Modelling and Forecasting Conditional Covariances: DCC and Multivariate GARCH

Massimo Guidolin
Dept. of Finance, Bocconi University

1. **Introduction**

In chapter 5 we have made additional progress in our *stepwise distribution modeling* (SDM) approach, i.e.:

1. Establish a variance forecasting model for each of the assets individually and introduce methods for evaluating the performance of these forecasts, which occurred in chapter 4;
2. Consider ways to model conditionally non-normal aspects of the return distribution of the assets in our portfolio—i.e., aspects that are not captured by time series models of conditional means and variances, which has been the focus of chapter 5.

The third and crucial step that we take in this chapter consists in

3. Linking individual variance forecasts with correlations forecasts, possibly by modelling the process of conditional variances themselves.

The simple fact is that most relevant (realistic) applications in empirical finance are actually *multivariate*: they involve $N \geq 2$ assets/securities/portfolios. If you collect returns on such N assets or portfolios in a $N \times 1$ vector $\mathbf{R}_t \equiv [R_t^1 \; R_t^2 \; ... \; R_t^N]'$, then the variance of a random vector turns out to be a matrix of second moments, i.e., variances and covariances:[1]

$$
\begin{aligned}
Var[\mathbf{R}_t] &\equiv Var\text{-}Cov[\mathbf{R}_t] \equiv E[(\mathbf{R}_t - E[\mathbf{R}_t])(\mathbf{R}_t - E[\mathbf{R}_t])'] \\
&= E\left\{ \begin{bmatrix} R_t^1 - E[R_t^1] \\ R_t^2 - E[R_t^2] \\ \vdots \\ R_t^N - E[R_t^N] \end{bmatrix} \begin{matrix} R_t^1 - E[R_t^1] & R_t^2 - E[R_t^2] & \dots & R_t^N - E[R_t^N] \end{matrix} \right\} \\
&= E\left\{ \begin{bmatrix} (R_t^1 - E[R_t^1])^2 & (R_t^1\text{-}E[R_t^1])(R_t^2\text{-}E[R_t^2]) & \dots & (R_t^1\text{-}E[R_t^1])(R_t^N\text{-}E[R_t^N]) \\ (R_t^1\text{-}E[R_t^1])(R_t^2\text{-}E[R_t^2]) & (R_t^2 - E[R_t^2])^2 & \dots & (R_t^2\text{-}E[R_t^2])(R_t^N\text{-}E[R_t^N]) \\ \vdots & \vdots & \ddots & \vdots \\ (R_t^1\text{-}E[R_t^1])(R_t^N\text{-}E[R_t^N]) & (R_t^2\text{-}E[R_t^2])(R_t^N\text{-}E[R_t^N]) & \dots & (R_t^N - E[R_t^N])^2 \end{bmatrix} \right\}
\end{aligned}
$$

[1] It is immaterial whether you want to call this a variance, a covariance, or a variance-covariance matrix. In this chapter we shall express a preference for the second term, *covariance matrix*. Moreover, this definition is easily extended from the unconditional covariance matrix, $Cov[\mathbf{R}_t]$ to the conditional covariance matrix, $Cov_t[\mathbf{R}_{t+1}] \equiv Cov[\mathbf{R}_{t+1}|\Im_t]$.

$$= \begin{bmatrix} Var[R_t^1] & Cov[R_t^1, R_t^2] & \dots & Cov[R_t^1, R_t^N] \\ Cov[R_t^1, R_t^2] & Var[R_t^2] & \dots & Cov[R_t^2, R_t^N] \\ \vdots & \vdots & \ddots & \vdots \\ Cov[R_t^1, R_t^N] & Cov[R_t^2, R_t^N] & \dots & Var[R_t^N] \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{12} & \sigma_2^2 & \dots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1n} & \sigma_{2n} & \dots & \sigma_N^2 \end{bmatrix}$$

Clearly, all variances are collected on the main diagonal, while all covariances are collected off the main diagonal. Moreover, because $Cov[R_t^i, R_t^j] = Cov[R_t^j, R_t^i]$ from simple properties of expectations, then $Cov[\mathbf{R}_t]$ is by construction a symmetric matrix. For instance, any portfolio choice methodology is clearly based on knowledge or estimation of $Cov[\mathbf{R}_t]$, as it is well known that optimal portfolio shares will also depend on the covariance of asset returns considered in pairs. Because risk management concerns either portfolios of securities or portfolios of investment projects, then risk management is also intrinsically a multivariate application. Although in many courses, pricing problems are mostly presented with reference to univariate applications only (i.e., we price one asset at the time, for instance a derivative written on an individual security), in reality this represents more the exception than the rule, as we are often called to price assets that concern several cash flows or underlying securities (think about compound or basket options). Also in this respect, one needs to develop useful *multivariate time series methods* to model and forecast quantities of interest and, among them, surely dynamic covariances and correlations.

In chapters 4 and 5 all of our attention has been directed to developing, estimating, testing, and forecasting univariate ($N = 1$) volatility models only. In this chapter, we broaden our interest to multivariate ($N \geq 2$) models that—as far as second moments are concerned—will necessarily also concern covariances and correlations besides variances. We therefore examine three approaches to multivariate estimation of conditional second moments. First, we deal with an approach that moves the core of the effort from the econometrics to the asset pricing, in the sense that covariances will predicted off factor pricing models (such as, but not exclusively, the CAPM). The advantage of this way of proceeding is that some of us prefer to do more economics and less econometrics (and this seems to be a good idea also to the Author of these notes). Unfortunately, most of the asset pricing theory currently circulating tends to be rejected (sometimes rather obviously, think of the CAPM, in other occasions only marginally) by most data sets. As a result, the majority of users of financial econometrics (risk and asset managers, some quantsy types of asset pricers and structurers) prefer to derive forecasts from econometric models, vs. incorrect, commonly rejected asset pricing models. Second, we propose models that directly model conditional covariances following a logic similar to chapter 4: these are in practice multivariate extensions of ARCH and GARCH models. As we shall see, the idea is similar to when in chapter 3 you did move from univariate time series models for the conditional mean to multivariate, vector models (such as vector autoregressions). However, in the case of covariance matrices, we shall see that extending univariate GARCH models to their multivariate counterparts will present many practical difficulties, unless a smart approach is adopted. Therefore the corresponding material is presented only in the final, but rather important Section 6. Third, such a smart approach—dynamic conditional correlations (DCC) models—represents the other important, key tool that is described in this chapter.

In spite of the difficulties we may encounter with a truly multivariate GARCH approach, its payoffs are obvious in terms of the questions such a framework make it possible to answer, besides whether or not correlations do change over time: Is the volatility of one specific market (say, the U.S.) leading the volatility of other markets? Is the volatility of an asset transmitted to another asset directly (through its conditional variance) or indirectly (through its conditional covariance)?

Section 2 presents the important distinction between passive and active risk management that motivates a need of a multivariate approach to the time series analysis of volatility and covariance. Section 3 investigates the special case in which there is no difference between passive and active estimation strategies, i.e., in which the econometrics of portfolio returns gives forecasts of variance that automatically incorporate forecasts of covariances between assets in pairs. Unfortunately, such an interesting result that could remarkably simplify variance forecasting obtains only when we assume rather specific asset pricing models that have a linear factor structure. Section 4 deals with simple, one would say naive, models used to forecast covariances. Section 5 presents the most imported and arguably best working set of methods to model and forecast dynamic correlations, Engle's (2002) DCC model. Section 6 finally extends our horizon to the full family of multivariate GARCH models, of which the DCC is in a one of the most recent and yet very successful members. Appendix A presents a few additional results concerning estimation methods, in particular the feasible GLS approach. Appendix B presents a fully worked out set of examples in Matlab® concerning DCC modelling.

6. **Multivariate GARCH Models**

In our introduction we have already emphasized that a full extension and generalization of simple, univariate GARCH methods to the multivariate case presents many issues and problems related to the large scale of the resulting models and their tendency to be over-parameterized. In this Section we take this task seriously and attempt to generalize the simple set-up of the first part of the course,

$$R_{t+1} = \sigma_{t+1} z_{t+1} \qquad z_{t+1} \text{ IID } \mathcal{N}(0,1),$$

to the case in which returns on N assets collected in $\mathbf{R}_{t+1}$, are described by

$$\mathbf{R}_{t+1} = \boldsymbol{\Omega}_{t+1}^{1/2} \mathbf{z}_{t+1} \qquad \mathbf{z}_{t+1} \text{ IID } \mathcal{N}(\mathbf{0}, \mathbf{I}_N), \tag{16}$$

where $\mathbf{I}_N$ is a $N \times N$ identity matrix, and (similarly to chapter 7), $\boldsymbol{\Omega}_{t+1}^{1/2}$ is the square-root, or Cholesky decomposition, of the covariance matrix, such that[24]

$$\boldsymbol{\Omega}_{t+1}^{1/2} (\boldsymbol{\Omega}_{t+1}^{1/2})' = \boldsymbol{\Sigma}_{t+1} \equiv Var[\mathbf{R}_{t+1} | \Im_t].$$

Even though in (16) we have specified $\mathbf{z}_{t+1}$ IID $\mathcal{N}(\mathbf{0}, \mathbf{I}_N)$, in certain situations it is desirable to search for a better distribution for the innovation process, $\mathbf{z}_{t+1}$. A natural alternative to the multivariate Gaussian density is the multivariate Student density, of which skewed versions exist. Moreover, note that $\boldsymbol{\Omega}_{t+1}^{1/2}$ is in no way the matrix of square roots of the elements of the full covariance matrix $\boldsymbol{\Sigma}_{t+1}$ (if so, how would you deal with potentially negative covariances?).[25] Our problem is then to write and estimate appropriate dynamic time series models for $\boldsymbol{\Sigma}_{t+1}$, knowing that this matrix contains $0.5N(N+1)$ distinct elements (because of symmetry these are less than N^2), which implies that in principle one would have to write and estimate dynamic models for each of these elements. However, as already discussed in Sections 4 and 5, constructing positive semi-definite (PSD) covariance matrix forecasts, which ensures that the portfolio variance is always non-negative, remains difficult. Appropriate structure needs to be imposed to guarantee the PSDness of the resulting forecast $\hat{\boldsymbol{\Sigma}}_{t+1}$. Here one thing needs to be appreciated: although much theoretical (econometrics) literature has focussed on relatively small multivariate cases of (16), for instance with $N = 2$ or 3, practioners need us to develop methods that apply to any value of the cross-sectional dimension N, including limit cases of N being large. In this respect—possibly with an exception of the diagonal BEKK model presented in Section 6.3 below—DCC remains the best option available. Therefore the models that are presented in the following are rather interesting on paper and for small-scale applications (up to $N = 4$ or 5) but rapidly become unwieldy or even impossible to estimate for realistic applications with hundreds of assets or securities to be modelled simultaneously.

This point is easily understood through the case of the straightforward, plain vanilla N-dimensional generalization of a GARCH(1,1) in VEC(H) form:

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

[24]In this section, to make the distinction starker, we denote as $\boldsymbol{\Omega}_{t+1}^{1/2}$ the Choleski factor of $\boldsymbol{\Sigma}_{t+1}$, also for analogy with the factors $\boldsymbol{\Omega}_{S_{t+1}}^{1/2}$ that will appear in chapter 7.

[25]In fact, $\boldsymbol{\Omega}_{t+1}^{1/2}$ is a lower triangular matrix appropriately defined according to an algorithm that is implemented in most software packages (sure enough, in Matlab). Section 10.1 of chapter 7 shows one example for the $N = 2$ case.

where $vech(\cdot)$ ("vector half") is the operator that converts the unique upper triangular elements of a symmetric matrix into a $0.5N(N+1)\times 1$ column vector. For instance

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

In this general "VEC model", each element of $\mathbf{\Sigma}_{t+1}$ is a linear function of the lagged squared errors and cross-products of errors and lagged values of the elements of $\mathbf{\Sigma}_{t+1}$. Note that while the $vec(\cdot)$ of a $N\times N$ symmetric matrix would simply be a $N^2\times 1$ vector, the $vech(\cdot)$ is instead a smaller, $0.5N(N+1)\times 1$ vector.[26] In the vech-GARCH(1,1) model above, $\mathbf{A}$ and $\mathbf{B}$ are $[0.5N(N+1)]\times[0.5N(N+1)]$ square matrices while $\mathbf{C}$ is a $N\times N$ symmetric matrix. In this vech-GARCH(1,1) framework, each element of $\mathbf{\Sigma}_t$ may affect each element of $\mathbf{\Sigma}_{t+1}$, and similarly for the outer product of past returns, $\mathbf{R}_t\mathbf{R}_t'$ (note that this is a $N\times N$ matrix because $\mathbf{R}_t$ is an N-dimensional vector). However, the structure of $\mathbf{C}$, $\mathbf{A}$, and $\mathbf{B}$ gives a total of[27]

$$\begin{aligned} 0.5N(N+1)+2[0.5N(N+1)]^2 &= 0.5N(N+1)[N^2+N+1] \\ &= 0.5N^4+0.5N^3+0.5N^2+0.5N^3+0.5N^2+0.5N \\ &= 0.5N^4+N^3+N^2+0.5N = O(N^4) \end{aligned}$$

parameters to be estimated. For instance, for $N=100$, which represents hardly a large portfolio or risk management problem, then the vech-GARCH(1,1) model has 51,010,050 parameters to be estimated. If you need to have at least 20 observations available, with $N=100$ assets this means $20\times 51,010,050/100 = 10,202,010$ observations per series, or a *daily* history of more than 40,484 years per series. This is clearly not feasible.[28] More generally, vech-GARCH models that naively generalize the GARCH models of chapter 4 to the multivariate case, tend to generate a serious "curse-of-dimensionality" problem, as estimating this many free parameters is obviously infeasible, both in terms of data availability and in numerical terms (try and propose your Matlab to estimate 51 million parameters and then you will see – you may take a 2,000-year vacation as well).[29] Moreover, this is not even the end of the bad news: these $O(N^4)$ parameters, need to be restricted for them to yield forecasts of the covariance matrix that are eventually SPD, as required. Such a restrictions are even too complex and involved to be presented here (see Gourieroux, 1997, section

[26] $vech(\cdot)$ denotes the operator that stacks the lower triangular portion of a $N\times N$ matrix as a $0.5N(N+1)\times 1$ vector.

[27] In what follows, as you may recall from your math classes, the notation $O(N^4)$ indicates that the quantity under examination grows at the same speed as N^4.

[28] This is the sense in which a textbook example with $N=3$, i.e., 78 parameters to be estimated based on, say, 2,600 observations per series, i.e., approximately 10 years of data is not that indicative of the feasibility of this model in practice.

[29] The fact over-parameterization represents the key obstacle in the generalization of GARCH to the multivariate case also explains why in what follows we entertain at most the (1,1) case. Of course, higher order GARCH is technically feasible but almost always unfeasible.

6.1).[30]

As you know, one often invoked trick to deal with the curse of dimensionality in GARCH, and also to make sure that the implied unconditional moments turn out to be consistent with what the model implies, consists of the so-called (co)*variance targeting*. As already mentioned in Section 5, the intuition is that the model-implied unconditional covariance matrix is constrained to equal a pre-calculated estimate from the simple sample covariance matrix by setting:

$$vech(\mathbf{C}_{VT}) = (\mathbf{I}_{0.5N(N+1)} - \mathbf{A} - \mathbf{B})vech\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{R}_t\mathbf{R}_t'\right). \tag{17}$$

Because by analogy to the univariate case, the unconditional, long-run covariance matrix from a vech-GARCH model is

$$vech(\bar{\mathbf{\Sigma}}) = (\mathbf{I}_{0.5N(N+1)} - \mathbf{A} - \mathbf{B})^{-1}vech(\mathbf{C}),$$

setting $vech(\mathbf{C})$ in the way reported above, gives

$$vech(\bar{\mathbf{\Sigma}}_{VT}) = vech\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{R}_t\mathbf{R}_t'\right),$$

where "VT" stands for variance targeting and the result is the desired one. This trick avoids cumbersome nonlinear estimation of $vech(\mathbf{C})$ and is also useful in a forecasting perspective to avoid that small perturbations in any of the elements of the matrices $\mathbf{A}$ and $\mathbf{B}$ may result in large changes in implied unconditional variances and covariances. However, even though setting $vech(\mathbf{C})$ as in (17), does reduce the number of estimable parameters by $0.5N(N+1)$, the residual number $2[0.5N(N+1)]^2$ remains $O(N^4)$ which means that there are still too many parameters to be estimated simultaneously in $\mathbf{A}$ and $\mathbf{B}$ when N is large. As a result, further ideas have been explored in the literature, besides covariance targeting.

6.1. *Diagonal and Scalar multivariate GARCH models*

One idea that has emerged early on (in the early 1990s) in this literature is that adequate restrictions on $\mathbf{A}$ and $\mathbf{B}$ would deliver a sensible reduction in the number of estimable parameters. One such possibility is offered by a *diagonal* multivariate GARCH(p,q), that we state in the general (p,q) form to emphasize that GARCH models may in principle be defined for cases more complex than the standard (1,1) framework, but also incorporating already covariance targeting:

$$\begin{aligned} vech(\mathbf{\Sigma}_{t+1}) &= \left(\mathbf{I}_{0.5N(N+1)} - \sum\nolimits_{i=1}^{p}\mathbf{A}_i - \sum\nolimits_{j=1}^{q}\mathbf{B}_j\right)\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{R}_t\mathbf{R}_t'\right) + \\ &\quad + \sum\nolimits_{i=1}^{p}\mathbf{A}_i vech(\mathbf{R}_{t+1-i}\mathbf{R}_{t+1-i}') + \sum\nolimits_{j=1}^{q}\mathbf{B}_j vech(\mathbf{\Sigma}_{t+1-j}), \end{aligned}$$

[30]For instance, to avoid estimating $\mathbf{C}$, $\mathbf{A}$ and $\mathbf{B}$ jointly, Ledoit et al. (2003) estimate each variance and covariance equation separately. The resulting estimates do not necessarily guarantee positive semi-definite $\mathbf{\Sigma}_{t+1}$. Therefore, in a second step, the estimates are transformed in order to achieve the requirement, keeping the disruptive effects as small as possible. The transformed estimates are still consistent with respect to the parameters of the diagonal VEC GARCH model.

where all the $[0.5N(N+1)]\times[0.5N(N+1)]$ matrices $\{\mathbf{A}_i\}_{i=1}^p$ and $\{\mathbf{B}_j\}_{j=1}^q$ are diagonal matrices, in the sense that all of their off-diagonal elements equal zero. However, although always useful because compact, in the case of diagonal M-GARCH, one does not really need vector and matrices to express the process. It is easy to see that each element of covariance matrix follows a simple dynamics:

$$\begin{aligned}\sigma_{kl,t+1} &= \left(1-\sum\nolimits_{i=1}^p \alpha_{kl,i}-\sum\nolimits_{j=1}^q \beta_{kl,j}\right)\frac{1}{T}\sum_{t=1}^T R_{k,t}R_{l,t}+\sum\nolimits_{i=1}^p \alpha_{kl,i}R_{k,t+1-i}R_{l,t+1-i}+\\ &\quad +\sum\nolimits_{j=1}^q \beta_{kl,j}\sigma_{kl,t+1-j}.\end{aligned}$$

This expression shows that conditional variances depend only on own lags and own lagged squared returns, and conditional covariances depend only on own lags and own lagged cross products of returns. Even the diagonal GARCH framework, however, results in $O(N^2)$ parameters to be jointly estimated, which is computationally infeasible with large to medium N; in fact, the number of parameters is

$$p0.5N(N+1)+q0.5N(N+1)=0.5(p+q)N(N+1).$$

We also know of another issue that is likely to show up in this case: because the coefficients $\alpha_{kl,i}$ and $\beta_{kl,j}$ are not restricted to be the same across different assets and pairs of assets, constraints will have to be imposed to keep the resulting $\mathbf{\Sigma}_{t+1}$ that collects the forecasts $\sigma_{kl,t+1}$ for $k,l=1,2,...,$ N PSD. In spite of the reduction of the number of parameters, such constraints may represent a considerable drag on the estimation speed and ease.

An even more drastic simplification, that we have in fact already examined before, is represented instead by a *scalar* GARCH(p,q):

$$\begin{aligned}\sigma_{kl,t+1} &= \left(1-\sum\nolimits_{i=1}^p \alpha_i-\sum\nolimits_{j=1}^q \beta_j\right)\frac{1}{T}\sum_{t=1}^T R_{k,t}R_{l,t}+\sum\nolimits_{i=1}^p \alpha_i R_{k,t+1-i}R_{l,t+1-i}+\\ &\quad +\sum\nolimits_{j=1}^q \beta_j\sigma_{kl,t+1-j},\end{aligned}$$

which means that ARCH and GARCH coefficients reduce to real scalar parameters common across assets. In matrix format, the model becomes:

$$\begin{aligned}vech(\mathbf{\Sigma}_{t+1}) &= \left(1-\sum\nolimits_{i=1}^p \alpha_i-\sum\nolimits_{j=1}^q \beta_j\right)\left(\frac{1}{T}\sum_{t=1}^T \mathbf{R}_t\mathbf{R}_t'\right)+\sum\nolimits_{i=1}^p \alpha_i vech(\mathbf{R}_{t+1-i}\mathbf{R}_{t+1-i}')+\\ &\quad +\sum\nolimits_{j=1}^q \beta_j vech(\mathbf{\Sigma}_{t+1-j}),\end{aligned}$$

As we know from Section 4, these strong restrictions do ensure that the resulting covariance matrix is SPD because all coefficients are restricted to be the same across different pairs i and j. Moreover, the parametric simplification is obvious as the number of parameters now simply becomes $p+q$, which is in fact independent of N. However, one is left to wonder about the exact meaning of a model in which the speed of mean reversion is restricted to be common across N different assets or portfolios.

6.2. *Constant Conditional Correlation (CCC) GARCH(p,q)*

In a way, you are already very familiar with CCC models from Section 5: if you consider a DCC model and you impose that the correlation matrix in the DCC representation, $\mathbf{\Sigma}_{t+1}^{DCC} \equiv \mathbf{D}_{t+1}\mathbf{\Gamma}_{t+1}\mathbf{D}_{t+1}$, is constant, so that $\mathbf{\Sigma}_{t+1}^{CCC} \equiv \mathbf{D}_{t+1}\mathbf{\Gamma}\mathbf{D}_{t+1}$, you obtain a CCC in which—indeed—the first letter of the acronym means "constant". Of course, tha assumption of constant correlations over time is unrealistic. However, it simply avoids all the business of defining and modelling with GARCH-type processes the $q_{ij,t+1}$ auxiliary variable. More generally, CCC and DCC models represents famous but special examples from a more general family that entertains non-linear combinations of univariate GARCH models and allows for models where one can specify separately, on the one hand, the individual conditional variances, and on the other hand, the conditional correlation matrix or another measure of dependence between the individual series (like a *copula* of the conditional joint density).[31] For models in this category, theoretical results on stationarity, ergodicity and moments may not be so straightforward to obtain as for models presented elsewhere in this Section. Nevertheless, they are less greedy in terms of number of estimated parameters than the models analyzed above and therefore they have been more successful in practice.

Analogously to the DCC case, a CCC model is based on a generalization to the vector/matrix case of the standard result that when correlations are constant, time-varying covariances may only derive from time variation in volatilities, $\sigma_{ij,t+1} \equiv \rho_{ij}\sigma_{i,t+1}\sigma_{j,t+1} = \sigma_{i,t+1}\rho_{ij}\sigma_{j,t+1}$:

$$\mathbf{\Sigma}_{t+1} \equiv \mathbf{D}_{t+1}\mathbf{\Gamma}\mathbf{D}_{t+1},$$

where $\mathbf{D}_{t+1}$ is a $N \times N$ matrix of standard deviations, $\sigma_{i,t+1}$, on the ith diagonal and zero everywhere else ($i = 1, 2, ..., N$), and $\mathbf{\Gamma}$ is a constant matrix of correlations ρ_{ij}, with ones on its main diagonal. For instance, in the $N = 2$ case:

$$\mathbf{\Sigma}_{t+1} \equiv \begin{bmatrix} \sigma^2_{1,t+1} & \sigma_{12,t+1} \\ \sigma_{12,t+1} & \sigma^2_{2,t+1} \end{bmatrix} = \begin{bmatrix} \sigma_{1,t+1} & 0 \\ 0 & \sigma_{2,t+1} \end{bmatrix} \begin{bmatrix} 1 & \rho_{12} \\ \rho_{12} & 1 \end{bmatrix} \begin{bmatrix} \sigma_{1,t+1} & 0 \\ 0 & \sigma_{2,t+1} \end{bmatrix}.$$

The key step of the DCC approach is based on the ability to disentangle the estimation and prediction of $\mathbf{D}_{t+1}$—to obtain $\hat{\mathbf{D}}_{t+1}$—and the estimation of $\mathbf{\Gamma}$, that will give $\hat{\mathbf{\Gamma}}$. In particular, also in the case of a DCC, we proceed in two steps:

1. The volatilities of each asset are estimated/predicted through a GARCH or one of the other methods considered in chapters 4 and 5. For instance, one can think of $\sigma^2_{i,t+1} = \omega + \alpha(R^i_t - \delta\sigma_{i,t})^2 + \beta\sigma^2_{i,t}$ for $i = 1, 2, ..., N$.

[31] The so-called copula-GARCH approach makes use of the theorem due to Sklar stating that any N-dimensional joint distribution function may be decomposed into its N marginal distributions, and a copula function that completely describes the dependence between the N variables. These models are specified by GARCH equations for the conditional variances (possibly with each variance depending on the lag of the other variances and of the other shocks), marginal distributions for each series (e.g. t-distributions) and a conditional copula function. The copula function may be time-varying through its parameters, which can be functions of past data. In this respect, like the DCC model of Engle (2002), copula-GARCH models can be estimated using a two-step QML approach.

2. Estimate constant correlations using a simple sample estimator based on the *standardized* returns, $\hat{z}_{t+1}^i \equiv R_{t+1}^i/\hat{\sigma}_{t+1}^i$, derived from the first step using GARCH-type models. Here, we exploit the fact that the conditional covariance of the z_{t+1}^i variables equals the conditional correlation of raw returns:
$$\hat{\rho}_{ij} = \frac{1}{T}\sum_{t=1}^{T} \hat{z}_{t+1}^i \hat{z}_{t+1}^j.$$
Such constant correlations are then inserted inside $\hat{\mathbf{\Gamma}}$ to estimate the constant correlation matrix.

6.3. *BEKK GARCH*

Given the picture provided above and the fact that DCC is a model popularized around the turn of the millenium, one may ask what was the state of multivariate GARCH modelling in practice before DCC became as popular as it is today. Apart from the uncomfortable case of CCC models that assume constant correlations over time, during the 1990s one of the most popular multivariate GARCH models had been Engle and Kroner's (1995) BEKK GARCH(p, q):[32]

$$\mathbf{\Sigma}_{t+1} = \mathbf{CC}' + \sum\nolimits_{i=1}^{p} \mathbf{A}_i(\mathbf{R}_{t+1-i}\mathbf{R}_{t+1-i}')\mathbf{A}_i' + \sum\nolimits_{j=1}^{q} \mathbf{B}_j\mathbf{\Sigma}_{t+1-j}\mathbf{B}_j',$$

where the matrices $\{\mathbf{A}_i\}_{i=1}^p$ and $\{\mathbf{B}_j\}_{j=1}^q$ are non-negative and *symmetric*. This special product-sandwich form that is used to write the BEKK ensures the PSD property without imposing further restrictions, which represent the key reason for the success of BEKK models. In fact, this full matrix BEKK is easier to estimate than vech-GARCH models, even though it remains rather complex to handle. In practice, the popular form of BEKK that many empirical analysts have come to appreciate is a simpler (1,1) *diagonal* BEKK that restricts the matrices $\mathbf{A}$ and $\mathbf{B}$ to be diagonal matrices. BEKK models possess three attractive properties:

1. A BEKK is a truncated, low-dimensional application of a theorem by which all non-negative, symmetric $N \times N$ matrices (say, $\mathbf{M}$) can be decomposed (for instance) as
$$\mathbf{M} = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} = \sum_{k=1}^{2N} \begin{bmatrix} \mathbf{m}_{k,1}'\mathbf{m}_{k,1} & \mathbf{m}_{k,1}'\mathbf{m}_{k,2} \\ \mathbf{m}_{k,2}'\mathbf{m}_{k,1} & \mathbf{m}_{k,2}'\mathbf{m}_{k,2} \end{bmatrix}$$
for appropriately selected vectors $\mathbf{m}_{k,j}$. In a sense, mathematically it is no surprise that BEKK models often offer a good fit to the dynamics of variance.

2. As already mentioned, it easily ensures PSDness of the covariance matrix.

3. BEKK is invariant to linear combinations: e.g., if $\mathbf{R}_{t+1}$ follows a BEKK GARCH(p, q), then any portfolio formed from the N securities or assets in $\mathbf{R}_{t+1}$ will also follow a BEKK.

[32] In case you wonder, BEKK means "Baba-Engle-Kraft-Kroner" and the acronym simply compacts the name of the four econometricians who contributed to its development.

However, the number of parameters in BEKK remains rather large:

$$0.5N(N+1)+0.5pN(N+1)+0.5qN(N+1)=0.5N(N+1)[1+p+q]=O(N^2).$$

Often, this has still made DCC models preferrable in practice. However, the number of parameters in BEKK is substantially inferior to those appearing in a full VEC specification. This happens because the parameters governing the dynamics of the covariance equation in BEKK models are the products of the corresponding parameters of the two corresponding variance equations in the same model.

The second and third properties of BEKK models can only be appreciated contrasting the features of BEKK under linear aggregation with the properties of alternative multivariate GARCH models, for instance even a simple diagonal vech ARCH. Not all multivariate GARCH models are invariant with respect to linear transformations.[33] For instance, for the case of two asset return series ($N=2$), consider as simple diagonal multivariate ARCH(1) model obtained from a simplification of the diagonal GARCH(p,q) introduced early on:

$$vech(\mathbf{\Omega}_t)=(\mathbf{I}_3-\mathbf{A})vech\left(T^{-1}\sum_{t=1}^{T}\mathbf{R}_t\mathbf{R}_t'\right)+\mathbf{A}vech\left(\mathbf{R}_{t-1}\mathbf{R}_{t-1}'\right), \tag{18}$$

where the helpful variance targeting restriction has already been imposed and $\mathbf{A}$ is a diagonal matrix. Because we have set $N=2$, $\mathbf{\Omega}_t$ will be a 2×2 matrix, $\mathbf{A}$ is a 3×3 diagonal matrix, $\mathbf{R}_t$ is 2×1 vector of asset returns, $vech(\mathbf{\Omega}_t)$ is a 3×1 vector of unique elements from $\mathbf{\Omega}_t$, $vech(T^{-1}\sum_{t=1}^{T}\mathbf{R}_t\mathbf{R}_t')$ is a 3×1 vector of unique elements from the sum of cross-product matrices $T^{-1}\sum_{t=1}^{T}\mathbf{R}_t\mathbf{R}_t'$, $vech\left(\mathbf{R}_{t-1}\mathbf{R}_{t-1}'\right)$ is a 3×1 vector of unique elements from the lagged cross-product matrix $\mathbf{R}_{t-1}\mathbf{R}_{t-1}'$. The number of coefficients to be estimated is of course 3, a^{11}, a^{22}, and a^{33} in the representation:

$$\begin{bmatrix}\sigma_{11,t}\\ \sigma_{12,t}\\ \sigma_{22,t}\end{bmatrix}=\left(\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}-\begin{bmatrix}a^{11}&0&0\\0&a^{22}&0\\0&0&a^{33}\end{bmatrix}\right)\begin{bmatrix}T^{-1}\sum_{t=1}^{T}R_{1t}^2\\ T^{-1}\sum_{t=1}^{T}R_{1t}R_{2t}\\ T^{-1}\sum_{t=1}^{T}R_{2t}^2\end{bmatrix}$$
$$+\begin{bmatrix}a^{11}&0&0\\0&a^{22}&0\\0&0&a^{33}\end{bmatrix}\begin{bmatrix}R_{1t-1}^2\\ R_{1t-1}R_{2t-1}\\ R_{2t-1}^2\end{bmatrix}=\begin{bmatrix}(1-a^{11})T^{-1}\sum_{t=1}^{T}R_{1t}^2+a^{11}R_{1t-1}^2\\ (1-a^{22})T^{-1}\sum_{t=1}^{T}R_{1t}R_{2t}+a^{22}R_{1t-1}R_{2t-1}\\ (1-a^{33})T^{-1}\sum_{t=1}^{T}R_{2t}^2+a^{33}R_{2t-1}^2\end{bmatrix}.$$

As for the conditions that guarantee that $\sigma_{11,t}>0$ and $\sigma_{22,t}>0$ at all times, i.e., that ensure PSDness of the model, clearly

$$\begin{aligned}(1-a^{11})T^{-1}\sum_{t=1}^{T}R_{1t}^2+a^{11}R_{1t-1}^2 &> 0 \text{ if and only if } a^{11}\in(0,1)\\ (1-a^{33})T^{-1}\sum_{t=1}^{T}R_{2t}^2+a^{33}R_{2t-1}^2 &> 0 \text{ if and only if } a^{33}\in(0,1).\end{aligned}$$

[33]By invariance of a model, we mean that it stays in the same class if a linear transformation is applied to $\tilde{\mathbf{R}}_{t+1}=\mathbf{F}\mathbf{R}_{t+1}$, where $\mathbf{F}$ is a square matrix of constants and $\tilde{\mathbf{R}}_{t+1}$ corresponds to new assets (portfolios combining the original assets). It seems sensible that a model should be invariant, otherwise the question arises which basic assets should be modelled.

At this point the filtered (predicted) correlation coefficient has expression

$$\rho_{12,t} = \frac{c^{22} + a^{22}R_{1t-1}R_{2t-1}}{\sqrt{c^{11} + a^{11}R_{1t-1}^2}\sqrt{c^{33} + a^{33}R_{2t-1}^2}},$$

and, as it is obvious, $\rho_{12,t}$ should belong to $[-1, 1]$ $\forall t \geq 1$. Here we have shortened the notation setting $c^{11} \equiv (1-a^{11})T^{-1}\sum_{t=1}^{T} R_{1t}^2$, $c^{33} \equiv (1-a^{33})T^{-1}\sum_{t=1}^{T} R_{2t}^2$, and $c^{22} \equiv (1-a^{22})T^{-1}\sum_{t=1}^{T} R_{1t}R_{2t}$. Focussing on the upper bound of the interval this means that

$$(c^{22} + a^{22}R_{1t-1}R_{2t-1})^2 \leq (c^{11} + a^{11}R_{1t-1}^2)(c^{33} + a^{33}R_{2t-1}^2)$$

or

$$(c^{22})^2+(a^{22})^2R_{1t-1}^2R_{2t-1}^2+2c^{22}a^{22}R_{1t-1}R_{2t-1} \leq c^{11}c^{33}+c^{33}a^{11}R_{1t-1}^2+c^{11}a^{33}R_{2t-1}^2+a^{11}a^{33}R_{1t-1}^2R_{2t-1}^2,$$

which is equivalent to

$$[a^{11}a^{33} - (a^{22})^2]R_{1t-1}^2R_{2t-1}^2 + [c^{11}c^{33} - (c^{22})^2] + c^{33}a^{11}R_{1t-1}^2 + c^{11}a^{33}R_{2t-1}^2 - 2c^{22}a^{22}R_{1t-1}R_{2t-1} \geq 0$$

which cannot hold for a continuous distribution for the asset return series as, even constraining $[a^{11}a^{33} - (a^{22})^2] \geq 0$ and $[c^{11}c^{33} - (c^{22})^2] \geq 0$,[34]

$$c^{33}a^{11}R_{1t-1}^2 + c^{11}a^{33}R_{2t-1}^2 - 2c^{22}a^{22}R_{1t-1}R_{2t-1} \geq 0$$

in general does not hold for $a^{22} \neq 0$. However, notice that if one sets $a^{22} = 0$, then the previous inequality simplifies to

$$a^{11}a^{33}R_{1t-1}^2R_{2t-1}^2 + \left\{\left[(1-a^{11})T^{-1}\sum_{t=1}^{T} R_{1t}^2\right]\left[(1-a^{33})T^{-1}\sum_{t=1}^{T} R_{2t}^2\right] - \left[T^{-1}\sum_{t=1}^{T} R_{1t}R_{2t}\right]^2\right\} + \\ + c^{33}a^{11}R_{1t-1}^2 + c^{11}a^{33}R_{2t-1}^2 \geq 0,$$

which has a chance to hold if a^{11} and a^{33} are such that

$$\left[(1-a^{11})T^{-1}\sum_{t=1}^{T} R_{1t}^2\right]\left[(1-a^{33})T^{-1}\sum_{t=1}^{T} R_{2t}^2\right] \geq \left[T^{-1}\sum_{t=1}^{T} R_{1t}R_{2t}\right]^2,$$

which also means that

$$\bar{\rho}_{12} = \frac{\bar{\sigma}_{12}}{\bar{\sigma}_{11}\bar{\sigma}_{22}} = \frac{T^{-1}\sum_{t=1}^{T} R_{1t}R_{2t}}{\sqrt{(1-a^{11})T^{-1}\sum_{t=1}^{T} R_{1t}^2}\sqrt{(1-a^{33})T^{-1}\sum_{t=1}^{T} R_{2t}^2}} \leq 1,$$

the unconditional correlation implied by the data and the diagonal bivariate ARCH(1) process is well-behaved. Therefore, if $a^{11} \in (0,1)$ and $a^{33} \in (0,1)$, then $a^{22} = 0$ (and some other restriction on a^{11} and a^{33}) must be imposed. This means that it is impossible to model the dynamics of volalities

[34] "At all times" here really means "for all possibile realizations of the continuous bivariate vector $\mathbf{R}_t$ which as domain $[-1, +\infty) \times [-1, +\infty)$", which alludes to the fact that even under limited responsibility, in finance asset returns may in principle take very large values

and covariances simultaneously while satisying the positivity requirement for the volatilities and keeping $\mathbf{\Omega}_t$ semi-positive definite at all times. Equivalently, if one wants to impose that the diagonal vech ARCH(1) model delivers a filtered covariance matrix $\mathbf{\Omega}_t$ that is semi-positive definite at all times, the diagonal model itself must be turned into a constant covariance multivariate ARCH model, as you understand that $a^{22} = 0$ implies $\sigma_{12,t} = T^{-1}\sum_{t=1}^{T} R_{1t}R_{2t} = \bar{\sigma}_{12}$ so that

$$\rho_{12,t} = \frac{\bar{\sigma}_{12}}{\sqrt{c^{11} + a^{11}R_{1t-1}^2}\sqrt{c^{33} + a^{33}R_{2t-1}^2}}$$

and dynamics in conditional correlations will exclusively come from dynamics in volatilities.[35]

Let's now examine the issues concerning the fact that while BEKK is "closed" under linear aggregation, a simpler diagonal vech-GARCH model is not. Consider a portfolio of the two assets, with weights w and $(1-w)$. We show that in spite of the fact that $\mathbf{R}_{t-1}$ is characterized by a diagonal bivariate ARCH(1), the portfolio returns $R_t^p = wR_{1t} + (1-w)R_{2t}$ has a variance process $\sigma_{pp,t} \equiv Var_{t-1}[R_t^p]$ that fails to display the typical "diagonal form", i.e., $(1-a^{kk})T^{-1}\sum_{t=1}^{T} R_t^p + a^{kk}(R_{t-1}^p)^2$. Note first that

$$\begin{aligned}\sigma_{pp,t} &\equiv Var_{t-1}[R_t^p] = Var_{t-1}[wR_{1t} + (1-w)R_{2t}] \\ &= w^2\sigma_{11,t} + (1-w)^2\sigma_{22,t} + 2w(1-w)\sigma_{12,t}\end{aligned}$$

$$\begin{aligned}= \; & w^2(1-a^{11})T^{-1}\sum_{t=1}^{T} R_{1t}^2 + w^2a^{11}R_{1t-1}^2 + (1-w)^2(1-a^{33})T^{-1}\sum_{t=1}^{T} R_{2t}^2 + \\ & +(1-w)^2a^{33}R_{2t-1}^2 + 2w(1-w)(1-a^{22})T^{-1}\sum_{t=1}^{T} R_{1t}R_{2t} + 2w(1-w)a^{22}R_{1t-1}R_{2t-1}\end{aligned}$$

which cannot be written in diagonal form, $(1-a^{pp})T^{-1}\sum_{t=1}^{T}[wR_{1t} + (1-w)R_{2t}]^2 + a^{pp}[wR_{1t} + (1-w)R_{2t}]^2$ because for no definition of a^{pp} it is possibile to show that

$$\begin{aligned}& w^2(1-a^{pp})T^{-1}\sum_{t=1}^{T} R_{1t}^2 + (1-w)^2(1-a^{pp})T^{-1}\sum_{t=1}^{T} R_{2t}^2 + 2w(1-w)(1-a^{pp})T^{-1}\sum_{t=1}^{T} R_{1t}R_{2t} = \\ & = w^2(1-a^{11})T^{-1}\sum_{t=1}^{T} R_{1t}^2 + (1-w)^2(1-a^{33})T^{-1}\sum_{t=1}^{T} R_{2t}^2 + 2w(1-w)(1-a^{22})T^{-1}\sum_{t=1}^{T} R_{1t}R_{2t}\end{aligned}$$

[35] In case you are curious, notice that the heuristic proof above is in itself sufficient to derive that $a^{22} = 0$ from $a^{11} \in (0,1)$ and $a^{33} \in (0,1)$ and that you do not need to deal with the lower bound of the filtered correlation coefficient derived from a^{11}, a^{22}, and a^{33}. Just for completess, let us also consider that case of imposing that $0 \geq \rho_{12,t} \geq -1$ $\forall t \geq 1$. This lower bound means that

$$-(c^{22} + a^{22}R_{1t-1}R_{2t-1}) \leq \sqrt{(c^{11} + a^{11}R_{1t-1}^2)(c^{33} + a^{33}R_{2t-1}^2)}$$

or

$$(c^{22} + a^{22}R_{1t-1}R_{2t-1})^2 \leq -(c^{11} + a^{11}R_{1t-1}^2)(c^{33} + a^{33}R_{2t-1}^2)$$

$$(c^{22})^2 + (a^{22})^2R_{1t-1}^2R_{2t-1}^2 + 2c^{22}a^{22}R_{1t-1}R_{2t-1} \leq c^{11}c^{33} + c^{33}a^{11}R_{1t-1}^2 + c^{11}a^{33}R_{2t-1}^2 + a^{11}a^{33}R_{1t-1}^2R_{2t-1}^2,$$

which is equivalent to

$$[a^{11}a^{33} - (a^{22})^2]R_{1t-1}^2R_{2t-1}^2 + [c^{11}c^{33} - (c^{22})^2] + c^{33}a^{11}R_{1t-1}^2 + c^{11}a^{33}R_{2t-1}^2 - 2c^{22}a^{22}R_{1t-1}R_{2t-1} \geq 0$$

which is the same condition used above.

and especially that

$$
\begin{aligned}
w^2 a^{11} R_{1t-1}^2 + (1-w)^2 a^{33} R_{2t-1}^2 + 2w(1-w) a^{22} R_{1t-1} R_{2t-1} &= w^2 a^{pp} R_{1t}^2 + (1-w)^2 a^{pp} R_{2t}^2 + \\
&\quad +2w(1-w) a^{pp} R_{1t-1} R_{2t-1}.
\end{aligned}
$$

This means that the Diagonal multivariate ARCH model fails to be invariant to linear combinations: if you start with N assets that follow a Diagonal multivariate ARCH model, the resulting portfolio of assets will fail to follow a similar Diagonal model, which is of course problematic if not confusing. As you should be reading in the paper by Bauwens et al. (2006), the problem of (18) that causes it to fail the invariance property is very simple to visualize: while in

$$
vech(\mathbf{\Omega}_t) = (\mathbf{I}_3 - \mathbf{A}) vech\left(T^{-1}\sum_{t=1}^{T} \mathbf{R}_t \mathbf{R}_t'\right) + \mathbf{A} vech\left(\mathbf{R}_{t-1}\mathbf{R}_{t-1}'\right)
$$

$\mathbf{A}$ is diagonal, R_t^p can be written as $[w\ 1-w]\mathbf{R}_t = \mathbf{w}'\mathbf{R}_t$ and $Var_{t-1}[R_t^p] = \mathbf{w}'\mathbf{\Omega}_t\mathbf{w}$ implies the need to use a vector of coefficients $\mathbf{w}'\mathbf{A}$ which is no longer a diagonal matrix (of course, it is not even a matrix).

It is also easy to see what you need to do in order for the invariance property to obtain: if you set $a^{11} = a^{22} = a^{33}$, then when $a^{pp} = a^{11}$

$$
\begin{aligned}
& w^2(1-a^{pp})T^{-1}\sum_{t=1}^{T} R_{1t}^2 + (1-w)^2(1-a^{pp})T^{-1}\sum_{t=1}^{T} R_{2t}^2 + 2w(1-w)(1-a^{pp})T^{-1}\sum_{t=1}^{T} R_{1t}R_{2t} = \\
& = w^2(1-a^{11})T^{-1}\sum_{t=1}^{T} R_{1t}^2 + (1-w)^2(1-a^{33})T^{-1}\sum_{t=1}^{T} R_{2t}^2 + 2w(1-w)(1-a^{22})T^{-1}\sum_{t=1}^{T} R_{1t}R_{2t}
\end{aligned}
$$

$$
\begin{aligned}
& w^2 a^{11} R_{1t-1}^2 + (1-w)^2 a^{33} R_{2t-1}^2 + 2w(1-w) a^{22} R_{1t-1} R_{2t-1} \\
& = w^2 a^{pp} R_{1t}^2 + (1-w)^2 a^{pp} R_{2t}^2 + 2w(1-w) a^{pp} R_{1t-1} R_{2t-1}
\end{aligned}
$$

will trivially hold. But this means that the only way for a Diagonal multivariate ARCH to possess the invariance property is for it to actually be a Scalar multivariate ARCH, in which the same ARCH coefficient applies to all conditional equations.

It remains natural to ask why and when researchers and practitioners alike should bother with complex and over-parameterized models of the multivariate GARCH type. On the one hand, this is no longer a timely question because we know that CCC and DCC models have been enjoying growing popularity also because they may be easily implemented in practice. On the other hand, there is another interesting reason: multivariate GARCH models speak to the heart of finance theory. To see one example of this feature, consider the case of an investor that maximizes $E_t[R_{t+1}^p]$ and minimizes $Var_t[R_{t+1}^p]$, with a trade-off coefficient λ, similarly to what we have seen in a few of our Matlab workouts:

$$
V(\mathbf{w}_t) = E_t[R_{t+1}^p] - \frac{1}{\lambda} Var_t[R_{t+1}^p] = \mathbf{w}_t' E_t[\mathbf{R}_{t+1}] - \frac{1}{\lambda}\mathbf{w}_t{}' Var_t[\mathbf{R}_{t+1}]\mathbf{w}_t,
$$

where $\mathbf{w}_t$ represents the vector of portfolio weights held by the investor and $V(\mathbf{w}_t)$ is an index of the satisfactor (happyness) of this investor. As you have seen in other courses and we have used in the our lab sessions a few times, the optimal portfolio weights (i.e., the demand function of securities by the investor) is:

$$\hat{\mathbf{w}}_t^d = \frac{1}{\lambda}\{Var_t[\mathbf{R}_{t+1}]\}^{-1}E_t[\mathbf{R}_{t+1}] = \frac{1}{\lambda}\boldsymbol{\Sigma}_{t+1}^{-1}E_t[\mathbf{R}_{t+1}].$$

At this point, equating demand to supply (say, a given $\bar{\mathbf{w}}_t^s$), we have

$$\frac{1}{\lambda}\boldsymbol{\Sigma}_{t+1}^{-1}E_t[\mathbf{R}_{t+1}] = \bar{\mathbf{w}}_t^s \Longrightarrow E_t[\mathbf{R}_{t+1}] = \lambda\boldsymbol{\Sigma}_{t+1}\bar{\mathbf{w}}_t^s,$$

which represents the mean-variance equilibrium vector of expected returns. At this point, if $\mathbf{R}_{t+1}$ follows (say) a BEKK GARCH(1,1) model and pricing errors have a multivariate IID distribution (not necessarily normal), we obtain that:

$$\begin{aligned}
\mathbf{R}_{t+1} &= E_t[\mathbf{R}_{t+1}] + \boldsymbol{\Omega}_{t+1}^{1/2}\mathbf{z}_{t+1} = \lambda\boldsymbol{\Sigma}_{t+1}\bar{\mathbf{w}}_t^s + \boldsymbol{\Omega}_{t+1}^{1/2}\mathbf{z}_{t+1} \\
&= \lambda\left[\mathbf{CC}' + \mathbf{A}(\mathbf{R}_t\mathbf{R}_t')\mathbf{A}' + \mathbf{B}\boldsymbol{\Sigma}_t\mathbf{B}'\right]\bar{\mathbf{w}}_t^s + \left[\mathbf{CC}' + \mathbf{A}(\mathbf{R}_t\mathbf{R}_t')\mathbf{A}' + \mathbf{B}\boldsymbol{\Sigma}_t\mathbf{B}'\right]^{1/2}\mathbf{z}_{t+1}.
\end{aligned}$$

At this point, a test of this simple asset pricing model is whether such a regression model may provide a high R-square thus explaining most of the variation of the N assets included in $\mathbf{R}_{t+1}$.

6.4. *Leverage Effects in multivariate GARCH*

The idea—especially befitting to stock returns—that negative shocks may have a larger impact on their volatility than positive shocks of the same absolute value already discussed in chapter 4 (and most often interpreted as a leverage effect) can be easily extended to multivariate models: both variances and covariances may react differently to a positive than to a negative shock. A useful and rather general model that takes explicitly the sign of the errors into account is the asymmetric dynamic covariance (ADC) model of Kroner and Ng (1998):

$$\begin{aligned}
\sigma_{ij,t+1} &= \rho_{ij,t+1}\sqrt{\theta_{ii,t+1}\theta_{jj,t+1}} + \phi_{ij,t+1}\theta_{ij,t+1}\ \forall i \neq j \qquad \sigma_{i,t+1}^2 = \theta_{ii,t+1} \\
\boldsymbol{\Theta}_{t+1} &= \mathbf{QQ}' + \mathbf{A}(\mathbf{R}_t\mathbf{R}_t')\mathbf{A}' + \mathbf{D}(\mathbf{v}_t\mathbf{v}_t')\mathbf{D}' + \mathbf{B}\boldsymbol{\Theta}_t\mathbf{B}' \\
\mathbf{v}_t &\equiv \max[\mathbf{0}, -\mathbf{R}_t]
\end{aligned}$$

where $\rho_{ij,t+1}$ comes from a DCC-type estimate of $\boldsymbol{\Gamma}_{t+1}$.

6.5. *Factor GARCH models*

In Section 3 we have investigated how factor models may be used to estimate and forecast correlations and in that when an exposure mapping approach is used, passive and active approaches to risk management become perfectly equivalent. The idea that factor models may greatly simplify the forecasting of conditional second moments may considerably generalized to the case of multivariate variance and covariance forecasting. The difficulty when estimating a VEC or even a BEKK model is the high number of unknown parameters, even after imposing several restrictions. It is thus not surprising that these models are rarely used when the number of series is larger than 3 or 4. Factor and orthogonal models circumvent this difficulty by imposing a common dynamic structure on all the elements of $\mathbf{\Sigma}_{t+1}$. However, we shall see that doing that within a multivariate framework is not much different from building and estimating special, constrained BEKK models. Suppose that the $N \times 1$ vector of returns $\mathbf{R}_{t+1}$ has a factor structure with 2 factors given by the 2×1 vector $\mathbf{f}_{t+1} \equiv [IP_{t+1}\ \text{Infl}_{t+1}]'$ (these are industrial production and CPI inflation) and time invariant factor loadings given by the $N \times 2$ matrix $\mathbf{B}$:

$$\mathbf{R}_{t+1} = \mathbf{B}\mathbf{f}_{t+1} + \boldsymbol{\epsilon}_{t+1}.$$

Although we consider the special case of two factors only, this example can be generalized to the case of any $K \geq 2$ factors, although the algebra becomes much more involved and challenging. Also assume that the idiosyncratic shocks in the vector $\boldsymbol{\epsilon}_{t+1}$ have conditional covariance matrix $Var_t[\boldsymbol{\epsilon}_{t+1}] = E_t[\boldsymbol{\epsilon}_{t+1}\boldsymbol{\epsilon}'_{t+1}] = \mathbf{\Psi}$ which is constant in time and semi-positive definite, and that the common factors are characterized by $E_t[\mathbf{f}_{t+1}] = \mathbf{0}$, $E_t[\mathbf{f}_{t+1}\boldsymbol{\epsilon}'_{t+1}] = \mathbf{O}$ (a matrix of zeros of the appropriate dimensions), and $E_t[\mathbf{f}_{t+1}\mathbf{f}'_{t+1}] = diag\{\sigma^2_{IP,t+1},\ \sigma^2_{Infl,t+1}\}$. Because $E_t[\boldsymbol{\epsilon}_{t+1}] = \mathbf{0},$ then $E_t[\mathbf{R}_{t+1}] = \mathbf{0}$, which means that the returns have also been de-meaned.

The expression for the conditional covariance matrix of $\mathbf{R}_{t+1}$ can be written by explicitly disantangling the role played by the risk exposures, the variance of the risk factors, and the variance of idiosyncratic risk:

$$\begin{aligned} Var_t[\mathbf{R}_{t+1}] &= \mathbf{\Sigma}_{t+1} = \mathbf{B}E_t[\mathbf{f}_{t+1}\mathbf{f}'_{t+1}]\mathbf{B}' + E_t[\boldsymbol{\epsilon}_{t+1}\boldsymbol{\epsilon}'_{t+1}] + E_t[\mathbf{f}_{t+1}\boldsymbol{\epsilon}'_{t+1}] \\ &= \mathbf{B}E_t[\mathbf{f}_{t+1}\mathbf{f}'_{t+1}]\mathbf{B}' + \mathbf{\Psi} = \mathbf{b}_{IP}\mathbf{b}'_{IP}\sigma^2_{IP,t+1} + \mathbf{b}_{Infl}\mathbf{b}'_{Infl}\sigma^2_{Infl,t+1} + \mathbf{\Psi}, \end{aligned} \tag{19}$$

where $\mathbf{b}_{IP}$ is the $N \times 1$ vector that collects the factor loadings of each of the N assets on the IP factor, and $\mathbf{b}_{Infl}$ is the $N \times 1$ vector that collects the factor loadings of each of the N assets on the inflation factor. We may highlight the role of variances and covariances of the assets through a simple $N = 2$ example:

$$\begin{aligned} \begin{bmatrix} \sigma_{11,t+1} & \sigma_{12,t+1} \\ \sigma_{12,t+1} & \sigma_{11,t+1} \end{bmatrix} &= \begin{bmatrix} (b_1^{IP})^2 & b_1^{IP}b_2^{IP} \\ b_1^{IP}b_2^{IP} & (b_2^{IP})^2 \end{bmatrix} \sigma^2_{IP,t+1} + \begin{bmatrix} (b_1^{Infl})^2 & b_1^{Infl}b_2^{Infl} \\ b_1^{Infl}b_2^{IP} & (b_2^{Infl})^2 \end{bmatrix} \sigma^2_{Infl,t+1} + \\ &\quad + \begin{bmatrix} \psi_{11} & \psi_{12} \\ \psi_{12} & \psi_{22} \end{bmatrix} \\ &= \begin{bmatrix} (b_1^{IP})^2\sigma^2_{IP,t+1} + (b_1^{Infl})^2\sigma^2_{Infl,t+1} + \psi_{11} & b_1^{IP}b_2^{IP}\sigma^2_{IP,t+1} + b_1^{Infl}b_2^{Infl}\sigma^2_{Infl,t+1} + \psi_{12} \\ b_1^{IP}b_2^{IP}\sigma^2_{IP,t+1} + b_1^{Infl}b_2^{Infl}\sigma^2_{Infl,t+1} + \psi_{12} & (b_2^{IP})^2\sigma^2_{IP,t+1} + (b_2^{Infl})^2\sigma^2_{Infl,t+1} + \psi_{22} \end{bmatrix} \end{aligned}$$

At this point, it is revealing to define the 2 *factor-mimicking* portfolios (with returns r_{kt}, $k = 1, 2$) with portfolio weights ($\boldsymbol{\phi}_k$, $k = 1, 2$) that are orthogonal to all but one set of factor loadings:

$$r_{IP,t+1} = \boldsymbol{\phi}'_{IP}\mathbf{R}_{t+1} \qquad r_{Infl,t+1} = \boldsymbol{\phi}'_{Infl}\mathbf{R}_{t+1}$$

such that $\boldsymbol{\phi}'_{IP}\mathbf{B} = [1\ \ 0]'$ and $\boldsymbol{\phi}'_{Infl}\mathbf{B} = [0\ \ 1]'$. The vector of factor-representing portfolios is then $\mathbf{r}_{t+1} = \mathbf{\Phi}'\mathbf{R}_{t+1}$, where $\mathbf{\Phi} \equiv [\boldsymbol{\phi}_{IP}\ \boldsymbol{\phi}_{Infl}]$. It is then possible to re-write the expression for the conditional covariance matrix of $\mathbf{r}_{t+1}$ and in particular for $Var_t[r^1_{t+1}]$ and $Var_t[r^2_{t+1}]$ in terms of the two factor mimicking portfolios:

$$\begin{aligned} Var_t[\mathbf{r}_{t+1}] &= \mathbf{\Phi}' E_t[\mathbf{R}_{t+1}\mathbf{R}'_{t+1}]\mathbf{\Phi} \\ &= \mathbf{\Phi}'\mathbf{\Sigma}_{t+1}\mathbf{\Phi} = \mathbf{\Phi}'\mathbf{b}_{IP}\mathbf{b}'_{IP}\mathbf{\Phi}\sigma^2_{IP,t+1} + \mathbf{\Phi}'\mathbf{b}_{Infl}\mathbf{b}'_{Infl}\mathbf{\Phi}\sigma^2_{Infl,t+1} + \mathbf{\Phi}'\mathbf{\Psi}\mathbf{\Phi}. \end{aligned}$$

In particular, notice that

$$\begin{aligned} V_t[r^1_{t+1}] &= \boldsymbol{\phi}'_{IP}\mathbf{b}_{IP}\mathbf{b}'_{IP}\boldsymbol{\phi}_{IP}\sigma^2_{IP,t+1} + \boldsymbol{\phi}'_{IP}\mathbf{b}_{Infl}\mathbf{b}'_{Infl}\boldsymbol{\phi}_{IP}\sigma^2_{Infl,t+1} + \boldsymbol{\phi}'_{IP}\boldsymbol{\phi}_{IP}\psi_{11} = \sigma^2_{IP,t+1} + \delta_1 \\ V_t[r^2_{t+1}] &= \boldsymbol{\phi}'_{Infl}\mathbf{b}_{IP}\mathbf{b}'_{IP}\boldsymbol{\phi}_{Infl}\sigma^2_{IP,t+1} + \boldsymbol{\phi}'_{Infl}\mathbf{b}_{Infl}\mathbf{b}'_{Infl}\boldsymbol{\phi}_{Infl}\sigma^2_{Infl,t+1} + \boldsymbol{\phi}'_{Infl}\boldsymbol{\phi}_{Infl}\psi_{22} = \sigma^2_{Infl,t+1} + \delta_2 \qquad (20) \end{aligned}$$

where δ_j is the $[j,j]$ element on the main diagonal of $\mathbf{\Phi}'\mathbf{\Psi}\mathbf{\Phi}$, $j = 1, 2$. Each factor-mimicking portfolio displays the exact time variation as the factor represented, which is why they are called factor-mimicking portfolios, plus some idiosynchratic risk which is due to the possible need to avoid complete diversification. At this point, it is possible to bring together the results in (19) and (20) to derive an expression that links

$$\mathbf{b}_{IP}\mathbf{b}'_{IP}Var_t[\mathbf{r}_{IP,t+1}] + \mathbf{b}_{Infl}\mathbf{b}'_{Infl}Var_t[\mathbf{r}_{IP,t+1}]$$

to the variance of the factors and terms of the type $\mathbf{b}_{IP}\mathbf{b}'_{IP}\delta_1 + \mathbf{b}_{Infl}\mathbf{b}'_{Infl}\delta_2$. Recall that

$$\begin{aligned} \mathbf{B}Var_t[\mathbf{r}_{t+1}]\mathbf{B}' &= \mathbf{B}\mathbf{\Phi}'\mathbf{b}_{IP}\mathbf{b}'_{IP}\mathbf{\Phi}\mathbf{B}'\sigma^2_{IP,t+1} + \mathbf{B}\mathbf{\Phi}'\mathbf{b}_{Infl}\mathbf{b}'_{Infl}\mathbf{\Phi}\mathbf{B}'\sigma^2_{Infl,t+1} + \mathbf{B}\mathbf{\Phi}'\mathbf{\Psi}\mathbf{\Phi}\mathbf{B}' \\ &= \mathbf{b}_{IP}\mathbf{b}'_{IP}\sigma^2_{IP,t+1} + \mathbf{b}_{Infl}\mathbf{b}'_{Infl}\sigma^2_{Infl,t+1} + \mathbf{b}_{IP}\mathbf{b}'_{IP}\delta_1 + \mathbf{b}_{Infl}\mathbf{b}'_{Infl}\delta_2. \end{aligned}$$

Therefore

$$\begin{aligned} \mathbf{b}_{IP}\mathbf{b}'_{IP}\sigma^2_{IP,t+1} + \mathbf{b}_{Infl}\mathbf{b}'_{Infl}\sigma^2_{Infl,t+1} &= \mathbf{B}Var_t[\mathbf{r}_{t+1}]\mathbf{B}' - \mathbf{b}_{IP}\mathbf{b}'_{IP}\delta_1 - \mathbf{b}_{Infl}\mathbf{b}'_{Infl}\delta_2 \\ &= \mathbf{b}_{IP}\mathbf{b}'_{IP}\theta_{IP,t} + \mathbf{b}_{Infl}\mathbf{b}'_{Infl}\theta_{Infl,t} - \mathbf{b}_{IP}\mathbf{b}'_{IP}\delta_1 - \mathbf{b}_{Infl}\mathbf{b}'_{Infl}\delta_2, \end{aligned}$$

where $\theta_{IP,t+1} \equiv Var_t[r_{IP,t+1}]$ and $\theta_{Infl,t+1} \equiv Var_t[r_{Infl,t+1}]$. Replacing this expression into $\mathbf{\Sigma}_{t+1} = \mathbf{b}_{IP}\mathbf{b}'_{IP}\sigma^2_{IP,t+1} + \mathbf{b}_{Infl}\mathbf{b}'_{Infl}\sigma^2_{Infl,t+1} + \mathbf{\Psi}$ found in (19), we have

$$\begin{aligned} \mathbf{\Sigma}_{t+1} &= \mathbf{b}_{IP}\mathbf{b}'_{IP}\theta_{IP,t+1} + \mathbf{b}_{Infl}\mathbf{b}'_{Infl}\theta_{Infl,t+1} - \mathbf{b}_{IP}\mathbf{b}'_{IP}\delta_1 - \mathbf{b}_{Infl}\mathbf{b}'_{Infl}\delta_2 + \mathbf{\Psi} \\ &= \mathbf{\Psi}^* + \mathbf{b}_{IP}\mathbf{b}'_{IP}\theta_{IP,t+1} + \mathbf{b}_{Infl}\mathbf{b}'_{Infl}\theta_{Infl,t+1}, \end{aligned}$$

where $\mathbf{\Psi}^* = \mathbf{\Psi} - \mathbf{b}_{IP}\mathbf{b}'_{IP}\delta_1 - \mathbf{b}_{Infl}\mathbf{b}'_{Infl}\delta_2$.

However, these labored (and boring) mathematical derivations simply show that the conditional covariance matrix of returns can be decomposed as a weighted sum of products of beta exposures to factor mimicking portfolio returns and conditional variance forecasts for each of the two portfolios. In practice, in order for the model to be implemented, one will need to parameterize $\theta_{k,t+1} \equiv Var_t[r_{k,t+1}]$ ($k = IP,\ Infl$), for instance as simple GARCH-type processes,

$$\theta_{k,t+1} \equiv E_t[r^2_{k,t+1}] = \omega_k + \alpha_k\epsilon^2_{k,t} + \beta_k\theta_{k,t} \quad k = IP,\ Infl.$$

Clearly, such a specification may be replaced with different ARCH specifications from chapter 4, without any qualitative differences. As a result, the conditional covariance matrix of returns $\mathbf{\Sigma}_{t+1}$ may be re-written in a BEKK form as:

$$\mathbf{\Sigma}_{t+1} = \mathbf{\Psi}^{**} + \mathbf{A}_{IP}(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t')\mathbf{A}_{IP}' + \mathbf{A}_{Infl}(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t')\mathbf{A}_{Infl}' + \mathbf{B}_{IP}\mathbf{\Sigma}_t\mathbf{B}_{IP}' + \mathbf{B}_{Infl}\mathbf{\Sigma}_t\mathbf{B}_{Infl}'.$$

This is a very interesting results: all factor GARCH models eventually may be written as special BEKK models in which the matrix of coefficients ($\mathbf{A}_{IP}$, $\mathbf{A}_{Infl}$, $\mathbf{B}_{IP}$, and $\mathbf{B}_{Infl}$) bear functional relationships to products of the beta exposures ($\mathbf{b}_{IP}$ and $\mathbf{b}_{Infl}$) with row vectors of portfolio weights ($\boldsymbol{\phi}_{IP}'$ and $\boldsymbol{\phi}_{Infl}'$) defining the mimicking relationships. This can be seen from the fact that

$$\begin{aligned}
\theta_{k,t+1} &= \omega_k + \alpha_k r_{k,t}^2 + \beta_k \theta_{k,t} \\
&= \omega_k + \alpha_k(\boldsymbol{\phi}_k'\boldsymbol{\epsilon}_t)^2 + \beta_k E_{t-1}[r_{k,t}^2] \\
&= \omega_k + \alpha_k\boldsymbol{\phi}_k'(\boldsymbol{\epsilon}_t'\boldsymbol{\epsilon}_t)\boldsymbol{\phi}_k + \beta_k\boldsymbol{\phi}_k' E_{t-1}[\mathbf{R}_t\mathbf{R}_t']\boldsymbol{\phi}_k \\
&= \omega_k + \alpha_k\boldsymbol{\phi}_k'(\boldsymbol{\epsilon}_t'\boldsymbol{\epsilon}_t)\boldsymbol{\phi}_k + \beta_k\boldsymbol{\phi}_k'\mathbf{\Sigma}_t\boldsymbol{\phi}_k.
\end{aligned}$$

As a result, because we have seen that $\mathbf{\Sigma}_{t+1} = \mathbf{\Psi}^* + \mathbf{b}_{IP}\mathbf{b}_{IP}'\theta_{IP,t+1} + \mathbf{b}_{Infl}\mathbf{b}_{Infl}'\theta_{Infl,t+1}$, we can write the conditional covariance matrix of returns as:

$$\begin{aligned}
\mathbf{\Sigma}_{t+1} &= \mathbf{\Psi}^* + \mathbf{b}_{IP}\mathbf{b}_{IP}'\left[\omega_{IP} + \alpha_{IP}\boldsymbol{\phi}_{IP}'(\boldsymbol{\epsilon}_t'\boldsymbol{\epsilon}_t)\boldsymbol{\phi}_{IP} + \beta_{IP}\boldsymbol{\phi}_{IP}'\mathbf{\Sigma}_t\boldsymbol{\phi}_{IP}\right] + \\
&\quad + \mathbf{b}_{Infl}\mathbf{b}_{Infl}'\left[\omega_{Infl} + \alpha_{Infl}\boldsymbol{\phi}_{Infl}'(\boldsymbol{\epsilon}_t'\boldsymbol{\epsilon}_t)\boldsymbol{\phi}_{Infl} + \beta_{Infl}\boldsymbol{\phi}_{Infl}'\mathbf{\Sigma}_t\boldsymbol{\phi}_{Infl}\right] \\
&= \left[\mathbf{\Psi}^* + \mathbf{b}_{IP}\mathbf{b}_{IP}'\omega_{IP} + \mathbf{b}_{Infl}\mathbf{b}_{Infl}'\omega_{Infl}\right] + \mathbf{b}_{IP}\mathbf{b}_{IP}'\alpha_{IP}\boldsymbol{\phi}_{IP}'(\boldsymbol{\epsilon}_t'\boldsymbol{\epsilon}_t)\boldsymbol{\phi}_{IP} + \\
&\quad + \mathbf{b}_{Infl}\mathbf{b}_{Infl}'\alpha_{Infl}\boldsymbol{\phi}_{Infl}'(\boldsymbol{\epsilon}_t'\boldsymbol{\epsilon}_t)\boldsymbol{\phi}_{Infl} + \mathbf{b}_{IP}\mathbf{b}_{IP}'\beta_{IP}\boldsymbol{\phi}_{IP}'\mathbf{\Sigma}_t\boldsymbol{\phi}_{IP} + \mathbf{b}_{Infl}\mathbf{b}_{Infl}'\beta_{Infl}\boldsymbol{\phi}_{Infl}'\mathbf{\Sigma}_t\boldsymbol{\phi}_{Infl} \\
&= \mathbf{\Psi}^{**} + \alpha_{IP}\mathbf{b}_{IP}\boldsymbol{\phi}_{IP}'(\boldsymbol{\epsilon}_t'\boldsymbol{\epsilon}_t)\boldsymbol{\phi}_{IP}\mathbf{b}_{IP}' + \alpha_{Infl}\mathbf{b}_{Infl}\boldsymbol{\phi}_{Infl}'(\boldsymbol{\epsilon}_t'\boldsymbol{\epsilon}_t)\boldsymbol{\phi}_{Infl}\mathbf{b}_{Infl}' + \\
&\quad + \beta_{IP}\mathbf{b}_{IP}\boldsymbol{\phi}_{IP}'\mathbf{\Sigma}_t\boldsymbol{\phi}_{IP}\mathbf{b}_{IP}' + \beta_{Infl}\mathbf{b}_{Infl}\boldsymbol{\phi}_{Infl}'\mathbf{\Sigma}_t\boldsymbol{\phi}_{Infl}\mathbf{b}_{Infl}' \\
&= \mathbf{\Psi}^{**} + \mathbf{A}_{IP}(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t')\mathbf{A}_{IP}' + \mathbf{A}_{Infl}(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t')\mathbf{A}_{Infl}' + \mathbf{B}_{IP}\mathbf{\Sigma}_t\mathbf{B}_{IP}' + \mathbf{B}_{Infl}\mathbf{\Sigma}_t\mathbf{B}_{Infl}'
\end{aligned}$$

where $\mathbf{\Psi}^{**} \equiv \mathbf{\Psi}^* + \mathbf{b}_{IP}\mathbf{b}_{IP}'\omega_{IP} + \mathbf{b}_{Infl}\mathbf{b}_{Infl}'\omega_{Infl}$, $\mathbf{A}_{IP} \equiv \sqrt{\alpha_{IP}}\mathbf{b}_{IP}\boldsymbol{\phi}_{IP}'$, $\mathbf{A}_{Infl} \equiv \sqrt{\alpha_{Infl}}\mathbf{b}_{Infl}\boldsymbol{\phi}_{Infl}'$, $\mathbf{B}_{IP} \equiv \sqrt{\beta_{IP}}\mathbf{b}_{IP}\boldsymbol{\phi}_{IP}'$, and $\mathbf{B}_{Infl} \equiv \sqrt{\beta_{Infl}}\mathbf{b}_{Infl}\boldsymbol{\phi}_{Infl}'$. In conclusion, this shows that so that the 2-factor GARCH model is a special case of the BEKK parametrization, although subject to restrictions.

6.6. *Estimation and diagnostic checks of multivariate GARCH models*

Multivariate GARCH estimation is performed using maximum likelihood to jointly estimate the parameters of the (conditional) mean and the variance equations in

$$\mathbf{R}_{t+1} = \boldsymbol{\mu}_{t+1} + \mathbf{\Omega}_{t+1}^{1/2}\mathbf{z}_{t+1} \qquad \mathbf{z}_{t+1} \text{ IID } \mathcal{N}(\mathbf{0}, \mathbf{I}_N),$$

where all the parameters characterizing $\boldsymbol{\mu}_{t+1}$ and $\mathbf{\Omega}_{t+1}^{1/2}$ are collected in some vector $\boldsymbol{\theta}$. Note that although the GARCH parameters do not affect the conditional mean, the conditional mean parameters generally enter the conditional variance specification through the residuals, $\mathbf{R}_{t+1} - \boldsymbol{\mu}_{t+1}(\boldsymbol{\theta})$.

Assuming multivariate normality, the log-likelihood contributions (i.e., the PDF values for each of the sample observations) for GARCH models are given by:[36]

$$l_{\mathcal{N}}(\mathbf{R}_{t+1};\boldsymbol{\theta}) \equiv -\frac{1}{2}N\ln(2\pi) - \frac{1}{2}\ln\det\boldsymbol{\Sigma}_{t+1}(\boldsymbol{\theta}) - \frac{1}{2}(\mathbf{R}_{t+1}-\boldsymbol{\mu}_{t+1}(\boldsymbol{\theta}))\boldsymbol{\Sigma}_{t+1}^{-1}(\boldsymbol{\theta})(\mathbf{R}_{t+1}-\boldsymbol{\mu}_{t+1}(\boldsymbol{\theta})).$$

In the case of a Student t-distribution, the contributions are of the form:

$$\begin{aligned} l_t(\mathbf{R}_{t+1};\boldsymbol{\theta},d) \quad &\equiv \quad \ln\frac{\Gamma\left(\frac{N+d}{2}\right)d^{\frac{N}{2}}}{(d\pi)^{\frac{N}{2}}\Gamma\left(\frac{d}{2}\right)(d-2)^{\frac{N}{2}}}\frac{1}{2}\ln\det\boldsymbol{\Sigma}_{t+1}(\boldsymbol{\theta}) - \frac{1}{2}(d+N)\times \\ &\times\ln\left[1+\frac{(\mathbf{R}_{t+1}-\boldsymbol{\mu}_{t+1}(\boldsymbol{\theta}))\boldsymbol{\Sigma}_{t+1}^{-1}(\boldsymbol{\theta})(\mathbf{R}_{t+1}-\boldsymbol{\mu}_{t+1}(\boldsymbol{\theta}))}{d-2}\right], \end{aligned}$$

where $d > 2$ is the "number of degrees of freedom". The asymptotic properties of ML (and QML) estimators in multivariate GARCH models are not yet firmly established, and are difficult to derive from low level assumptions. While consistency has been proven by Jeantheau (1998), asymptotic normality of the QMLE is not established generally. However, applied researchers who use MGARCH models have generally proceeded as if asymptotic normality holds in all cases.[37]

As usual, you may hesitate before introducing a specific parametric assumption on the distribution of the (standardized) residuals and may want to proceed instead under the weaker assumption that

$$\mathbf{R}_{t+1} = \boldsymbol{\mu}_{t+1} + \boldsymbol{\Omega}_{t+1}^{1/2}\mathbf{z}_{t+1} \qquad \mathbf{z}_{t+1}\ \text{IID}\ \mathcal{D}(\mathbf{0},\mathbf{I}_N),$$

where $\mathcal{D}$ is some distribution that is not specified. In this case you will be able to obtain QML estimates using the same logic illustrated in chapter 4 in the case of univariate GARCH models. In sum, even though the conditional joint distribution of the shocks $\mathbf{z}_{t+1}$ is **not** normal (i.e.,$\mathbf{z}_{t+1}$ IID $\mathcal{D}(\mathbf{0},\mathbf{I}_N)$ and $\mathcal{D}$ does not reduce to a $\mathcal{N}$), *under some conditions*, an application of MLE based on $\mathbf{z}_{t+1}$ ~IID $\mathcal{N}(\mathbf{0},\mathbf{I}_N)$ will yield estimators of the mean and variance parameters which converge to the true parameters as the sample gets infinitely large, i.e. that are *consistent*. What are the conditions mentioned above? You will need that:

- The conditional variance function, $\boldsymbol{\Sigma}_{t+1}$ seen as a function of the information at time t, $\mathcal{F}_t$, must be correctly specified.
- The conditional mean function, $\boldsymbol{\mu}_{t+1}$ seen as a function of the information at time t, $\mathcal{F}_t$, must be correctly specified.

Because estimating M-GARCH models is time-consuming, it is desirable to check ex ante whether the data present evidence of multivariate (G)ARCH effects. This is done both on the individual series by testing whether squared returns are serially correlated for each individual series, but also testing

[36] The conditional mean and covariance functions are denoted as $\boldsymbol{\mu}_{t+1}(\boldsymbol{\theta})$ and $\boldsymbol{\Omega}_{t+1}(\boldsymbol{\theta})$ to emphasize their dependence on the parameter vector $\boldsymbol{\theta}$.

[37] Gourieroux (1997, section 6.3) proves it for a general formulation using high level assumptions. Comte and Lieberman (2003) prove it for the BEKK formulation.

whether squared returns appear to display any significant cross-correlations, $Corr[R_{i,t}^2, R_{j,t-k}^2] \neq 0$ for $i \neq j$ and $k \neq 0$. See chapter 4 for examples of how this may be done and how one tests for the significance of (cross-) serial correlations.

Ex post, it is also of crucial importance to check the adequacy of the M-GARCH specification. However, few tests are specific to multivariate models. Univariate tests applied independently to each series of (standardized residuals) remain very common, but not completely appropriate. For instance, as seen in chapter 4, it is typical—when $\mathbf{z}_{t+1}$ ~IID $\mathcal{N}(\mathbf{0}, \mathbf{I}_N)$ has been assumed—to applystandard *univariate tests of normality* to the standardized model residuals defined as $\hat{z}_{i,t+1} \equiv R_{i,t+1}/\hat{\sigma}_{i,t+1}$, where $\hat{\sigma}_{i,t+1}$ denotes the time series of filtered standard deviations derived from the estimated volatility model, $\hat{\sigma}_{i,t+1} = \mathbf{e}_i'\hat{\mathbf{\Sigma}}_{t+1}\mathbf{e}_i$ ($i = 1, 2, ..., N$, i.e., the ith element on the main diagonal of $\hat{\mathbf{\Sigma}}_{t+1}$). Here, we are clearly exploiting the fact that $\mathbf{z}_{t+1}$ ~IID $\mathcal{N}(\mathbf{0}, \mathbf{I}_N)$ implies that each of the elements $\mathbf{z}_{t+1}$ must have a marginal normal distrubution.[38] As you know, one commonly used test is Jarque and Bera's and measures departures from normality in terms of the skewness and kurtosis of standardized residuals. A second method exploits the fact that even though normality has not been assumed (this is the case of QMLE) so that the assumed model for returns is $\mathbf{z}_{t+1}$ IID $\mathcal{D}(\mathbf{0}, \mathbf{I}_N)$ and $\mathcal{D}(\mathbf{0}, \mathbf{I}_N)$ is not $\mathcal{N}(\mathbf{0}, \mathbf{I}_N)$, a correctly specified anyway implies

$$\mathbf{z}_{t+1} \sim \text{IID}.$$

As we know, independence implies that $\hat{Q}_k^g(z_i) \simeq 0$ for all $k \geq 1$ where

$$\hat{Q}_k^g(z_i) \equiv T\sum_{\tau=1}^{k}(\hat{\rho}_{i,\tau}^g)^2 \overset{a}{\sim} \chi_k^2 \qquad \hat{\rho}_\tau^g \equiv \frac{\sum_{t=1}^{T-\tau}(g(z_{i,t}) - \overline{g(z_{i,t})})(g(z_{i,t+\tau}) - \overline{g(z_{i,t})})}{\sum_{t=1}^{T-\tau}(g(z_{i,t}) - \overline{g(z_{i,t})})^2}$$

and $g(\cdot)$ is any (measurable) function. Because we are testing the correct specification of a conditional volatility model, it is typical to set $g(x) = x^2$, i.e., we test whether the squared standardized residuals, $\hat{z}_{i,t+1}^2 \equiv R_{i,t+1}^2/\hat{\sigma}_{i,t+1}^2$, display any systematic autocorrelation patterns.[39]

Although univariate tests can provide some guidance, contemporaneous correlation of disturbances entails that statistics from individual equations are not independent. Therefore, truly multivariate tests have been developed and are routinely applied in practice. Recalling the framework $\mathbf{z}_{t+1}$ IID $\mathcal{D}(\mathbf{0}, \mathbf{I}_N)$, it is typical to also test ex-post cross-serial correlations of functions of standardized residuals, e.g., (i) $Corr[z_{i,t}, z_{j,t-k}^2] = 0$ for $i \neq j$, (ii) $Corr[z_{i,t}^2, z_{j,t-k}^2] = 0$ for $i \neq j$, (iii) $Var[z_{i,t}^2] = 0$ for $i = 1, 2, ..., N$, and (iv) $Corr[z_{i,t}, z_{j,t-k}] = 0$ for $i \neq j$. These zero cross-serial correlations are tested as usual using sample correlograms (here, cross-correlograms involving pairs of series) and Portmanteau Box-Pierce tests, as seen in chapter 4. Moreover, a generalization of the

[38] As you will recall from your statistics courses, the opposite does not hold: $z_{i,t+1} \sim \mathcal{N}(0,1)\ \forall i \not\Longrightarrow \mathbf{z}_{t+1}$ ~IID $\mathcal{N}(\mathbf{0}, \mathbf{I}_N)$.

[39] We face one additional problem when $\mathbf{z}_{t+1}$ ~IID is tested by sequentially applying tests for each of the resulting N times series of standardized residuals because $\hat{\mathbf{z}}_{t+1}$ derives from a multivariate model and therefore after the first test has been implemented, the subsequent tests may be affected by the previous inferential methods employed.

standard Box-Pierce/Ljung-Box test,

$$Q_{LB}(k) \equiv T(T+2)\sum_{j=1}^{k}\frac{\rho_j^2}{T-k} \overset{a}{\sim} \chi_k^2$$

exist, such as Hosking's (1980) (here "HM" stands for Hosking's multivariate test):

$$Q_{HM}(k) \equiv T^2\sum_{j=1}^{k}(T-j)tr[C_Y^{-1}(0)C_Y(j)C_Y^{-1}(0)C_Y'(j)] \overset{a}{\sim} \chi^2_{(\max\{p,q\})^2k}$$

where $\mathbf{Y}_t \equiv vech(\mathbf{z}_t\mathbf{z}_t')$ and $C_Y(j)$ is the sample autocovariance matrix of order j for the series $\mathbf{Y}_t$.

6.7. *One easily implemented multivariate model: PC GARCH*

If one leaves the DCC model of Section 5 aside, it seems legitimate to ask whether multivariate GARCH models only give occasions for pain and sorrow. The simple answer is that unless one makes smart attempts at getting multivariate estimates by only using univariate GARCH estimates—as DCC and CCC models accomplish—this tends to be the case, at least when N exceeds low values such as 3 or 4. In fact, the literature features several such attempts, e.g., the *orthogonal* GARCH and the *principal component* (PC) GARCH.[40] In a PC GARCH model, the observed data are assumed to be generated by an orthogonal transformation of N (or a smaller number of) univariate GARCH processes. The matrix of the linear transformation is the orthogonal matrix (or a selection) of eigenvectors of the population unconditional covariance matrix of the standardized returns.[41] In a PC GARCH, estimation is organized in 7 steps. The input is a matrix ($\mathbf{R}$) of returns with rows representing T points in time and columns representing N assets. The steps are as follows:

1. Estimate a univariate GARCH model for each of the assets or portfolios in $\mathbf{R}$, i.e., for the N columns in the matrix; the GARCH models for each of the assets could be different, as in Section 5; the parameters are estimated independently by ML or QML.

2. Standardize the residuals with the estimated variance for each asset, obtaining the $z_{i,t}$ to be collected in a $T \times N$ matrix $\mathbf{Z}$.

[40] Principal component analysis models the variance structure of a set of observed variables using linear combinations of the variables. While we generally require as many PCs as variables to reproduce the original variance structure, we usually hope to account for most of the original variability using a relatively small number of components ("data reduction"). The PCs of a set of variables are obtained by computing the eigenvalue decomposition of the sample variance matrix: $\mathbf{\Omega} = \mathbf{L}\mathbf{\Lambda}\mathbf{L}'$, where $\mathbf{L}$ is the matrix of eigenvectors and $\mathbf{\Lambda}$ is the diagonal matrix with eigenvalues on the diagonal. The first PC is the unit-length linear combination of the original variables with maximum variance; subsequent PCs maximize variance among unit-length linear combinations that are orthogonal to the previous PCs. PCs may be computed starting from either covariance matrices or correlation matrices; correlations may also be computed in nonparametric fashion (e.g., Spearman rank-order or Kendall's tau measures).

[41] Of course, PC/orthogonal models can also be considered as factor models, where the factors are univariate GARCH-type processes. Therefore PC/orthogonal models are nested in the BEKK family. In particular, the PC/orthogonal-GARCH model is covariance-stationary if the m univariate GARCH processes built from m principal compoents are themselves stationary.

3. Compute the N principal components of the matrix of standardized residuals $\mathbf{Z}$, obtaining the PC matrix $\mathbf{P}$, $\mathbf{P} = \mathbf{ZL}$, where $\mathbf{L}$ is the matrix of loadings of the vectors of standardized returns on each of the eigenvectors.

4. Estimate a univariate GARCH model for each of the N principal components (that is, for each column of $\mathbf{P}$); a GARCH(1,1) is generally recommended.

5. Use the loading matrix $\mathbf{L}$ to rotate the PC variances back to variable space; at each point in time compute: $\mathbf{C}_t = \mathbf{LD}_t\mathbf{L}'$, where $\mathbf{D}_t$ is the time t diagonal matrix of the estimated variances of the PCs at time t. At this point, the matrix $\mathbf{C}_t$ is an approximate correlation matrix for the original variables at time t; however, there is no guarantee that the elements on the diagonal of $\mathbf{C}_t$ are equal to 1.

6. Standardize $\mathbf{C}_t$ so that it is a correlation matrix with all of its diagonal elements equal to 1, call the result $\mathbf{Corr}_t$; practically, this step is simply performed by using any software to compute the correlation matrix of $\mathbf{C}_t$.

7. At this point, we scale $\mathbf{Corr}_t$ with the estimated variances of the original GARCH models in $\mathbf{D}_t$ to get the covariance matrix:

$$\boldsymbol{\Sigma}_t = \mathbf{D}_t^{1/2}\mathbf{Corr}_t\mathbf{D}_t^{1/2}.$$

PC GARCH, although there are no compelling reasons for why it may work or accurately forecast variances and covariances, can handle any practically interesting value for N: computationally, a problem would need to have several thousand variables/assets before computing time becomes a serious issue. In the case of the PC GARCH, experiments have been performed to see if not performing the GARCH estimates for the smallest—those that explain a smaller percentage of total variance—PCs may be a good thing, in the sense that not only fitting the time variation of variances and covariances is not seriously impaired, but there is actually evidence that forecasting accuracy may benefit. Therefore, in practice PCs with very small contributions to variance may be skipped. For instance, Alexander (2001, section 7.4.3) illustrates the use of the PC GARCH model (that she also calls orthogonal GARCH). She emphasizes that using a small number of principal components compared to the number of assets is the strength of the approach (in one example, she fixes m at 2 for $N = 12$ assets). However, note that the conditional variance matrix has then reduced rank (if $m < N$), which may be a problem for applications and for diagnostic tests which depend on the inverse of $\boldsymbol{\Sigma}_t$.

Appendix C — Another Matlab® Workout

You are a European investor with the Euro as a reference currency. Using *daily* data in STOCK-INT2013.XLS, construct monthly returns (*in Euros*) using the three price indices DS Market-PRICE indices for Germany, the US and the UK.

1. For the sample period January 1, 2007 - December 31, 2010, compute and plot the return series of each of the three indices expressed in Euros.

2. With reference to the same sample period, compute the unconditional covariance matrix of the 3×1 vector of index returns as well as the unconditional variance of an *equally-weighted* portfolio. Based on these sample estimates of variances, compute and plot the 1% unconditional (Gaussian) VaR measures for each of the three national indices as well as for the equally weighted portfolio. Moreover, compute an equally weighted average of the three 1% VaR measures and compare it with the 1% normal VaR of the equally weighted portfolio: why are they not the same? Link this finding to the concepts of passive and active risk management.

3. With reference to the post-financial crisis period January 3, 2011 - December 31, 2012 and to your equally weighted portfolio returns, compute and plot the following 3 recursive, daily 1% VaR measures: (i) a constant, unconditional Gaussian VaR; (ii) a Gaussian GARCH(1,1) VaR directly estimated on your portfolio returns (i.e., a passive VaR measure); (iii) a trivariate Gaussian Constant Conditional Correlation (CCC) GARCH (1,1) VaR in which correlations are assumed to be constant and equal to the unconditional pair-wise correlations between first-stage standardized residuals from appropriately defined GARCH(1,1) models. (iii) is an active risk-management measure because it depends on your specific portfolio weights. [*Hint*: Notice that this question implies that you will have to estimate three simple GARCH(1,1) processes for the three indices and also for your own portfolio over the assigned 6-month sample]

4. Repeat question 3 with reference to the same sample period, but this time comparing the daily, recursive, 1% VaR measures for: (i) a Gaussian GARCH(1,1) VaR directly estimated on your portfolio returns (i.e., a passive VaR measure); (ii) a trivariate Gaussian Constant Conditional Correlation GARCH (1,1) VaR in which correlations are assumed to be constant and equal to the unconditional pair-wise correlations between first-stage standardized residuals from appropriately defined GARCH(1,1) models, (iii) a trivariate Gaussian Dynamic Conditional Correlation GARCH (1,1) VaR in which correlations are estimated from an Exponential Smoothing, RiskMetrics-style model for the elements of the auxiliary matrix $\mathbf{Q}_t$ as discussed in Lecture 4 of the second part of the course.

5. Estimate over the 2007-2010 sample a simple constant-mean GARCH(1,1)-DCC(1) model to filter the dynamics of the correlations of returns at daily frequency. In essence, the model is:

$$\begin{aligned} r_{ger,t+1} &= \mu_{ger} + \epsilon_{ger,t+1} \\ r_{us,t+1} &= \mu_{us} + \epsilon_{us,t+1} \\ r_{uk,t+1} &= \mu_{uk} + \epsilon_{uk,t+1} \end{aligned} \qquad \begin{bmatrix} \epsilon_{us,t+1} \\ \epsilon_{uk,t+1} \\ \epsilon_{ger,t+1} \end{bmatrix} \sim N\left(\mathbf{0}, \mathbf{\Omega}_{t+1}^{GC/DCC}\right),$$

where $r_{i,t+1}$ denotes daily returns ($i =$ EU/Ger, US, UK). Also extract the dynamic, conditional correlation matrix implied by the model and plot the (predicted) correlations during the period January 2011 - December 2012.

6. Compute and plot the 1% VaR from the DCC model of question 5 over the out-of-sample period January 1, 2011 - December 31, 2012 and compare it with the 1% VaR computed from the CCC and the RiskMetrics, exponentially smoothed DCC of questions 3 and 4.

7. Estimate over the 2007-2010 sample a constant-mean Principal Component (also called Orthogonal) GARCH(1,1) model. Plot the predicted, one-day ahead dynamic volatilities and correlations resulting from the PC/Orthogonal GARCH model. In the case of volatilities, it is easier if you compute and plot the volatility over time of your portfolio. Compare such a

dynamic, conditional volatility with the time series you should have derived from the DCC GARCH of question 5.

8. Recursively compute and plot 1% VaR over the out-of-sample period January 1, 2011 - December 31, 2012 using both historical and weighted historical simulations with a rolling window of $m = 252$ days and—in the case of weighted historical simulations—a decay factor of $\eta = 0.99$. Apply your calculations to each index return series individually as well as to your portfolio returns. Check whether a simple, weighted-combination of individual asset historical VaRs equals the historical VaR for your portfolio.

9. Estimate a BEKK-GARCH(1,1) model. In case you fail, try again using a longer sample Jan. 1, 2003 - Dec. 31, 2010. How long does it take on your computer to estimate a truly multivariate (and yet, already simplified in some ways, as commented in the lectures) BEKK model? [*Hint*: You need to use Kevin Shepard's *full_bekk_mvgarch* function; in case your first attempt at estimating the BEKK model fails, you will need to change the beginning and end dates of the sample and then F9 to re-estimate the BEKK, since the code will have stopped at that point] WARNING: on not-so-good, not-so-new laptops this point of the code may take up to 15 minutes to run. You can always stop execution by pressing the combination CRTL+C.

Solution

This solution is a commented version of the MATLAB code Ex_Multi_GARCH_2013.m posted on the course web site. Note that in this case, all the Matlab functions needed for the correct functioning of the code have been included. The loading and pre-processing of the data is similar to Appendix B and therefore it will not be repeated here. The same applies to the exchange rate transformations that have now become customary in the first part of our Matlab workouts.

1. Figure C1 shows the plots of the daily data and shows no surprises.

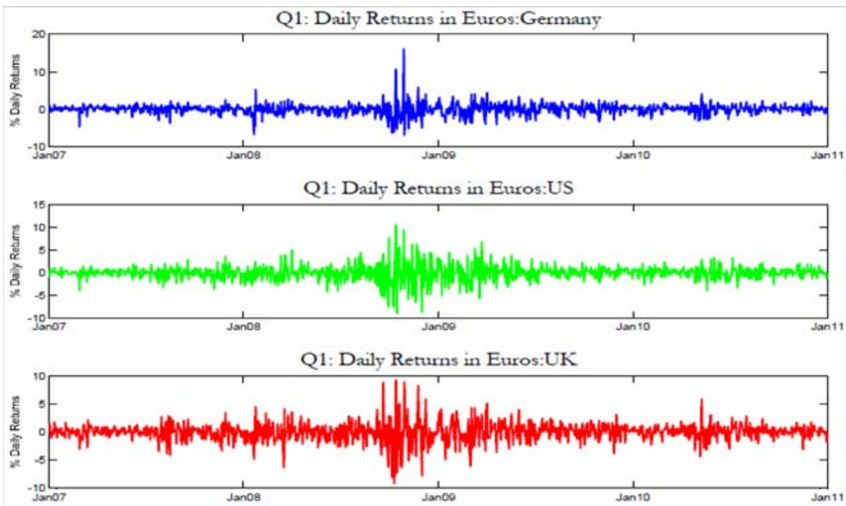

Figure C1: Daily index returns (expressed in euros) for the sample 2007-2010

2. Next, we compute the unconditional covariance matrix of the 3×1 vector of index returns as well as the unconditional variance of an *equally-weighted* portfolio. Based on these sample estimates of variances, we compute and plot the 1% unconditional (Gaussian) VaR measures for each of the three national indices as well as for the equally weighted portfolio. We compare it to an equally weighted average of the three 1% VaR measures and compare it with the 1% normal VaR of the equally weighted portfolio. This is accomplished by the following simple lines of code:

```
Cov_matrix=cov(ret(first:last,:));
Corr_matrix=corr(ret(first:last,:));
p=0.01;
VaR_port_unc=-norminv(p,0,sqrt(w'*Cov_matrix*w));
VaR_ger=-norminv(p,0,sqrt(Cov_matrix(1,1)));
VaR_us=-norminv(p,0,sqrt(Cov_matrix(2,2)));
VaR_uk=-norminv(p,0,sqrt(Cov_matrix(3,3)));
VaR_sum=[VaR_ger, VaR_us, VaR_uk]*w;
```

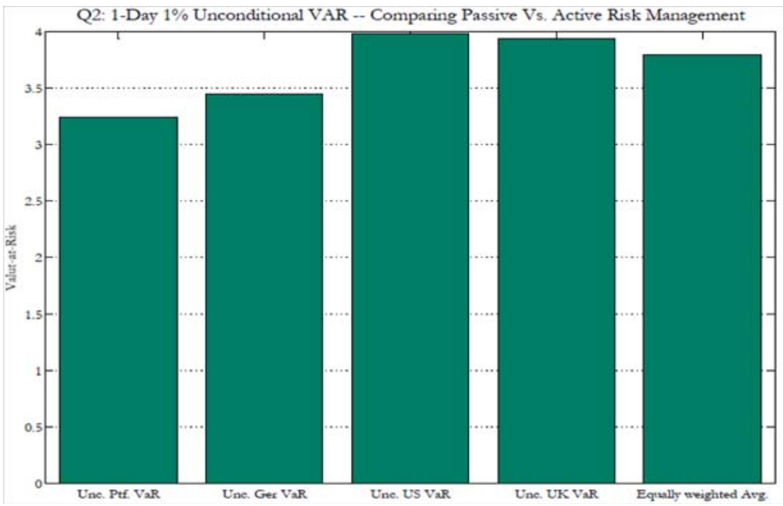

Figure C2: Comparing 1% VaR measures based on unconditional correlation estimates

Figure C2 shows the results. Interestingly, the VaR of the equally weighted portfolio is not the same as—it is in fact considerably lower (3.2% per day vs. 3.7%)—the equally weighted average of the VaRs of each individual market. This is caused by the fact that if VaR is computed on the basis of $\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}$, then it becomes a highly complex (non-linear) function of the portfolio weights, which is not the case when one simply sums and weights the VaR measures obtained for each individual market. In fact, the figure clearly shows that diversification reduces—as you would expect—risk not only when the latter is measured by portfolio variance, but also when you measure risk as VaR: this is the difference between the first bar concerning the VaR of $\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}$ and the remaining values, in which either diversification is not applied (when you put 100% of your wealth in each of the markets) or the calculation is incorrectly performed, as *the VaR of a portfolio is not the same as the portfolio of the VaRs*.

3. With reference to the post-financial crisis period January 3, 2011 - December 31, 2012 and the equally weighted portfolio returns, next we compute and plot 3 *recursive*, daily (i.e., in correspondence on each single day in this out-of-sample period) 1% VaR measures: (i) a constant, unconditional Gaussian VaR computed as

```
VaR_port_unc=-norminv(p,0,sqrt(w'*Cov_matrix*w));
```

(ii) a Gaussian GARCH(1,1) VaR directly estimated on portfolio returns (i.e., a passive VaR measure)

```
VaR_garch_port=-vol_Port*norminv(p,0,1);
```

(iii) a trivariate Gaussian Constant Conditional Correlation (CCC) GARCH (1,1) VaR in which correlations are assumed to be constant and equal to the unconditional pair-wise correlations between first-stage standardized residuals from appropriately defined GARCH(1,1) models:[56]

```
% Standardized returns to be used in CCC calculations
ret_std_GER=ret_ger(first:last,1)./sigma_GER;
ret_std_US=ret_us(first:last,1)./sigma_US;
ret_std_UK=ret_uk(first:last,1)./sigma_UK;
ret_std_Port=port_ret(first:last,1)./sigma_Port;
% Estimates correlations from a Constant Conditional Correlation (CCC)
Multivariate ARCH
Gamma=corr([ret_std_GER ret_std_US ret_std_UK]);
VaR_CCC=NaN(final-last,1);
for i=1:final-last
D=diag([vol_GER(i) vol_US(i) vol_UK(i)]);
VaR_CCC(i,1)=-norminv(p,0,sqrt(w'*D*Gamma*D*w));
end
```

Figure C3 shows the resulting VaR estimates from the three models, plus realized daily returns (the blue time series). Obviously the unconditional VaR gives a constant 1% VaR that is clearly and repeatedly violated only in the Summer of 2011. The remaining two models give similar VaR estimates that are visibly time-varying and—which debunks a myth often entertained—most of

[56]This question implies that we estimate three GARCH(1,1) processes for the three indices. We omit those Matlab commands as they are identical to those already commented in chapter 4.

the time less restrictive than the unconditional VaR. However, between the Summer and the Fall of 2011, in correspondence to the first bout of the European sovereign debt crisis, the univariate GARCH-based and the CCC VaRs drastically decline (i.e., VaR becomes larger in absolute value) and as a result a few of the violations recorded with respect to the unconditional, normal-based VaR are avoided.

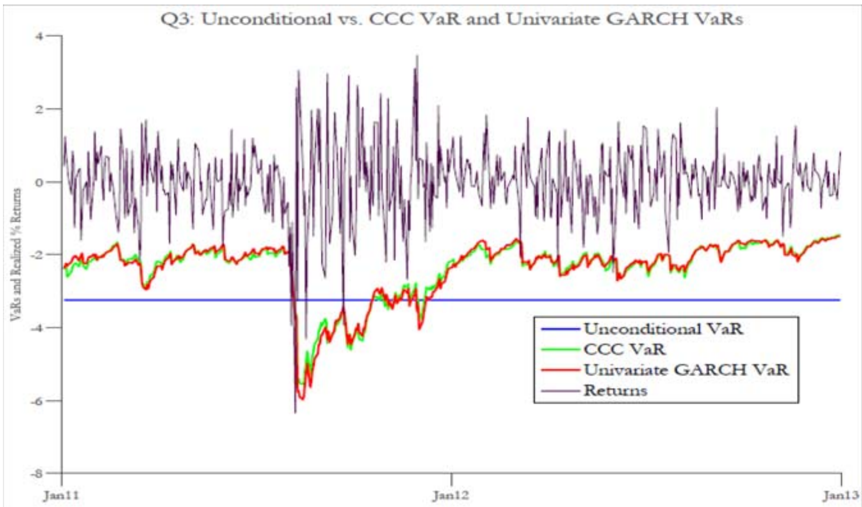

Figure C3: Recursive daily 1% VaR estimates from different alternative models

4. As instructed by the text of the workout, we have repeated question 3 with reference to the same sample period, but this time comparing the daily, recursive, 1% VaR measures for: (i) a Gaussian GARCH(1,1) VaR directly estimated on portfolio returns (i.e., a passive VaR measure); (ii) a trivariate Gaussian Constant Conditional Correlation GARCH (1,1) VaR in which correlations are assumed to be constant and equal to the unconditional pair-wise correlations between first-stage standardized residuals from appropriately defined GARCH(1,1) models—and these are the same models already employed in Figure C3—and (iii) a trivariate Gaussian Dynamic Conditional Correlation GARCH (1,1) VaR in which correlations are estimated from an Exponential Smoothing, RiskMetrics-style model for the elements of the auxiliary matrix $\mathbf{Q}_t$:

options = optimset('fmincon');

options.Display = 'iter';

parm=0.5; LB = 0; UB = 1; A = 1; b = 0.998;

lambda = fmin-con(@dcc_ES_3assets,parm,A,b,[],[],LB,UB,[],options,ret_std_GER,ret_std_US,ret_std_UK);

These lines of code perform manual fmincon-type estimation of a RiskMetrics-based DCC model that uses as an objective function the routine @dcc_ES_3assets that comes with Kevin Sheppard's GARCH package. Here LB is the lower bound of the region over which the search for the parameter lambda, and UB is the upper bound.[57] As printed on the Matlab screen, the estimated RiskMetrics

[57] A = 1 means that the constraint is linear and proportional.

parameter λ is 0.998. Figure C4 shows the recursive daily 1% VaR estimated under the three models. The three models now all give roughly similar estimates, in spite of the important qualitative difference between the CCC that assumes a constant correlation and the DCC that estimates a time-varying conditional correlation.

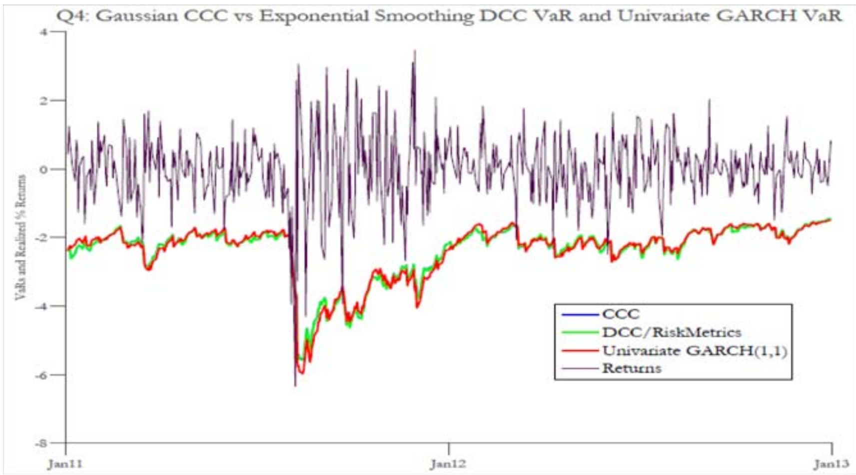

Figure C4: Recursive daily 1% VaR estimates from different alternative models including RiskMetrics-DCC

Iter	F-count	f(x)	Max constraint	Line search steplength	Directional derivative	First-order optimality
0	3	784.457	-0.009998			
1	13	783.692	-0.01773	0.00781	-222	496
2	18	780.762	-0.07064	0.25	-162	236
3	21	777.926	-0.0482	1	-103	68.4
4	24	777.763	-0.04028	1	-35.1	10.5
5	27	777.724	-0.04246	1	-7.64	8.18
6	30	777.686	-0.04402	1	-4.27	15.8
7	33	777.621	-0.04629	1	-3.63	18.9
8	36	777.589	-0.04685	1	-2.49	12.5
9	39	777.576	-0.04662	1	-1.73	3.95
10	42	777.575	-0.04642	1	-0.613	0.492
11	45	777.575	-0.04638	1	-0.0863	0.0232

```
==============================
GARCH parameters:Omega_GER
0.062  with SE:  0.001
GARCH parameters:Alpha_GER

0.115  with SE:  0.001
GARCH parameters:Beta_GER
0.853  with SE:  0.000
GARCH parameters:Omega_US
0.034  with SE:  0.000
GARCH parameters:Alpha_US
0.085  with SE:  0.000
GARCH parameters:Beta_US
0.899  with SE:  0.001
GARCH parameters:Omega_UK
0.040  with SE:  0.000
GARCH parameters:Alpha_UK
0.125  with SE:  0.001
GARCH parameters:Beta_UK
0.865  with SE:  0.001
==============================
==============================
DCC parameters:Alpha_DCC
0.046  with SE:  0.001
DCC parameters:Beta_DCC
0.689  with SE:  0.088
==============================
```

Figure C5: Estimation output from GARCH(1,1)-DCC(1,1) model

5. We estimate over the 2007-2010 sample a simple constant-mean GARCH(1,1)-DCC(1) model to filter the dynamics of the correlations of returns at daily frequency. The lines of code accomplishing these instructions are performed using functions that were already used in Appendix B. The estimates of the parameters are printed on the screen and are shown in Figure C5.The estimates obtained are rather typical, although the sum of the conditional covariance ($q_{ij,t}$) DCC coefficients (0.735) is rather low and surely lower than what we have already reported in Appendix B. Such a low persistence of covariance also explains why in Figure C4 the differences between CCC and DCC estimates are modest at best. The dynamic, predicted correlations implied by the model are plotted for the period January 2011 - December 2012 in Figure C6.

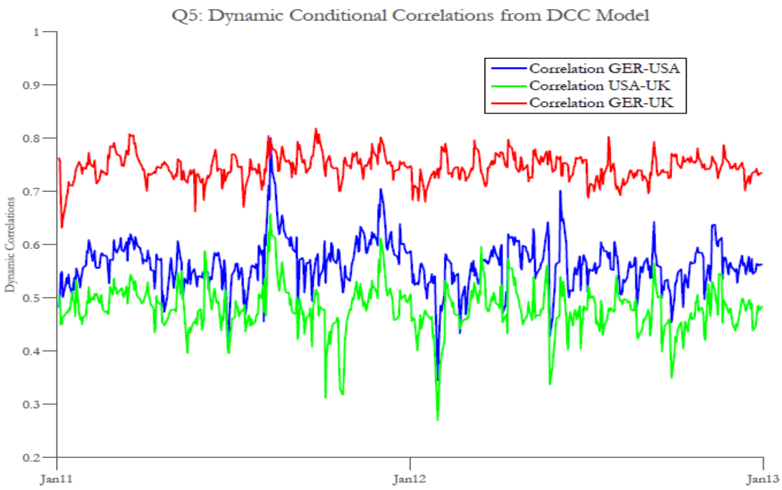

Figure C6: Predicted dynamic conditional correlations from DCC model

The plot confirms—in spite of some peaks and troughs—the approximate constancy of pairwise correlations from the estimated DCC model, even over the out-of-sample 2011-2012 period. These predicted correlations are obtained from the lines of code:

```
%A function designed to compute conditional variances and standardize returns
[sigma2 z]=garchfor(garch_p,ret1);
[rho12_for rho23_for rho13_for Gamma_dyn_garch]=
dcc_mvgarch_for(par_dcc,sret_GER,sret_US,sret_UK,Corr_matrix(1,2),Corr_matrix(1,3),Corr_matr
```

6. We compute and plot the 1% VaR from the DCC model of question 5 over the out-of-sample period January 1, 2011 - December 31, 2012 and compare it with the 1% VaR computed from the CCC and the RiskMetrics, exponentially smoothed DCC of questions 3 and 4. Figure C7 shows the 1% VaR from CCC and from two alternative implementations of DCC. Apart from the very early part of the out-of-sample period, when the RiskMetrics-based DCC risk measures are lower (in absolute value) than those under CCC and GARCH-DCC, the model give very similar results. As seen in Figure C3, the real differences can be computed with

respect to the unconditional VaR and passive measures of VaR that are based on the univariate time series of portfolio returns.

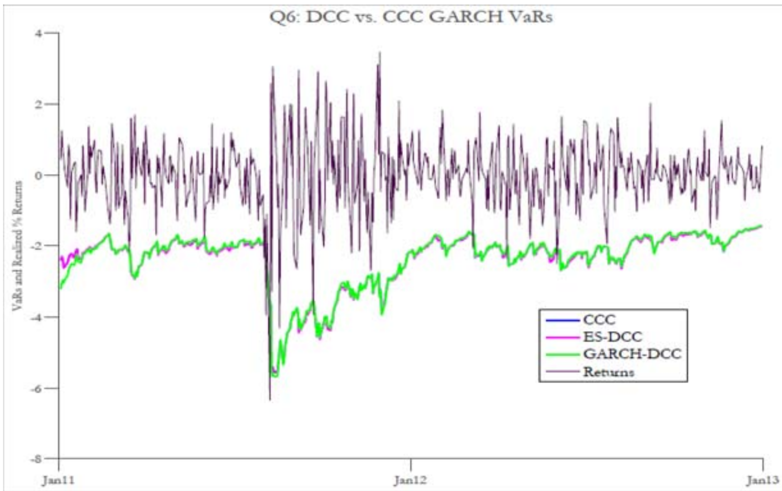

Figure C7: Recursive daily 1% VaR estimates from alternative CCC and DCC models

7. We estimate over a 2007-2010 sample a constant-mean Principal Component (also called Orthogonal) GARCH(1,1) model:

% ORTHOGONAL GARCH (or PC-GARCH) - estimation

z_pc=[z_GER(:,1) z_US(:,1) z_UK(:,1)];

[H_orth, parameters_orth, Ht_orth, stdresid_orth, stderrors_orth, A_orth, B_orth, weights_orth, principalcomponets, cumR2_orth] =...

o_mvgarch(z_pc,3,1,1);

This uses a toolbox function, o_mvgarch for which help information appears in the code. Figure C8 plots the predicted, one-day ahead dynamic volatilities and correlations resulting from the PC/Orthogonal GARCH model.

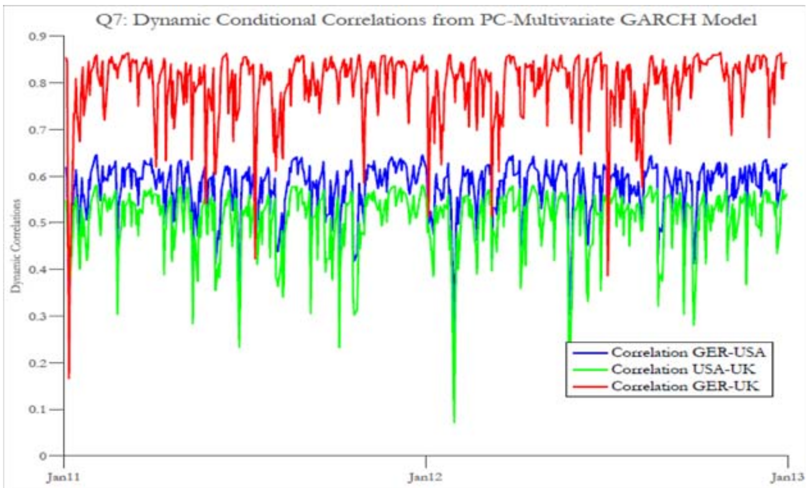

Figure C8: Predicted dynamic conditional correlations from PC/Orthogonal GARCH model

Figure C9 compares instead the the dynamic, conditional volatility with the time series we have derived from the DCC GARCH in question 5. Of course, the difference is given by the fact that the predicted volatilities in the case of a PC GARCH are obtained from the principal components of returns. The figure shows tha two series—with the exception of the first month of 2011—are quite similar, which indicates that an application of GARCH to principal components tends to give results that are close to those obtained from each of the return series individually.

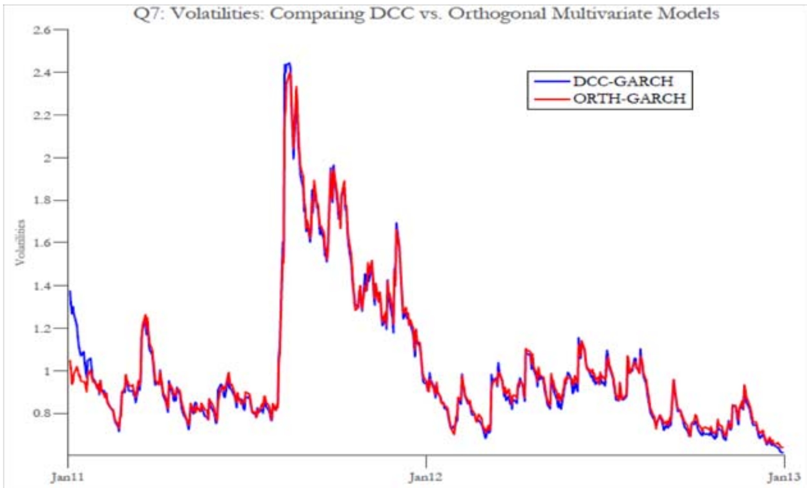

Figure C9: Predicted dynamic volatilities from PC/Orthogonal vs. DCC GARCH models

8. We have recursively computed and plotted 1% VaR over the out-of-sample period January 1, 2011 - December 31, 2012 using both historical and weighted historical simulations with a rolling window of $m = 252$ days and—in the case of weighted historical simulations—a decay factor of $\eta = 0.99$. We apply calculations to each index return series individually as well as to your portfolio returns:

```
m=253; %Length of the rolling window (approximately one year)
eta=0.99; %Decay parameter in weighted historical simulation
VaR_HS=NaN(final-last,4);
VaR_WHS=NaN(final-last,4);
for jj=1:4
for k=1:final-last
%Historical Simulation
VaR_HS(k,jj)=-quantile(RET_all(last+k-m:last+k,jj),p);
%Weighted Historical Simulation
[ret_sort I]=sort(RET_all(last+k-m:last+k,jj));
weight=(eta.^(m-I)*(1-eta)/(1-eta^m));
csum=cumsum(weight);
ind=rows(csum(csum<p));
```

```
if ind==0
VaR_WHS(k,jj)=-ret_sort(ind+1);
else
lambda=(p-csum(ind,1))./(csum(ind+1,1)-csum(ind));
VaR_WHS(k,jj)=-(lambda*ret_sort(ind+1,1)+(1-lambda)*ret_sort(ind,1));
end
end
```

Figure C10 shows results for individual stock markets, while Figure C11 for the equally-weighted portfolio. Clearly for all national stock markets, the late Summer and Fall of 2011 turned difficult for these rather simple methods of VaR calculation. In both pictures we also notice that the VaR estimated by simulation tend to be rather "generous", i.e., to imply high levels for the risk measure, generally higher than what is required by 99% of all realized returns over long periods of time.

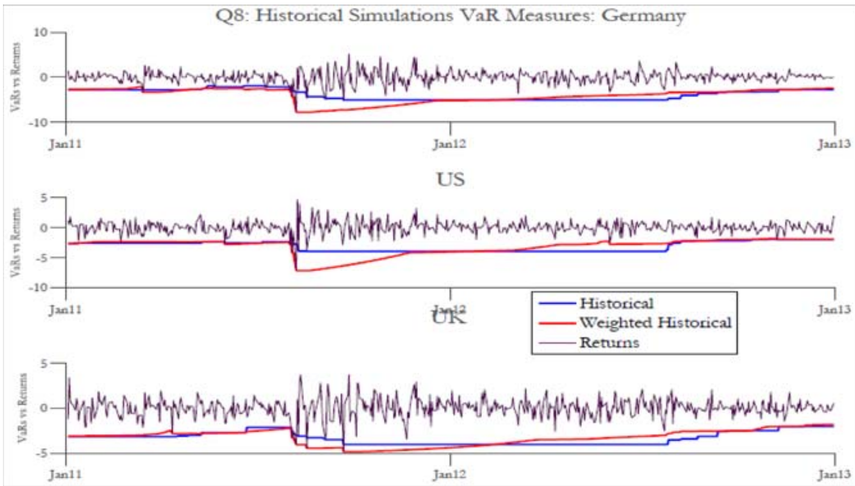

Figure C10: Recursive daily 1% VaR estimates from historical and weighted historical simulations

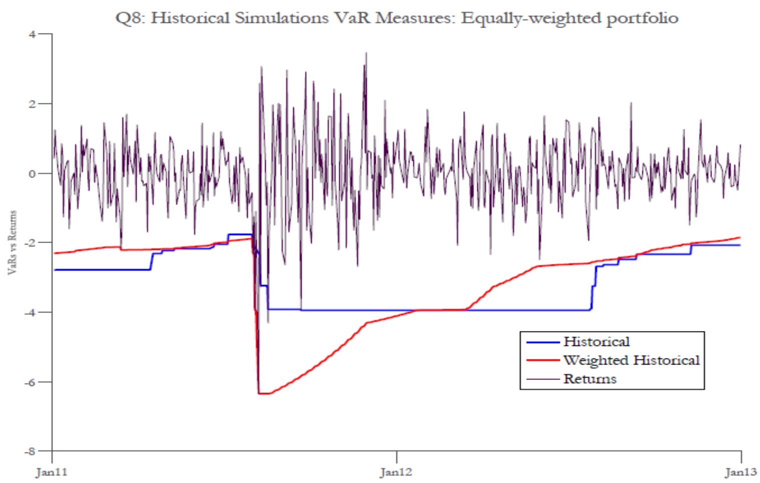

Figure C11: Recursive daily 1% VaR estimates from historical and weighted historical simulations

9. We estimate a BEKK-GARCH(1,1) model using the code:

```
first=datefind(datenum('01/03/2006'),date);
last=datefind(datenum('12/31/2007'),date);
[par_bekk, llk_bekk, Ht_bekk, likelihoods_bekk, stdresid_bekk, stderrors_bekk, A, B, scores_bekk] =...
full_bekk_mvgarch([ret_ger(first:last,1) ret_us(first:last,1) ret_uk(first:last,1)],1,1);
% Display estimation results
disp(['FULL BEKK-GARCH(1,1) PARAMETERS']);
disp([' Estimate Std. Error Robust t-stat']);
disp([par_bekk diag(stderrors_bekk) (par_bekk./diag(stderrors_bekk))]);
```

As advised, we use Kevin Shepard's *full_bekk_mvgarch* function obtaining

Iteration	Func-count	f(x)	Step-size	First-order optimality
0	25	5327.05		699
1	150	5325.47	2.98377e-06	434
2	175	5325.28	1	563
3	200	5324.7	1	232
4	225	5324.13	1	205
149	3850	5273.9	1	7.29
150	3875	5273.9	1	11.9
151	3900	5273.89	1	17.6
152	3925	5273.89	1	19.3
153	3950	5273.88	1	14.4
154	3975	5273.88	1	5.64
155	4000	5273.88	1	2.03
156	4025	5273.88	1	1.96
157	4050	5273.88	1	1.95

```
FULL BEKK-GARCH(1,1)  PARAMETERS
   Estimate Std. Error  Robust t-stat
   1.0e+03 *

    0.0001    0.0003    0.0005
    0.0000    0.0000    0.0152
    0.0001    0.0009    0.0001
    0.0001    0.0001    0.0010
   -0.0001    0.0000   -0.0048
   -0.0000    0.0000   -0.0004
    0.0002    0.0000    0.0405
   -0.0000    0.0005   -0.0001
   -0.0001    0.0002   -0.0004
   -0.0002    0.0000   -0.0164
    0.0001    0.0001    0.0019
   -0.0003    0.0000   -0.0843
    0.0001    0.0000    0.0190
    0.0002    0.0005    0.0003
    0.0004    0.0002    0.0023
    0.0010    0.0000    1.4305
    0.0000    0.0003    0.0001
    0.0001    0.0000    0.0022
    0.0001    0.0000    0.0478
    0.0010    0.0000    0.3050
    0.0001    0.0000    0.0032
   -0.0001    0.0000   -0.0051
   -0.0001    0.0006   -0.0002
    0.0008    0.0000    0.2506
```

Figure C12: Parameter estimates from BEKK GARCH(1,1) model

As a matter of fact, even on a relatively new and decent laptop, because of its many parameters, estimation of a BEKK model may take up to 5 minutes.

References

[1] Alexander, C., 2001. *Market Models*. John Wiley: New York.

[2] Andersen T., T. Bollerslev, P. Christoffersen, and F. Diebold, 1998. "Practical Volatility and Correlation Modeling for Financial Market Risk Management," in Risks of Financial Institutions, University of Chicago Press, pp. 513-548.

[3] Bauwens, L., S., Laurent, and J., Rombouts, 2006. "Multivariate GARCH Models: A Survey", *Journal of Applied Econometrics*, 21, 79-109.

[4] Comte, F., and O., Lieberman, 2003. "Asymptotic theory for multivariate GARCH processes", *Journal of Multivariate Analysis*, 84, 61-84.

[5] Engle, R., 2002. "Dynamic Conditional Correlation: A Simple Class of Multivariate GARCH Models", *Journal of Business and Economic Statistics*, 20, 339-350.

[6] Engle R., and F., Kroner, 1995. "Multivariate Simultaneous Generalized ARCH", *Econometric Theory*, 11, 122-150.

[7] Fama, E., and K., French, 1992. "The Cross-Section of Expected Stock Returns", *Journal of Financem* 47, 427-465.

[8] Gourieroux, C., 1997. *ARCH Models and Financial Applications*. Springer-Verlag: New York.

[9] Green, W., 2008. *Econometric Analysis*, 6th Edition, Prentice Hall.

[10] Jeantheau T., 1998. "Strong Consistency of Estimators for Multivariate ARCH Models", *Econometric Theory,* 14, 70-86.

[11] Hamilton, J., 2004. *Time Series Analysis*, Princeton University Press.

[12] Kroner F., K., and V., Ng, 1998. "Modelling Asymmetric Comovements of Asset Returns", *Review of Financial Studies*, 11, 817-844.

[13] Ledoit O., P., Santa-Clara, and M., Wolf, 2003. "Flexible Multivariate GARCH Modeling with an Application to International Stock Markets", *Review of Economics and Statistics*, 85, 735-747.

[14] Litterman R., and K. Winkelmann. (1998). "Estimating Covariance Matrices", Goldman Sachs, Risk Management Series.

[15] Tse Y., K., and A., KC, Tsui, 2002. "A Multivariate GARCH Model with Time-Varying Correlations", *Journal of Business and Economic Statistics*, 20, 351-362.