVec=%xt(z,t),$
  %logdensity(sigma,zVec)
*
* Do initial AR regression.
* Copy initial values for regression parameters, and
* Initialize GARCH variance series
  do i=1,nSeries
    LINREG(noPrint) y(i) / z(i)# CONSTANT y(i){1}
    COMPUTE phi0(i) = %BETA(1), phi1(i) = %BETA(2)
    set h(i) = %seesq**2
  end do i
*
* Initialize GARCH estimates
  comp omega = alpha = beta = %mscalar(0.05)
*
* Estimate GARCH parameters
  MAXIMIZE(parmset=arParms+garchparms,$
    METHOD=SIMPLEX,ITERS=10) GLOGL GSTART GEND ;* Improve starting val
  MAXIMIZE(parmset=arParms+garchparms,$
    METHOD=BFGS,ITERS=100,robust) GLOGL GSTART GEND

* CORRELATION PART
* * Standardized variables
  * do i=1,nSeries
    set z(i) = z(i)/sqrt(h(i))
  end do i
  decl symm[series] r(nSeries,nSeries)
  decl symm[series] q(nSeries,nSeries)
  decl symm qMat(nSeries,nSeries) qMat1(nSeries,nSeries)
  decl symm rMat(nSeries,nSeries) zMat1(nSeries,nSeries)
  decl frml[symm] qf
  nonlin(parmset=corrParms) alpha_c beta_c
  * * Define the q-function
    frml qf = (zMat1 = %outerxx(%xt(z,t-1))-rMat),$
      (qMat1 = %xt(q,t-1)-rMat),$
      rMat + alpha_c*zMat1 + beta_c*qMat1
  * * Initialize values
    VCV(MATRIX=rMat,NOPRINT)$
      # z
      * do i=1,nSeries
        do j=1,i
          set q(i,j) = rMat(i,j)
        end do j
      end do i
* * Define correlation part of the log likelihood
  FRML CLOGL = qMat = qf(t), zVec = %xt(z,t), %pt(q,t,qMat),$
    %pt(r,t,%mqform(qMat,inv(%diag(%sqrt(%xdiag(qMat))))),$
    qMat = %xt(r,t),$
    %logdensity(qMat,zVec)
  comp alpha_c = 0.05, beta_c = 0.80
  MAXIMIZE(parmset=corrParms,$
    METHOD=SIMPLEX,ITERS=10) CLOGL GSTART+3 GEND
  MAXIMIZE(parmset=corrParms,$
    METHOD=BFGS,ITERS=100) CLOGL GSTART+3 GEND
For further details regarding DCC, see Engle (2002) and Engle and Sheppard (2001) (Engle's home page).