Multivariate GARCH models

CIDE-Bertinoro-Courses for PhD students - June 2015

Malvina Marchese

Universita' degli Studi di Genova

m.marchese@lse.ac.uk

Suggested readings:

- Bauwens, L., Laurent, S., Rombouts, J.V.K. (2006), *Multivariate GARCH models*, Journal of Applied Econometrics, 21, 79-109.
- Silvennoinen, A, Terasvirta, T. (2008) *Multivariate GARCH models*, in T. G. Andersen, R. A. Davis, J.-P. Kreiss and T. Mikosch, eds. Handbook of Financial Time Series, New York, Springer.

Portfolio returns

- Assume to have a portfolio of n assets. The portfolio return is given by

$$r_t^{(p)} = \sum_{i=1}^{n} w_{t,i} r_{t,i}$$

where $r_{t,i}$ is the (close-to-close) return on the i-th portfolio asset and $w_{i,t}$ is the associated *portfolio weights*. The $w_{i,t}$ are such that

1. $w_{i,t} \geq 0$
2. $\sum_{i=1}^{k} n_{i,t} = 1$

- The first assumption could be relaxed allowing for negative weights (related to short-selling operations).

Portfolio volatility (1)

From standard properties of the variance, it follows that portfolio volatility is given by

$$\sigma^2_{(p),t} = \sum_{i=1}^{n} w^2_{t,i}\sigma_{t,ii} + \sum_{i \neq j} w_{t,i}w_{t,j}\sigma_{t,ij}$$
$$= \sum_{i=1}^{k} w^2_{t,i}\sigma_{t,ii} + \sum_{i \neq j} w_{t,i}w_{t,j}\sqrt{\sigma_{t,ii}\sigma_{t,jj}}\rho_{ij,t}$$

where

$$\sigma_{t,ij} = cov(r_{t,i}, r_{t,j}|I^{t-1}) \quad \text{for } i, j = 1, \ldots, n.$$

and

$$\rho_{ij,t} = \frac{\sigma_{t,ij}}{\sqrt{\sigma_{t,ii}\sigma_{t,jj}}} = corr(r_{t,i}, r_{t,j}|I^{t-1})$$

is the *conditional correlation* between returns on assets i and j.

Portfolio volatility (2)

- In matrix terms, portfolio volatility can be written as

$$\sigma^2_{(p),t} = \mathbf{w}_t^{'} \Sigma_t \mathbf{w}_t$$

where $\mathbf{w}_t = (w_{t,1}, \ldots, w_{t,n})^{'}$ and $\Sigma_t = var(\mathbf{r}_t | I^{t-1})$ is the *conditional variance-covariance matrix* of returns

$$\underline{r}_t = (r_{t,1}, \ldots, r_{t,n})^{'}$$

- In terms of correlations, $\sigma^2_{(p),t}$ can be equivalently expressed as

$$\sigma^2_{(p),t} = \mathbf{w}_t^{'} D_t D_t^{-1} \Sigma_t D_t^{-1} D_t \mathbf{w}_t = \mathbf{w}_t^{'} D_t R_t D_t \mathbf{w}_t$$

where D_t is a $(k \times k)$ diagonal matrix whose i-th diagonal element is given by $\sqrt{\sigma_{t,ii}}$ and R_t is the conditional correlation matrix of $\mathbf{r}_t$.

A general class of multivariate CH models

- The main aim of multivariate CH models is to predict future values of Σ_t.
- A wide class of multivariate CH models can be obtained as a special case of the following general model scheme

$$\mathbf{r}_t = \boldsymbol{\mu}_t + \Sigma_t^{1/2}\mathbf{z}_t = \boldsymbol{\mu}_t + \mathbf{u}_t \quad (1)$$
$$\Sigma_t = \Sigma(I^{t-1}; \theta_\sigma). \quad (2)$$

  where
  1. $\boldsymbol{\mu}_t = E(\mathbf{r}_t | I^{t-1})$
  2. $\mathbf{z}_t \underset{iid}{\sim} (\mathbf{0}, I_k)$
  3. $\Sigma_t^{1/2}$ is any p.d. $(k \times k)$ matrix such that $\Sigma_t^{1/2}(\Sigma_t^{1/2})' = \Sigma_t$
- The model also implies that

$$var(\mathbf{r}_t | I^{t-1}) = var(\mathbf{u}_t | I^{t-1}) = \Sigma_t^{1/2} var(\mathbf{z}_t)(\Sigma_t^{1/2})' = \Sigma_t$$

  where $\mathbf{u}_t = \mathbf{r}_t - \boldsymbol{\mu}_t$. It follows that equation (2) defines the dynamic behaviour of the conditional covariance matrix of returns Σ_t.

Multivariate GARCH (MGARCH) models

- For ease of reference we will denote as *Multivariate GARCH (MGARCH)* the models belonging to the class defined by equations (1-2). Different MGARCH models will be characterized by different specifications of the dynamic equation (2).
- Bauwens, Laurent and Rombouts (2006, JAE) classify MGARCH models into three different categories
  1. direct generalizations of the univariate GARCH model: VECH(Bollerslev, Engle, and Wooldridge, 1988, JPE), BEKK (Engle and Kroner, 1995, ET), RiskMetrics and factor models.
  2. linear combinations of univariate GARCH models: generalized orthogonal models and latent factor models.
  3. nonlinear combinations of univariate GARCH models: Dynamic Conditional Correlation (DCC) models (Engle, 2002, JBES)
- In this course we will focus on categories (1) and (3).

Main issues in MGARCH modelling

- Identify appropriate conditions to be imposed on θ_σ for guaranteeing the PDness of Σ_t.
- In order to make the estimation feasible, we need to find parsimonious parameterization without paying a too high price in terms of flexibility of the dynamics of Σ_t.
- Find $\Sigma = E(\Sigma_t) = Var(\mathbf{u}_t)$ and identify appropriate conditions to be imposed on θ_σ for guarateeing week stationarity of the model and existence of Σ.

For ease of exposition we will focus on 1-lag dynamic models (which is the most widely diffused choice in practical applications).

Financial applications of MGARCH models

MGARCH models have several applications in different fields of finance

- Prediction of VaR and ES (eg Christoffersen, 2008, HoFTS)
- Hedging (eg Storti, 2008, SMA)
- Portfolio Optimization (eg Engle and Colacito, 2006, JBES)
- Option Pricing (eg Rombouts, Stentoft and Violante, 2012, WP)
- Analysis of contagion (eg Billio and Caporin, 2005, SMA) and volatility spillovers (eg Chang and McAleer, 2011,WP)

Focus: VaR prediction via MGARCH models (1)

- VaR prediction is one of the main applications of multivariate GARCH models
- Assume that $\mathbf{w}_t = \mathbf{w}(I^{t-1})$, which is a natural assumption: on a daily scale an hypothetical investor decides the allocation of his portfolio using information available at market closure on the previous day.
- Remind that portfolio returns are given by

$$r_t^{(p)} = \mathbf{w}_t^{'} \boldsymbol{\mu}_t + \mathbf{w}_t^{'} \Sigma_t^{1/2} \mathbf{z}_t$$

- The availability of an analytical expression for VaR depends on the shape of the distribution of $\mathbf{z}_t$

Focus: VaR prediction via MGARCH models (2)

- Normal errors: $\mathbf{z}_t \sim MVN(\mathbf{0}, I_n)$. Since linear transformations of MVN distributions are still normal, we have

$$(r_t^{(p)} | I^{t-1}) \sim MVN(\mathbf{w}_t^{'}\mu_t, \mathbf{w}_t^{'}\Sigma_t\mathbf{w}_t)$$

The one-step-ahead VaR at level $(1-p)$ is then given by:

$$\text{VaR}_{t,p,1} = \mathbf{w}_t^{'}\boldsymbol{\mu}_t + \sqrt{\mathbf{w}_t^{'}\Sigma_t\mathbf{w}_t}N_p$$

where N_p is the order p quantile of a standardized normal distribution.

Focus: VaR prediction via MGARCH models (3)

- Multivariate Student's t errors: $\mathbf{z}_t \sim \mathbf{t}_n(\mathbf{0}, I_n, \nu)$. As for the Normal distribution, linear transformations of MV t distributions are still t with the same number of degrees of freedom. The one-step-ahead VaR at level $(1-p)$ is then given by:

$$\text{VaR}_{t,p,1} = \mathbf{w}_t^{'}\boldsymbol{\mu}_t + \sqrt{\mathbf{w}_t^{'}\Sigma_t \mathbf{w}_t} t^*_{p,\nu}$$

  where $t^*_{p,\nu} = \sqrt{(\nu-2)/\nu} t_{p,\nu}$ is the order p quantile of an univariate standardized Student's t distribution with ν degrees of freedom.

- In general, it is not always possible to derive the exact form of VaR. In this case, and for horizons $k > 1$, simulation techniques should be used, as in the univariate case.

VECH models

- The general VECH model (Bollerslev, Engle and Wooldridge, 1988, JPE) is defined as

$$\mathbf{r}_t = \boldsymbol{\mu}_t + \Sigma_t^{1/2}\mathbf{z}_t = \boldsymbol{\mu}_t + \mathbf{u}_t \quad (3)$$
$$\mathbf{h}_t = c + A\boldsymbol{\eta}_{t-1} + G\mathbf{h}_{t-1} \quad (4)$$

where

$$\mathbf{h_t} = \text{vech}(\Sigma_t)$$
$$\boldsymbol{\eta}_t = \text{vech}(\mathbf{u}_t\mathbf{u}_{t-1}^{'})$$

A and G are $(n(n+1)/2 \times n(n+1)/2)$ parameter matrices and c is a $(n(n+1)/2 \times 1)$ parameter vector

- Heavily parameterized: The total number of parameters is $n(n+1)(n(n+1)+1)/2$ that, for $n=2$ gives 21 parameters, 78 for $n=3$ and 210 for $n=4$!
- Conditions on c, A and G for positive definiteness of Σ_t are difficult to derive.

Diagonal VECH models (1)

- In order to reduce the number of parameters to a tractable number, some simplifying assumptions must be imposed. One solution is to assume that A and G are diagonal matrices reducing the number of parameters to $n(n+5)/2$ (e.g. 7,12,18 for n=2,3,4).
- The *diagonal* VECH model can be also represented in terms of Hadamard products as

$$\Sigma_t = C^{\circ} + A^{\circ} \odot (\mathbf{u}_{t-1}\mathbf{u}_{t-1}^{'}) + G^{\circ} \odot \Sigma_{t-1}$$

  where $C = diag(\text{vech}(C^{\circ}))$, $A = diag(\text{vech}(A^{\circ}))$ and $G = diag(\text{vech}(G^{\circ}))$
- It is easy to show that Σ_t will be PD $\forall t$ if C° is PD while and Σ_0, A° and G° are PSD. The PDness of C° can be easily imposed reparameterizing the model in terms of the Cholesky decompositions of the original system matrices.

Diagonal VECH models (2)

- In order to obtain a more parsimonious model, an alternative strategy is to constrain the parameter matrices to be of rank one.

$$A = \mathbf{aa}^{'} \qquad G = \mathbf{gg}^{'} \qquad C = \mathbf{cc}^{'}$$

with a, b, c being $(n \times 1)$ vectors. In this case, Σ_t will be only PSD unless we impose PDness of C.

- For vast dimensional systems, A° and G° are usually constrained to be given by matrices of ones multiplied by a positive scalar (*scalar* VECH model).

$$A = a \times \mathbf{uu}^{'} \qquad G = g \times \mathbf{uu}^{'}$$

with $u_i = 1$, for $i = 1, \ldots, n$.

Statistical properties of VECH models

- The VECH model in equations (3-4) is covariance stationary if and only if the eigenvalues of (A+G) are in modulus less than one

$$max(|\text{eig}(A+G)|) < 1$$

- The second unconditional moment of a stationary VECH process is

$$\text{vech}\Sigma = (I_{n*} - A - G)^{-1} c$$

where $\Sigma = E(\Sigma_t)$ and $n^* = n(n+1)/2$.

Covariance targeting

- Assume that $\mathbf{r}_t$ is a scalar VECH process.

$$\Sigma_t = C^{\circ} + a(\mathbf{u}_{t-1}\mathbf{u}_{t-1}^{'}) + g\Sigma_{t-1} \tag{5}$$

The above expression for the unconditional covariance matrix can be reformulated as

$$\Sigma = (1-a-g)^{-1}C^{\circ}$$

that can be inverted to give

$$C^{\circ} = (1-a-g)\Sigma \tag{6}$$

- Substituting (6) into (5) allows to further reduce to 2 the number of parameters to be simultaneously estimated in the scalar VECH model (Σ can be estimated as the sample covariance matrix of filtered returns).
- This technique is known as *covariance targeting* and generalizes to the multivariate case the variance targeting of Engle and Mezrich (1996, Risk).

The multivariate RiskMetrics (EWMA) predictor

- The JP Morgan (1996) has proposed a multivariate extension of the univariate RiskMetrics volatility predictor.
- This can be represented as a special case of the scalar VECH model

$$h_t = \lambda \boldsymbol{\eta}_{t-1} + \lambda \mathbf{h}_{t-1}$$

  where $0 \leq \lambda \leq 1$.
- The *decay factor* λ proposed by RiskMetrics is 0.94, for daily data, and 0.97 for monthly data.

Focus: the bivariate VECH(1,1) model

$$
\begin{aligned}
h_{t,11} &= c_1 + a_{11}u_{t-1,1}^2 + a_{12}u_{t-1,1}u_{t-1,2} + a_{13}u_{t-1,2}^2 + \\
&+ g_{11}h_{t-1,11} + g_{12}h_{t-1,12} + g_{13}h_{t-1,22} \\
h_{t,12} &= c_2 + a_{21}u_{t-1,1}^2 + a_{22}u_{t-1,1}u_{t-1,2} + a_{23}u_{t-1,2}^2 + \\
&+ g_{21}h_{t-1,11} + g_{22}h_{t-1,12} + g_{23}h_{t-1,22} \\
h_{t,22} &= c_3 + a_{31}u_{t-1,1}^2 + a_{32}u_{t-1,1}u_{t-1,2} + a_{33}u_{t-1,2}^2 + \\
&+ g_{31}h_{t-1,11} + g_{32}h_{t-1,12} + g_{33}h_{t-1,22}
\end{aligned}
$$

where $h_{t,ij} = cov(r_{t,i}, r_{t,j}|\mathcal{I}_{t-1})$.

BEKK models (1)

- The class of BEKK models was proposed by Engle and Kroner (1995, ET). Differently from VECH models, BEKK models guarantee PDness of Σ_t without imposing costraints on the model parameters.
- In a BEKK(1,1,K) the dynamic updating equation for Σ_t is given by

$$\Sigma_t = C^{'}C + \sum_{k=1}^{K} A_k^{'}(\mathbf{u_{t-1}}\mathbf{u_{t-1}}^{'})A_k + \sum_{k=1}^{K} G_k^{'}(\Sigma_{t-1})G_k$$

  where A_k, G_k are $(n \times n)$ matrices (for $k = 1, \ldots, K$) and C is upper triangular.
- The BEKK model can be shown to be a special case of the general VECH model.

BEKK models (2)

- In practical applications BEKK models with $K = 1$ are usually considered.
- For a BEKK(1,1,1) model the number of parameters is $n(5n+1)/2$ (e.g. 11,24,42 for n=2,3,4). In order to reduce this number the A_1 and G_1 matrices can be constrained to be diagonal or scalar. It is immediate to see that the scalar BEKK coincides with the scalar VECH model.
- The stationarity conditions can be obtained by deriving the equivalent VECH formulation of the model (see Engle and Kroner, 1995)

Focus: the bivariate BEKK(1,1,1) model

The implied models for the conditional variances and covariances are constrained versions of those implied by the VECH(1,1) model

$$
\begin{aligned}
h_{t,11} &= \omega_{11} + a^*_{11} u^2_{t-1,1} + 2a^*_{11} a_{21} u_{t-1,1} u_{t-1,2} + a^*_{21} u^2_{t-1,2} + \\
&+ g^*_{11} h_{t-1,11} + 2g^*_{11} g^*_{21} h_{t-1,12} + g^*_{21} h_{t-1,22} \\
h_{t,12} &= \omega_{21} + (a^*_{11} a^*_{12}) u^2_{t-1,1} + (a^*_{11} a^*_{22} + a^*_{21} a^*_{12}) u_{t-1,1} u_{t-1,2} + \\
&+ (a^*_{22} a^*_{21}) u^2_{t-1,2} + (g^*_{11} g^*_{12}) h_{t-1,11} + (g^*_{11} g^*_{22} + g^*_{21} g^*_{12}) h_{t-1,12} + \\
&+ (g^*_{22} g^*_{21}) h_{t-1,22}
\end{aligned}
$$

where $h_{t,ij} = cov(r_{t,i}, r_{t,j} | \mathcal{I}_{t-1})$; a^*_{ij} and g^*_{ij} are the elements of the A_1 and G_1 matrices in the BEKK model formulation, ω_{ij} are the elements of the $\Omega = C'C$ matrix.

Estimation of BEKK and VECH models

- The estimation of BEKK and VECH model parameters can be performed maximizing the Gaussian QML function

$$\ell_T(\boldsymbol{\theta}_{\boldsymbol{\mu}}, \boldsymbol{\theta}_{\boldsymbol{\sigma}}) = -\frac{1}{2}\sum_{t=1}^{T} log|\Sigma_t| - \frac{1}{2}\sum_{t=1}^{T}(\mathbf{r}_t - \boldsymbol{\mu}_t)'\Sigma_t^{-1}(\mathbf{r}_t - \boldsymbol{\mu}_t) \quad (7)$$

- Alternatively, the model can be estimated by ML under the assumption that the standardized errors $\mathbf{z}_t$ follow a multivariate t distribution with density

$$f(\mathbf{z}_t|\boldsymbol{\theta}_{\boldsymbol{\mu}}, \boldsymbol{\theta}_{\boldsymbol{\sigma}}, \nu) = \frac{\Gamma((\nu+n)/2)}{\Gamma(\nu/2)[\pi(\nu-2)]^{n/2}}[1 + \frac{\mathbf{z}_t'\mathbf{z}_t}{\nu-2}]^{-(n+\nu)/2}$$

- Other distributions have been considered: eg Bauwens and Laurent (2005, JBES) assume a multivariate skewed t distribution for $\mathbf{z}_t$.

Conditional Correlation models

- Conditional correlation models separate the modelling of conditional variances from that of conditional correlations in two different steps
  1. we separately define and estimate n univariate models for the conditional variances
  2. we estimate the conditional correlation matrix.
- This approach has two important advantages
  1. *computational simplicity*: the number of parameters to be simultaneously estimated is reduced since a complex optimization problem is disaggregated into simpler ones
  2. *flexibility*: it allows for more flexible model structure since conditional variances and correlations can be separately modelled.

Constant Conditional Correlation (CCC) models

- In the CCC model (Bollerslev, 1990, RES) the conditional correlation matrix is assumed to be constant. This is equivalent to impose that the conditional covariances are proportional to the product of conditional standard deviations.
- The conditional covariance matrix is modelled as

$$\Sigma_t = D_t R D_t$$

  with $D_t = diag(\sqrt{\sigma_{t,11}}, \ldots, \sqrt{\sigma_{t,nn}})'$ where the conditional variances $\sigma_{t,jj}$ can be generated by any GARCH type model and R is the conditional correlation matrix of returns

$$R = corr(\mathbf{r_t} | I^{t-1}) \qquad \rho_{i,i} = 1, \forall i$$

- It is easy to show that the element of place (i, j) in Σ_t is given by

$$\sigma_{t,ij} = \rho_{i,j} \sqrt{\sigma_{t,ii} \sigma_{t,jj}}$$

Dynamic Conditional Correlation (DCC) models

- The constant conditional correlation assumption is often inadequate. Empirical evidence suggests that the level of conditional correlations is time varying (e.g. a higher correlation is usually detected in high volatility periods).
- DCC models are also based on a two-step model building strategy. Differently from CCC models, the conditional correlation matrix is time varying (R_t) as a function of a vector of unknown parameters

$$R_t = R(I^{t-1}; \theta_c)$$

- Several different versions of the DCC model have been proposed. In this course we focus on
  1. the DCC model proposed by Engle (2002, JBES), DCC-E, and its variants
  2. the DCC model proposed by Tse and Tsui (2002, JBES), DCC-T.

The DCC-E model: general formulation

- The DCC-E(1,1) model is defined by the following set of equations (for simplicity assume $\boldsymbol{\mu}_t = \mathbf{0}$)

$$
\begin{aligned}
H_t &= D_t R_t D_t \\
\sigma_{t,ii} &= \omega_i + \alpha_i r^2_{t-1,i} + \beta_i \sigma_{t-1,ii} \quad i = 1, ..., n \\
D_t &= diag(\sqrt{\sigma_{t,11}}, \ldots, \sqrt{\sigma_{t,n}})' \\
\boldsymbol{\epsilon}_t &= D_t^{-1} \mathbf{r}_t \\
Q_t &= C'C + A \odot \boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t' + B \odot Q_{t-1} \\
R_t &= (diag(Q_t))^{-1/2} Q_t (diag(Q_t))^{-1/2}
\end{aligned}
$$

  where C is upper triangular, A and B are $(n \times n)$ PSD parameter matrices.
- The last equation is needed in order to guarantee that R_t is a well defined correlation matrix.
- The specification of the $\sigma_{t,ii}$ can be easily changed to allow for other different univariate volatility models.

The DCC-E model: scalar formulation

- For vast dimensional systems the general DCC-E model is not feasible due to the high-number of parameters and so it is replaced by a restricted *version* with scalar parameter matrices

$$Q_t = S(1 - a - b) + a\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^{'} + bQ_{t-1} \tag{8}$$

  where $a, b \geq 0$, $a + b < 1$ and S is PD. These restrictions imply that Q_t is PD.

- In order to reduce the number of parameters to be simultaneously estimated, Engle (2002) concentrates out the matrix S setting $S = E(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^{'}) = E(R_t) = \bar{R}$. In practical applications $\bar{R}$ is replaced by the sample covariance matrix of standardized returns $\hat{\boldsymbol{\epsilon}}_t$:

$$\hat{\bar{R}} = (1/T)\sum_{t=1}^{T} \hat{\boldsymbol{\epsilon}}_t\hat{\boldsymbol{\epsilon}}_t^{'}$$

  This technique is called *correlation targeting*.

The DCC-E model: Aielli's critique (1)

- Aielli (2011) shows that $S = E(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^{'})$ if and only if $E(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^{'}) = E(Q_t) = \bar{Q}$
- This equality in general does not hold (except for the constant conditional correlation case). By the law of iterated expectation:

$$\begin{aligned} E(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^{'}) &= E[E(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^{'} | I^{t-1})] \\ &= E(R_t) = E((diag(Q_t))^{-1/2} Q_t (diag(Q_t))^{-1/2}) \neq \bar{Q}. \end{aligned}$$

- This motivates a new variant of the DCC-E model called the *corrected DCC (cDCC)* that is not affected by this bias (although we must remark that the empirical performances of cDCC and DCC-E models are very close).

The DCC-E model: Aielli's critique (2)

- The cDCC model replaces equation (8) by the following recursion

$$Q_t = \bar{Q}(1-a-b) + a\mathbf{e}_t\mathbf{e}_t^{'} + bQ_{t-1}$$

where $\mathbf{e}_t = (diag(Q_t))^{1/2}\boldsymbol{\epsilon}_t$. It is easy to show that

$$E(\mathbf{e}_t\mathbf{e}_t^{'}) = E(\bar{Q})$$

- If the $\mathbf{e}_t$ were observable, $\bar{Q}$ could have been estimated as

$$\hat{\bar{Q}} = (1/T)\sum_{t=1}^{T}\mathbf{e}_t\mathbf{e}_t^{'}$$

but this is not feasible since $\mathbf{e}_t$ is dependent on the unknown parameters a and b. So Aielli's cDCC rules out the possibility of correlation targeting

The DCC-T model

- The DCC-T model was proposed by Tse and Tsui (2002). The main difference with respect to the DCC-E model is that the conditional correlation matrix R_t is directly generated by the linear dynamic equation

$$R_t = (1 - a - b)S + a\Psi_{t-1} + bR_{t-1}$$

where S is a PD matrix with diagonal entries equal to 1, Ψ_{t-1} is the sample correlation matrix of $\boldsymbol{\epsilon}_t$ computed over a moving window of length M

$$\psi_{ij,t-1} = \frac{\sum_{m=1}^{M} \epsilon_{t-m,i}\epsilon_{t-m,j}}{\sqrt{\sum_{m=1}^{M} \epsilon_{t-m,i}^2 \sum_{m=1}^{M} \epsilon_{t-m,j}^2}}$$

where $\epsilon_{t,i} = r_{t,i}/\sqrt{\sigma_{t,ii}}$. A necessary condition for PDness of Ψ_t, and then of R_t, is that $M \geq n$.

- Aielli (2011) shows that, even for this model, the targeting matrix R is not easy to estimate.

Estimation of conditional correlation models (1)

- The estimation of conditional correlation models is based on a two-step procedure
  1. Estimation of univariate conditional variance ($\boldsymbol{\theta}_\sigma$) parameters
  2. Estimation of conditional correlation parameters ($\boldsymbol{\theta}_c$), given first stage estimates of $\boldsymbol{\theta}_\sigma$
- Assuming (for simplicity) $\boldsymbol{\mu}_t = \mathbf{0}$, the log-likelihood function can be decomposed as the sum of a volatility and a correlation component

$$\ell(\mathbf{r}; \boldsymbol{\theta}_\sigma, \boldsymbol{\theta}_c) = \ell_v(\mathbf{r}; \boldsymbol{\theta}_\sigma) + \ell_c(\mathbf{r}; \boldsymbol{\theta}_\sigma, \boldsymbol{\theta}_c)$$

- The volatility part can be written as

$$\begin{aligned}\ell_v(\mathbf{r};\boldsymbol{\theta}_\sigma) &= -\frac{1}{2}\sum_{t=1}^{T}\Big(log(|D_t|^2) + \mathbf{r}_t^{'} D_t^{-2}\mathbf{r}_t\Big) \\ &= -\frac{1}{2}\sum_{i=1}^{n}\sum_{t=1}^{T}\Big(log(\sigma_{t,ii}) + r_{t,i}^2\sigma_{t,ii}^{-1}\Big) \\ &= \sum_{i=1}^{n}\ell_{v,i}(r_i;\theta_{\sigma,i})\end{aligned}$$

which is the sum of the univariate likelihoods of the 1st stage volatility models.

- The correlation part is then given by

$$\ell_c(\mathbf{r}_t;\boldsymbol{\theta}_\sigma,\boldsymbol{\theta}_c) = -\frac{1}{2}\sum_{t=1}^{T}(log|R_t|) + \epsilon_t^{'} R_t^{-1}\epsilon_t - \epsilon_t^{'}\epsilon_t)$$

Estimation of conditional correlation models (2)

- The consistency of 1st ($\hat{\boldsymbol{\theta}}_v$) and 2nd ($\hat{\boldsymbol{\theta}}_c$) stage estimators follows from standard likelihood theory and asymptotic results on two-stage estimation (White, 1994, Theorem 3.10)[1]. Asymptotic normality of $\hat{\boldsymbol{\theta}}_v$ also follows from standard likelihood theory results.
- The whole estimation problem can be represented as a two stage GMM estimation problem. Theorem 6.1 in Newey and McFadden (1994, HoE vol. IV, chap. 36) can be applied in order to prove the asymptotic normality of $\hat{\theta}_c$

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_{c,T} - \boldsymbol{\theta}_{c,0}) \xrightarrow{d} N(\mathbf{0}, V_c)$$

where the analytical expression for V_c can be found in Engle(2002).

[1] *Estimation, Inference and Specification Analysis*, Econometric Society Monographs no. 22, CUP.

Main challenges in multivariate volatility modelling

- *Curse of Dimensionality*: in practical applications the dimension of the portfolio (k) is usually very high and this leads to a very large number of parameters to be estimated (unless severe constraints are imposed on the dynamics of H_t).
- *Model uncertainty*. Several alternative models and approaches are available: is it possible to improve their predictive performance by considering combinations of different models (forecast combinations, model averaging)?
- *Inference* in very large dimensional models, even if the number of parameters is manageable, presents some relevant statistical and computational problems.