Università Commerciale
Luigi Bocconi

Lecture 3: Multivariate GARCH and Conditional Correlation Models

20541– Advanced Quantitative Methods for Asset Pricing and Structuring

Spring 2020

Overview

- Three issues in multivariate modelling of CH covariances
- Naïve models that extend univariate models to covariances
- Full Multivariate GARCH, part 1: VECH GARCH
- Full Multivariate GARCH, part 2: BEKK GARCH
- Constant Conditional Correlations (CCC) models
- Dynamic Conditional Correlations (DCC) models
- Hints to Estimation and Inference

Motivation: Finance is Multivariate

- Most relevant and realistic applications in empirical finance are multivariate, that is, they involve N ⍰ 2 assets/securities/portfolios
  - Collect the N returns in the Nx1 vector, $\boldsymbol{R}_{t+1} \equiv [R^1_{t+1} R^2_{t+1} \ldots R^N_{t+1}]$
  - Then CH is about a matrix of second moments, that is, variances and covariances:

$$Var[\mathbf{R}_{t+1}] \equiv E[(\mathbf{R}_{t+1} - E[\mathbf{R}_{t+1}])(\mathbf{R}_{t+1} - E[\mathbf{R}_{t+1}])']$$

$$= E\left\{\begin{bmatrix} (R^1_{t+1} - E[R^1_{t+1}])^2 & (R^1_{t+1} - E[R^1_{t+1}])(R^2_{t+1} - E[R^2_{t+1}]) & \cdots & (R^1_{t+1} - E[R^1_{t+1}])(R^N_{t+1} - E[R^N_{t+1}]) \\ (R^1_{t+1} - E[R^1_{t+1}])(R^2_{t+1} - E[R^2_{t+1}]) & (R^2_{t+1} - E[R^2_{t+1}])^2 & \cdots & (R^2_{t+1} - E[R^2_{t+1}])(R^N_{t+1} - E[R^N_{t+1}]) \\ \vdots & \vdots & \ddots & \vdots \\ (R^1_{t+1} - E[R^1_{t+1}])(R^N_{t+1} - E[R^N_{t+1}]) & (R^2_{t+1} - E[R^2_{t+1}])(R^N_{t+1} - E[R^N_{t+1}]) & \cdots & (R^N_{t+1} - E[R^N_{t+1}])^2 \end{bmatrix}\right\}$$

  - All variances are collected on the main diagonal (they are N), while all covariances are collected off the main diagonal (they are N(N -1)/2)
  - Because $Cov[R^i_{t+1}, R^j_{t+1}] = Cov[R^j_{t+1}, R^i_{t+1}]$, $Var[\boldsymbol{R}_{t+1}]$ is NxN symmetric
  - See 20135, ptf. choice lectures for examples of applications

Motivation: Finance is Multivariate

- In the same way in which we have developed CH models for $\sigma^2_{t+1|t}$ we now perform the same operation for $Var_t[\boldsymbol{R}_{t+1}]$
  - The models will be multivariate in nature
  - They imply a need to model and forecast conditional covariances

- This goal raises 3 issues:

  ① Because there are N + N(N – 1)/2 = N(N + 1)/2 moments to be estimated, the availability of sufficiently long-time series may be an issue

  - E.g., with 15 assets in a ptf.—as common in asset mgmt—need (*i*) 15 volatility and (*ii*) 15x14/2=105 covariance forecasts, for a total of 120
  - Even when variances and covariances are constant, with 15 series of returns and 120 parameters ⟹ 120/15 = 8 data points per series

Three Issues with Multivariate VarCov Modelling

- Yet, this is a unit saturation ratio! To get a ratio of 20, you need 160 data points per series, i.e., 7+ months of daily data, 3+ years of weekly data, 13+ years of monthly data, 40 years of quarterly data
- These data requirements are moderate, but not negligible for OTC instruments or recently floated stocks in the aftermath of IPOs
- Try the example before with 100 assets… you will need to estimate 100 + 100x99/2=5,050 parameters
- Because the size of $Var[\boldsymbol{R}_{t+1}]$ grows as a function of N^2, the size of the estimation problem and the data requirements grow as a quadratic function of the number of assets (i.e., very quickly)
- To some extent models help in reducing the estimation burden, through multivariate CH models, concerning $Var_t[\boldsymbol{R}_{t+1}]$

Three Issues with Multivariate VarCov Modelling

② But models face a unique problem: unless restrictions are imposed, it is **difficult to guarantee that $Var_t[\boldsymbol{R}_{t+1}]$ be a positive definite (PD) matrix**

- A PD matrix $Var_t[\boldsymbol{R}_{t+1}]$ is one such that for any Nx1 vector (of weights) $\boldsymbol{\omega}$, the ptf. Return variance $\boldsymbol{\omega}'Var_t[\boldsymbol{R}_{t+1}]\boldsymbol{\omega} > 0$
- A PD matrix $\Longrightarrow$ positive variances and all correlations < 1

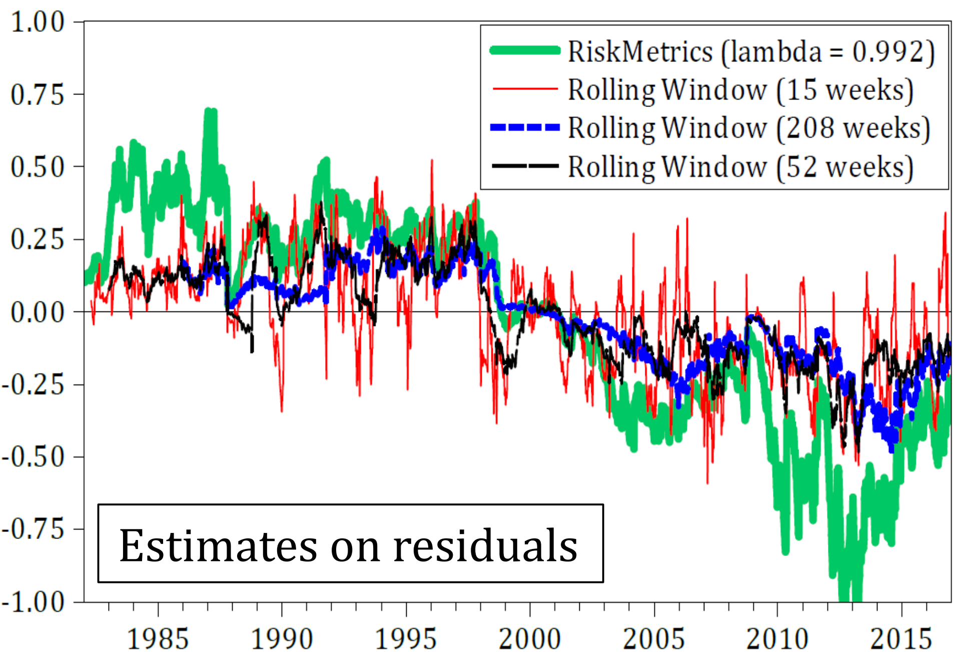
③ Many **CH models themselves are often over-parameterized and characterized by low saturation ratios**

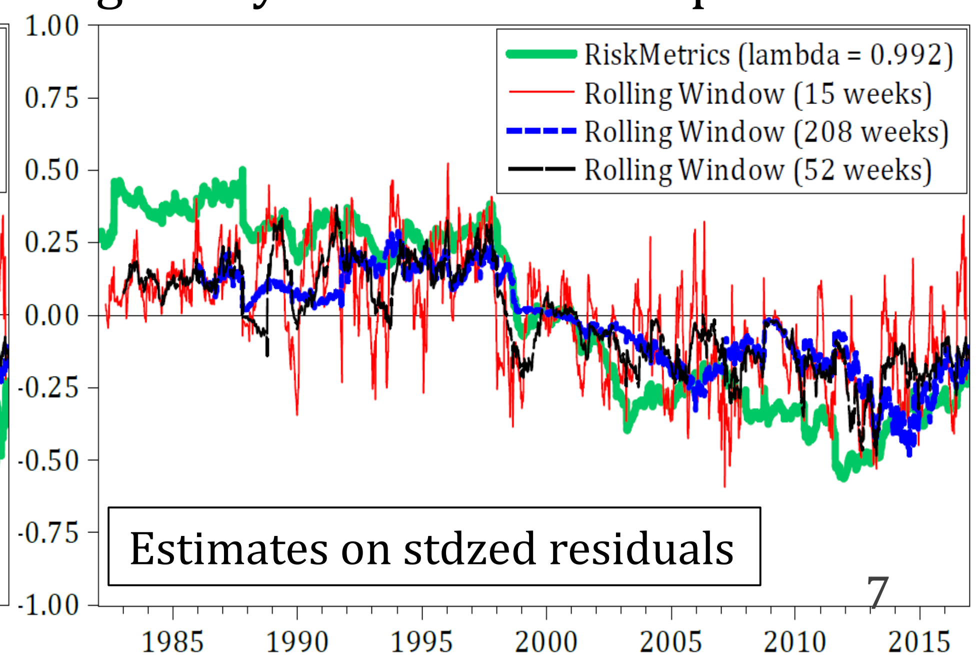
- Models react to 1), however face issue 2) unless they are rich, issue 3)

- Early idea (late 1980s): each element of $Var_t[\boldsymbol{R}_{t+1}]$ can modelled separately and using the same models as in lecture 6
  - For instance, **rolling window moving avg.** $\sigma_{ij,t+1|t}(R) = \frac{1}{R}\sum_{\tau=1}^{R}\varepsilon_{i,t+1-\tau}\varepsilon_{j,t+1-\tau}$ the avg. of the cross-residual products for a pair i ? j
  - Alternative idea that has had some impact on the practice of risk mgmt consists of extending **RiskMetrics**, with ? possibly set at 0.94 $\sigma_{ij,t+1|t} = (1-\lambda)\varepsilon_{i,t}\varepsilon_{j,t} + \lambda\sigma_{ij,t|t-1}$

Simple Models Compared: Stock-Bond Correlation

- Why not GARCH for covariances? $\sigma_{ij,t+1|t} = \omega_{ij} + \alpha_{ij}\varepsilon_{i,t}\varepsilon_{j,t} + \beta_{ij}\sigma_{ij,t|t-1}$
- RiskMetrics is a non-stationary GARCH, unconditional covariance fails to exist, and if tomorrow's covariance is high (low) then it is predicted to remain high (low), rather than reverting back to its mean
- Suppose you are in charge of modelling and forecasting the dynamics of both covariances and correlations between weekly returns on US value-weighted stock index and 10-year Treasury notes
- The sample is January 1982-December 2016
- Claim that stocks and bonds are negatively correlated is simplistic





Simple Models Compared: Stock-Bond Correlation

- In all cases, correlations computed for each asset using predicted standard deviations generated by a GARCH(1,1) $\hat{\rho}_{ij,t+1|t} \equiv \hat{\sigma}_{ij,t+1|t} / (\hat{\sigma}_{ii,t+1|t} \times \hat{\sigma}_{jj,t+1|t})$
- The estimate of λ is higher than the classical 0.94, in excess of 0.99
- Without restrictions (left panel), the RiskMetrics forecast goes below -1 in a few weeks of 2013 ⍰ the resulting $Var_t[\boldsymbol{R}_{t+1}]$ is not PD
- One way to correct problem is to estimate RiskMetrics on stdzed residuals (right panel) + predict correlations in RW case with RW stdev

- The example shows already one case of non-PD predicted $Var_t[\boldsymbol{R}_{t+1}]$
- Unfortunately, unless $\alpha_{ij} = \alpha$ and $\beta_{ij} = \beta$ for all possible pairs *i, j* (ω_{ij} is allowed to depend on the pair *i* and *j*), even though $\sigma_{ij,t+1|t}$ can be anyway predicted, when one organizes such predictions into a covariance matrix $Var_t[\boldsymbol{R}_{t+1}]$, this is not guaranteed to be PD
  - Lecture notes provide a heuristic proof referred to RiskMetrics
  - Note that the restriction applies to both variances and covariances, e.g., in the GARCH(1,1) case:

$$\sigma^2_{i,t+1|t} = \omega_{ii} + \alpha\varepsilon^2_{i,t} + \beta\sigma^2_{i,t|t-1} \quad i = 1, 2, \ldots, N$$
$$\sigma_{ij,t+1|t} = \omega_{ij} + \alpha\varepsilon_{j,t}\varepsilon_{j,t} + \beta\sigma_{ij,t|t-1} \quad i \neq j$$

Stock-Bond Correlation under Restricted GARCH

- There is nothing special about a GARCH(1,1), and this can be extended to more general GARCH(p, q) or RiskMetrics(q) structures
- Unfortunately, the restrictions $\alpha_{ij} = \alpha$ and $\beta_{ij} = \beta$ are often contrary to the evidence in the data
- For the stock-bond case, fitting a **threshold GARCH(1,1)** to capture any asymmetries (also in dynamic covariances), we obtain

$$MKT_{t+1} = \underset{(0.000)}{0.535} + \underset{(0.004)}{0.093} Bond_t + \varepsilon_{MKT,t+1} \qquad \varepsilon_{t+1} \text{ IID N}(0,\ \sigma^2_{MKT,t+1|t})$$

$$\sigma^2_{MKT,t+1|t} = \underset{(0.000)}{0.083} + \underset{(0.000)}{0.065}\varepsilon^2_{MKT,t} + \underset{(0.000)}{0.036} I_{\{\varepsilon_{MKT,t}<0\}}\varepsilon^2_{MKT,t} + \underset{(0.000)}{0.902}\sigma^2_{MKT,t|t-1}$$

$$Bond_{t+1} = \underset{(0.708)}{-0.008} + \underset{(0.000)}{0.231} Bond_t + \varepsilon_{Bond,t+1} \qquad \varepsilon_{t+1} \text{ IID N}(0,\ \sigma^2_{Bond,t+1|t})$$

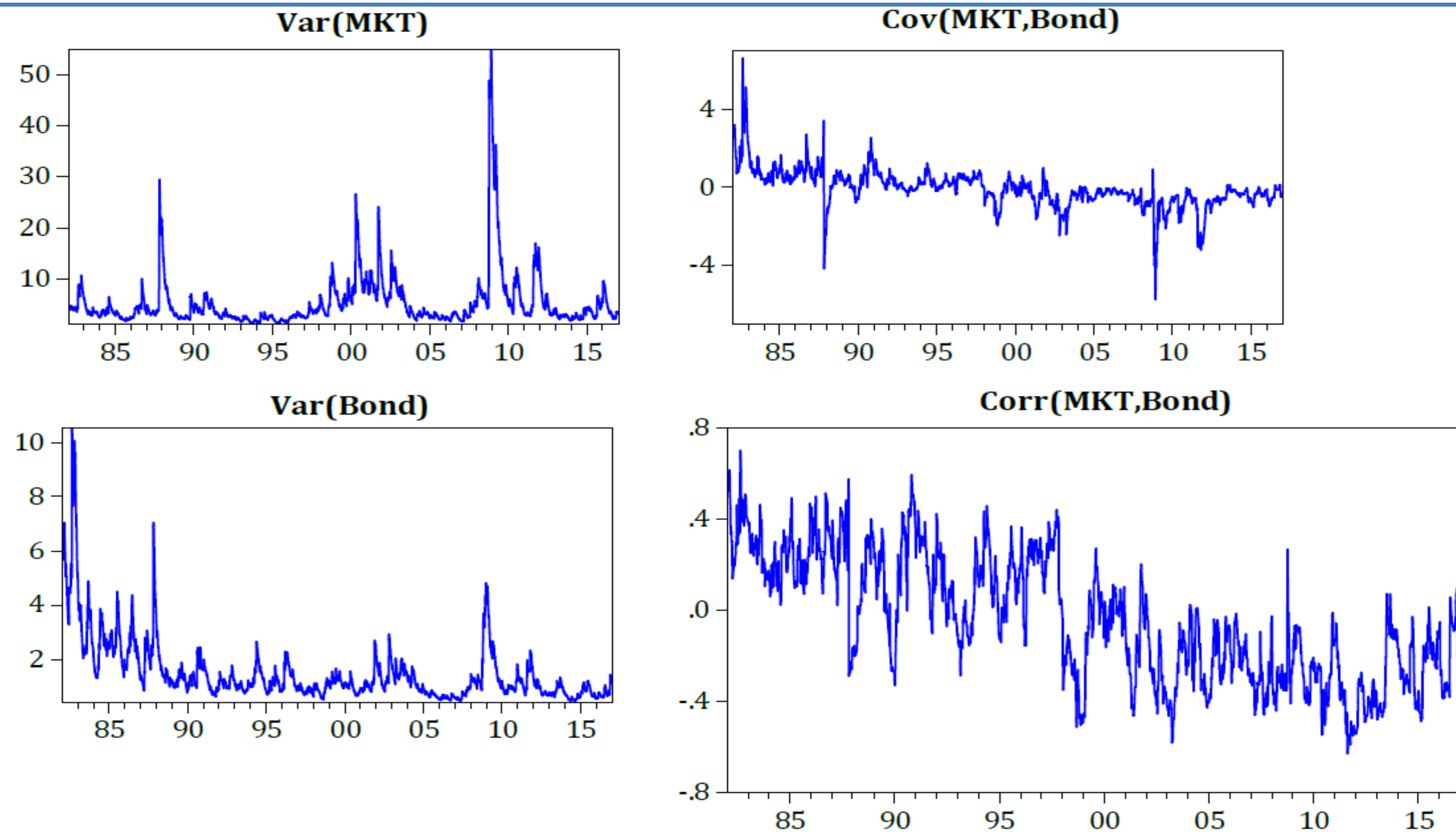
$$\sigma^2_{Bond,t+1|t} = \underset{(0.000)}{0.025} + \underset{(0.000)}{0.065}\varepsilon^2_{Bond,t} + \underset{(0.000)}{0.036} I_{\{\varepsilon_{Bond,t}<0\}}\varepsilon^2_{Bond,t} + \underset{(0.000)}{0.902}\sigma^2_{Bond,t|t-1}$$

$$\sigma_{MKT-Bond,t+1|t} = \underset{(0.000)}{-0.018} + \underset{(0.000)}{0.065}\varepsilon_{MKT,t}\varepsilon_{Bond,t} + \underset{(0.000)}{0.036} I_{\{\varepsilon_{MKT,t}\varepsilon_{Bond,t}<0\}}\varepsilon_{MKT,t}\varepsilon_{Bond,t} + \underset{(0.000)}{0.902}\sigma_{MKT-Bond,t|t-1}$$

Asymmetries

- The common persistence index is 0.985
- The asymmetry term means that when shocks had a different sign one period ago, this increases the negative impact of the current product of shocks on the predicted covariance, a sort of accelerator effect

Stock-Bond Correlation under Restricted GARCH



- Need to impose restrictions paves the way to adoption of complex fully-fledged multivariate GARCH models
- Key concept: as *N* grows, such models become over-parameterized and difficult to estimate, even though they allow PD-restrictions

Fully-Fledged Multivariate GARCH

- In this lecture only a pair of examples to be treated as such
- In a *N*-dimensional generalization, our framework is:

$$\mathbf{R}_{t+1} = \boldsymbol{\mu}_{t+1|t} + \boldsymbol{\varepsilon}_{t+1} = \boldsymbol{\mu}_{t+1|t} + \boldsymbol{\Sigma}_{t+1|t}^{1/2}\mathbf{z}_{t+1} \quad \mathbf{z}_{t+1}\text{IID } MD(\mathbf{0}, \mathbf{I}_N; \boldsymbol{\theta}_D)$$

  - MD indicates a generic multivariate density parameterized by $\boldsymbol{\theta}_D$, such that the multivariate normal or the multivariate t-Student
  - $\boldsymbol{\Sigma}_{t+1|t}^{1/2}$ is the square-root, or Cholesky decomposition, of the covariance matrix, such that $\boldsymbol{\Sigma}_{t+1|t}^{1/2}(\boldsymbol{\Sigma}_{t+1|t}^{1/2})' = \boldsymbol{\Sigma}_{t+1|t} \equiv Var[\mathbf{R}_{t+1}|\Im_t]$.
  - This matrix is in no way the matrix of square roots of elements of the full covariance matrix (how would deal with negative covariances?)
- A first, highly parameterized multi-GARCH is the **VECH-GARCH**:

$$vech(\boldsymbol{\Sigma}_{t+1|t}) = vech(\mathbf{C}) + \mathbf{A}vech(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t') + \mathbf{B}vech(\boldsymbol{\Sigma}_{t|t-1})$$

where *vech*(⍰) ("vector half") converts the unique upper triangular elements of a symmetric matrix into a $0.5N(N-1)$ column vector that removes any duplicates, e.g.:

$$vech\left(\begin{bmatrix} \sigma^2_{1,t+1|t} & \sigma_{12,t+1|t} \\ \sigma_{12,t+1|t} & \sigma^2_{2,t+1|t} \end{bmatrix}\right) = \begin{bmatrix} \sigma^2_{1,t+1|t} \\ \sigma_{12,t+1|t} \\ \sigma^2_{2,t+1|t} \end{bmatrix}$$

VECH GARCH Model

- In an unrestricted VECH model, each element of $\boldsymbol{\Sigma}_{t+1|t}$ is a linear function of the lagged squared errors, the cross-products of past errors for all assets, and the lagged values of the elements of $\boldsymbol{\Sigma}_{t|t-1}$
- **A** and **B** are 0.5N(N+1)-dimensional square matrices, whereas **C** is an NxN symmetric matrix, for a total number of parameters equal to:

$$0.5N(N+1) + 2[0.5N(N+1)]^2 = 0.5N(N+1)[N^2+N+1]$$
$$= 0.5N^4 + N^3 + N^2 + 0.5N = O(N^4)$$

Grows at the same speed as

- E.g., for N = 100, a VECH-GARCH(1,1) model has 51,010,050 parameters to be estimated!
- If you need a saturation ratio of at least 20 $\Longrightarrow$ 20x51,010,050/100 = 10,202,010 obs. per series or a daily history >> 40,484 yrs per series!
- More generally, VECH-GARCH(p, q) models,

$$vech(\boldsymbol{\Sigma}_{t+1|t}) = vech(\mathbf{C}) + \sum_{i=1}^{p} \mathbf{A}_i vech(\boldsymbol{\varepsilon}_{t+1-i}\boldsymbol{\varepsilon}'_{t+1-i}) + \sum_{j=1}^{q} \mathbf{B}_j vech(\boldsymbol{\Sigma}_{t+1-j|t-j})$$

that naively generalize GARCH to the multivariate case, tend to generate a serious curse of dimensionality problem

VECH GARCH Model

- In an unrestricted VECH model, each element of $\boldsymbol{\Sigma}_{t+1|t}$ is a linear function of the lagged squared errors, the cross-products of past errors for all assets, and the lagged values of the elements of $\boldsymbol{\Sigma}_{t|t-1}$
- Tricks—such as covariance targeting—are often invoked to deal with the curse of dimensionality in VECH GARCH:

$$vech(\mathbf{C}_{VT}) = (\mathbf{I}_{0.5N(N+1)} - \mathbf{A} - \mathbf{B})vech\left(\frac{1}{T}\sum_{t=1}^{T} \varepsilon_t \varepsilon'_t\right)$$

- Appropriate restrictions on **A** and **B** $\Longrightarrow$ reduction in the number of parameters, e.g., a diagonal multivariate VECH GARCH, **A** and **B** diagonal
- Algebra shows that then each element of the varcov matrix follows:

$$\sigma_{kl,t+1|t} = \left(1 - \sum_{i=1}^{p} \alpha_{kl,i} - \sum_{j=1}^{q} \beta_{kl,j}\right) \frac{1}{T}\sum_{t=1}^{T} \varepsilon_{k,t}\varepsilon_{l,t} + \sum_{i=1}^{p} \alpha_{kl,i}\varepsilon_{k,t+1-i}\varepsilon_{l,t+1-i} + \sum_{j=1}^{q} \beta_{kl,j}\sigma_{kl,t+1-j|t-j}$$

- But then conditional variances depend only on own lags and own lagged squared residuals, and conditional covariances depend only on own lags and own lagged cross products of residuals

VECH GARCH Model

- This is restrictive because it prevents the detection of any causality in variance, that past shocks to some variable forecast variances of others
- Even the diagonal GARCH framework results in $O(N^2)$ parameters
- Other issue: the coefficients are not restricted to be the same across different assets and pairs $\Longrightarrow$ constraints will have to be imposed
- Appendix A gives a taste for such conditions in a VECH ARCH(1) case
- We now apply VECH-GARCH models to investigate the dynamics of the correlation between stock and bond returns
- We specify a restricted bivariate VAR(1) as a conditional mean model, in which only lagged bond returns forecast both stocks and bonds

Stock-Bond Correlation under VECH GARCH

- Using BIC, a full C diagonal t-Student VECH GARCH(1,1) is preferred:

$$MKT_{t+1} = \underset{(0.000)}{0.574} + \underset{(0.003)}{0.098}\,Bond_t + \varepsilon_{MKT,t+1}$$

$$\sigma^2_{MKT,t+1|t} = \underset{(0.001)}{0.092} + \underset{(0.000)}{0.074}\,\varepsilon^2_{MKT,t} + \underset{(0.000)}{0.906}\,\sigma^2_{MKT,t|t-1},$$

$$Bond_{t+1} = -\underset{(0.888)}{0.003} + \underset{(0.000)}{0.247}\,Bond_t + \varepsilon_{Bond,t+1} \quad [\varepsilon_{MKT,t+1} \;\; \varepsilon_{Bond,t+1}]' \text{ IID } Mt(0, \mathbf{\Sigma}_{t+1|t}; \underset{(0.000)}{9.631})$$

$$\sigma^2_{Bond,t+1|t} = \underset{(0.008)}{0.020} + \underset{(0.000)}{0.062}\,\varepsilon^2_{Bond,t} + \underset{(0.000)}{0.923}\,\sigma^2_{Bond,t|t-1}$$

$$\sigma_{MKT-Bond,t+1|t} = -\underset{(0.895)}{0.0002} + \underset{(0.000)}{0.028}\,\varepsilon_{MKT,t}\varepsilon_{Bond,t} + \underset{(0.000)}{0.963}\,\sigma_{MKT-Bond,t|t-1}.$$

t-Student parameter v

not a problem, uncond. covariance can be <0

- Conditional variances react to shocks much more (0.074 and 0.062) than conditional covariances do (0.028)

- After printing ML estimates and their standard errors, E-Views warns us that "Coefficient matrix is not SPD," and this is a reason for concern