

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 $\mathbf{R}_t \equiv [R_t^1 \ R_t^2 \ \dots \ R_t^N]'$, then the variance of a random vector turns out to be a matrix of second moments, i.e., variances and covariances:¹

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

¹It 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, $\text{Cov}[\mathbf{R}_t]$ to the conditional covariance matrix, $\text{Cov}_t[\mathbf{R}_{t+1}] \equiv \text{Cov}[\mathbf{R}_{t+1}|\mathfrak{S}_t]$.

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

Clearly, all variances are collected on the main diagonal, while all covariances are collected off the main diagonal. Moreover, because $\text{Cov}[R_t^i, R_t^j] = \text{Cov}[R_t^j, R_t^i]$ from simple properties of expectations, then $\text{Cov}[\mathbf{R}_t]$ is by construction a symmetric matrix. For instance, any portfolio choice methodology is clearly based on knowledge or estimation of $\text{Cov}[\mathbf{R}_t]$, 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 be 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 quant 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.

2. Motivation: Passive vs. Active Risk Management

Suppose you are a risk manager in charge of measuring and controlling risk for a given portfolio, whose return we call R_{t+1}^{ptf} . Although it is easy to read what follows with reference to a portfolio of securities, more generally, this could be a portfolio of loans or other OTC positions/exposures. Such a portfolio *ptf* is composed of N positions such that:

$$\begin{aligned} R_{t+1}^{ptf} &\equiv w_1 R_{t+1}^1 + w_2 R_{t+1}^2 + \dots + w_p R_{t+1}^p \\ &= \sum_{i=1}^N w_i R_{t+1}^i = \mathbf{w}' \mathbf{R}_{t+1}, \end{aligned} \quad (1)$$

where the vector $\mathbf{w} \equiv [w_1 \ w_2 \ \dots \ w_N]'$ represents the weights that apply between time t and $t + 1$. Formally, one could even be more precise and write that $R_{t+1}^{ptf} \equiv \sum_{i=1}^N w_t^i R_{t+1}^i = \mathbf{w}_t' \mathbf{R}_{t+1}$ to emphasize that $\mathbf{w}_t \equiv [w_t^1 \ w_t^2 \ \dots \ w_t^N]'$ has been selected at time t . The first, simplest choice, is to ignore the underlying structure and origins of R_{t+1}^{ptf} in (1): because once the summation on the right-hand side of (1) has been performed, we are likely to have available a time series $\{R_{t+1}^{ptf}\}_{t=0}^T$ of data on portfolio returns, one possibility is to just use such returns and apply the univariate methods covered in chapters 4 and 5. For instance, under the assumption of multivariate normality

of the vector of returns \mathbf{R}_{t+1} , $\mathbf{R}_{t+1} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, one may compute

$$VaR_t^{ptf}(p) = -\hat{\sigma}_{t+1}^{ptf} \Phi^{-1}(p) - \hat{\mu}_{t+1}^{ptf},$$

where $\hat{\sigma}_{t+1}^{ptf}$ represents a forecast of the quantity $\sqrt{Var_t[R_{t+1}^{ptf}]}$, i.e., of volatility at time $t+1$ obtained from some conditional heteroskedastic model estimated on (applied to) the time series $\{R_{t+1}^{ptf}\}_{t=0}^T$, for instance $\hat{\sigma}_{t+1}^{ptf} = \sqrt{\hat{\omega} + \hat{\alpha}(R_t^{ptf})^2 + \hat{\beta}(\hat{\sigma}_t^{ptf})^2}$. This is called a *passive* risk management approach: it is passive because all of these steps are just fine of course, but have a considerable counterindication, they condition on the vector of portfolio weights between time t and $t+1$ being given and constant, $\mathbf{w}_{t+1} = \mathbf{w}_t \equiv [w_t^1 \ w_t^2 \ \dots \ w_t^N]'$. Formally, we should then write²

$$VaR_t^{ptf}(p; \mathbf{w}_t) = -\hat{\sigma}_{t+1}^{ptf}(\mathbf{w}_t) \Phi^{-1}(p)$$

because both $\hat{\sigma}_{t+1}^{ptf}$ and $\hat{\mu}_{t+1}^{ptf}$ condition on the time t portfolio weights, \mathbf{w}_t . Passive risk management implies that the quant steps (modelling, estimation, forecasting, etc.) are simple because they are based on univariate tools. However, the results produced by passive portfolio management are fragile by construction. For instance, even though we may be able to say that for all the assets ($i = 1, 2, \dots, N$) $\hat{\sigma}_{t+2}^i \simeq \hat{\sigma}_{t+1}^i$, notice that $VaR_{t+1}^{ptf}(p)$ may be completely different from $VaR_t^{ptf}(p)$ for an obvious reason: if the weights \mathbf{w}_{t+1} differ from the weights \mathbf{w}_t , then $\hat{\sigma}_{t+2}^{ptf}(\mathbf{w}_{t+1}) \neq \hat{\sigma}_{t+1}^{ptf}(\mathbf{w}_t)$ even though $\hat{\sigma}_{t+2}^i \simeq \hat{\sigma}_{t+1}^i$ for all assets in the portfolio. Probably, in practice, this means that a passive risk management user will be forced to repeat estimation and all calculations at each point in time, or at least every time the structure of the portfolio is modified. Moreover, suppose that your interest lies in understanding how your risk measures could change if you alter the structure of the portfolio weights, i.e., something like

$$\frac{\partial VaR_t^{ptf}(p; \mathbf{w}_t)}{\partial w_t^i} \text{ for some } i = 1, 2, \dots, N.$$

Clearly, econometric methods simply applied to the aggregate time series $\{R_{t+1}^{ptf}\}_{t=0}^T$ will be incapable of accomplishing that.

A risk manager may indeed resort to *active* instead of passive risk management methods. Adopting active methods is equivalent to using *multivariate* econometric modelling. The advantage of active methods is that the individual, asset- or security-specific contributions to risk (or portfolio performance, in the case of asset allocation applications) can be estimated, like $\partial VaR_t^{ptf}(p; \mathbf{w}_t) / \partial w_t^i$ in the example above. To adopt a multivariate model means to switch focus from modelling and forecasting $Var_t[R_{t+1}^{ptf}]$ to $Cov_t[\mathbf{R}_{t+1}]$ or

$$\text{from } Var_t[R_{t+1}^{ptf}] \text{ to } Var_t[\mathbf{w}'_t \mathbf{R}_{t+1}] = \mathbf{w}'_t Cov_t[\mathbf{R}_{t+1}] \mathbf{w}_t,$$

where (as we have already noted) there is no difference between $Cov_t[\mathbf{R}_{t+1}]$ and $Var_t[\mathbf{R}_{t+1}]$. Moreover, remember that for any random vector \mathbf{X}_{t+1} , $Cov_t[\mathbf{w}'_t \mathbf{X}_{t+1}] = \mathbf{w}'_t Cov_t[\mathbf{X}_{t+1}] \mathbf{w}_t$ in a way similar

²From now on, we resume assuming that $\mu_{t+1}^{ptf}(\boldsymbol{\omega}_t) = 0$ and we omit it.

to the fact that when $N = 1$, $Var_t[w_t X_{t+1}] = w_t^2 Var_t[X_{t+1}]$. However, to come up with models and estimation methods for $Cov_t[\mathbf{R}_{t+1}]$ is a much more serious endeavour than for $Var_t[R_{t+1}^{ptf}]$.

We have already stated in the Introduction that modelling $Cov_t[\mathbf{R}_{t+1}]$ means to model covariances and/or correlations. Note that the statements are not equivalent, because³

$$Corr_t[R_{t+1}^i, R_{t+1}^j] \equiv \rho_{ij,t+1} = \frac{Cov_t[R_{t+1}^i, R_{t+1}^j]}{\sqrt{Var_t[R_{t+1}^i]} \sqrt{Var_t[R_{t+1}^j]}} = \frac{\sigma_{ij,t+1}}{\sigma_{i,t+1} \sigma_{j,t+1}},$$

so that while modelling and forecasting correlations requires modelling and forecasting covariances, it also implies that one can model and forecast variances, which we can do already using the methods developed in chapters 4 and 5.

The true nature of active risk management and the fact that it involves correlations and covariances emerges from an example for the case $N = 2$:

$$\begin{aligned} Var_t[\mathbf{w}'_t \mathbf{R}_{t+1}] &= \mathbf{w}'_t Cov_t[\mathbf{R}_{t+1}] \mathbf{w}_t \\ &= [w_t^1 \ w_t^2] \begin{bmatrix} \sigma_{1,t+1}^2 & \sigma_{12,t+1} \\ \sigma_{12,t+1} & \sigma_{2,t+1}^2 \end{bmatrix} \begin{bmatrix} w_t^1 \\ w_t^2 \end{bmatrix} \\ &= (w_t^1)^2 \sigma_{1,t+1}^2 + 2w_t^1 w_t^2 \sigma_{12,t+1} + (w_t^2)^2 \sigma_{2,t+1}^2 \\ &= (w_t^1)^2 \sigma_{1,t+1}^2 + (w_t^2)^2 \sigma_{2,t+1}^2 + 2w_t^1 w_t^2 \rho_{12,t+1} \sigma_{1,t+1} \sigma_{2,t+1}. \end{aligned} \quad (2)$$

The last line shows that under an active risk management approach ($Var_t[\mathbf{w}'_t \mathbf{R}_{t+1}]$), also dynamic forecasts of either covariance ($\sigma_{12,t+1}$) or, equivalently, correlation ($\rho_{12,t+1}$) are required. Obviously, the very last line derives from the definition of a correlation, $\sigma_{12,t+1} = \rho_{12,t+1} \sigma_{1,t+1} \sigma_{2,t+1}$. For a general $N \geq 2$, the expression in (2) generalizes to:

$$\begin{aligned} Var_t[\mathbf{w}'_t \mathbf{R}_{t+1}] &= \mathbf{w}'_t Cov_t[\mathbf{R}_{t+1}] \mathbf{w}_t = \sum_{i=1}^N \sum_{j=1}^N w_t^i w_t^j \sigma_{ij,t+1} \\ &= \sum_{i=1}^N \sum_{j=1}^N w_t^i w_t^j \rho_{ij,t+1} \sigma_{i,t+1} \sigma_{j,t+1} \\ &= \sum_{i=1}^N (w_t^i)^2 \sigma_{i,t+1}^2 + 2 \sum_{i=1}^N \sum_{j=i+1}^N w_t^i w_t^j \rho_{ij,t+1} \sigma_{i,t+1} \sigma_{j,t+1}, \end{aligned} \quad (3)$$

which still emphasizes that not all the N variance forecasts will matter, but also the $N(N-1)/2$ correlation forecasts.⁴ For instance, in simple risk-management applications, under active risk management we shall have:

$$VaR_t^{ptf}(p) = -\{\mathbf{w}'_t Cov_t[\mathbf{R}_{t+1}] \mathbf{w}_t\} \Phi^{-1}(p) = - \left[\sum_{i=1}^N \sum_{j=1}^N w_t^i w_t^j \hat{\rho}_{ij,t+1} \hat{\sigma}_{i,t+1} \hat{\sigma}_{j,t+1} \right] \Phi^{-1}(p),$$

³In this chapter $\sigma_{ij,t+1}$ and σ_{t+1}^{ij} , $\sigma_{i,t+1}$ and σ_{t+1}^i mean the same. Where one places the indices is irrelevant, provided a Reader is alerted of the meaning. You should also recall that correlations simply represent measure of *linear dependence* between pairs of random variables, meaning that more complex form of dependence may exist that correlations will not necessarily capture.

⁴Relevant correlation forecasts are only $n(n-1)/2$ because of the symmetry of the covariance matrix, i.e., $\rho_{ij,t+1} = \rho_{ji,t+1}$ for all $i \neq j$.

which clearly leads to expressions for the partial derivatives mentioned above, of the type

$$\frac{\partial VaR_t^{ptf}(p)}{\partial w_t^i} = \left[-2w_t^i \sigma_{i,t+1}^2 + 2 \sum_{j=1, j \neq i}^N w_t^j \hat{\rho}_{ij,t+1} \hat{\sigma}_{i,t+1} \hat{\sigma}_{j,t+1} \right] \Phi^{-1}(p).$$

This gives the contribution of the second moments of the i th asset to the p -percent VaR of the portfolio.

The expression in (3) makes it clear that in general, an active risk management problem will involve the forecasts of N variances and of $N(N - 1)/2$ covariances (or correlations). While up to this point we have generally assumed that given a conditional heteroskedastic model, we always have sufficient observations to proceed to estimation, we immediately note that when it comes to multivariate covariance matrix estimation and forecasting, the availability of sufficiently long time series may become an issue that requires attention.⁵ For instance, with only 15 assets in a portfolio—which is a rather sensible and commonly seen portfolio—you will need: (i) 15 variance forecasts; (ii) $(15 \times 14/2) = 105$ correlation forecasts, for a total of 120 parameters or moments to forecasts. Suppose, for simplicity, that variances and covariances are constant over time. Then the 120 objects that you care for in this example, simply become parameters to estimate, $\{\hat{\sigma}_i\}_{i=1}^{15}$ and $\{\hat{\sigma}_{ij}\}_{i=1, j>i}^N$. At this point, with 15 series of return data (because with 15 assets you will have at least these 15 time series), note that a total of 120 parameters to be estimated on 15 series, gives you $120/15 = 8$ data points per series. Even though you may think that 15 time series are a lot, for each of them you will need at least 8 observations in order to proceed. However: would you ever estimate 120 parameters using exactly 120 observations? Hopefully not. In fact, time series econometricians normally use a simple rule-of-thumb by which one should always have 20 observations *per parameter* before proceeding to any econometric analysis. The ratio between the total number of observations and the number of parameters to be estimated is called *saturation ratio*. In this case, $20 \times 120 = 2,400$ observations. This means that for each series, you should have $2400/15 = 160$ observations per series before seriously thinking of tackling this problem. 160 observations per series mean that you should recover almost 14 years of monthly data; or 32 weeks of daily data. These requirements are moderate, but already not completely negligible when you deal with over-the-counter instruments or newly floated stocks in the aftermath of IPOs.

If you worked through this example afresh after having increased the number of assets to something even more realistic such as 100 or so assets, you will come to realize that there is a new dimension of multivariate time series problems that was unknown before: because the size of the covariance matrix grows as a function of N^2 (formally, $N + N(N - 1)/2 = O(N^2)$), the size of the estimation problem and the corresponding data requirements grow quadratically in the number of

⁵What follows assumes that you have realized prior to today that if you have M parameters to estimate, then you will need at least $T > M$ observations. However, it is common to expect that you will have many more observations than parameters to estimate, say $T = \kappa M$ where κ is at least 10 or 20 and will be called *saturation ratio* in what follows.

assets (i.e., very quickly).⁶ In fact, you can read most of the material that follows not only as an attempt to develop good multivariate econometric methods that may accurately forecast variances and covariances, but also as a way to deal with the issue of excessively high number of parameters that estimating covariance matrices implies.

3. Exposure Mapping (Factor-Based) Approach

A simple way to reduce the dimensionality of the problem of estimating and forecasting portfolio variance is to impose a factor structure using observed returns as factors. Although, as we shall see below, the method is considerably more general, let's start from an analysis of the CAPM. Assume that the CAPM *holds exactly*, i.e., that it perfectly describes portfolio returns:

$$R_{t+1}^{ptf} = R^f + \beta_{ptf}[R_{t+1}^M - R^f], \quad (4)$$

where R^f is the riskless rate (assumed to be constant, just to simplify) and R_{t+1}^M is the return on the market portfolio (you know what this means from your asset pricing courses). Also suppose that you have already managed to estimate the beta of your portfolio, for instance, using simple OLS methods:

$$\hat{\beta}_{ptf} = \frac{\widehat{Cov}[R_{t+1}^{ptf} - R^f, R_{t+1}^M - R^f]}{\widehat{Var}[R_{t+1}^M - R^f]},$$

where “hats” refer to sample estimates, i.e., obtained from the data. Once you are through with that, then

$$\widehat{Var}_t[R_{t+1}^{ptf}] = \hat{\beta}_{ptf}^2 \widehat{Var}[R_{t+1}^M - R^f]$$

which is simple enough. At this point, you should be confused (for a short time only) because the two previous formulas seemingly depend on portfolio returns only (besides the time series of market portfolio returns). In Section 2 we said that if you just use realized portfolio returns then your approach will be a passive one, with all its limitations. However, you would be incorrect in your confusion, because under the exposure mapping approach, it turns out that the passive and active approaches are identical. Equivalently (and these are excellent news), it turns out that a passive exposure-based approach to variance forecasting gives the same result as an active approach; therefore the easier, passive approach is preferred. The reason for this surprising result is that—as you should recall from your asset pricing courses—CAPM (more generally, factor) betas are *linear* in portfolio weights, i.e.,

$$\beta_{ptf}(\mathbf{w}_t) = \sum_{i=1}^N w_t^i \beta_i, \quad (5)$$

⁶Although a covariance matrix contains n^2 elements, it is symmetric and collects variances on its main diagonal. This means that the number of distinct elements (parameters) collected in $\Sigma \equiv Cov_t[\mathbf{R}_{t+1}]$ is $n(n+1)/2$, of which n are variances and $n(n+1)/2 - n = n(n-1)/2$ correlations.

and $\hat{\beta}_{ptf}(\mathbf{w}_t) = \sum_{i=1}^N w_t^i \hat{\beta}_i$, i.e., the beta of a portfolio is the weighted sum of the individual betas, with weights equal to portfolio weights. (5) derives from a well-known properties of covariances:

$$\begin{aligned}\beta_{ptf}(\mathbf{w}_t) &= \frac{Cov[R_{t+1}^{ptf} - R^f, R_{t+1}^M - R^f]}{Var[R_{t+1}^M - R^f]} \\ &= \frac{Cov[\sum_{i=1}^N w_t^i R_{t+1}^i - R^f, R_{t+1}^M - R^f]}{Var[R_{t+1}^M - R^f]} = \frac{\sum_{i=1}^N w_t^i Cov[R_{t+1}^i - R^f, R_{t+1}^M - R^f]}{Var[R_{t+1}^M - R^f]} \\ &= \sum_{i=1}^N w_t^i \frac{Cov[R_{t+1}^i - R^f, R_{t+1}^M - R^f]}{Var[R_{t+1}^M - R^f]} = \sum_{i=1}^N w_t^i \beta_i.\end{aligned}$$

Therefore, it is easy to see why passive and active risk management need to give the same result:

$$\begin{aligned}Var_t[R_{t+1}^{ptf}] &= \left[\sum_{i=1}^N w_t^i \hat{\beta}_i \right]^2 Var[R_{t+1}^M - R^f] \quad (\text{passive}) \\ &= Var_t \left[\sum_{i=1}^N w_t^i R_{t+1}^i \right] = \sum_{i=1}^N \sum_{j=1}^N w_t^i w_t^j Cov_t [R_{t+1}^i, R_{t+1}^j] \\ &= \sum_{i=1}^N \sum_{j=1}^N w_t^i w_t^j Cov_t [R^f + \beta_i(R_{t+1}^M - R^f), R^f + \beta_j(R_{t+1}^M - R^f)] \\ &= \sum_{i=1}^N \sum_{j=1}^N w_t^i w_t^j Cov_t [\beta_i(R_{t+1}^M - R^f), \beta_j(R_{t+1}^M - R^f)] \\ &= \sum_{i=1}^N \sum_{j=1}^N w_t^i w_t^j \beta_i \beta_j Cov_t [R_{t+1}^M - R^f, R_{t+1}^M - R^f] \\ &= \sum_{i=1}^N \sum_{j=1}^N w_t^i w_t^j \beta_i \beta_j Var_t [R_{t+1}^M - R^f].\end{aligned}$$

At this point, exploiting our assumption that the CAPM holds exactly, we have:

$$\begin{aligned}Var_t[R_{t+1}^{ptf}] &= \sum_{i=1}^N (w_t^i)^2 \underbrace{\beta_i^2 Var_t [R_{t+1}^M - R^f]}_{Var_t[R_{t+1}^i] \text{ from CAPM}} + 2 \sum_{i=1}^N \sum_{j=i+1}^N w_t^i w_t^j \underbrace{\beta_i \beta_j Var_t [R_{t+1}^M - R^f]}_{Cov[R_{t+1}^i, R_{t+1}^j] \text{ from CAPM}} \\ &= \sum_{i=1}^N (w_t^i)^2 \sigma_{i,t+1}^2 + 2 \sum_{i=1}^N \sum_{j=i+1}^N w_t^i w_t^j \sigma_{ij,t+1} = \mathbf{w}_t' Cov_t[\mathbf{R}_{t+1}] \mathbf{w}_t \quad (\text{active}),\end{aligned}$$

which shows that starting from a seemingly passive approach, one can get to an active one thanks to the properties of the CAPM.⁷ In this case, assuming normality,

$$\frac{\partial VaR_t^{ptf}(p)}{\partial w_t^i} = \left[-2w_t^i \beta_i^2 Var_t(R_{t+1}^M - R^f) + 2 \sum_{j=1, j \neq i}^N w_t^j \beta_i \beta_j Var_t(R_{t+1}^M - R^f) \right] \Phi^{-1}(p),$$

i.e., the typical partial derivatives of interest in active risk management applications can all be re-expressed in terms of individual asset betas.

⁷Make sure to understand why under the CAPM, $Cov[R_{t+1}^i, R_{t+1}^j] = \beta_i \beta_j Var_t [R_{t+1}^M - R^f]$.

Of course, these results obtain only when the CAPM applies exactly, or equivalently, when the returns of none of the assets in the portfolio contains any firm- or security-specific *idiosyncratic risk*. In other words, asset returns must be entirely explained by systematic risks. To see what happens in case there is any residual, idiosyncratic risk left in the process of individual asset returns, note that if

$$R_{t+1}^i = R^f + \beta_i[R_{t+1}^M - R^f] + \epsilon_{t+1}^i \quad (i = 1, 2, \dots, N),$$

where ϵ_{t+1}^i captures such idiosyncratic risk, then

$$\begin{aligned} \text{Var}_t[R_{t+1}^{ptf}] &= \left\{ \sum_{i=1}^N (w_t^i)^2 \beta_i^2 \text{Var}_t [R_{t+1}^M - R^f] + \sum_{i=1}^N (w_t^i)^2 \text{Var}_t[\epsilon_{t+1}^i] \right\} + \\ &+ 2 \left\{ \sum_{i=1}^N \sum_{j=i+1}^N w_t^i w_t^j \beta_i \beta_j \text{Var}_t [R_{t+1}^M - R^f] + \sum_{i=1}^N \sum_{j=i+1}^N w_t^i w_t^j \text{Cov}_t[\epsilon_{t+1}^i, \epsilon_{t+1}^j] \right\} \\ &= \left[\sum_{i=1}^N w_t^i \beta_i \right]^2 \text{Var}_t [R_{t+1}^M - R^f] + \sum_{i=1}^N (w_t^i)^2 \text{Var}_t[\epsilon_{t+1}^i] \\ &= \text{passive mode} + \text{active mode applied to } \sum_{i=1}^N (w_t^i)^2 \text{Var}_t[\epsilon_{t+1}^i]. \end{aligned}$$

Even though the definition of idiosyncratic risk implies that these risks must be uncorrelated across assets so that $\text{Cov}_t[\epsilon_{t+1}^i, \epsilon_{t+1}^j] = 0 \quad \forall i \neq j$, fact remains that the terms $\text{Var}_t[\epsilon_{t+1}^i]$ for $i = 1, 2, \dots, N$ need to be predicted on the basis of some model that cannot be the CAPM itself (the CAPM is silent about idiosyncratic risk by construction, being only about systematic risk). This means at this stage one will be forced to go back to her Matlab copy to perform estimation and forecasting of $\{\text{Var}_t[\epsilon_{t+1}^i]\}_{i=1}^N$ using econometric methods, for instance N different GARCH models for idiosyncratic risk,

$$\sigma_{\epsilon_i, t+1}^2 = \delta_0^i + \delta_1^i \epsilon_{i, t+1}^2 + \delta_2^i \sigma_{\epsilon_i, t}^2,$$

one for each of the assets in our portfolio. This is of course very active but at the same time also rather painful which goes to show that the advantages of active and passive management are lost because of the very existence of idiosyncratic risk.

3.1. Applications to risk management

Assuming the CAPM in (4) is subject to IID shocks from a Gaussian distribution with zero cross-sectional correlations (i.e., $\text{Cov}[\epsilon_{i, t+1}, \epsilon_{j, t+1}] = 0$ for $i \neq j = 1, 2, \dots, N$),

$$R_{i, t+1} = R^f + \beta_i(R_{Mkt, t+1} - R^f) + \epsilon_{i, t+1} \quad \epsilon_{i, t+1} \text{ IID } N(0, \sigma_{i, t+1}^2), \quad (6)$$

we now derive an expression for the 1% VaR of a portfolio characterized by weights $\{w_1, w_2, \dots, w_N\}$ and that wants to emphasize that such an expression involves only quantities that are specific to each of the N assets, i.e., their portfolio weights, their betas, their (estimated) idiosyncratic risk levels, etc.

Because $R_{PF,t+1} \equiv \sum_{i=1}^N w_i R_{i,t+1}$ and the linear properties of the covariance (i.e., $Cov[\delta_1 X + \delta_2 Y, Z] = \delta_1 Cov[X, Z] + \delta_2 Cov[Y, Z]$), we know that the estimate of β_{PF} in

$$R_{PF,t+1} = R^f + \hat{\beta}_{PF}(R_{Mkt,t+1} - R^f) + \hat{\epsilon}_{PF,t+1},$$

is $\hat{\beta}_{PF} = \sum_{i=1}^N w_i \hat{\beta}_i$ while

$$\begin{aligned} Var[\hat{\epsilon}_{PF,t+1}] &= Var \left[\sum_{i=1}^N w_i (R_{i,t+1} - R^f) - \sum_{i=1}^N w_i \hat{\beta}_i (R_{Mkt,t+1} - R^f) \right] \\ &= Var \left[\sum_{i=1}^N w_i (R_{i,t+1} - R^f - \hat{\beta}_i (R_{Mkt,t+1} - R^f)) \right] \\ &= Var \left[\sum_{i=1}^N w_i \hat{\epsilon}_{i,t+1} \right] = \sum_{i=1}^N w_i^2 \hat{\sigma}_{i,t+1}^2, \end{aligned}$$

which requires assuming that $Cov[\epsilon_{i,t+1}, \epsilon_{j,t+1}] = 0$ for $i \neq j = 1, 2, \dots, N$ to also obtain $Cov[\hat{\epsilon}_{i,t+1}, \hat{\epsilon}_{j,t+1}] = 0$. Therefore

$$\sigma_{PF,t+1}^2 = \hat{\beta}_{PF}^2 \sigma_{Mkt,t+1}^2 + \sigma_{\epsilon_{PF,t+1}}^2 = \left(\sum_{i=1}^N w_i \hat{\beta}_i \right)^2 \sigma_{Mkt,t+1}^2 + \sum_{i=1}^N w_i^2 \hat{\sigma}_{i,t+1}^2$$

and under normality,

$$\begin{aligned} VaR_{t+1}(p) &= - \left\{ \left(\sum_{i=1}^N w_i \hat{\beta}_i \right)^2 \sigma_{Mkt,t+1}^2 + \sum_{i=1}^N w_i^2 \hat{\sigma}_{i,t+1}^2 \right\}^{1/2} \Phi^{-1}(p) + \\ &\quad - \left\{ R^f + \sum_{i=1}^N w_i \hat{\beta}_i \hat{E}(R_{Mkt,t+1} - R^f) \right\}, \end{aligned}$$

where $\hat{E}(R_{Mkt,t+1} - R^f)$ can be simply estimated as the sample mean excess return on the market portfolio. Notice that if you compute VaRs assuming that the CAPM holds, then

$$\mu_{PF,t+1} \equiv R^f + \sum_{i=1}^N w_i \hat{\beta}_i \hat{E}(R_{Mkt,t+1} - R^f) = R^f \left\{ 1 - \sum_{i=1}^N w_i \hat{\beta}_i \right\} + \sum_{i=1}^N w_i \hat{\beta}_i \hat{E}(R_{Mkt,t+1})$$

will generally be non-zero, unless $\sum_{i=1}^N w_i \hat{\beta}_i \hat{E}(R_{Mkt,t+1}) = 0$ and $\sum_{i=1}^N w_i \hat{\beta}_i = \hat{\beta}_{PF} = 1$, which are rather special restrictions that in general will not be satisfied. In particular, in the case of a 1% VaR, we will have that under the assumption of $\epsilon_{i,t+1}$ IID $N(0, \sigma_i^2)$,

$$\epsilon_{PF,t+1} \equiv \sum_{i=1}^N w_i \epsilon_{i,t+1} \sim \text{IID } N \left(0, \sum_{i=1}^N w_i^2 \sigma_{i,t+1}^2 \right) \quad (7)$$

so that

$$VaR_{t+1}(0.01) = 2.33 \left\{ \left(\sum_{i=1}^N w_i \hat{\beta}_i \right)^2 \sigma_{Mkt,t+1}^2 + \sum_{i=1}^N w_i^2 \hat{\sigma}_{i,t+1}^2 \right\}^{1/2} - \left\{ R^f + \sum_{i=1}^N w_i \hat{\beta}_i \hat{E}(R_{Mkt,t+1} - R^f) \right\},$$

and clearly the risk exposure will entirely depend on the portfolio composition, the betas of the securities in the portfolio, and their idiosyncratic risk coefficients. Finally, the result in (7) shows

the key role played by the assumption that $\epsilon_{i,t+1}$ IID $N(0, \sigma_{\epsilon_i}^2)$ for $i = 1, 2, \dots, N$. As for the shocks, their importance needs little emphasizing: while the CAPM as an asset pricing model, $E[R_{i,t+1}] = R^f + \beta_i E[R_{Mkt,t+1} - R^f]$, has no empirical implications because it just pins down expected returns, (6) can be used in risk management but it requires us making some assumptions— as we did—on the distribution of the $\epsilon_{i,t+1}$ as well as on $Cov[\epsilon_{i,t+1}, \epsilon_{j,t+1}]$.

Because our result has emphasized the specific role of individual asset portfolio weights, we can also quantify the VaR (total risk exposure) loss of changing the portfolio weights in the direction of realizing an optimal degree of diversification such that the condition

$$\sum_{i=1}^N \check{w}_i^2 \hat{\sigma}_{i,t+1}^2 \simeq 0,$$

eventually holds as N grows larger and larger. In fact, note that if the portfolio is not well-diversified, then

$$\begin{aligned} VaR_{t+1}(p) = & - \left\{ \left(\sum_{i=1}^N w_i \hat{\beta}_i \right)^2 \sigma_{Mkt,t+1}^2 + \sum_{i=1}^N w_i^2 \hat{\sigma}_{i,t+1}^2 \right\}^{1/2} \Phi^{-1}(p) + \\ & - \left\{ R^f + \sum_{i=1}^N w_i \hat{\beta}_i \hat{E}(R_{Mkt,t+1} - R^f) \right\}. \end{aligned}$$

If the portfolio is well diversified, then

$$VaR_{t+1}^{well-d}(p) = - \left(\sum_{i=1}^N \check{w}_i \hat{\beta}_i \right) \sigma_{Mkt,t+1} \Phi^{-1}(p) - \left\{ R^f + \sum_{i=1}^N \check{w}_i \hat{\beta}_i \hat{E}(R_{Mkt,t+1} - R^f) \right\}.$$

Clearly, because $\Phi^{-1}(p) < 0$ for $p < 0.5$, the suspicion is that

$$VaR_{t+1}(p) \geq VaR_{t+1}^{well-d}(p)$$

and the decline can be quantified as the disappearance of the term

$$- \left\{ \sum_{i=1}^N w_i^2 \hat{\sigma}_{i,t+1}^2 \right\}^{1/2} \Phi^{-1}(p) > 0.$$

However to compute the actual difference is a bit of a mess.⁸

Define $\mu_{PF,t+1} = R^f + \sum_{i=1}^N w_i \hat{\beta}_i \hat{E}(R_{Mkt,t+1} - R^f)$ and $\mu_{PF,t+1}^{well-d} = R^f + \sum_{i=1}^N \check{w}_i \hat{\beta}_i \hat{E}(R_{Mkt,t+1} - R^f)$. Notice that

$$\begin{aligned} [VaR_{t+1}(p)]^2 = & \left(\sum_{i=1}^N w_i \hat{\beta}_i \right)^2 \sigma_{Mkt,t+1}^2 [\Phi^{-1}(p)]^2 + \sum_{i=1}^N w_i^2 \hat{\sigma}_{i,t+1}^2 [\Phi^{-1}(p)]^2 + \\ & + \mu_{PF,t+1}^2 + 2\mu_{PF,t+1} \left\{ \left(\sum_{i=1}^N w_i \hat{\beta}_i \right)^2 \sigma_{Mkt,t+1}^2 + \sum_{i=1}^N w_i^2 \hat{\sigma}_{i,t+1}^2 \right\}^{1/2} \Phi^{-1}(p), \end{aligned}$$

⁸Those with a weak heart are advised to skip the algebra that follows and to go directly to conclusions to prevent permanent damage.

while

$$\left[VaR_{t+1}^{well-d}(p) \right]^2 = \left(\sum_{i=1}^N \check{w}_i \hat{\beta}_i \right)^2 \sigma_{Mkt,t+1}^2 \left[\Phi^{-1}(p) \right]^2 + (\mu_{PF,t+1}^{well-d})^2 + 2\mu_{PF,t+1}^{well-d} \left(\sum_{i=1}^N \check{w}_i \hat{\beta}_i \right) \sigma_{Mkt,t+1} \Phi^{-1}(p).$$

The difference is then

$$\begin{aligned} [VaR_{t+1}(p)]^2 - [VaR_{t+1}^{well-d}(p)]^2 &= \left\{ \left(\sum_{i=1}^N w_i \hat{\beta}_i \right)^2 - \left(\sum_{i=1}^N \check{w}_i \hat{\beta}_i \right)^2 \right\} \sigma_{Mkt,t+1}^2 \left[\Phi^{-1}(p) \right]^2 + \\ & \quad [(\mu_{PF,t+1})^2 - (\mu_{PF,t+1}^{well-d})^2] + \sum_{i=1}^N w_i^2 \hat{\sigma}_{i,t+1}^2 \left[\Phi^{-1}(p) \right]^2 + \\ & \quad + 2 \left\{ \mu_{PF,t+1} \left[\left(\sum_{i=1}^N w_i \hat{\beta}_i \right)^2 \sigma_{Mkt,t+1}^2 + \sum_{i=1}^N w_i^2 \hat{\sigma}_{i,t+1}^2 \right]^{1/2} \right. \\ & \quad \left. - \mu_{PF,t+1}^{well-d} \left(\sum_{i=1}^N \check{w}_i \hat{\beta}_i \right) \sigma_{Mkt,t+1} \right\} \Phi^{-1}(p) \end{aligned}$$

Now notice that if $[VaR_{t+1}(p)]^2 - [VaR_{t+1}^{well-d}(p)]^2 \geq 0$ holds, then it must be that $VaR_{t+1}(p) \geq VaR_{t+1}^{well-d}(p)$. However, $[VaR_{t+1}(p)]^2 - [VaR_{t+1}^{well-d}(p)]^2 \geq 0$ would be guaranteed by the fact that

$$\begin{aligned} \sum_{i=1}^N w_i^2 \hat{\sigma}_{i,t+1}^2 \left[\Phi^{-1}(p) \right]^2 + \left\{ \left(\sum_{i=1}^N w_i \hat{\beta}_i \right)^2 - \left(\sum_{i=1}^N \check{w}_i \hat{\beta}_i \right)^2 \right\} &\geq 0 \\ \left\{ \mu_{PF,t+1} \left[\left(\sum_{i=1}^N w_i \hat{\beta}_i \right)^2 \sigma_{Mkt,t+1}^2 + \sum_{i=1}^N w_i^2 \hat{\sigma}_{i,t+1}^2 \right]^{1/2} \right. & \\ \left. - \mu_{PF,t+1}^{well-d} \left(\sum_{i=1}^N \check{w}_i \hat{\beta}_i \right) \sigma_{Mkt,t+1} \right\} &\geq 0 \end{aligned}$$

and $(\mu_{PF,t+1}) - (\mu_{PF,t+1}^{well-d}) \geq 0$, but in general any combination of these three condition may deliver the result. Importantly, even if it is intuitive to think that setting $\sum_{i=1}^N \check{w}_i^2 \hat{\sigma}_{i,t+1}^2 \simeq 0$ should bring a reduction in VaR and this remains likely, formally it is possible that setting the weights in such a way may cause a reduction in the expected portfolio return so large to over-turn the effect. Indeed, observe that if we were able to say that $\mu_{PF,t+1} = \mu_{PF,t+1}^{well-d} \simeq 0$ as often assumed in our lectures as well as in chapters 4-5, then

$$\begin{aligned} [VaR_{t+1}(p)]^2 - [VaR_{t+1}^{well-d}(p)]^2 &= \sum_{i=1}^N w_i^2 \hat{\sigma}_{i,t+1}^2 \left[\Phi^{-1}(p) \right]^2 + \\ & \quad + \left\{ \left(\sum_{i=1}^N w_i \hat{\beta}_i \right)^2 - \left(\sum_{i=1}^N \check{w}_i \hat{\beta}_i \right)^2 \right\} \sigma_{Mkt,t+1}^2 \left[\Phi^{-1}(p) \right]^2 \geq 0, \end{aligned}$$

which is likely to hold if the individual security betas are all finite because $\sum_{i=1}^N \check{w}_i^2 \hat{\sigma}_{i,t+1}^2 \simeq 0$ must require that $\sum_{i=1}^N \check{w}_i^2 \rightarrow 0$ as the degree of diversification in the portfolio increases (i.e., as $N \rightarrow \infty$), while $\sum_{i=1}^N w_i \hat{\beta}_i$ remains unrestricted.⁹

⁹This proof is not sufficiently “tight” to be called a proof. However to make the proof compelling would require imposing assumptions on the asymptotic behavior of the weights what would just increase the formal burden. The statement is highly likely to hold in most realistic circumstances.

3.2. Multi-factor exposure mappings

As we have stated while introducing the topic, the exposure mapping approach also works beyond the simple case of the CAPM. As usual, although the intuition and mechanics remains the same, the details are a bit more tedious. Consider the following (empirical) asset pricing model for a generic asset/security/portfolio $i = 1, 2, \dots, N$ (due to Fama and French, 1992):

$$E[R_{i,t+1}] = R^f + \beta_i E[R_{Mkt,t+1} - R^f] + \gamma_i E[R_{SMB,t+1}] + \xi_i E[R_{HML,t+1}], \quad (8)$$

where R^f is the constant riskless rate of return, $R_{SMB,t+1}$ is the rate of return from a special long-short (zero net investment) portfolio that goes long in small capitalization stocks and short in large capitalization stocks, and $R_{HML,t+1}$ is the rate of return from a special long-short (zero net investment) portfolio that goes long in high book-to-market ratio (also called value) stocks and short in low book-to-market ratio (growth) stocks.¹⁰ Notice that this is an asset pricing model and not (yet) a ready-to-use econometric framework to be applied in risk management because the model only imposes restrictions on expected returns, i.e., it is not a model for returns but of their expectations.¹¹

For instance, how does the expression of the 1% VaR of a portfolio characterized by weights $\{w_1, w_2, \dots, w_N\}$ look like in the case of (8), assuming the asset pricing model is subject to IID shocks from a Gaussian distribution with zero cross-sectional correlations (i.e., $Cov[\epsilon_{i,t+1}, \epsilon_{j,t+1}] = 0$ for $i \neq j = 1, 2, \dots, N$),

$$R_{i,t+1} = R^f + \beta_i (R_{Mkt,t+1} - R^f) + \gamma_i R_{SMB,t+1} + \xi_i R_{HML,t+1} + \epsilon_{i,t+1} \quad \epsilon_{i,t+1} \text{ IID } N(0, \sigma_{\epsilon_i}^2) ? \quad (9)$$

In this case it may be simpler to express (9) in matrix form as $\mathbf{R}_{t+1} - R^f \mathbf{1}_N = \mathbf{B} \boldsymbol{\lambda}_{t+1} + \boldsymbol{\epsilon}_{t+1}$, where \mathbf{R}_{t+1} is a $N \times 1$ vector of asset returns, $\mathbf{1}_N$ is a $N \times 1$ vector of ones, $\boldsymbol{\lambda}_{t+1} \equiv [(R_{Mkt,t+1} - R^f) R_{SMB,t+1} R_{HML,t+1}]'$ is a 3×1 vector, and \mathbf{B} is $N \times 3$ a matrix that collects in each of its N rows the exposure coefficients $[\beta_i \gamma_i \xi_i]'$. Finally $\boldsymbol{\epsilon}_{t+1} \text{ IID } N(\mathbf{0}, \boldsymbol{\Omega}_\epsilon)$, where $\boldsymbol{\Omega}_\epsilon$ is a $N \times N$ diagonal matrix that collects the variance coefficients $\sigma_{\epsilon_1}^2, \sigma_{\epsilon_2}^2, \dots, \sigma_{\epsilon_N}^2$.

We start by noting that $R_{PF,t+1} \equiv \sum_{i=1}^N w_i R_{i,t+1}$ can be re-written as $R_{PF,t+1} \equiv \mathbf{R}'_{t+1} \mathbf{w}$ and the $T \times 1$ vector \mathbf{x}_{PF} that collects the *excess return* observations $R_{PF,1} - R^f, R_{PF,2} - R^f, \dots, R_{PF,T} - R^f$, can be written as

$$\mathbf{x}_{PF} = \mathbf{X} \mathbf{w},$$

where \mathbf{X} is a $T \times N$ matrix $[\mathbf{x}_1 \mathbf{x}_2 \dots \mathbf{x}_N]$ and \mathbf{w} a $N \times 1$ vector of portfolio weights. Hence the estimate of (the 3×1 column vector) $\mathbf{b}_{PF} \equiv [\beta_{PF} \gamma_{PF} \xi_{PF}]'$ in the stacked regression

$$\mathbf{x}_{PF} = \mathbf{\Lambda} \hat{\mathbf{b}}_{PF} + \hat{\boldsymbol{\epsilon}}_{PF}, \quad (10)$$

¹⁰SMB is the acronym for “Small minus Big” and HML is the acronym for “High minus Low” (referred to the book-to-market ratio). The fact that SMB and HML are two zero net investment portfolios explains why we do not need to subtract the constant riskless rate from $R_{SMB,t+1}$ and $R_{HML,t+1}$. You must have already encountered this factor models in at least three of your MSc. courses.

¹¹Moreover, under the restrictions $\gamma_i = \xi_i = 0$ for $i = 1, 2, \dots, n$ (i.e., for all assets under consideration), this model becomes the CAPM used in the lectures, $E[R_{i,t+1}] = R^f + \beta_i E[R_{Mkt,t+1} - R^f]$.

will be

$$\hat{\mathbf{b}}_{PF} = (\mathbf{\Lambda}'\mathbf{\Lambda})^{-1}\mathbf{\Lambda}'\mathbf{x}_{PF} = (\mathbf{\Lambda}'\mathbf{\Lambda})^{-1}\mathbf{\Lambda}'(\mathbf{X}\mathbf{w}) = ((\mathbf{\Lambda}'\mathbf{\Lambda})^{-1}\mathbf{\Lambda}'\mathbf{X})\mathbf{w} = \hat{\mathbf{B}}\boldsymbol{\omega}$$

In (10) T is the sample size, $\mathbf{\Lambda} \equiv [\boldsymbol{\lambda}'_1 \ \boldsymbol{\lambda}'_2 \ \dots \ \boldsymbol{\lambda}'_T]'$ is a $T \times 3$ matrix, and $\hat{\boldsymbol{\epsilon}}_{PF}$ is the $T \times 1$ vector that collects the observations $\hat{\epsilon}_{PF,1}, \hat{\epsilon}_{PF,2}, \dots, \hat{\epsilon}_{PF,T}$.¹² However, the expression for $\hat{\mathbf{b}}_{PF}$ indicates that this vector of estimated coefficients may be written as weighted combination of the columns of the $3 \times N$ matrix $\hat{\mathbf{B}} \equiv ((\mathbf{\Lambda}'\mathbf{\Lambda})^{-1}\mathbf{\Lambda}'\mathbf{X})$, where each column has an interpretation similar to a vector of ratios of covariance-type terms between the returns on the factor portfolios and the returns on each of the N assets ($\mathbf{\Lambda}'\mathbf{X}$) “divided” (loosely speaking) by a 3×3 matrix of variance-type terms concerning ($\mathbf{\Lambda}'\mathbf{\Lambda}$) returns on the three factor portfolios. The weights of the combinations are provided by the portfolio weights. Moreover

$$Var[\hat{\epsilon}_{PF,t+1}] = Var[\hat{\boldsymbol{\epsilon}}'_{t+1}\mathbf{w}] = \mathbf{w}'E[\hat{\boldsymbol{\epsilon}}_{t+1}\hat{\boldsymbol{\epsilon}}'_{t+1}]\mathbf{w} = \mathbf{w}'\hat{\boldsymbol{\Omega}}_{\epsilon}\mathbf{w}.$$

Therefore

$$\begin{aligned} \sigma_{PF,t+1}^2 &= Var[R_{PF,t+1}] = Var[(R^f\boldsymbol{\iota}_N + \hat{\mathbf{B}}\boldsymbol{\lambda}_{t+1} + \boldsymbol{\epsilon}_{t+1})'\mathbf{w}] \\ &= \mathbf{w}'Var[(\hat{\mathbf{B}}\boldsymbol{\lambda}_{t+1} + \boldsymbol{\epsilon}_{t+1})']\mathbf{w} \\ &= \mathbf{w}'\hat{\mathbf{B}}Var[\boldsymbol{\lambda}_{t+1}]\hat{\mathbf{B}}'\mathbf{w} + \mathbf{w}'Var[\boldsymbol{\epsilon}_{t+1}]\mathbf{w} \\ &= \mathbf{w}'\hat{\mathbf{B}}Var[\boldsymbol{\lambda}_{t+1}]\hat{\mathbf{B}}'\mathbf{w} + \mathbf{w}'\hat{\boldsymbol{\Omega}}_{\epsilon}\mathbf{w} = \mathbf{w}'\left\{\hat{\mathbf{B}}Var[\boldsymbol{\lambda}_{t+1}]\hat{\mathbf{B}}' + \hat{\boldsymbol{\Omega}}_{\epsilon}\right\}\mathbf{w}, \end{aligned}$$

where $Var[\boldsymbol{\lambda}_{t+1}]$ is the 3×3 covariance matrix of the returns on the factor portfolios, i.e.,

$$Var[\boldsymbol{\lambda}_{t+1}] \equiv \begin{bmatrix} \sigma_{Mkt,t+1}^2 & Cov[R_{Mkt,t+1}, R_{SMB,t+1}] & Cov[R_{Mkt,t+1}, R_{HML,t+1}] \\ Cov[R_{Mkt,t+1}, R_{SMB,t+1}] & \sigma_{SMB,t+1}^2 & Cov[R_{SMB,t+1}, R_{HML,t+1}] \\ Cov[R_{Mkt,t+1}, R_{HML,t+1}] & Cov[R_{SMB,t+1}, R_{HML,t+1}] & \sigma_{HML,t+1}^2 \end{bmatrix}.$$

Finally, it is easy to see that assuming normality,

$$\begin{aligned} VaR_{t+1}(0.01) &= 2.33 \left\{ \mathbf{w}' \left[\hat{\mathbf{B}}Var[\boldsymbol{\lambda}_{t+1}]\hat{\mathbf{B}}' + \hat{\boldsymbol{\Omega}}_{\epsilon} \right] \mathbf{w} \right\}^{1/2} - (R^f\boldsymbol{\iota}_N + \hat{\mathbf{B}}\hat{E}[\boldsymbol{\lambda}_{t+1}])'\mathbf{w} \\ &= 2.33 \{ \hat{\mathbf{B}}Var[\boldsymbol{\lambda}_{t+1}]\hat{\mathbf{B}}' + Var[\hat{\boldsymbol{\epsilon}}_{PF,t+1}] \} - E[R_{PF,t+1}] \end{aligned} \quad (11)$$

where $\hat{E}[\boldsymbol{\lambda}_{t+1}]$ can be simply estimated as the sample mean of (excess) returns on the factor portfolios. Once more, the risk exposure will entirely depend on the portfolio composition, the exposures of the securities in the portfolio vs. each of the three priced risk factors as measured by $\hat{\mathbf{B}}$, the covariance matrix of the factors themselves, $Var[\boldsymbol{\lambda}_{t+1}]$, and the idiosyncratic risk coefficients of all the securities in the portfolio, as captured by the diagonal matrix $\hat{\boldsymbol{\Omega}}_{\epsilon}$. Although it is algebraically more involved, also in this case we see that a seemingly passive expression (second line of (11)) gives the same answer as a perfectly active one (first line of (11)).

¹²In case this sounds unfamiliar, please review your notes from the first semester of Econometrics to see how the multivariate regression $R_{PF,t+1} - R^f = \beta_{PF}(R_{Mkt,t+1} - R^f) + \gamma_{PF}R_{SMB,t+1} + \xi_{PF}R_{HML,t+1} + \epsilon_{PF,t+1}$ can be written in stacked form.

4. Naive Models of Covariance Prediction

What if the asset pricing models discussed in Section 3 are rejected by the data? Unfortunately, you may have learned from your courses that this tends to be the case for most models, data sets, and sample periods. The key alternative idea that populates the financial econometrics literature is that concepts and tools introduced in chapters 4 and 5 with reference to volatility forecasts can now be extended to covariances as well (and hence to correlations). This of course starts with rather simple, naive techniques and models that we have already commented in chapter 4. The simplest idea is to build time-varying estimates of covariances using rolling (moving) averages:

$$\sigma_{ij,t+1} = \frac{1}{m} \sum_{\tau=1}^m R_{t+1-\tau}^i R_{t+1-\tau}^j,$$

where m is the window length and we have assumed a zero mean for the returns on both assets i and j . Clearly, when $i = j$ this becomes the rolling window variance estimator already analyzed in chapter 4. The problem of this rolling window covariance estimator/predictor remains the same one that we have already encountered in chapter 4: how should one pick the window parameter m ? Obviously, its choice is critical for the estimator that can be obtained. Too long a window makes the estimator rather smooth but also risks of including in the calculation returns that may have been originated from a possibly different period or regime, in either a statistical or in an economic sense. The choice of a small m leads to a jagged and quickly changing estimator. Moreover, the “box-shaped” spurious effects already discussed in chapter 4 would also characterize this rolling window covariance estimator, with the risk of covariance prediction at time t changing not because of events recorded at time t , but because returns recorded m periods ago drop out of the rolling window. Finally, the rolling window estimator attaches equal weights on past cross products of returns, which may be highly questionable. Figure 1 reports one such example. The appearance of some “box-shaped” is obvious.

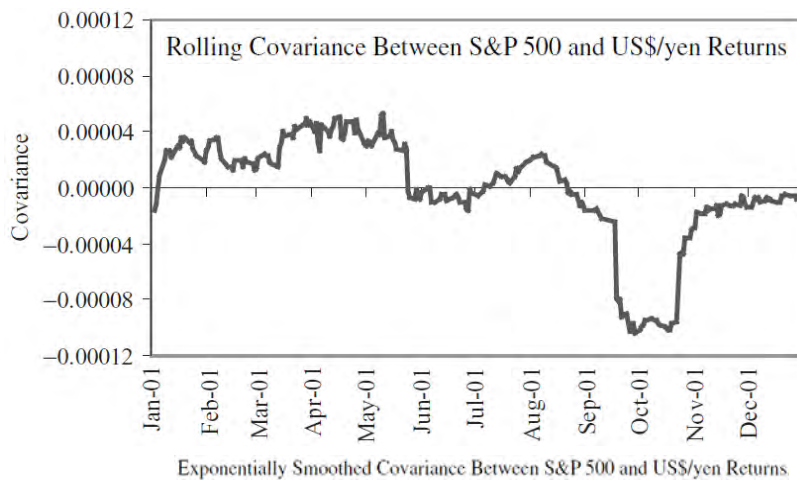


Figure 1: S&P500 vs. USD/Yen return moving average covariance estimate, $m = 25$

An alternative idea that has had some impact on the practice of risk and asset management consists of extending the RiskMetrics variance estimator to covariance, i.e., the idea is that of an

exponential smoother applied to covariances:

$$\sigma_{ij,t+1} = (1 - \lambda)R_t^i R_t^j + \lambda \sigma_{ij,t}, \quad (12)$$

with $\lambda \in (0, 1)$.¹³ As already discussed, JP Morgan had originally popularized a choice $\lambda = 0.94$ which turns out to work rather well also for covariances. However, the restriction that the coefficient $(1 - \lambda)$ on the cross product of returns and the coefficient λ on past covariance sum to one is not necessarily desirable. To understand this, consider the model $\sigma_{ij,t+1} = \delta_0 + \delta_1 R_t^i R_t^j + \delta_2 \sigma_{ij,t}$ of which (12) represents a special case. Because when means are zero, then $\bar{\sigma}_{ij} = E[R_t^i R_t^j]$, we have that

$$\begin{aligned} \bar{\sigma}_{ij} &\equiv E[\sigma_{ij,t+1}] = \delta_0 + \delta_1 E[R_t^i R_t^j] + \delta_2 E[\sigma_{ij,t}] \\ &= \delta_0 + \delta_1 \bar{\sigma}_{ij} + \delta_2 \bar{\sigma}_{ij} \implies \bar{\sigma}_{ij} = \frac{\delta_0}{1 - \delta_1 - \delta_2}. \end{aligned}$$

However, because in (12), $\delta_0 = 0$ and $\delta_1 + \delta_2 = 1$, we have that under the RiskMetrics model, $\bar{\sigma}_{ij} = 0/0$, i.e., that long-run, unconditional covariance actually fails to exist. This implies that there is no mean-reversion in covariance: based on the closing price today, if tomorrow's covariance is high then it will remain high, rather than revert back to its mean. Equivalently, we say that under (12) covariance follows a non-stationary, unit root process. Figure 2 shows an example of predicted covariance dynamics generated from a RiskMetrics model for the same data as in Figure 1. Clearly, RiskMetrics covariance reacts more to shocks to return cross-products than a rolling window covariance estimator does.



Figure 2: S&P500 vs. USD/Yen return RiskMetrics covariance estimate ($\lambda = 0.94$)

In fact, (12) can be easily re-written in the equally familiar format

$$\sigma_{ij,t+1} = (1 - \lambda) \sum_{k=0}^{\infty} \lambda^k R_{i,t-k} R_{j,t-k},$$

¹³As we shall explain below, we need to set λ to be independent of i and j to ensure that the covariance matrix predicted from the model is semi-positive definite.

which shows that that this forecast corresponds to an exponentially weighted, infinite moving average. To see this, we just need to re-write the model in recursive fashion moving backwards in time:

$$\begin{aligned}
\sigma_{ij,t+1} &= (1 - \lambda)R_{i,t}R_{j,t} + \lambda\sigma_{ij,t} \\
&= (1 - \lambda)R_{i,t}R_{j,t} + \lambda(1 - \lambda)R_{i,t-1}R_{j,t-1} + \lambda^2\sigma_{ij,t-1} \\
&= (1 - \lambda)[R_{i,t}R_{j,t} + \lambda R_{i,t-1}R_{j,t-1}] + \lambda^2(1 - \lambda)R_{i,t-2}R_{j,t-2} + \lambda^3\sigma_{ij,t-2} \\
&= (1 - \lambda)[R_{i,t}R_{j,t} + \lambda R_{i,t-1}R_{j,t-1} + \lambda^2 R_{i,t-2}R_{j,t-2}] + \lambda^3\sigma_{ij,t-2} \\
&= \dots \\
&= (1 - \lambda) \sum_{k=0}^{\infty} \lambda^k R_{i,t-k}R_{j,t-k}
\end{aligned}$$

as $\lim_{k \rightarrow \infty} \lambda^k \sigma_{ij,t-k} = 0$ for $\lambda \in (0, 1)$.

The model $\sigma_{ij,t+1} = \delta_0 + \delta_1 R_t^i R_t^j + \delta_2 \sigma_{ij,t}$ used above to work on the non-stationarity of the RiskMetrics estimator shows already an obvious direction in which we ought to be looking at, i.e., extending GARCH(1,1) style models to predict covariances besides variances:

$$\sigma_{ij,t+1} = \omega_{ij} + \alpha_{ij} R_t^i R_t^j + \beta_{ij} \sigma_{ij,t}, \quad (13)$$

where α_{ij} and β_{ij} in principle depend on the couple of assets i and j under examination. Similarly to a GARCH(1,1), one needs $\alpha_{ij} + \beta_{ij} < 1$ for the process to be stationary, as

$$\bar{\sigma}_{ij} = \frac{\omega_{ij}}{1 - \alpha_{ij} - \beta_{ij}}$$

is finite if and only if $\alpha_{ij} + \beta_{ij} < 1$. However, because covariances can be negative, in this case one does not need to restrict any of the parameters to be estimated to be positive (or non-negative). Unfortunately, it is possible to show that unless $\alpha_{ij} = \alpha$ and $\beta_{ij} = \beta$ for all possible pairs $i \neq j$, even though $\sigma_{ij,t+1}$ can be anyway estimated/predicted, when one organizes such estimates/predictions into a covariance matrix predicted at time t for time $t + 1$,

$$\hat{\Sigma}_{t+1} \equiv \widehat{Cov}_t[\mathbf{R}_{t+1}] = \begin{bmatrix} \hat{\sigma}_{1,t+1}^2 & \hat{\sigma}_{12,t+1} & \dots & \hat{\sigma}_{1n,t+1} \\ \hat{\sigma}_{12,t+1} & \hat{\sigma}_{2,t+1}^2 & \dots & \hat{\sigma}_{2n,t+1} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\sigma}_{1n,t+1} & \hat{\sigma}_{2n,t+1} & \dots & \hat{\sigma}_{N,t+1}^2 \end{bmatrix},$$

unfortunately the resulting $\hat{\Sigma}_{t+1}$ is not guaranteed to be *semi-positive definite* (SPD), while it should be.¹⁴ Why do we need so desperately that $\hat{\Sigma}_{t+1}$ be SPD? As often in this course, the reason is purely an economic one, not a statistical one. First, let's recall that a $N \times N$ square symmetric matrix \mathbf{A} is SPD if and only if $\forall \mathbf{x} \in \mathcal{R}^N$, then $\mathbf{x}'\mathbf{A}\mathbf{x} \geq 0$. Second, when applied to our problem this definition implies that for any vector of portfolio weights $\mathbf{w}_t \in \mathcal{R}^N$, $\mathbf{w}_t' \hat{\Sigma}_{t+1} \mathbf{w}_t \geq 0$. So far we

¹⁴Notice that instead ω_{ij} is allow to depend on the pair $i \neq j$.

have only applied a mathematical definition. The beauty of this result is that $\mathbf{w}'_t \hat{\Sigma}_{t+1} \mathbf{w}_t$ has a very precise meaning to us:

$$\widehat{Var}_t[R_{t+1}^{ptf}] = \mathbf{w}'_t \hat{\Sigma}_{t+1} \mathbf{w}_t \geq 0,$$

i.e., the SPD nature of $\hat{\Sigma}_{t+1}$ is necessary (as well as sufficient) to ensure that the variance of any portfolio is non-negative, as it should be.

To get a feeling for why we need to prevent the GARCH-type coefficients to be a function of the pair of assets to ensure a good behavior of the covariance matrix, let's consider the RiskMetrics case, which as we know it is just a case of zero-intercept, non-stationary GARCH model. We deal with RiskMetrics because to think about this problem with reference to one parameter only (λ) delivers ready intuition with less algebra. Assume the exponential smoothing model is applied to both variances and covariances in the case of two assets, $N = 2$, i.e.:

$$\sigma_{ij,t+1} = (1 - \lambda)R_{i,t}R_{j,t} + \lambda\sigma_{ij,t} \quad i, j = 1, 2,$$

so that the dynamic model also applies to variances when $i = j$. The exponential smoothing estimator of the entire conditional covariance matrix is then:

$$\Sigma_{t+1} = \begin{bmatrix} \sigma_{11,t+1} & \sigma_{12,t+1} \\ \sigma_{12,t+1} & \sigma_{22,t+1} \end{bmatrix} = \begin{bmatrix} (1 - \lambda)R_{1,t}^2 + \lambda\sigma_{11,t} & (1 - \lambda)R_{1,t}R_{2,t} + \lambda\sigma_{12,t} \\ (1 - \lambda)R_{1,t}R_{2,t} + \lambda\sigma_{12,t} & (1 - \lambda)R_{2,t}^2 + \lambda\sigma_{22,t} \end{bmatrix}.$$

Assume now, by the sake of contradiction that the smoothing parameters that apply to conditional variances and covariances are allowed to be different:¹⁵

$$\begin{aligned} \sigma_{ii,t+1} &= (1 - \lambda_V)R_{i,t}^2 + \lambda_V\sigma_{ij,t} & i = 1, 2, \\ \sigma_{12,t+1} &= (1 - \lambda_C)R_{1,t}R_{2,t} + \lambda_C\sigma_{12,t}, \end{aligned}$$

with $\lambda_V \neq \lambda_C$ (but both of them still belong to the interval $(0, 1)$). It is relatively easy to use a simple example to show that $\lambda_V \neq \lambda_C$ may lead to a conditional correlation between returns on assets 1 and 2 that fails to be in the interval $[-1, 1]$. The idea is to work by focussing on few returns setting all other returns to zero, which may of course happen only by accident in reality—recall that you just need one example, not a general proof. Suppose you have available a sample of $t + 1$ paired returns, i.e., $\{R_{1,\tau}, R_{2,\tau}\}_{\tau=0}^t$. Then the RiskMetrics processes can be re-written in exponential smoothing form as:

$$\begin{aligned} \sigma_{ii,t+1} &= (1 - \lambda_V) \sum_{k=0}^t \lambda^k R_{i,t-k}^2 + \lambda_V^{t+1} \sigma_{ii,0} & i = 1, 2, \\ \sigma_{12,t+1} &= (1 - \lambda_C) \sum_{k=0}^t \lambda^k R_{1,t-k} R_{2,t-k} + \lambda_C^{t+1} \sigma_{12,0} \end{aligned}$$

Assume that at time 0, $\sigma_{ii,0} = \sigma_{12,0} = 0$ for $i = 1, 2$ and that—just by accident—while $R_{1,t-k} = R_{2,t-k} = 0$ for $k = 0, 1, \dots, t - 1$, at time t the returns are potentially non-zero, call them $R_{1,t}$ and

¹⁵Here $\sigma_{ii,t+1}$ is the same as $\sigma_{i,t+1}^2$. This derives from the fact that the covariance of a random variable with itself is the same as the variance. Therefore, this means that λ is allowed to differ according to the fact that $i = j$ or not.

$R_{2,t}$. Then

$$\begin{aligned}\sigma_{ii,t+1} &= (1 - \lambda_V)R_{i,t}^2 & i = 1, 2, \\ \sigma_{12,t+1} &= (1 - \lambda_C)R_{1,t}R_{2,t}\end{aligned}$$

and the corresponding conditional correlation is

$$\rho_{12,t+1} = \frac{(1 - \lambda_C)R_{1,t}R_{2,t}}{\sqrt{(1 - \lambda_V)R_{1,t}^2(1 - \lambda_V)R_{2,t}^2}} = \frac{(1 - \lambda_C)R_{1,t}R_{2,t}}{(1 - \lambda_V)|R_{1,t}||R_{2,t}|} = \frac{(1 - \lambda_C)}{(1 - \lambda_V)} \text{sign}(R_{1,t}R_{2,t}),$$

where the sign function $\text{sign}(R_{1,t}R_{2,t})$ takes a value of +1 when the sign of $R_{1,t}R_{2,t}$ is positive and -1 otherwise. At this point, note that if $\lambda_V > \lambda_C$, then

$$\rho_{12,t+1} \begin{cases} > 1 & \text{if } \text{sign}(R_{1,t}R_{2,t}) = +1 \\ < -1 & \text{if } \text{sign}(R_{1,t}R_{2,t}) = -1 \end{cases}$$

which is clearly inadmissible. As argued above, this shows (although it is just a very special example) that $\lambda_V = \lambda_C$ is sufficient for $\rho_{12,t+1}$ to be in $[-1, 1]$ because in that case

$$\rho_{12,t+1} = \text{sign}(R_{1,t}R_{2,t}) = \begin{cases} +1 & \text{if } \text{sign}(R_{1,t}R_{2,t}) = +1 \\ -1 & \text{if } \text{sign}(R_{1,t}R_{2,t}) = -1 \end{cases},$$

which is again admissible. How does an example of $\rho_{12,t+1} \notin [-1, 1]$ map into our claim that when the coefficients are allowed to depend on either the assets or the moments (as in this case), then Σ_{t+1} may fail to be SPD so that it cannot be a covariance matrix used in financial applications? Recall that if Σ_{t+1} is semi-positive definite, then $\det(\Sigma_{t+1}) \geq 0$. In our example, we have

$$\Sigma_{t+1} = \begin{bmatrix} (1 - \lambda_V)R_{1,t}^2 & (1 - \lambda_C)R_{1,t}R_{2,t} \\ (1 - \lambda_C)R_{1,t}R_{2,t} & (1 - \lambda_V)R_{2,t}^2 \end{bmatrix}.$$

The determinant of this matrix is simply:

$$\begin{aligned}\det(\Sigma_{t+1}) &= [(1 - \lambda_V)R_{1,t}^2(1 - \lambda_V)R_{2,t}^2] - [(1 - \lambda_C)R_{1,t}R_{2,t}(1 - \lambda_C)R_{1,t}R_{2,t}] \\ &= (1 - \lambda_V)^2 R_{1,t}^2 R_{2,t}^2 - (1 - \lambda_C)^2 R_{1,t}^2 R_{2,t}^2 \\ &= R_{1,t}^2 R_{2,t}^2 [(1 - \lambda_V)^2 - (1 - \lambda_C)^2]\end{aligned}$$

which is non-negative if and only if $[(1 - \lambda_V)^2 - (1 - \lambda_C)^2] \geq 0$ or

$$\frac{(1 - \lambda_V)^2}{(1 - \lambda_C)^2} \geq 1 \implies \frac{1 - \lambda_V}{1 - \lambda_C} \geq 1 \implies \lambda_C \geq \lambda_V.$$

Once more, should we be setting $\lambda_V > \lambda_C$, the outcome is that $\det(\Sigma_{t+1}) < 0$ which would show that Σ_{t+1} is negative definite and hence it cannot be a covariance matrix.

Therefore, going through a re-cap of the points above, only setting $\alpha_{ij} = \alpha$ and $\beta_{ij} = \beta$ for all possible pairs $i \neq j$, i.e., setting the coefficients of (13) to be identical across *all* pairs of assets and therefore across models for conditional variance vs. covariance, one guarantees that for any portfolio that can be formed, $\widehat{Var}_t[R_{t+1}^{ptf}] \geq 0$. Unfortunately, empirically it is unclear that the persistence

parameters α and β should be the same for all variances and covariances across assets. Often the data reject this restriction. Moreover, this business of having to set $\alpha_{ij} = \alpha$ and $\beta_{ij} = \beta$ for all possible pairs $i \neq j$, has another critical implication. While in principle we would like to use data to proceed to the estimation of (13) for each possible pair of assets, imposing constraints such as the ones discussed above, implies that all the dynamics processes underlying the elements of Σ should be estimated jointly, in one single pass. If this were not enough to scare you, this means that given N assets, we should jointly estimate the parameters of $N(N-1)/2$ different processes, one for the covariance of each possible pair of assets. For instance, in the case of (13), this implies the need to estimate $N+2$ parameters (N different constant coefficients ω_{ij} plus one single α and one single β) that enter $N(N-1)/2$ different econometric models. At this point, hoping that you are scared by the perspective of actually implementing this task for a large N , it becomes obvious that we need to develop better, multivariate econometric models of the conditional covariance matrix.

4.1. Comparing the properties of RiskMetrics and GARCH models for covariances

This is one of our traditional stops in the flow of our arguments. What are the effects of selecting a conditional model for covariance to be of a GARCH vs. of a RiskMetrics type? In this subsection, we therefore compare the RiskMetrics-style model,

$$\sigma_{ij,t+1}^{ES} = (1 - \lambda)R_{i,t}R_{j,t} + \lambda\sigma_{ij,t},$$

(that we know it is easy to re-write as $\sigma_{ij,t+1}^{ES} = (1 - \lambda) \sum_{k=0}^{\infty} \lambda^k R_{i,t-k}R_{j,t-k}$, where ES stands for exponential smoother) and

$$\sigma_{ij,t+1}^{GARCH} = \omega + \alpha R_{i,t}R_{j,t} + \beta\sigma_{ij,t}, \quad (14)$$

where ω has been simplified not to depend on i and j for the purposes of this subsection. As a starting point, we show that the GARCH model in (14) can be written in a format similar to the ES one. We just need to re-write the GARCH in a recursive fashion moving backwards in time:

$$\begin{aligned} \sigma_{ij,t+1}^{GARCH} &= \omega + \alpha R_{i,t}R_{j,t} + \beta\sigma_{ij,t} \\ &= \omega + \alpha R_{i,t}R_{j,t} + \beta[\omega + \alpha R_{i,t-1}R_{j,t-1} + \beta\sigma_{ij,t-1}] \\ &= \omega(1 + \beta) + \alpha(R_{i,t}R_{j,t} + \beta R_{i,t-1}R_{j,t-1}) + \beta^2\sigma_{ij,t-1} \\ &= \omega(1 + \beta + \beta^2) + \alpha(R_{i,t}R_{j,t} + \beta R_{i,t-1}R_{j,t-1} + \beta^2 R_{i,t-2}R_{j,t-2}) + \beta^3\sigma_{ij,t-2} \\ &= \dots \\ &= \omega \sum_{k=0}^{\infty} \beta^k + \alpha \sum_{k=0}^{\infty} \beta^k R_{i,t-k}R_{j,t-k} = \frac{\omega}{1 - \beta} + \alpha \sum_{k=0}^{\infty} \beta^k R_{i,t-k}R_{j,t-k}, \end{aligned}$$

which is the desired expression. Clearly, this process for $\sigma_{ij,t+1}^{GARCH}$ simplifies to $\sigma_{ij,t+1}^{ES} = (1 - \lambda) \sum_{k=0}^{\infty} \lambda^k R_{i,t-k}R_{j,t-k}$ when $\omega = 0$, $\alpha = 1 - \lambda$, and $\beta = \lambda$. However, because you now perfectly understand that the persistence of a GARCH(1,1) process is measured by the sum $\alpha + \beta$, it is also clear that $\alpha = 1 - \lambda$ and $\beta = \lambda$ implies that $\alpha + \beta = 1 - \lambda + \lambda = 1$ which means that an exponentially

smoothed process for conditional covariance implies that covariance is integrated of order 1, i.e., that $\sigma_{ij,t+1}^{ES}$ has “infinite memory”, in the sense any shock to $\sigma_{ij,t}^{ES}$ will have an impact on future, subsequent $\sigma_{ij,t+H}^{ES}$ that lasts forever ($H > 0$).

We call $\bar{\sigma}_{ij} \equiv E[R_{i,t}R_{j,t}] = \omega/(1 - \alpha - \beta)$ the unconditional covariance implied by a stationary GARCH(1,1) model in which $\alpha + \beta < 1$. As usual, this is obtained as

$$\bar{\sigma}_{ij} \equiv E[R_{i,t}R_{j,t}] = \omega + \alpha E[R_{i,t}R_{j,t}] + \beta E[R_{i,t}R_{j,t}] \implies \bar{\sigma}_{ij} \equiv E[R_{i,t}R_{j,t}] = \frac{\omega}{1 - \alpha - \beta}.$$

Therefore (14) can be re-written as

$$\sigma_{ij,t+1}^{GARCH} = \bar{\sigma}_{ij} + \alpha(R_{i,t}R_{j,t} - \bar{\sigma}_{ij}) + \beta(\sigma_{ij,t} - \bar{\sigma}_{ij}).$$

This follows from

$$\begin{aligned} \sigma_{ij,t+1}^{GARCH} &= \omega + \alpha R_{i,t}R_{j,t} + \beta \sigma_{ij,t} \\ &= \bar{\sigma}_{ij}(1 - \alpha - \beta) + \alpha R_{i,t}R_{j,t} + \beta \sigma_{ij,t} \\ &= \bar{\sigma}_{ij} + \alpha(R_{i,t}R_{j,t} - \bar{\sigma}_{ij}) + \beta(\sigma_{ij,t} - \bar{\sigma}_{ij}), \end{aligned}$$

as desired. This expression shows that forecasts of future covariance depend on three “ingredients”: (i) the baseline forecast is represented by the unconditional covariance $\bar{\sigma}_{ij}$, which depends on all the parameters, ω , α , and β ; (ii) the deviation of the current cross-product of asset returns $R_{i,t}R_{j,t}$ from the unconditional covariance $\bar{\sigma}_{ij}$, weighted by the coefficient α ; (iii) the deviation of the current conditional covariance $\sigma_{ij,t}$ from the unconditional covariance $\bar{\sigma}_{ij}$, weighted by the coefficient β .

Interestingly, (14) can be equivalently re-written as

$$\sigma_{ij,t+1}^{GARCH} = \bar{\sigma}_{ij} + (\alpha + \beta)(\sigma_{ij,t} - \bar{\sigma}_{ij}) + \alpha \sigma_{ij,t} \left(\frac{z_{i,t}z_{j,t}}{\rho_{ij,t}} - 1 \right)$$

because (using a simple trick, i.e., adding and subtracting $\alpha(\sigma_{ij,t} - \bar{\sigma}_{ij})$ at the right stage and exploiting the fact that $\sigma_{ij,t} = \rho_{ij,t}\sqrt{\sigma_{ii,t}}\sqrt{\sigma_{jj,t}}$):

$$\begin{aligned} \sigma_{ij,t+1}^{GARCH} &= \bar{\sigma}_{ij} + \alpha(R_{i,t}R_{j,t} - \bar{\sigma}_{ij}) + \beta(\sigma_{ij,t} - \bar{\sigma}_{ij}) \\ &= \bar{\sigma}_{ij} + \alpha(R_{i,t}R_{j,t} - \bar{\sigma}_{ij}) - \alpha(\sigma_{ij,t} - \bar{\sigma}_{ij}) + \beta(\sigma_{ij,t} - \bar{\sigma}_{ij}) + \alpha(\sigma_{ij,t} - \bar{\sigma}_{ij}) \\ &= \bar{\sigma}_{ij} + (\alpha + \beta)(\sigma_{ij,t} - \bar{\sigma}_{ij}) + \alpha(R_{i,t}R_{j,t} - \sigma_{ij,t}) \\ &= \bar{\sigma}_{ij} + (\alpha + \beta)(\sigma_{ij,t} - \bar{\sigma}_{ij}) + \alpha \sigma_{ij,t} \left(\frac{R_{i,t}R_{j,t}}{\sigma_{ij,t}} - 1 \right) \\ &= \bar{\sigma}_{ij} + (\alpha + \beta)(\sigma_{ij,t} - \bar{\sigma}_{ij}) + \alpha \sigma_{ij,t} \left(\frac{R_{i,t}R_{j,t}}{\rho_{ij,t}\sqrt{\sigma_{ii,t}}\sqrt{\sigma_{jj,t}}} - 1 \right) \\ &= \bar{\sigma}_{ij} + (\alpha + \beta)(\sigma_{ij,t} - \bar{\sigma}_{ij}) + \alpha \sigma_{ij,t} \left(\frac{z_{i,t}z_{j,t}}{\rho_{ij,t}} - 1 \right). \end{aligned}$$

At this point the H -step ahead forecast for the variance can be found as:

$$E_t[\sigma_{ij,t+H}^{GARCH}] = \bar{\sigma}_{ij} + (\alpha + \beta)(E_t[E_{t+H-2}[\sigma_{ij,t+H-1}]] - \bar{\sigma}_{ij}) +$$

$$\begin{aligned}
& +\alpha E_t \left[E_{t+H-2} \left(\sigma_{ij,t+H-1} \frac{z_{i,t+H-1} z_{j,t+H-1}}{\rho_{ij,t+H-1}} - 1 \right) \right] \\
& = \bar{\sigma}_{ij} + (\alpha + \beta)(E_t[\bar{\sigma}_{ij} + (\alpha + \beta)(\sigma_{ij,t+H-2} - \bar{\sigma}_{ij})] - \bar{\sigma}_{ij}) \\
& = \bar{\sigma}_{ij} + (\alpha + \beta)^2(E_t[\sigma_{ij,t+H-2}] - \bar{\sigma}_{ij}) \\
& = \bar{\sigma}_{ij} + (\alpha + \beta)^2(E_t[E_{t+H-3}[\sigma_{ij,t+H-2}]] - \bar{\sigma}_{ij}) \\
& = \bar{\sigma}_{ij} + (\alpha + \beta)^3(E_t[\sigma_{ij,t+H-3}] - \bar{\sigma}_{ij}) \\
& = \dots = \bar{\sigma}_{ij} + (\alpha + \beta)^H(\sigma_{ij,t} - \bar{\sigma}_{ij}),
\end{aligned}$$

where we have exploited the law of iterated expectations by which $E_t[X_{t+H-1}] = E_t[E_{t+H-1}[X_{t+H-1}]]$ and

$$E_{t+H-2} \left(\sigma_{ij,t+H-1} \frac{z_{i,t+H-1} z_{j,t+H-1}}{\rho_{ij,t+H-1}} - 1 \right) = E_{t+H-2}(\sigma_{ij,t+H-1}) E_{t+H-2} \left(\frac{z_{i,t+H-1} z_{j,t+H-1}}{\rho_{ij,t+H-1}} - 1 \right) = 0$$

as $E_{t+H-2}(z_{i,t+H-1} z_{j,t+H-1}) = \rho_{ij,t+H-1}$ and $\sigma_{ij,t+H-1}$ is independent by construction of $z_{i,t+H-1}$ and $z_{j,t+H-1}$ (think of it: this is a GARCH process and the filtered GARCH covariance just depends on past shocks).¹⁶ Similarly to what found in chapter 4, it is then rather simple to compute forecasts of future, H -step ahead covariances from $E_t[\sigma_{ij,t+H}^{GARCH}] = \bar{\sigma}_{ij} + (\alpha + \beta)^H(\sigma_{ij,t} - \bar{\sigma}_{ij})$.

As for the comparison, notice that from

$$\sigma_{ij,t+1}^{ES} = (1 - \lambda)R_{i,t}R_{j,t} + \lambda\sigma_{ij,t}$$

we have that (exploiting that in a RiskMetrics model $\alpha + \beta = 1$ and $E_t[\sigma_{ij,t+H}^{ES}] = \sigma_{ij,t}^{ES}$)

$$\begin{aligned}
E_t[\sigma_{ij,t+H}^{ES}] & = \bar{\sigma}_{ij} + \underbrace{(\alpha + \beta)}_{=1} E_t[E_{t+H-2}(\sigma_{ij,t+H-1} - \bar{\sigma}_{ij})] \\
& = E_t[E_{t+H-2}(\sigma_{ij,t+H-1})] \\
& = E_t[\sigma_{ij,t+H-2}] = E_t[E_{t+H-3}(\sigma_{ij,t+H-2})] = E_t[\sigma_{ij,t+H-3}] = \dots = E_t[\sigma_{ij,t}] = \sigma_{ij,t}.
\end{aligned}$$

This means that the forecast H -step ahead in an exponential smoothing model is simply the current estimate of covariance. This is what we have alluded to early on when we have claimed that under (12) covariance follows a non-stationary, unit root process. On the contrary, because

$$E_t[\sigma_{ij,t+H}^{GARCH}] = (\alpha + \beta)^H \sigma_{ij,t} + [1 - (\alpha + \beta)^H] \bar{\sigma}_{ij},$$

in the case of the GARCH(1,1) the current covariance receives a weight $(\alpha + \beta)^H$ which declines to zero as $H \rightarrow \infty$ under covariance stationarity, while the complement to one weight (which therefore increases to 1 as $H \rightarrow \infty$) is assigned to the unconditional covariance $\bar{\sigma}_{ij}$. In fact, because the RiskMetrics model can be obtained by setting $\alpha + \beta = 1$, in that case the GARCH forecast reduces to the exponential smoothing one.

¹⁶Here you need to pay attention to the notation: $E_{t+H-2}(z_{i,t+H-1} z_{j,t+H-1})$ is the time $t + H - 2$ predicted conditional covariance between $z_{i,t+H-1}$ and $z_{j,t+H-1}$ that we have denoted $\rho_{ij,t+H-1}$. Notice that properties such as

$$E_{t+H-l} \left(\frac{z_{i,t+H-l+1} z_{j,t+H-l+1}}{\rho_{ij,t+H-l+1}} - 1 \right) = 0$$

with $l \geq H$ have to be used repeatedly in the proof.

5. Dynamic Conditional Correlation (DCC) Modelling

If you listen to trading desk and asset management lingo, they will hardly talk about covariances: instead the focus will be on correlations, besides volatility. For instance, one interesting (worrisome) phenomenon is that all correlations tend to “skyrocket” during market crisis (bear regimes) and this—as well shall see in detail in chapter 7—has recently attracted considerable attention.¹⁷ Even though it is obvious that given any type (e.g., GARCH) models for covariances ($\hat{\sigma}_{ij,t+1}$) and variances ($\hat{\sigma}_{i,t+1}$ and $\hat{\sigma}_{j,t+1}$) separately, one can always compute the implied dynamic (prediction of) correlation,

$$\hat{\rho}_{ij,t+1} = \frac{\hat{\sigma}_{ij,t+1}}{\hat{\sigma}_{i,t+1}\hat{\sigma}_{j,t+1}},$$

one generation of financial econometricians has tried to actually write and estimate models for dynamic conditional correlations *directly*, i.e., when the model directly concerns the behavior of $\rho_{ij,t+1}$ over time. However, that task appears to be far from trivial for one obvious reason: $\rho_{ij,t+1} \in [-1, 1]$ and so a dynamic estimator would imply a need to constrain parameter estimates to avoid that for all times t , $|\rho_{ij,t+1}|$ exceeds 1. For instance, consider a model $\rho_{ij,t+1} = \delta_0 + \delta_1(R_t^i/\sigma_{i,t})(R_t^j/\sigma_{j,t}) + \delta_2\rho_{ij,t}$ and try to ask yourself what kind of restrictions on δ_1 and δ_2 may keep $|\rho_{ij,t+1}|$ from exceeding 1 given that $R_t^i \in [-1, +\infty)$.¹⁸

A more fruitful approach is the one that leads to specify and estimate DCC models. This approach is based on the idea that it is more appealing to model an appropriate auxiliary variable, $q_{ij,t+1}$. The DCC approach is based on a generalization to the vector/matrix case of the standard result that $\sigma_{ij,t+1} \equiv \rho_{ij,t+1}\sigma_{i,t+1}\sigma_{j,t+1} = \sigma_{i,t+1}\rho_{ij,t+1}\sigma_{j,t+1}$:

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

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

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

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 and prediction of $\mathbf{\Gamma}_{t+1}$, that will give $\hat{\mathbf{\Gamma}}_{t+1}$. In particular, 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 $\sigma_{i,t+1}^2 = \omega + \alpha(R_t^i - \delta\sigma_{i,t})^2 + \beta\sigma_{i,t}^2$ for $i = 1, 2, \dots, N$.

¹⁷The verb skyrocket appears in quotation marks because you should recall that correlations are defined in $[-1, 1]$ and they cannot diverge to $+\infty$.

¹⁸Please also ask yourself why $R_t^i \in [-1, +\infty)$, i.e., why $R_t^i < -1$ is impossible.

2. Model conditional covariances of *standardized* returns, $z_{t+1}^i \equiv R_{t+1}^i/\sigma_{t+1}^i$, 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:

$$\begin{aligned} \text{Cov}_t[z_{t+1}^i, z_{t+1}^j] &= \text{Cov}_t\left[\frac{R_{t+1}^i}{\sigma_{i,t+1}}, \frac{R_{t+1}^j}{\sigma_{j,t+1}}\right] = \frac{1}{\sigma_{i,t+1}\sigma_{j,t+1}}\text{Cov}_t\left[R_{t+1}^i, R_{t+1}^j\right] \\ &= \frac{\sigma_{ij,t+1}}{\sigma_{i,t+1}\sigma_{j,t+1}} = \rho_{ij,t+1} \end{aligned}$$

However, the GARCH-type modelling will not concern directly the standardized returns, but instead an auxiliary variable $q_{ij,t+1}$ to be estimated/predicted to be able to compute conditional correlations. Typically, the most popular model used in this second DCC step is:

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

which is a GARCH(1,1) for the auxiliary variable, written in deviations from the unconditional, long-run mean, usually set to be equal to the unconditional sample covariance of the data. An alternative that has been sometimes used is of a RiskMetrics type:

$$q_{ij,t+1} = (1 - \lambda)z_{t+1}^i z_{t+1}^j + \lambda q_{ij,t} \quad \forall i, j,$$

with $\lambda \in (0, 1)$. Note that these models apply to all pairs i and j , even when $i = j$, which is used in (15) below. How do you go now from a forecast of the auxiliary variable $q_{ij,t+1}$ to a forecast of correlations? Here one uses the transformation

$$\rho_{ij,t+1} = \frac{q_{ij,t+1}}{\sqrt{q_{ii,t+1}}\sqrt{q_{jj,t+1}}}. \quad (15)$$

Note that (15) guarantees by construction that $\rho_{ij,t+1} \in [-1, 1]$.¹⁹ Therefore, in the GARCH(1,1) case, $\bar{\rho}_{ij} = \omega_{ij}/(1 - \alpha - \beta)$. Interestingly, only the intercept parameter ω_{ij} is allowed to differ across different pairs of assets: also in this case, α and β are common across different pairs of assets; the same occurs for the single parameter λ in the RiskMetrics-type model. This implies that the persistence of the correlation between any two assets in the portfolio is the same, which is obviously unrealistic.²⁰ With reference to same two series of S&P 500 returns and the USD/yen exchange rate log-changes featured in Figures 1 and 2, Figures 3 and 4 show the predicted, one-step correlations

¹⁹Other transformations have been explored in the literature, for instance the Fisher transformation of the correlation coefficient: $\rho_{ij,t+1} = [\exp(2\nu_{t+1}) - 1]/[\exp(2\nu_{t+1}) + 1]$, where ν_{t+1} can be defined as any GARCH model using $z_{t+1}^1 z_{t+1}^2$ as innovations. This model is easy to implement because the positive definiteness of the conditional correlation matrix is guaranteed by the Fisher transformation. However, it is only a bivariate model and it is not clear how to generalize it to any $N \geq 2$.

²⁰In any event, because ω_{ij} is allowed to differ across pairs of assets, the fact that α and β or λ must be common across pairs of assets, does not imply that the level of the correlations at any time is the same across pairs of assets.

estimated from a DCC RiskMetrics and a DCC GARCH(1,1) model, respectively.

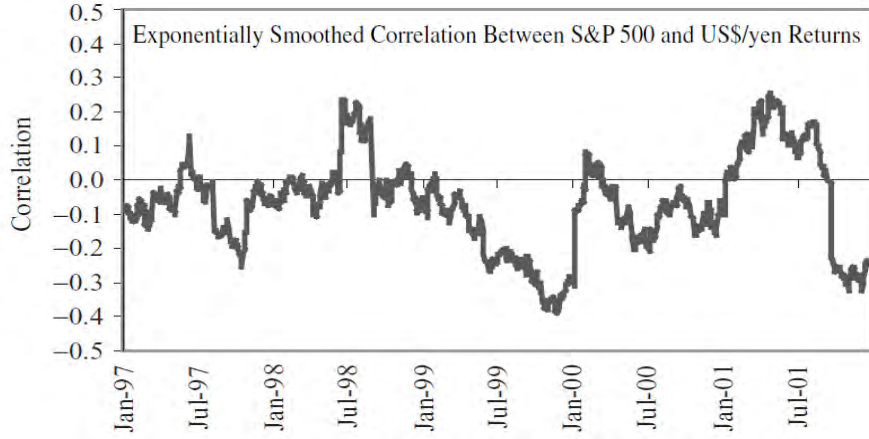


Figure 3: S&P500 vs. USD/Yen return DCC/RiskMetrics covariance estimate ($\lambda = 0.94$)

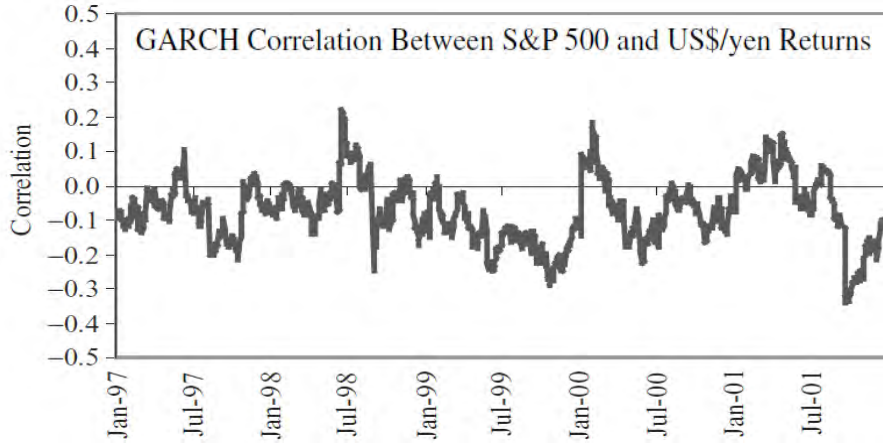


Figure 4: S&P500 vs. USD/Yen return DCC/GARCH(1,1) covariance estimate

Clearly, the two figures are quite different, even though the general dynamics of the correlation forecasts is qualitatively similar.

The RiskMetrics and GARCH/DCC models can also be written in matrix form as

$$\begin{aligned}\mathbf{Q}_{t+1} &= (1 - \lambda)\mathbf{z}_t\mathbf{z}'_t + \lambda\mathbf{Q}_t \\ \mathbf{Q}_{t+1} &= \mathbf{\Pi} + \alpha\mathbf{z}_t\mathbf{z}'_t + \beta\mathbf{Q}_t,\end{aligned}$$

where $\mathbf{z}_t \equiv [z_{t+1}^1 \ z_{t+1}^2 \ \dots \ z_{t+1}^N]'$, \mathbf{Q}_{t+1} is a $N \times N$ symmetric matrix that collects the values/predictions of the auxiliary variables $q_{ij,t+1}$,

$$\mathbf{Q}_{t+1} \equiv \begin{bmatrix} q_{11,t+1} & q_{12,t+1} & \dots & q_{1n,t+1} \\ q_{12,t+1} & q_{22,t+1} & \dots & q_{2n,t+1} \\ \vdots & \vdots & \ddots & \vdots \\ q_{1n,t+1} & q_{2n,t+1} & \dots & q_{nn,t+1} \end{bmatrix}$$

for $i, j = 1, 2, \dots, N$, and $\mathbf{\Pi}$ is a symmetric SPD matrix.²¹ \mathbf{Q}_{t+1} is SPD because it is a weighted average of positive semi-definite and positive definite matrices. This will in turn ensure that the correlation matrix $\mathbf{\Gamma}_{t+1}$ and the covariance matrix, $\mathbf{\Sigma}_{t+1}$, will be positive semi-definite. One often used variation of the DCC GARCH model features a *covariance targeting* variant:

$$\mathbf{Q}_{t+1} = (1 - \alpha - \beta)E[\mathbf{z}_t\mathbf{z}'_t] + \alpha\mathbf{z}_t\mathbf{z}'_t + \beta\mathbf{Q}_t,$$

which guarantees that the unconditional correlation will be identical to the sample unconditional correlation of the data.

DCC models are enjoying a massive popularity because they are easy to implement in 3 steps:

1. All the individual variances are estimated one by one;
2. The returns are standardized and the unconditional correlation matrix is estimated;
3. The correlation persistence parameters α and β are estimated.²²

One point that is worthwhile to make (see Bauwens et al., 2006, for additional details, as well as Section 6.2 that follows) is that the class of models that is based on non-linear combinations of univariate GARCH models in which one can specify separately, on the one hand, the individual conditional variances, and on the other hand, the conditional correlation matrix (this is called a *hierarchical* specification strategy) is not composed of DCC models only. For instance, Tse and Tsui (2002) specify their multivariate DCC as $\mathbf{\Sigma}_{t+1} \equiv \mathbf{D}_{t+1}\mathbf{\Lambda}_{t+1}\mathbf{D}_{t+1}$, where $\mathbf{\Lambda}_{t+1}$ follows a GARCH-like process but it is not necessarily a time-varying correlation matrix and has a factor structure. In practice, in this case the DCC formulates the conditional correlation directly as a weighted sum of past correlations.

DCC models are estimated by QMLE by construction: because the model is implemented in three different steps, even though in each of these stages a log-likelihood function is written and estimated, the overall outcome only represents a QMLE. In each of the stages, only few parameters are estimated simultaneously using numerical optimization. This feature makes DCC models extremely tractable for risk management of large portfolios: in the first step, one only estimates univariate GARCH-type models; the resulting GARCH inferences are then used to compute the standardized returns, $z_{t+1}^i \equiv R_{t+1}^i/\sigma_{t+1}^i$; in the final step, one only estimates two parameters— α and β —that apply to all pairs of assets. For instance, in the case of $N = 2$, in the first stage one solves (in the simple GARCH(1,1) case)

$$\max_{\omega_i, \alpha_i, \beta_i} \left\{ -\frac{T}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^T \log [\omega_i + \alpha_i R_{i,t-1}^2 + \beta_i \sigma_{i,t-1}^2] - \frac{1}{2} \sum_{t=1}^T \frac{R_{i,t}^2}{\omega_i + \alpha_i R_{i,t-1}^2 + \beta_i \sigma_{i,t-1}^2} \right\},$$

²¹Note that $q_{ii,t+1}$ is in general different from $\sigma_{i,t+1}^2$ obtained in the first step, $i = 1, 2, \dots, n$. This is important and it represents the logical sacrifice on which the two-step DCC estimation is based: the use of first-step predictions for variances that potentially differ from those that are then used to go from predictions of the $q_{ij,t+1}$ to predictions of correlations constrained to be in $[-1, 1]$.

²²This last claim and the fact that ω_{ij} has not been mentioned assumes that covariance targeting has been applied as we shall see in our Matlab session.

where $\sigma_{i,0}^2$ is initialized at

$$\sigma_{i,0}^2 = \frac{\omega_i}{1 - \alpha_i - \beta_i}.$$

In the second step, given the pair of time series $z_{t+1}^i \equiv R_{t+1}^i/\sigma_{t+1}^i$ and $z_{t+1}^j \equiv R_{t+1}^j/\sigma_{t+1}^j$, one solves

$$\max_{\alpha, \beta} -\frac{T}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^T \ln(1 - \rho_{12,t-1}^2) - \frac{1}{2} \sum_{t=1}^T \frac{(z_{t-1}^1)^2 + (z_{t-1}^2)^2 - 2\rho_{12,t-1} z_{t-1}^1 z_{t-1}^2}{(1 - \rho_{12,t-1}^2)},$$

where $\rho_{12,t-1} = q_{12,t+1}/[\sqrt{q_{11,t+1}}\sqrt{q_{22,t+1}}]$ and $q_{ij,t+1}$ follows either a GARCH or a RiskMetrics model, involving the parameters α and β in the former case (under covariance targeting), and λ in the latter. Notice that the variables that enter the log-likelihood are the standardized returns, z_t^i , and not the original raw returns themselves. We are essentially treating the standardized returns as actual observations here and this is an alternative way to appreciate the loss of estimation efficiency that QMLE implies in this application. In theory, we could obtain more precise estimators by estimating all the volatility models and the correlation model simultaneously, which would yield a ML estimator. In practice, this is not feasible for large portfolios, i.e. for cases in which N exceeds 3 or 4 different assets.

5.1. One detailed bivariate example

Consider the GARCH(1,1) version of the dynamic conditional correlation model

$$\begin{aligned} q_{ij,t+1}^{GARCH} &= E[z_{i,t}z_{j,t}] + \alpha(z_{i,t}z_{j,t} - E[z_{i,t}z_{j,t}]) + \beta(q_{ij,t}^{GARCH} - E[z_{i,t}z_{j,t}]) \\ \rho_{ij,t+1}^{(\cdot)} &\equiv \frac{q_{ij,t+1}^{(\cdot)}}{\sqrt{q_{ii,t+1}^{(\cdot)}q_{jj,t+1}^{(\cdot)}}} \end{aligned}$$

for the case of $N = 2$. We show that in the GARCH case, the covariance matrix of returns can be written as:

$$\begin{aligned} \Sigma_{t+1}^{GARCH} &= \mathbf{D}_{t+1}^{GARCH} \Gamma_{t+1,t+1}^{GARCH} \mathbf{D}_{t+1}^{GARCH} \\ \mathbf{D}_{t+1}^{GARCH} &= \sqrt{\text{diag}(\mathbf{Q}_{t+1}^{GARCH})} \quad \Gamma_{t+1}^{GARCH} = (\mathbf{D}_{t+1}^{GARCH})^{-1} \mathbf{Q}_{t+1}^{GARCH} (\mathbf{D}_{t+1}^{GARCH})^{-1} \\ \mathbf{Q}_{t+1}^{GARCH} &\equiv \begin{bmatrix} q_{11,t+1}^{GARCH} & q_{12,t+1}^{GARCH} \\ q_{12,t+1}^{GARCH} & q_{22,t+1}^{GARCH} \end{bmatrix} = E[\mathbf{z}_t \mathbf{z}_t'] (1 - \alpha - \beta) + \alpha \mathbf{z}_t \mathbf{z}_t' + \beta \mathbf{Q}_t^{GARCH}, \end{aligned}$$

where $\mathbf{z}_t \equiv [z_{1t} \ z_{2t}]'$ is the vector that collects standardized residuals from a GARCH(1,1) model and under the special assumption that $q_{ii,t+1}^{GARCH} = \sigma_{ii,t+1}^{GARCH}$ $i = 1, 2$.²³ The point simply consists in the patient verification of a few matrix expressions. Let's work backwards:

$$\mathbf{Q}_{t+1}^{GARCH} \equiv \begin{bmatrix} q_{11,t+1}^{GARCH} & q_{12,t+1}^{GARCH} \\ q_{12,t+1}^{GARCH} & q_{22,t+1}^{GARCH} \end{bmatrix} = (1 - \alpha - \beta) E \left(\begin{bmatrix} z_{1t}^2 & z_{1t}z_{2t} \\ z_{1t}z_{2t} & z_{2t}^2 \end{bmatrix} \right) +$$

²³Notice that in this case the same matrix \mathbf{D}_{t+1}^{GARCH} appears in the definitions of Σ_{t+1}^{GARCH} and of Γ_{t+1}^{GARCH} . In general this does not need to be the case, i.e., as stated in the subsection title, this represents a special example.

$$\begin{aligned}
& +\alpha \begin{bmatrix} z_{1t}^2 & z_{1t}z_{2t} \\ z_{1t}z_{2t} & z_{2t}^2 \end{bmatrix} + \beta \begin{bmatrix} q_{11,t}^{GARCH} & q_{12,t}^{GARCH} \\ q_{12,t}^{GARCH} & q_{22,t}^{GARCH} \end{bmatrix} \\
& = \begin{bmatrix} (1-\alpha-\beta)E(z_{1t}^2) + \alpha z_{1t}^2 + \beta q_{11,t}^{GARCH} & (1-\alpha-\beta)E(z_{1t}z_{2t}) + \alpha z_{1t}z_{2t} + \beta q_{12,t}^{GARCH} \\ (1-\alpha-\beta)E(z_{1t}z_{2t}) + \alpha z_{1t}z_{2t} + \beta q_{12,t}^{GARCH} & (1-\alpha-\beta)E(z_{2t}^2) + \alpha z_{2t}^2 + \beta q_{22,t}^{GARCH} \end{bmatrix}
\end{aligned}$$

which is a matrix collecting standard GARCH dynamic models for the $q_{ij,t+1}$ as defined above and in which

$$E(q_{12,t+1}^{GARCH}) = (1-\alpha-\beta)E(z_{1t}z_{2t}) + \alpha E(z_{1t}z_{2t}) + \beta E(q_{12,t}^{GARCH}) = E(z_{1t}z_{2t}).$$

Next, we need to show that $\mathbf{\Gamma}_{t+1}^{GARCH}$ is the correlation matrix implied by the process:

$$\begin{aligned}
\mathbf{\Gamma}_{t+1}^{GARCH} &= (\mathbf{D}_{t+1}^{GARCH})^{-1} \mathbf{Q}_{t+1}^{GARCH} (\mathbf{D}_{t+1}^{GARCH})^{-1} \\
&= \begin{bmatrix} \frac{1}{\sqrt{(1-\alpha-\beta)E(z_{1t}^2) + \alpha z_{1t}^2 + \beta q_{11,t}^{GARCH}}} & 0 \\ 0 & \frac{1}{\sqrt{(1-\alpha-\beta)E(z_{2t}^2) + \alpha z_{2t}^2 + \beta q_{22,t}^{GARCH}}} \end{bmatrix} \times \\
&\times \begin{bmatrix} (1-\alpha-\beta)E(z_{1t}^2) + \alpha z_{1t}^2 + \beta q_{11,t}^{GARCH} & (1-\alpha-\beta)E(z_{1t}z_{2t}) + \alpha z_{1t}z_{2t} + \beta q_{12,t}^{GARCH} \\ (1-\alpha-\beta)E(z_{1t}z_{2t}) + \alpha z_{1t}z_{2t} + \beta q_{12,t}^{GARCH} & (1-\alpha-\beta)E(z_{2t}^2) + \alpha z_{2t}^2 + \beta q_{22,t}^{GARCH} \end{bmatrix} \times \\
&\times \begin{bmatrix} \frac{1}{\sqrt{(1-\alpha-\beta)E(z_{1t}^2) + \alpha z_{1t}^2 + \beta q_{11,t}^{GARCH}}} & 0 \\ 0 & \frac{1}{\sqrt{(1-\alpha-\beta)E(z_{2t}^2) + \alpha z_{2t}^2 + \beta q_{22,t}^{GARCH}}} \end{bmatrix} \\
&= \begin{bmatrix} \frac{(1-\alpha-\beta)E(z_{1t}^2) + \alpha z_{1t}^2 + \beta q_{11,t}^{GARCH}}{\sqrt{[(1-\alpha-\beta)E(z_{1t}^2) + \alpha z_{1t}^2 + \beta q_{11,t}^{GARCH}][(1-\alpha-\beta)E(z_{1t}^2) + \alpha z_{1t}^2 + \beta q_{11,t}^{GARCH}]}} & \frac{(1-\alpha-\beta)E(z_{1t}z_{2t}) + \alpha z_{1t}z_{2t} + \beta q_{12,t}^{GARCH}}{\sqrt{[(1-\alpha-\beta)E(z_{1t}^2) + \alpha z_{1t}^2 + \beta q_{11,t}^{GARCH}][(1-\alpha-\beta)E(z_{2t}^2) + \alpha z_{2t}^2 + \beta q_{22,t}^{GARCH}]}} \\ \frac{(1-\alpha-\beta)E(z_{1t}z_{2t}) + \alpha z_{1t}z_{2t} + \beta q_{12,t}^{GARCH}}{\sqrt{[(1-\alpha-\beta)E(z_{1t}^2) + \alpha z_{1t}^2 + \beta q_{11,t}^{GARCH}][(1-\alpha-\beta)E(z_{2t}^2) + \alpha z_{2t}^2 + \beta q_{22,t}^{GARCH}]}} & \frac{(1-\alpha-\beta)E(z_{2t}^2) + \alpha z_{2t}^2 + \beta q_{22,t}^{GARCH}}{\sqrt{[(1-\alpha-\beta)E(z_{2t}^2) + \alpha z_{2t}^2 + \beta q_{22,t}^{GARCH}][(1-\alpha-\beta)E(z_{2t}^2) + \alpha z_{2t}^2 + \beta q_{22,t}^{GARCH}]}} \end{bmatrix} \\
&= \begin{bmatrix} 1 & \rho_{12,t+1}^{GARCH} \\ \rho_{12,t+1}^{GARCH} & 1 \end{bmatrix}.
\end{aligned}$$

Finally, if we re-assemble everything, we have:

$$\begin{aligned}
\mathbf{\Sigma}_{t+1}^{GARCH} &= \mathbf{D}_{t+1}^{GARCH} \mathbf{\Gamma}_{t+1}^{GARCH} \mathbf{D}_{t+1}^{GARCH} \\
&= \begin{bmatrix} \sqrt{q_{11,t}^{GARCH}} & 0 \\ 0 & \sqrt{q_{22,t}^{GARCH}} \end{bmatrix} \begin{bmatrix} 1 & \rho_{12,t+1}^{GARCH} \\ \rho_{12,t+1}^{GARCH} & 1 \end{bmatrix} \begin{bmatrix} \sqrt{q_{11,t}^{GARCH}} & 0 \\ 0 & \sqrt{q_{22,t}^{GARCH}} \end{bmatrix} \\
&= \begin{bmatrix} \sqrt{q_{11,t}^{GARCH}} & \sqrt{q_{11,t}^{GARCH}} \rho_{12,t+1}^{GARCH} \\ \sqrt{q_{22,t}^{GARCH}} \rho_{12,t+1}^{GARCH} & \sqrt{q_{22,t}^{GARCH}} \end{bmatrix} \begin{bmatrix} \sqrt{q_{11,t}^{GARCH}} & 0 \\ 0 & \sqrt{q_{22,t}^{GARCH}} \end{bmatrix} \\
&= \begin{bmatrix} q_{11,t}^{GARCH} & \sqrt{q_{11,t}^{GARCH}} \sqrt{q_{22,t}^{GARCH}} \rho_{12,t+1}^{GARCH} \\ \sqrt{q_{11,t}^{GARCH}} \sqrt{q_{22,t}^{GARCH}} \rho_{12,t+1}^{GARCH} & q_{22,t}^{GARCH} \end{bmatrix} \\
&= \begin{bmatrix} \sigma_{11,t+1}^{GARCH} & \sigma_{12,t+1}^{GARCH} \\ \sigma_{12,t+1}^{GARCH} & \sigma_{22,t+1}^{GARCH} \end{bmatrix}
\end{aligned}$$

Just one final remark: in this case we have assumed that α and β are taken to characterize the dynamic process for \mathbf{Q}_{t+1}^{GARCH} as a whole under the assumption of $q_{ii,t+1}^{GARCH} = \sigma_{ii,t+1}^{GARCH}$ or, in matrix

form, that $\mathbf{D}_{t+1}^{GARCH} = \mathbf{F}_{t+1}^{GARCH}$ in the more general structure:

$$\begin{aligned}\boldsymbol{\Sigma}_{t+1}^{GARCH} &= \mathbf{D}_{t+1}^{GARCH} \boldsymbol{\Gamma}_{t+1}^{GARCH} \mathbf{D}_{t+1}^{GARCH} \\ \mathbf{F}_{t+1}^{GARCH} &= \sqrt{\text{diag}(\mathbf{Q}_{t+1}^{GARCH})} \\ \boldsymbol{\Gamma}_{t+1}^{GARCH} &= (\mathbf{F}_{t+1}^{GARCH})^{-1} \mathbf{Q}_{t+1}^{GARCH} (\mathbf{F}_{t+1}^{GARCH})^{-1} \\ \mathbf{Q}_{t+1}^{GARCH} &= E[\mathbf{z}_t \mathbf{z}_t'] (1 - \alpha - \beta) + \alpha \mathbf{z}_t \mathbf{z}_t' + \beta \mathbf{Q}_t^{GARCH}.\end{aligned}$$

As you know from the lectures this need not be the case and in applied work it is typical to first estimate “some” GARCH models (it does not need to be GARCH(1,1), as it could be N(A)GARCH, GJR, etc.) for volatilities and then use the corresponding standardized residuals to estimate \mathbf{Q}_{t+1}^{GARCH} . This implies additional complexity but the advantage is that the α and β that characterize the dynamic process for correlations needs not to be the same as the one for volatility.

5.2. Asymmetric (GJR-type) DCC models

So far we have considered only symmetric correlation models where the effect of two positive shocks (i.e., standardized returns) is the same as the effect of two negative shocks of the same magnitude. But, just as we modeled the asymmetry in volatility (the leverage effect) in the univariate case, we may want to allow for a down-market effect in correlations. This can be achieved using an asymmetric DCC model where, for instance (also imposing correlation targeting):

$$\mathbf{Q}_{t+1} = (1 - \alpha - \beta) \hat{E}[\mathbf{z}_t \mathbf{z}_t'] + \alpha \mathbf{z}_t \mathbf{z}_t' + \beta \mathbf{Q}_t + \gamma (\mathbf{g}_t \mathbf{g}_t' - \hat{E}[\mathbf{g}_t \mathbf{g}_t']),$$

where “hatted” expectations will be estimated using sample moments, for instance

$$\hat{E}[\mathbf{z}_t \mathbf{z}_t'] = \frac{1}{T} \sum_{t=1}^T \mathbf{z}_t \mathbf{z}_t',$$

and the vectors \mathbf{g}_t are defined as the negative part of \mathbf{z}_t as follows:

$$g_{i,t} \equiv \begin{cases} z_{i,t} & \text{if } z_{i,t} < 0 \\ 0 & \text{if } z_{i,t} \geq 0 \end{cases}, \quad i = 1, \dots, N.$$

The term $\gamma (\mathbf{g}_t \mathbf{g}_t' - \hat{E}[\mathbf{g}_t \mathbf{g}_t'])$ captures a leverage effect in correlations: When $\gamma > 0$, then the correlation for asset i and j will increase more when $z_{i,t}$ and $z_{j,t}$ are negative than in any other case. This captures a phenomenon often observed in markets for risky assets: Their correlation increases more in down markets ($z_{i,t}$ and $z_{j,t}$ both negative) than in up markets ($z_{i,t}$ and $z_{j,t}$ both positive).

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,

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

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

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

where \mathbf{I}_N is a $N \times N$ identity matrix, and (similarly to chapter 7), $\mathbf{\Omega}_{t+1}^{1/2}$ is the square-root, or Cholesky decomposition, of the covariance matrix, such that²⁴

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

Even though in (16) we have specified $\mathbf{z}_{t+1} \text{ IID } \mathcal{N}(\mathbf{0}, \mathbf{I}_N)$, 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 $\mathbf{\Omega}_{t+1}^{1/2}$ is in no way the matrix of square roots of the elements of the full covariance matrix $\mathbf{\Sigma}_{t+1}$ (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 $\mathbf{\Sigma}_{t+1}$, 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 $\hat{\mathbf{\Sigma}}_{t+1}$. 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:

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

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

²⁵In fact, $\mathbf{\Omega}_{t+1}^{1/2}$ 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

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

In this general “VEC model”, each element of $\boldsymbol{\Sigma}_{t+1}$ is a linear function of the lagged squared errors and cross-products of errors and lagged values of the elements of $\boldsymbol{\Sigma}_{t+1}$. 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 $\boldsymbol{\Sigma}_t$ may affect each element of $\boldsymbol{\Sigma}_{t+1}$, 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 of²⁷

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

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 Gouriou, 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.

²⁷In 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 .

²⁸This 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.

²⁹The 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:

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

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

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

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

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

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:

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

³⁰For 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 Σ_{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:

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

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 $\boldsymbol{\Sigma}_{t+1}$ that collects the forecasts $\sigma_{kl,t+1}$ for $k, l = 1, 2, \dots, N$ 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):

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

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

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

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, $\Sigma_{t+1}^{DCC} \equiv \mathbf{D}_{t+1}\mathbf{\Gamma}_{t+1}\mathbf{D}_{t+1}$, is constant, so that $\Sigma_{t+1}^{CCC} \equiv \mathbf{D}_{t+1}\mathbf{\Gamma}\mathbf{D}_{t+1}$, you obtain a CCC in which—indeed—the first letter of the acronym means “constant”. Of course, the 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 represent 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, $\sigma_{ij,t+1} \equiv \rho_{ij}\sigma_{i,t+1}\sigma_{j,t+1} = \sigma_{i,t+1}\rho_{ij}\sigma_{j,t+1}$:

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

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

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

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 $\sigma_{i,t+1}^2 = \omega + \alpha(R_t^i - \delta\sigma_{i,t})^2 + \beta\sigma_{i,t}^2$ for $i = 1, 2, \dots, N$.

³¹The 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

$$\Sigma_{t+1} = \mathbf{C}\mathbf{C}' + \sum_{i=1}^p \mathbf{A}_i(\mathbf{R}_{t+1-i}\mathbf{R}'_{t+1-i})\mathbf{A}'_i + \sum_{j=1}^q \mathbf{B}_j\Sigma_{t+1-j}\mathbf{B}'_j,$$

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.

³²In 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 preferable 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:

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

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 $T^{-1}\sum_{t=1}^T \mathbf{R}_t \mathbf{R}'_t$, $vech(\mathbf{R}_{t-1} \mathbf{R}'_{t-1})$ 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:

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

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

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

³³By 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

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

which is equivalent to

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

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

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

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 volatilities

³⁴“At all times” here really means “for all possible 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 satisfying 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

$$\begin{aligned} \sigma_{pp,t} &\equiv Var_{t-1}[R_t^p] = Var_{t-1}[wR_{1t} + (1-w)R_{2t}] \\ &= w^2\sigma_{11,t} + (1-w)^2\sigma_{22,t} + 2w(1-w)\sigma_{12,t} \\ &= w^2(1-a^{11})T^{-1} \sum_{t=1}^T R_{1t}^2 + w^2a^{11}R_{1t-1}^2 + (1-w)^2(1-a^{33})T^{-1} \sum_{t=1}^T R_{2t}^2 + \\ &\quad + (1-w)^2a^{33}R_{2t-1}^2 + 2w(1-w)(1-a^{22})T^{-1} \sum_{t=1}^T R_{1t}R_{2t} + 2w(1-w)a^{22}R_{1t-1}R_{2t-1} \end{aligned}$$

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 possible to show that

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

³⁵In 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 completeness, 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

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

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

$$w^2 a^{11} R_{1t-1}^2 + (1-w)^2 a^{33} R_{2t-1}^2 + 2w(1-w)a^{22} R_{1t-1} R_{2t-1} = w^2 a^{pp} R_{1t}^2 + (1-w)^2 a^{pp} R_{2t}^2 + 2w(1-w)a^{pp} R_{1t-1} R_{2t-1}.$$

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

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

\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 $a^{pp} = a^{11}$

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

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

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

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:

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

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):

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

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 be 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 $\boldsymbol{\Sigma}_{t+1}$. 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 $\text{Var}_t[\boldsymbol{\epsilon}_{t+1}] = E_t[\boldsymbol{\epsilon}_{t+1}\boldsymbol{\epsilon}'_{t+1}] = \boldsymbol{\Psi}$ which is constant in time and semi-positive definite, and that the common factors are characterized by $E_t[\mathbf{f}_{t+1}] = \mathbf{0}$, $E_t[\mathbf{f}_{t+1}\boldsymbol{\epsilon}'_{t+1}] = \mathbf{0}$ (a matrix of zeros of the appropriate dimensions), and $E_t[\mathbf{f}_{t+1}\mathbf{f}'_{t+1}] = \text{diag}\{\sigma_{IP,t+1}^2, \sigma_{Infl,t+1}^2\}$. 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-meanned.

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

$$\begin{aligned} \text{Var}_t[\mathbf{R}_{t+1}] &= \boldsymbol{\Sigma}_{t+1} = \mathbf{B}E_t[\mathbf{f}_{t+1}\mathbf{f}'_{t+1}]\mathbf{B}' + E_t[\boldsymbol{\epsilon}_{t+1}\boldsymbol{\epsilon}'_{t+1}] + E_t[\mathbf{f}_{t+1}\boldsymbol{\epsilon}'_{t+1}] \\ &= \mathbf{B}E_t[\mathbf{f}_{t+1}\mathbf{f}'_{t+1}]\mathbf{B}' + \boldsymbol{\Psi} = \mathbf{b}_{IP}\mathbf{b}'_{IP}\sigma_{IP,t+1}^2 + \mathbf{b}_{Infl}\mathbf{b}'_{Infl}\sigma_{Infl,t+1}^2 + \boldsymbol{\Psi}, \end{aligned} \quad (19)$$

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:

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

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

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

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 $\mathbf{r}_{t+1} = \boldsymbol{\Phi}'\mathbf{R}_{t+1}$, where $\boldsymbol{\Phi} \equiv [\boldsymbol{\phi}_{IP} \ \boldsymbol{\phi}_{Infl}]$. It is then possible to re-write the expression for the

conditional covariance matrix of \mathbf{r}_{t+1} and in particular for $Var_t[r_{t+1}^1]$ and $Var_t[r_{t+1}^2]$ in terms of the two factor mimicking portfolios:

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

In particular, notice that

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

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 idiosyncratic 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

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

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

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

Therefore

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

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 $\mathbf{\Sigma}_{t+1} = \mathbf{b}_{IP} \mathbf{b}'_{IP} \sigma_{IP,t+1}^2 + \mathbf{b}_{Infl} \mathbf{b}'_{Infl} \sigma_{Infl,t+1}^2 + \mathbf{\Psi}$ 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,

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

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 Σ_{t+1} may be re-written in a BEKK form as:

$$\Sigma_{t+1} = \Psi^{**} + \mathbf{A}_{IP}(\epsilon_t \epsilon_t') \mathbf{A}'_{IP} + \mathbf{A}_{Infl}(\epsilon_t \epsilon_t') \mathbf{A}'_{Infl} + \mathbf{B}_{IP} \Sigma_t \mathbf{B}'_{IP} + \mathbf{B}_{Infl} \Sigma_t \mathbf{B}'_{Infl}.$$

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 (ϕ'_{IP} and ϕ'_{Infl}) defining the mimicking relationships. This can be seen from the fact that

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

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

$$\begin{aligned} \Sigma_{t+1} &= \Psi^* + \mathbf{b}_{IP} \mathbf{b}'_{IP} [\omega_{IP} + \alpha_{IP} \phi'_{IP} (\epsilon_t' \epsilon_t) \phi_{IP} + \beta_{IP} \phi'_{IP} \Sigma_t \phi_{IP}] + \\ &\quad + \mathbf{b}_{Infl} \mathbf{b}'_{Infl} [\omega_{Infl} + \alpha_{Infl} \phi'_{Infl} (\epsilon_t' \epsilon_t) \phi_{Infl} + \beta_{Infl} \phi'_{Infl} \Sigma_t \phi_{Infl}] \\ &= [\Psi^* + \mathbf{b}_{IP} \mathbf{b}'_{IP} \omega_{IP} + \mathbf{b}_{Infl} \mathbf{b}'_{Infl} \omega_{Infl}] + \mathbf{b}_{IP} \mathbf{b}'_{IP} \alpha_{IP} \phi'_{IP} (\epsilon_t' \epsilon_t) \phi_{IP} + \\ &\quad + \mathbf{b}_{Infl} \mathbf{b}'_{Infl} \alpha_{Infl} \phi'_{Infl} (\epsilon_t' \epsilon_t) \phi_{Infl} + \mathbf{b}_{IP} \mathbf{b}'_{IP} \beta_{IP} \phi'_{IP} \Sigma_t \phi_{IP} + \mathbf{b}_{Infl} \mathbf{b}'_{Infl} \beta_{Infl} \phi'_{Infl} \Sigma_t \phi_{Infl} \\ &= \Psi^{**} + \alpha_{IP} \mathbf{b}_{IP} \phi'_{IP} (\epsilon_t' \epsilon_t) \phi_{IP} \mathbf{b}'_{IP} + \alpha_{Infl} \mathbf{b}_{Infl} \phi'_{Infl} (\epsilon_t' \epsilon_t) \phi_{Infl} \mathbf{b}'_{Infl} + \\ &\quad + \beta_{IP} \mathbf{b}_{IP} \phi'_{IP} \Sigma_t \phi_{IP} \mathbf{b}'_{IP} + \beta_{Infl} \mathbf{b}_{Infl} \phi'_{Infl} \Sigma_t \phi_{Infl} \mathbf{b}'_{Infl} \\ &= \Psi^{**} + \mathbf{A}_{IP}(\epsilon_t \epsilon_t') \mathbf{A}'_{IP} + \mathbf{A}_{Infl}(\epsilon_t \epsilon_t') \mathbf{A}'_{Infl} + \mathbf{B}_{IP} \Sigma_t \mathbf{B}'_{IP} + \mathbf{B}_{Infl} \Sigma_t \mathbf{B}'_{Infl} \end{aligned}$$

where $\Psi^{**} \equiv \Psi^* + \mathbf{b}_{IP} \mathbf{b}'_{IP} \omega_{IP} + \mathbf{b}_{Infl} \mathbf{b}'_{Infl} \omega_{Infl}$, $\mathbf{A}_{IP} \equiv \sqrt{\alpha_{IP}} \mathbf{b}_{IP} \phi'_{IP}$, $\mathbf{A}_{Infl} \equiv \sqrt{\alpha_{Infl}} \mathbf{b}_{Infl} \phi'_{Infl}$, $\mathbf{B}_{IP} \equiv \sqrt{\beta_{IP}} \mathbf{b}_{IP} \phi'_{IP}$, and $\mathbf{B}_{Infl} \equiv \sqrt{\beta_{Infl}} \mathbf{b}_{Infl} \phi'_{Infl}$. 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

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

where all the parameters characterizing $\boldsymbol{\mu}_{t+1}$ and $\boldsymbol{\Omega}_{t+1}^{1/2}$ 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:

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

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

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

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 $\mathbf{z}_{t+1} \sim \text{IID } \mathcal{N}(\mathbf{0}, \mathbf{I}_N)$ 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

³⁶The 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 $\boldsymbol{\theta}$.

³⁷Gourieroux (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 $\mathbf{z}_{t+1} \sim \text{IID } \mathcal{N}(\mathbf{0}, \mathbf{I}_N)$ has been assumed—to apply standard *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, $\hat{\sigma}_{i,t+1} = \mathbf{e}_i' \hat{\Sigma}_{t+1} \mathbf{e}_i$ ($i = 1, 2, \dots, N$, i.e., the i th element on the main diagonal of $\hat{\Sigma}_{t+1}$). Here, we are clearly exploiting the fact that $\mathbf{z}_{t+1} \sim \text{IID } \mathcal{N}(\mathbf{0}, \mathbf{I}_N)$ implies that each of the elements \mathbf{z}_{t+1} must have a marginal normal distribution.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 $\mathbf{z}_{t+1} \text{ IID } \mathcal{D}(\mathbf{0}, \mathbf{I}_N)$ 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

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

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 $\mathbf{z}_{t+1} \text{ IID } \mathcal{D}(\mathbf{0}, \mathbf{I}_N)$, 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 $i = 1, 2, \dots, N$, 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

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

³⁹We face one additional problem when $\mathbf{z}_{t+1} \sim \text{IID}$ is tested by sequentially applying tests for each of the resulting N time 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,

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

exist, such as Hosking’s (1980) (here “HM” stands for Hosking’s multivariate test):

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

where $\mathbf{Y}_t \equiv \text{vech}(\mathbf{z}_t \mathbf{z}_t')$ 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} .

⁴⁰Principal 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).

⁴¹Of 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 components are themselves stationary.

3. Compute the N principal components of the matrix of standardized residuals \mathbf{Z} , obtaining the PC matrix \mathbf{P} , $\mathbf{P} = \mathbf{Z}\mathbf{L}$, 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: $\mathbf{C}_t = \mathbf{L}\mathbf{D}_t\mathbf{L}'$, 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:

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

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 $\mathbf{\Sigma}_t$.

Appendix A — A Quick Review of Basic Estimation Methods

1. A1. Where the OLS World Ends...

Consider two time series $x_{1:T} = \{x_1, x_2, \dots, x_T\}$ and $y_{1:T} = \{y_1, y_2, \dots, y_T\}$. At this stage of your studies in statistics and econometrics you are familiar with linear regression models:

$$y_t = \beta_0 + \beta_1 x_t + \epsilon_t.$$

In order to estimate this model by ordinary least squares (OLS), several assumptions need to be “in place”. Typically, one resorts to the so-called *weak OLS hypotheses*, i.e. $(\forall t, i, j \in \{1, 2, \dots, T\})$:

1. $E(\epsilon_t) = 0$;
2. $Cov(\epsilon_t, x_t) = 0$, which can also be written $E(\epsilon_t | x_t) = E(\epsilon_t)$ (deterministic regressors);
3. $Var(y_t | x_t) = Var(\epsilon_t) = \sigma^2 < \infty$ (homoskedasticity);
4. $Cov(\epsilon_i, \epsilon_j) = 0$ (no autocorrelation in residuals).

Hypothesis 3 and 4 can be summarized by saying that $Var(\epsilon_{1:T}) = \mathbf{I}_T \sigma^2$. The *strong OLS Hypotheses set* implies the four conditions above plus an assumption on the distribution of residuals:

$$\epsilon_t \sim N(0, \sigma^2) \quad t = 1, 2, \dots, T.$$

However, data from many real-life problems often fail to fulfill the OLS conditions, even in their weak form. In fact:

- Hypothesis 1 is trivially verified if you just add an intercept to your model.
- However, Hypothesis 2 seems considerably more problematic. For instance, suppose that you observe your independent variable with a random, white noise measurement error: $\tilde{x}_t = x_t + v_t$. In this case your regressor is no longer deterministic, i.e., fixed in repeated samples. Even if you know that $Cov(\epsilon_t, x_t) = 0$, what about $Cov(\epsilon_t, v_t)$? There is a wide literature dealing with stochastic regressors, which uses Two Stage Least Squares (2SLS) and Instrumental Variables.⁴²
- The potential issues with Hypotheses 3 and 4 are on the contrary particularly relevant in finance. As you now know so well, for many financial time series, the conditional variance $Var(y_{t+1} | \mathfrak{S}_t)$ (where \mathfrak{S}_t is an information set or an information structure determined by the financial application at hand) is not constant over time. In addition, when fitting OLS regression models to financial time series, the null of $Cov(\epsilon_i, \epsilon_j) = 0$ is often rejected. We shall

⁴²However, we will not discuss these estimators in this Appendix because these problems are not specific to finance applications, but you should be aware of the fact that Hypothesis 2 may be often problematic in applications.

see below that if Hypotheses 3 and 4 are violated, the OLS estimator is not any longer the best linear unbiased estimator (BLUE), even if it is still consistent. In fact, there is another unbiased estimator whose variance is smaller than the one of OLS: it is the Generalized Least Squares (GLS) estimator described below.

In addition, there is a further point that deserves to be discussed in depth. Consider a matrix of N time series $\mathbf{Y}_t = [y_{1,t}, y_{2,t}, \dots, y_{N,t}]'$ (also called a “Panel”) and a set of regressors $x_{a,1:T}, x_{b,1:T}, x_{c,1:T}, \dots$. How would you estimate the relationship between dependent variables and regressors in this case?

A first way to do that (since you are familiar with OLS) is to run a separate OLS regression for each time series $y_{i,t}$, $i = 1, \dots, N$. In fact, you know that the OLS estimator is BLUE among the estimators which are a linear function of $y_{i,t}$. However, your problem involves more than just one dependent variable. Is it possible to build an unbiased estimator which is a function of both $y_{1,t}, y_{2,t}, \dots, y_{N,t}$ and is more efficient than the OLS? The answer is yes, and we will talk more extensively later about models with multiple equations.

In the following subsections we will quickly review some estimators used in the econometrics literature, which you might find helpful when dealing with problems where OLS are no longer a feasible (or at least, an efficient) solution.

A2. Generalized (and Feasible Generalized) Least Squares

As already noted, in the OLS framework we assume that $Var(\epsilon_{1:T}) = Var(\epsilon) = \mathbf{I}\sigma^2$. Now, let's abandon this hypothesis and let's say that $Var(\epsilon) = \Sigma$, where Σ is any valid (i.e., symmetric and positive definite) variance-covariance matrix. If the matrix Σ is known (we will come back to this point later), it is possible to derive an estimator for regression coefficients in a way similar to what we normally do under OLS. As you shall recall, if the covariance matrix is positive definite (notice: this does not hold in case Σ is semi-positive definite), you can always invert it and write $\Sigma^{-1} = \mathbf{D}\mathbf{D}'$. You can then write your standard Least Squares problem in matrix form as:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

Where \mathbf{Y} is the $T \times 1$ vector collecting observations for the N dependent variables, \mathbf{X} is a $T \times k$ matrix of the k regressors, $\boldsymbol{\beta}$ is a $k \times 1$ vector of coefficients to be estimated, and $\boldsymbol{\epsilon}$ is a vector of residuals.⁴³ If we pre-multiply the regression model by the matrix \mathbf{D} we get:

$$\mathbf{D}'\mathbf{Y} = \mathbf{D}'\mathbf{X}\boldsymbol{\beta} + \mathbf{D}'\boldsymbol{\epsilon}$$

or

$$\tilde{\mathbf{Y}} = \tilde{\mathbf{X}}\boldsymbol{\beta} + \tilde{\boldsymbol{\epsilon}},$$

⁴³If needed, \mathbf{X} may be expanded to also include a vector of ones, that will then represent the constant of the regression model to be estimated as the corresponding element of the $k \times 1$ vector of coefficients $\boldsymbol{\beta}$.

where $\tilde{\mathbf{Y}} \equiv \mathbf{D}'\mathbf{Y}$, $\tilde{\mathbf{X}} \equiv \mathbf{D}'\mathbf{X}$, and $\tilde{\boldsymbol{\epsilon}} \equiv \mathbf{D}'\boldsymbol{\epsilon}$. Note that $Var(\tilde{\boldsymbol{\epsilon}}) = E(\tilde{\boldsymbol{\epsilon}}\tilde{\boldsymbol{\epsilon}}') = \mathbf{D}'E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}')\mathbf{D} = \mathbf{D}'(\mathbf{D}\mathbf{D}')^{-1}\mathbf{D} = \mathbf{D}'(\mathbf{D}')^{-1}\mathbf{D}^{-1}\mathbf{D} = \mathbf{I}$ because $\boldsymbol{\Sigma}^{-1} = \mathbf{D}\mathbf{D}'$ implies that $E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}') = \boldsymbol{\Sigma} = (\mathbf{D}\mathbf{D}')^{-1}$.⁴⁴ Coefficients can now be estimated by OLS, since any heteroskedasticity has been removed. The estimator will be:

$$\begin{aligned}\hat{\boldsymbol{\beta}}^{GLS} &= (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{Y}} = (\mathbf{X}'\mathbf{D}\mathbf{D}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{D}\mathbf{D}'\mathbf{Y} \\ &= (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Y}.\end{aligned}$$

This estimator is unbiased:

$$\begin{aligned}E(\hat{\boldsymbol{\beta}}^{GLS}) &= (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}E(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}) \\ &= (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\beta} + (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}E(\boldsymbol{\epsilon}) = \boldsymbol{\beta}.\end{aligned}$$

In the same way we can derive the variance-covariance matrix of the GLS estimators:

$$\begin{aligned}Var(\hat{\boldsymbol{\beta}}^{GLS}) &= (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'E(\tilde{\boldsymbol{\epsilon}}\tilde{\boldsymbol{\epsilon}}')\tilde{\mathbf{X}} \left[(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})' \right]^{-1} \\ &= (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\mathbf{I}\tilde{\mathbf{X}} \left[(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})' \right]^{-1} = (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1} \\ &= (\mathbf{X}'\mathbf{D}\mathbf{D}'\mathbf{X})^{-1} = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}.\end{aligned}$$

Note that if we were not taking into consideration heteroskedasticity and we were using OLS, our estimator would have been inefficient, but still unbiased. In fact:

$$\hat{\boldsymbol{\beta}}^{OLS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

implies that $E(\hat{\boldsymbol{\beta}}^{OLS}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}) = \boldsymbol{\beta}$. However,

$$Var(\hat{\boldsymbol{\beta}}^{OLS}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}')\mathbf{X}(\mathbf{X}'\mathbf{X})'^{-1} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X}(\mathbf{X}'\mathbf{X})'^{-1}$$

which can be proven to always exceed $Var(\hat{\boldsymbol{\beta}}^{GLS}) = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}$. In fact, if $\boldsymbol{\Sigma} = \mathbf{I}\sigma^2$ the two estimators are identical and also the two variances are: $Var(\hat{\boldsymbol{\beta}}^{OLS}) = Var(\hat{\boldsymbol{\beta}}^{GLS}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$. When $\boldsymbol{\Sigma} \neq \mathbf{I}\sigma^2$, we can write their ratio as:

$$\begin{aligned}Var(\hat{\boldsymbol{\beta}}^{OLS})/Var(\hat{\boldsymbol{\beta}}^{GLS}) &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X}(\mathbf{X}'\mathbf{X})'^{-1}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}) \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{D}\mathbf{D}'\mathbf{X}(\mathbf{X}'\mathbf{X})'^{-1}(\mathbf{X}'\mathbf{D}^{-1}\mathbf{D}^{-1}\mathbf{X}) \\ &= \mathbf{D}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{D}\mathbf{D}'\mathbf{X}(\mathbf{X}'\mathbf{X})'^{-1}\mathbf{X}'\mathbf{D}^{-1}.\end{aligned}$$

Because $\mathbf{X}(\mathbf{X}'\mathbf{X})'^{-1}\mathbf{X}'$ is an idempotent matrix, it follows that

$$\begin{aligned}Var(\hat{\boldsymbol{\beta}}^{OLS})/Var(\hat{\boldsymbol{\beta}}^{GLS}) &= \mathbf{D}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{D}\mathbf{D}'\mathbf{X}(\mathbf{X}'\mathbf{X})'^{-1}\mathbf{X}'\mathbf{D}^{-1} \\ &= \mathbf{D}^{-1}\mathbf{D}^{-1}\mathbf{D}\mathbf{D}'\mathbf{X}(\mathbf{X}'\mathbf{X})'^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})'^{-1}\mathbf{X}' = \mathbf{X}(\mathbf{X}'\mathbf{X})'^{-1}\mathbf{X}',\end{aligned}$$

which is always greater than 1. This is why the OLS estimator is less efficient than the GLS estimator.

⁴⁴ $Var(\tilde{\boldsymbol{\epsilon}}) = E(\tilde{\boldsymbol{\epsilon}}\tilde{\boldsymbol{\epsilon}}')$ because $E(\tilde{\boldsymbol{\epsilon}}) = E(\mathbf{D}'\boldsymbol{\epsilon}) = \mathbf{D}'E(\boldsymbol{\epsilon}) = \mathbf{0}$.

We now have to make an important remark. When estimating GLS in actual applications, the structure of Σ will never be known. You will rather have at your disposal some estimator $\hat{\Sigma}$ which is an estimator of Σ and that—assuming you want to focus only on consistent estimators, as you should—will converge to the “true” covariance matrix as the sample becomes large (for those of you who know more, we can say that $\text{plim } \hat{\Sigma} = \Sigma$). Our estimator will therefore be the *Feasible Generalized Least Squares* (FGLS) estimator:

$$\hat{\beta}^{FGLS} = (\mathbf{X}'\hat{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\Sigma}^{-1}\mathbf{Y}.$$

The specification of a peculiar structure/model for the variance-covariance matrix becomes therefore a problem of key importance. In a finite sample $\hat{\Sigma}$ is different from Σ . This might cause a bias in the GLS estimator, since we are not ex ante guaranteed (especially if the sample is very small) that $E[(\mathbf{X}'\hat{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\Sigma}^{-1}\epsilon] = \mathbf{0}$.

A3. Completing the Picture: the (Generalized) Method of Moments

The estimators that we have reviewed so far (OLS and GLS) and the ones that we will consider in the following pages can be derived using Maximum Likelihood Estimation (MLE), which is based on an assumption (called parametric) on the density of distribution of the residuals. In the case of OLS, this assumption is part of the *Strong OLS Hypotheses* set, and can be written as (see also A1): $\epsilon_t \sim N(0, \sigma^2)$. Given this hypothesis, it is possible to estimate parameters maximizing the natural logarithm of the likelihood function. In the OLS case:

$$\ln L(\mathbf{Y}; \beta, \sigma^2) = -\frac{N}{2} \ln 2\pi - \frac{N}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} (\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta).$$

However, OLS and GLS estimators do not require any parametric assumption on residuals; they can just be derived using orthogonality conditions. Parametric assumptions might be used in a second step, in order to determine the probabilistic properties of the estimators. This is why they are called *semiparametric* estimators. The underlying idea is that the sample statistics of a given sample converge to the true moments of the population if the size of the sample gets larger and larger. Sample statistics can then be used to impose conditions in order to estimate regression parameters. As the size of the sample becomes large, parameter estimates will converge to their true values. OLS and GLS can therefore be included in a wider group of estimation methods, which is called *Generalized Method of Moments (GMM)*.

To formalize the problem, suppose that we want to identify k parameters β_i ($i = 1, \dots, k$). We have to define $q \geq k$ sample moments

$$\hat{m}_j = \frac{1}{T} \sum_{t=1}^T f_{jt}(\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k) \quad j = 1, \dots, q,$$

knowing (or assuming) that $\text{plim } \hat{m}_i = m_i$, where $m_i = E[f_i(\beta_1, \beta_2, \dots, \beta_k)]$ is the true but unknown population moment. Then parameter estimates may be estimated by imposing the set of m

conditions:

$$\begin{aligned}
\bar{m}_1 - f_1(\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k) &= 0 \\
\bar{m}_2 - f_2(\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k) &= 0 \\
&\dots = 0 \\
\bar{m}_q - f_q(\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k) &= 0,
\end{aligned} \tag{21}$$

where \bar{m}_j is the sample value of a statistic of interest, $j = 1, \dots, q$. Basically, from a mathematical point of view, all that you are doing is to look for the set of parameter values $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$ which jointly minimizes the distance between the conditions imposed on parameters and sample moments. Notice that the number of moment conditions q may exceed the number of parameters to be estimated, k . When $q = k$, we speak of *just identified* GMM or simply method of moment estimates: in this case you impose the minimal number of conditions for the system (21) to be solved—assuming a solution exists—to find $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$. This is the case we have dealt with in chapter 5, when t-Student shocks have been introduced. Often in practice, one elects to set $q > k$ and in this case one additional problem is how to weight the different moment conditions, because in a mathematical perspective, it is clear that only by chance $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$ will exist such that all the conditions/equations in (21) may find a solution. Although this is a rather advanced topic to be developed in the appropriate context, one solution, which in fact also delivers a number of optimal properties consists of estimating $\hat{\beta}^{GMM}$ by solving:

$$\min_{\beta} [\bar{\mathbf{m}} - \mathbf{f}(\beta)]' \tilde{\mathbf{S}}^{-1} [\bar{\mathbf{m}} - \mathbf{f}(\beta)], \tag{22}$$

where $\bar{\mathbf{m}} \equiv [m_1 \ m_2 \ \dots \ m_q]'$,

$$\begin{aligned}
\mathbf{f}(\beta) &\equiv \left[T^{-1} \sum_{t=1}^T f_{1t}(\beta_1, \dots, \beta_k) \ T^{-1} \sum_{t=1}^T f_{2t}(\beta_1, \dots, \beta_k) \ \dots \ T^{-1} \sum_{t=1}^T f_{qt}(\beta_1, \dots, \beta_k) \right]' \\
\tilde{\mathbf{S}} &= \text{Var}[\bar{\mathbf{m}} - \mathbf{f}(\tilde{\beta})]
\end{aligned}$$

and $\tilde{\beta}$ is a first-step estimator that simply minimizes the quadratic form $[\hat{\mathbf{m}} - \mathbf{f}(\beta)]' [\hat{\mathbf{m}} - \mathbf{f}(\beta)]$. Basically, (22) means that you pick $\hat{\beta}^{GMM}$ to minimize a set of q moment conditions weighted by the inverse of their covariance matrix, so that the moment conditions that are estimated more precisely in the data sample will receive a higher weight. Clearly, a just identified GMM estimation problem corresponds to (22) when $\tilde{\mathbf{S}} = \mathbf{I}$.

How can we write an OLS or a GLS estimator as a GMM one? Here recall that one of the Weak OLS Hypotheses stated that $Cov(\epsilon_t, x_t) = 0$. Thus, the GMM condition for estimating β is:

$$E[\epsilon' \mathbf{X}] = E[(\mathbf{Y} - \mathbf{X}\beta)' \mathbf{X}] = \mathbf{0}.$$

The same happens for GLS, with the only difference that we have to take into account the structure of the variance-covariance matrix:

$$E[(\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\beta)' \tilde{\mathbf{X}}] = E[(\mathbf{Y} - \mathbf{X}\beta)' \Sigma^{-1} \mathbf{X}] = \mathbf{0}.$$

A4. Dealing with More than One Equation: Seemingly Unrelated Regressions (SUR)

Let's now consider a situation where there are two or more dependent variables, say $N \geq 2$.

The regression model is

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

$$\begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \dots \\ \mathbf{Y}_N \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & 0 & \dots & 0 \\ 0 & \mathbf{X}_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \mathbf{X}_k \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \dots \\ \boldsymbol{\beta}_N \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \\ \dots \\ \boldsymbol{\epsilon}_N \end{bmatrix},$$

where $\boldsymbol{\beta}_i$ is the vector of coefficients for the i th equation, $i = 1, \dots, N$. The covariance matrix of the residuals is:

$$\begin{aligned} \text{Var}(\boldsymbol{\epsilon}) &= E \left(\begin{bmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \\ \vdots \\ \boldsymbol{\epsilon}_N \end{bmatrix} \begin{bmatrix} \boldsymbol{\epsilon}'_1 & \boldsymbol{\epsilon}'_2 & \dots & \boldsymbol{\epsilon}'_N \end{bmatrix} \right) \\ &= E \begin{bmatrix} \boldsymbol{\epsilon}_1 \boldsymbol{\epsilon}'_1 & \boldsymbol{\epsilon}_1 \boldsymbol{\epsilon}'_2 & \dots & \boldsymbol{\epsilon}_1 \boldsymbol{\epsilon}'_N \\ \boldsymbol{\epsilon}_2 \boldsymbol{\epsilon}'_1 & \boldsymbol{\epsilon}_2 \boldsymbol{\epsilon}'_2 & \dots & \boldsymbol{\epsilon}_2 \boldsymbol{\epsilon}'_N \\ \vdots & \dots & \ddots & \vdots \\ \boldsymbol{\epsilon}_N \boldsymbol{\epsilon}'_1 & \boldsymbol{\epsilon}_N \boldsymbol{\epsilon}'_2 & \dots & \boldsymbol{\epsilon}_N \boldsymbol{\epsilon}'_N \end{bmatrix} = \begin{bmatrix} \mathbf{I}\sigma_{11} & \mathbf{I}\sigma_{12} & \dots & \mathbf{I}\sigma_{1n} \\ \mathbf{I}\sigma_{21} & \mathbf{I}\sigma_{22} & \dots & \mathbf{I}\sigma_{2n} \\ \vdots & \dots & \ddots & \vdots \\ \mathbf{I}\sigma_{n1} & \mathbf{I}\sigma_{n2} & \dots & \mathbf{I}\sigma_{nn} \end{bmatrix} = \boldsymbol{\Omega}, \end{aligned}$$

where $\sigma_{ij} \equiv \text{Cov}(\epsilon_{it}, \epsilon_{jt})$ is the covariance between the residuals of regressions i and j . As we said early on, a first approach to the problem would be to estimate a separate OLS model for each regression equation. However, as we have seen when introducing GLS methods, this might not be the most efficient solution. If we look at the model above, the equations for the different dependent variables are apparently unrelated as far as the conditional mean is concerned. However, relationships exist through the covariance matrix $\boldsymbol{\Omega}$. Working through an algebra very similar to the one shown in Section 2, it is possible to derive the Seemingly Unrelated Regressions estimator:

$$\hat{\boldsymbol{\beta}}^{SUR} = (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{Y}.$$

This estimator is BLUE with variance

$$\text{Var}(\hat{\boldsymbol{\beta}}^{SUR}) = (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}.$$

Also for the SUR estimator we have to bear in mind that usually we do not know the matrix $\boldsymbol{\Omega}$. This may pose problems in small samples. Interestingly, the OLS equation by equation estimator and the SUR estimator are equivalent in only two very special situations:

- $\boldsymbol{\Omega}$ is a diagonal matrix, or in other words $\sigma_{ij} = 0$ if $i \neq j$.
- The set of regressors is the same for all the equations: $\mathbf{X}_i = \mathbf{X}$, $i = 1, 2, \dots, N$.

When you are dealing with a single regression model ($N = 1$), the key issue is obviously only the choice of the right set of regressors. However, as you can understand after our discussion in this section, in the case of a system of regression models ($N \geq 2$), there is another choice you have to make: you have to decide whether to model a time series “on its own” or along with other time series. If you choose the right set of time series and regressors, you will be able to exploit cross-sectional information and improve the efficiency of your estimates by obtaining a GLS-type estimator.

This is the appropriate point in which we can introduce some ideas on estimation of models with fixed and random effects.⁴⁵ These models are useful when you are analyzing a population (e.g., the set of EU countries) and you want to check whether there are any structural differences among units in the population (e.g., countries) conditionally on a common set of explanatory variables (GDP, current account, inflation, etc.). Consider the regression of different vectors of dependent variables, $\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_N$, on a matrix of k independent variables \mathbf{X} . We can write:

$$y_{i,t} = \beta_{0,i} + \sum_{j=1}^k \beta_{j,i} x_{j,t} + \epsilon_{i,t} \quad i = 1, \dots, N.$$

Assume that in this model $\beta_{j,i} = \beta_j \forall j$, $E(\epsilon_i) = 0$, $E(\epsilon_i \epsilon_i') = \mathbf{I} \sigma_{ii}$, $E(\epsilon_i \epsilon_j') = \mathbf{0}$. Summing up, in this model cross-sectional covariances between errors are null, while the set of regressors and regression slopes are the same for each dependent variable. The only estimator measuring cross sectional differences is the intercept $\beta_{0,i}$ ($i = 1, \dots, N$). In practice we can think of the intercept as a constant common to all models plus a fixed component that changes according to the dependent variable. In formulas: $\beta_{0,i} = \bar{\beta}_0 + \gamma_i$. This model is known as a model with *fixed effects*. The name comes from the fact that the cross-sectional effects captured by the intercept (the γ_i s) are “fixed” for each dependent variable. This model (as usual) has its own BLUE estimator for the regression coefficients.

In order to use a model with *fixed effects*, you need to assume that your model encloses the whole cross-sectional dimension of the population. If you think about it, when you are introducing fixed effects, you completely neglect any conditional (cross-sectional) volatility of the intercept. This can be done only if you observe all the individuals (e.g., countries) that are part of the population and you can perfectly identify the intercept effect for each one of them. If your population is a sub-sample of a larger set, you need to take into account this conditional volatility and use *random effects*. A model with *random effects* is very similar to a model with *fixed effects*, even though there is one key difference. In a random effect model, the intercept for the dependent variable i is

$$\beta_{0,i} = \bar{\beta}_0 + \mu_i,$$

⁴⁵This is a rather advanced topic and its appearance (while it has not emerged in the lectures and it will not) reflect from my discomfort with the large number of MSc. Finance students that deal with panels in their theses and in their jobs: some quick introduction is most needed.

where μ_i is a random error such that: $E(\mu_i) = 0$, $E(\mu_i^2) = \sigma_\mu^2$, $E(\mu_i\mu_j) = 0$, $E(\mu_i\epsilon_j) = \mathbf{0}$. So the model can be re-written as:

$$y_{i,t} = \bar{\beta}_0 + \sum_{j=1}^k \beta_{j,i} x_{j,t} + \mu_i + \epsilon_{i,t} \quad i = 1, \dots, N.$$

Also *random effects* have their own BLUE estimator. Greene (2008) represents a great starting point for additional details and the formulas for the BLUE estimators mentioned above.

A5. Simultaneous Equations and Vector Autoregressive Models

In our discussion so far we have considered very general econometric frameworks. When modeling multiple time series in economics, a key question is the study of contemporaneous and lagged casual relationships between variables. *Simultaneous Equations Models* are very popular in macro-economics, where *Structural Models* are used to bring theoretical frameworks with large number of equilibrium equations to the data. A consistent modeling of causality becomes of key importance. One of the most basic (and most popular) frameworks in this area are *Vector Autoregressive Models* (VAR).

VARs are based on the assumption that, given a vector of time series, there is a feedback relationship between the current value of each time series and lagged values of the same or other series. In practice this model can be seen as an autoregressive model for multiple time series. In formulas, a VAR model with p lags, also called VAR(p) can be written as:

$$\mathbf{Z}_t = \Phi_0 + \Phi_1 \mathbf{Z}_{t-1} + \Phi_2 \mathbf{Z}_{t-2} + \dots + \Phi_p \mathbf{Z}_{t-p} + \epsilon_t,$$

where \mathbf{Z}_t and ϵ_t are both $N \times 1$ vectors. Note that this model can always be written as a VAR(1) by re-arranging the equations in the following state-space form:

$$\begin{bmatrix} \mathbf{Z}_t \\ \mathbf{Z}_{t-1} \\ \vdots \\ \mathbf{Z}_{t-p+1} \end{bmatrix} = \begin{bmatrix} \Phi_0 \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} \Phi_1 & \Phi_2 & \dots & \Phi_p \\ \mathbf{I} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \dots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{I} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{Z}_{t-1} \\ \mathbf{Z}_{t-2} \\ \vdots \\ \mathbf{Z}_{t-p} \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}.$$

Because of this result, let's focus now on the simple framework: $\mathbf{Z}_t = \Phi_0 + \Phi_1 \mathbf{Z}_{t-1} + \epsilon_t$. First, this model is estimated by OLS, because the set of regressors is the same in each equation, and the SUR estimator coincides with the OLS one. Second, the estimation of this model involves a large number of parameters. Consider an example in which \mathbf{Z}_t has four elements. Even if you neglect intercepts you need to fill the full matrix Φ_1 (and the variance-covariance matrix of residuals):

$$\Phi_1 = \begin{bmatrix} \phi_{11} & \phi_{12} & \phi_{13} & \phi_{14} \\ \phi_{21} & \phi_{22} & \phi_{23} & \phi_{24} \\ \phi_{31} & \phi_{32} & \phi_{33} & \phi_{34} \\ \phi_{41} & \phi_{42} & \phi_{43} & \phi_{44} \end{bmatrix}$$

The presence of a large number of parameters may pose tricky risks of “over-fitting” the data, lowering saturation ratios below what is sensible to tolerate. By this, we mean that some of the

parameters in the matrix Φ_1 might come out to be significantly different from zero, even if they do not explain any meaningful relationship in the data. This is why in estimating VAR matrices econometricians usually impose constraints and set ex-ante some entries to equal zero. The stationarity condition for a VAR(1) model is somehow similar to the one that we impose for an AR(1) model. Unconditional moments are:

$$\begin{aligned} E(\mathbf{Z}) &= (I - \Phi_1)^{-1} \Phi_0 \\ \text{vec}(\text{Var}(\mathbf{Z})) &= (I - \Phi_1 \otimes \Phi_1)^{-1} \text{vec}(\Sigma), \end{aligned}$$

where \otimes is the Kronecker product of two matrices.⁴⁶ Conditional k -steps ahead moments can be derived as:

$$\begin{aligned} E(\mathbf{Z}_{t+k}|\mathbf{Z}_t) &= \Phi_0 + \Phi_1 \Phi_0 + \Phi_1^2 \Phi_0 + \dots + \Phi_1^{k-1} \Phi_0 + \Phi_1^k \mathbf{Z}_t \\ \text{Var}(\mathbf{Z}_{t+k}|\mathbf{Z}_t) &= \Sigma + \Phi_1 \Sigma \Phi_1' + \Phi_1^2 \Sigma \Phi_1^{2'} + \dots + \Phi_1^{k-1} \Sigma (\Phi_1^{k-1})'. \end{aligned}$$

Unconditional moments can then be obtained as the limit of the conditional moments as $k \rightarrow \infty$. Remember that we can always write the eigendecomposition: $\Phi_1 = \mathbf{C} \mathbf{\Lambda} \mathbf{C}$; then: $\Phi_1^k = \mathbf{C} \mathbf{\Lambda}^k \mathbf{C}$. To be sure that Φ_1^k is well defined for $k \rightarrow \infty$, we need to ask that $\mathbf{\Lambda}^k$ is well defined because

$$\mathbf{\Lambda}^k = \begin{bmatrix} \lambda_1^k & 0 & 0 & 0 \\ 0 & \lambda_2^k & 0 & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \lambda_N^k \end{bmatrix}.$$

This is equivalent to requiring that all the eigenvalues of Φ_1 fall in the unit circle. If this was not enough and you want to know more, it time to consult Hamilton (1994), which remains the most complete references on the market for specification and estimation of vector linear models.

⁴⁶The $\text{vec}(\cdot)$ operator simply stacks the elements of a $n \times n$ matrix by concatenating them vertically, i.e., if the matrix \mathbf{A} is written as $\mathbf{A} \equiv [\mathbf{a}_1 \ \mathbf{a}_2 \ \dots \ \mathbf{a}_n]$ where \mathbf{a}_j is a $n \times 1$ vector ($j = 1, 2, \dots, n$), then $\text{vec}(\mathbf{A}) = [\mathbf{a}_1' \ \mathbf{a}_2' \ \dots \ \mathbf{a}_n']'$ or

$$\text{vec}(\mathbf{A}) = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_n \end{bmatrix}.$$

The Kronecker product is instead the element-wise multiplication operator. An example will suffice:

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \otimes \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} & a_{11}b_{12} & a_{12}b_{11} & a_{12}b_{12} \\ a_{11}b_{21} & a_{11}b_{22} & a_{12}b_{21} & a_{12}b_{22} \\ a_{21}b_{11} & a_{21}b_{12} & a_{22}b_{11} & a_{22}b_{12} \\ a_{21}b_{21} & a_{21}b_{22} & a_{22}b_{21} & a_{22}b_{22} \end{bmatrix}.$$

Appendix B — A Matlab[®] Workout

You are a European investor with the Euro as a reference currency. Using *monthly* data in STOCKINT2013.XLS, construct monthly *excess* returns (*in Euros*) using the three price indices DS Market-PRICE indices for Germany, the US and the UK.

1. For the sample period April 1977 - December 2009, plot the values of each of the three individual indices (in logarithmic terms) and excess returns for the equally weighted portfolio denominated in Euros. Make sure to include the dividends paid by each of the three indices in each of the monthly return series. Notice that the monthly data made available on the course web site also include data on the dividend yield on index i ($i = \text{GER, US, UK}$), $I_{i,t}$, defined as:⁴⁷

$$dy_{i,t} \equiv \frac{D_{i,t}}{I_{i,t-1}}.$$

2. Over the same sample, use a GARCH(1,1)-DCC(1) model with constant mean to “estimate” (filter) the dynamics of the covariance matrix of *excess* returns at monthly frequency. The model is:

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

where $r_{i,t+1}$ denotes excess returns ($i = \text{US, UK, Ger}$). Make sure to extract and print parameter estimates, their robust standard errors, and their t-ratios. Also extract the dynamic, conditional covariance matrix implied by the model and use a three-panel plot to graph the one-month ahead (predicted) expected excess returns, volatilities, and correlations. [*Hint*: You have to use Kevin Sheppard’s MFE toolbox; you can find a link on the course web page; before blindly proceeding, please make sure to read and understand what the toolbox allows you to do]

3. Use the dynamic variance-covariance matrix filtered from question 2 and the unconditional historical means of excess returns to build an in-sample, recursive dynamic Markowitz portfolio based on the simple expression

$$\mathbf{w}_t^{Markow} = [\hat{\Sigma}_{t+1}^{GC/DCC}]^{-1} \hat{\boldsymbol{\mu}},$$

⁴⁷As a result, the log- (continuously compounded) return on the stock index i between $t - 1$ and t will be defined as:

$$R_{i,t} \equiv \ln\left(\frac{I_{i,t}}{I_{i,t-1}} + dy_{i,t}\right).$$

In the following, while $R_{i,t}$ denotes returns, $r_{i,t}$ will denote returns in excess of the Euro riskless rate.

where $\hat{\boldsymbol{\mu}} \equiv [\hat{\mu}_{us} \ \hat{\mu}_{uk} \ \hat{\mu}_{ger}]'$ and implicitly you have imposed a unit coefficient of risk aversion. Plot these recursive portfolio weights.

- Now we turn instead to build and use models for the conditional mean, something we have avoided so far. As a starter, estimate by OLS a “contagion regression model” for the three stock indexes:

$$\begin{aligned} r_{us,t+1} &= \beta_0^{us} + \beta_1^{us} \ln dy_{us,t} + \beta_2^{us} p_{us,t} + \beta_3^{us} p_{uk,t} + \beta_4^{us} p_{ger,t} + \epsilon_{us,t+1} \\ r_{uk,t+1} &= \beta_0^{uk} + \beta_1^{uk} \ln dy_{us,t} + \beta_2^{uk} p_{us,t} + \beta_3^{uk} p_{uk,t} + \beta_4^{uk} p_{ger,t} + \epsilon_{uk,t+1} \\ r_{ger,t+1} &= \beta_0^{ger} + \beta_1^{ger} \ln dy_{us,t} + \beta_2^{ger} p_{us,t} + \beta_3^{ger} p_{uk,t} + \beta_4^{ger} p_{ger,t} + \epsilon_{ger,t+1} \end{aligned}$$

$$\begin{bmatrix} \epsilon_{us,t+1} \\ \epsilon_{uk,t+1} \\ \epsilon_{ger,t+1} \end{bmatrix} \sim N(\mathbf{0}, \text{diag} \{ \sigma_{us}^2, \sigma_{uk}^2, \sigma_{ger}^2 \}).$$

where $p_{us,t}$, $p_{uk,t}$, and $p_{ger,t}$ are log-index levels expressed in euros.⁴⁸ Make sure to extract and print conditional mean parameter estimates, their standard errors, and their t-ratios. Use a three-panel plot to graph the one-month ahead (predicted) expected excess returns, volatilities, and correlations.

- Use the dynamic conditional means predicted from question 4 and the unconditional historical variances and covariances of excess returns to build an in-sample, recursive dynamic Markowitz portfolio based on

$$\mathbf{w}_t^{Markow} = [\text{diag} \{ \hat{\sigma}_{us}^2, \hat{\sigma}_{uk}^2, \hat{\sigma}_{ger}^2 \}]^{-1} \hat{\boldsymbol{\mu}}_t,$$

where $\hat{\boldsymbol{\mu}}_t \equiv [E_t(r_{us,t+1}) \ E_t(r_{uk,t+1}) \ E_t(r_{ger,t+1})]'$. Plot these recursive portfolio weights. If needed, you may decide to cut excessively large and small weights (for instance, exceeding 400% in absolute value) and produce a parallel plot in which visibility may be enhanced. Are these weights economically sensible?

- Bring now the models of questions 2-3 and 4-5 together, by estimating by iterated MLE (see below for details), the following restricted VAR(1) model that jointly captures the presence of time variation in conditional means, conditional variances, and conditional covariances:⁴⁹

$$\begin{aligned} r_{us,t+1} &= \beta_0^{us} + \beta_2^{us} r_{us,t} + \epsilon_{us,t+1} \\ r_{uk,t+1} &= \beta_0^{uk} + \beta_3^{uk} r_{uk,t} + \epsilon_{uk,t+1} \\ r_{ger,t+1} &= \beta_0^{ger} + \beta_4^{ger} r_{ger,t} + \epsilon_{ger,t+1} \end{aligned} \quad \begin{bmatrix} \epsilon_{us,t+1} \\ \epsilon_{uk,t+1} \\ \epsilon_{ger,t+1} \end{bmatrix} \sim N\left(\mathbf{0}, \boldsymbol{\Omega}_{t+1}^{GC/DCC}\right).$$

Notice that by estimating this model you are generalizing the model in question 4 not only because you are allowing variances to vary over time, but also (or especially) because you

⁴⁸Estimation by OLS implies that you will not be able to estimate the covariance of the shocks to different stock indices, which are therefore simply set to be zero throughout the exercise.

⁴⁹This VAR is restricted because each equation simply features a dependence of future excess returns in market j from past excess returns in the same market. Please make sure to understand what a full, unrestricted VAR would imply.

are no longer constraining the correlations to be always zero. Make sure to extract and print parameter estimates, their standard errors, and their t-ratios. Use a three-panel plot to graph the one-month ahead (predicted) expected excess returns, volatilities, and correlations. [*Hint*: the reference above to “iterated MLE” refers to the fact that Kevin Sheppard’s GARCH/DCC utilities will *not* allow you to jointly estimate the regression coefficients and also the GARCH/DCC dynamic covariance matrix; therefore what you are advised to do is to first estimate the residuals from the regression models, then fit on them your GARCH/DCC model, and finally proceed to re-estimate the regression model by GLS when the covariance matrix is the one estimated by GARCH/DCC].⁵⁰

- Using the dynamic conditional means, variances, and covariances predicted in question 6 compute in-sample, recursive dynamic Markowitz portfolios based on

$$\mathbf{w}_t^{Markow} = [\hat{\Omega}_{t+1}^{GC/DCC}]^{-1} \hat{\boldsymbol{\mu}}_t^{2FGLS},$$

where $\hat{\boldsymbol{\mu}}_t^{2FGLS}$ is now estimated from the second, GLS-type step of question 6. Plot these recursive portfolio weights. If needed, you may decide to cut excessively large and small weights (for instance, exceeding 500% in absolute value) and produce a parallel plot in which visibility may be enhanced. Are these weights economically sensible?

- With reference to the out-of-sample period January 2010 - December 2012, proceed to compute optimal weights for two models: the unconditional mean GARCH/DCC model of question 2; the VAR/GARCH/DCC of question 6. Perform the calculation in the following way: for both models use the same estimated conditional mean (the intercept in the first case, the restricted VAR parameters in the latter) and the GARCH/DCC parameters estimated in questions 2 and 6, that you should have saved; compute the dynamic covariance matrix on the basis of those parameter performing the updating on the basis of the out-of-sample forecast errors over the out-of-sample period. They weights will come from the classical Markowitz formula. After obtaining the weights, compute the realized Sharpe ratios over the out of sample period. Compare these realized Sharpe ratios with those that you would have achieved by simply investing all of your wealth in each of the three stock indices under consideration. [*Hint*: In this case it is a very good idea to use Kevin Sheppard’s `dcc_mvgarch_cov` function]
- Load the daily data set already used on many occasions before. Construct the monthly non-overlapping cumulative returns and monthly non-overlapping realized variance (RV) series for the US and German stock indices. Repeat the same exercise to construct monthly non-overlapping realized covariance (RCOV) series for the US and German stock indices. Plot the

⁵⁰For our purposes, it will be sufficient to iterate these two estimation steps only once, which is what we normally do with feasible two-stage GLS. However, one may want to iterate these two steps as long as it takes to reach convergence in parameter estimates. However, under the assumption of correct specification of the dynamic model, notice that also the two-stage MLE is consistent. Please see the lecture notes that are available on the class web site.

resulting monthly realized variances, covariances, and the autocorrelograms of these series. Are the natural logarithms of the resulting series normally distributed?

- Use the first 100 observations of the monthly realized variance sample you have obtained in point 9 to estimate two univariate AR(1) models for the US and German RVs. Repeat the same estimation on the basis of the initial 100 observations of the monthly sample to estimate AR(1) models of the US-German RCOV. Compute a one-step ahead forecast of the US, German and US-German realized variance and covariance from observation 100 to the end of the sample by using the estimates obtained in the previous points. At this point, construct an equally weighted portfolio, compute its returns and the one-step ahead predicted variance and covariances. Plot the corresponding Q-Q plots for German, US, and the equally weighted portfolio standardized returns.

Solution

This solution is a commented version of the MATLAB code `Ex_Modeling_correlations_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. This means that no “Set Path” instructions should be used (only for this workout Matlab). The loading of the *monthly* data is performed by the lines of code:

```
filename=uigetfile('*.txt');
data=dlmread(filename);
```

The above two lines import only the numbers, not the strings, from a `.txt` file.⁵¹ The usual lines of code take care of the strings and are not repeated here. The same applies to the exchange rate transformations that have now become customary in the first part of our Matlab workouts.

- Figure A1 plots the values of each of the three individual indices (in logarithmic terms) and excess returns for the equally weighted portfolio denominated in Euros. Here the message is that you should always take a solid look at your data before venturing in any analysis. Note that the lines of code

```
dy_ger_m = data(:,4)/(100);
dy_ger_m = dy_ger_m/(12); % Monthly dividend yield used in return calculation
lret_ger_m = log((p_ger./lag(p_ger))+dy_ger_m);
dy_us_m = data(:,6)/(100);
dy_us_m = dy_us_m/(12); % Monthly dividend yield used in return calculation
```

⁵¹The reason for loading from a `.txt` file in place of the usual Excel is to favor usage from Mac computers that sometimes have issues with reading directly from Excel, because of copyright issues with shareware spreadsheets.

```

lret_us_dol = log((p_us./lag(p_us))+dy_us_m);
dy_uk_m = data(:,8)/(100);
dy_uk_m = dy_uk_m/(12); % Monthly dividend yield used in return calculation
lret_uk_str = log((p_uk./lag(p_uk))+dy_uk_m);

```

make sure that total index returns include dividends.

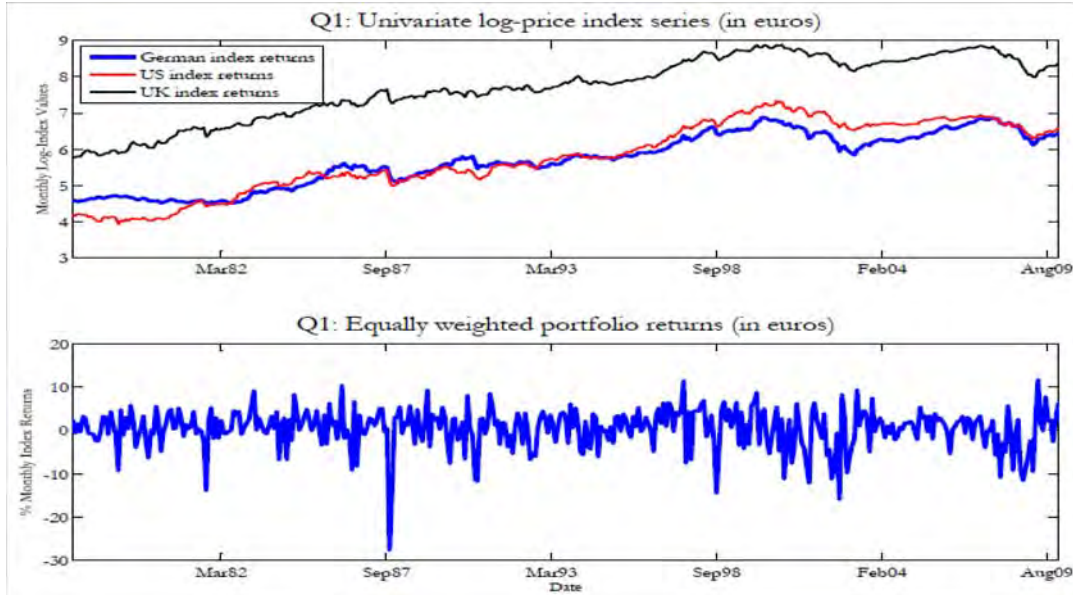


Figure A1: Monthly portfolio indices and exchange rates

2. We now use the *dcc* function in Kevin Sheppard's MFE toolbox

```

R_eq_dm = [R_eq(:,1)-mean(R_eq(:,1)) R_eq(:,2)-mean(R_eq(:,2))
           R_eq(:,3)-mean(R_eq(:,3))];
[parameters, loglikelihood, SIGMA_1, stderrors, jointscores, diagnostics]=
           dcc(R_eq_dm,[],1,0,1);

```

to estimate over the monthly sample 1977-2009 a GARCH(1,1)-DCC(1,1) model with constant mean to and sample correlation (matrix targeting):

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

where $r_{i,t+1}$ denotes *excess* returns ($i = \text{US, UK, Ger}$, i.e., in excess over 3-month euro riskless rates). Parameter estimates, with QMLE standard errors are printed on the Matlab screen as shown in Figure A2. Interestingly, also on these monthly return data we find rather precisely estimated and rather persistent GARCH dynamics. However, the implied persistence indices ($\hat{\alpha}_i + \hat{\beta}_i$ using our standard notation) tend to be (at 0.96-0.98) slightly (but only slightly) smaller than those typical of monthly data.⁵² More interestingly, the estimated coefficients for the DCC(1,1) model are 0.028

⁵²The only oddity is that the estimates of constant coefficients (ω_i , $i = \text{US, UK, Ger}$) in the GARCH(1,1) models are not statistically significant.

and 0.969, respectively, and both are estimated at a relatively high level of precision. The reason for why only two coefficients are estimated in the DCC(1,1) model is that a sample correlation targeting model—as examined in the lectures—has been specified and estimated:

$$q_{ij,t+1} = \left[\frac{1}{T} \sum_{t=1}^T z_{i,t} z_{j,t} \right] (1 - \alpha - \beta) + \alpha z_{i,t} z_{j,t} + \beta q_{ij,t}.$$

GARCH (1,1) -DCC (1)	PARAMETERS			
omega US returns	1.4563	with robust S.E.	0.8654	t-stat 1.6828
alpha US returns	0.0548	with robust S.E.	0.0007	t-stat 78.9502
beta US returns	0.8971	with robust S.E.	0.0011	t-stat 789.4175
omega UK returns	0.6630	with robust S.E.	0.8665	t-stat 0.7651
alpha UK returns	0.0853	with robust S.E.	0.0039	t-stat 21.6041
beta UK returns	0.9009	with robust S.E.	0.0013	t-stat 690.4052
omega German returns	1.0814	with robust S.E.	1.4481	t-stat 0.7467
alpha German returns	0.1341	with robust S.E.	0.0049	t-stat 27.5341
beta German returns	0.8399	with robust S.E.	0.0104	t-stat 80.9029
alpha DCC equation	0.0277	with robust S.E.	0.0000	t-stat 926.7436
beta DCC equation	0.9686	with robust S.E.	0.0001	t-stat 14532.1494
MaxLikelihood:	-3346.7898			

Figure A2:GARCH/DCC estimates from monthly international excess return indices

The “extraction” of the dynamic, conditional covariance matrix implied by the DCC model is performed automatically by Sheppard’s *dcc* function,

**[parameters, loglikelihood, SIGMA_1, stderrors, jointscores, diagnostics]=
dcc(R_eq_dm,[],1,0,1);**

where the sequence of (one-month ahead) predictions of the covariance matrix $\{\hat{\Sigma}_t\}_{t=1}^T$ corresponds to **SIGMA_1**. When we plot conditional means, volatilities, and correlations (all at a monthly frequency) using standard instructions, we obtain Figure A3.

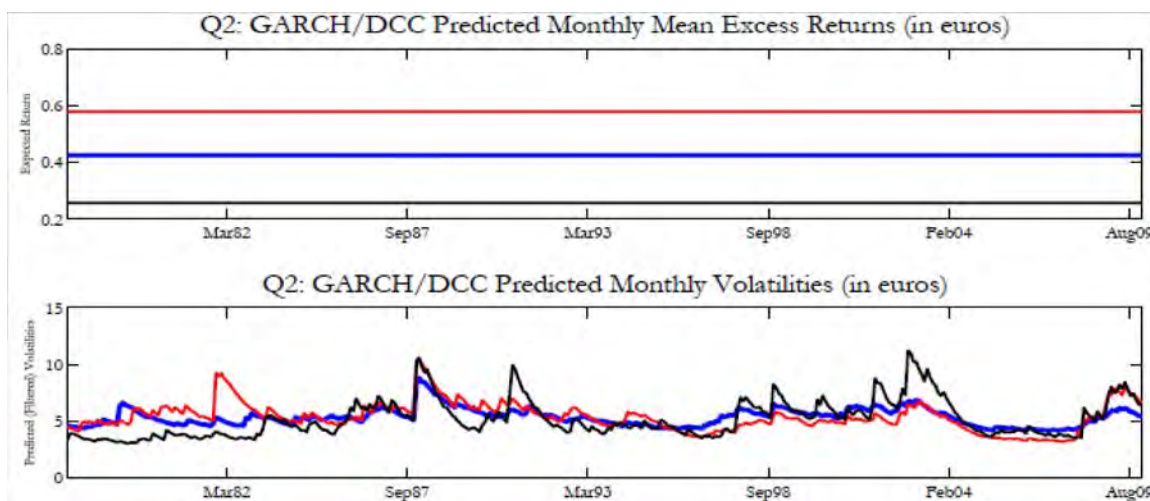


Figure A3-I: Conditional means and volatilities from a GARCH DDC model

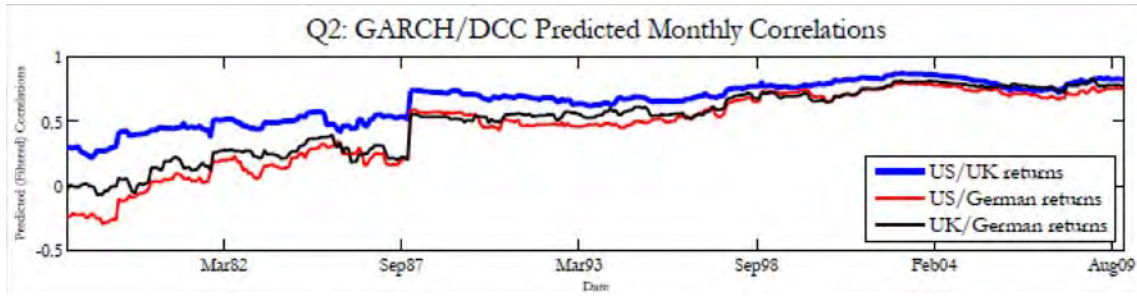


Figure A3-II: Conditional correlations from a GARCH DCC model

The conditional means, as per the very instructions of the workout, are in fact constant over time, which implies that unconditional and conditional means are constant in this case. On the contrary, conditional volatilities are time-varying in ways that are typical of GARCH models, with spikes that may receive a rather natural economic interpretation in the light of well-known events. Conditional correlations as estimated by the model appear to be largely increasing over time, after undergoing a discrete upward jump in the Fall of 1987. Although, as one may expect, the pairwise conditional correlation between U.S. and U.K. (excess) equity returns is larger than the other two pair-wise correlations, after 2005 all these correlations converge to similar levels, generally in excess of 0.7. Such high correlations levels imply that the diversification opportunities across these major, developed stock markets are limited at best.

3. The mean-variance portfolio weights obtained from the classical formula

$$\mathbf{w}_t^{Markow} = [\hat{\Sigma}_{t+1}^{GC/DCC}]^{-1} \hat{\boldsymbol{\mu}},$$

where $\hat{\boldsymbol{\mu}} \equiv [\hat{\mu}_{us} \ \hat{\mu}_{uk} \ \hat{\mu}_{ger}]'$ and implicitly you have imposed a unit coefficient of risk aversion, are computed by the lines of code

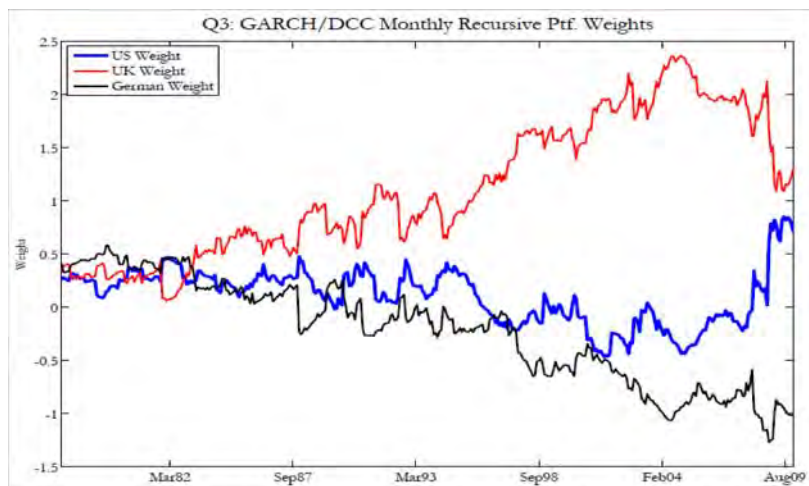


Figure A4: Optimal mean-variance portfolio weights from a GARCH DCC model for conditional covariance

```

for t=1:n
    At_1(:,t) = inv(SIGMA_1(:, :, t))*FIT_1';

```

```

Wt_1(t,1) = At_1(1,t)/sum(At_1(:,t));
Wt_1(t,2) = At_1(2,t)/sum(At_1(:,t));
Wt_1(t,3) = At_1(3,t)/sum(At_1(:,t));
end

```

where the three ratios defining **Wt_1(t,i)** ($i = \text{US, UK, Ger}$) are simply enforcing the fact that weights must be summing to one. Figure A4 plots these recursive portfolio weights.

4. At this point, we replace the constant mean function of questions 2-3 with a regression-based mean function, that makes the conditional mean time-varying:

$$\begin{aligned}
 r_{us,t+1} &= \beta_0^{us} + \beta_1^{us} \ln dy_{us,t} + \beta_2^{us} p_{us,t} + \beta_3^{us} p_{uk,t} + \beta_4^{us} p_{ger,t} + \epsilon_{us,t+1} \\
 r_{uk,t+1} &= \beta_0^{uk} + \beta_1^{uk} \ln dy_{us,t} + \beta_2^{uk} p_{us,t} + \beta_3^{uk} p_{uk,t} + \beta_4^{uk} p_{ger,t} + \epsilon_{uk,t+1} \\
 r_{ger,t+1} &= \beta_0^{ger} + \beta_1^{ger} \ln dy_{us,t} + \beta_2^{ger} p_{us,t} + \beta_3^{ger} p_{uk,t} + \beta_4^{ger} p_{ger,t} + \epsilon_{ger,t+1}
 \end{aligned}$$

$$\begin{bmatrix} \epsilon_{us,t+1} \\ \epsilon_{uk,t+1} \\ \epsilon_{ger,t+1} \end{bmatrix} \sim N(\mathbf{0}, \text{diag} \{ \sigma_{us}^2, \sigma_{uk}^2, \sigma_{ger}^2 \}).$$

where $p_{us,t}$, $p_{uk,t}$, and $p_{ger,t}$ are log-index levels expressed in euros. This is called a “contagion regression model” because it implies that lagged stock index levels in country j may forecast subsequent excess returns in other markets $i \neq j$. Moreover, note that in such a model, the lagged dividend yield in each equity market forecasts subsequent excess stock returns. Because the covariance matrix is assumed to be diagonal and constant over time (homoskedastic), in this case (see your Financial Econometrics I notes and readings) simple, classical OLS will be unbiased and efficient. Interestingly, estimation is performed using

```

[beta,sigma,resid,cov_b] = mvregress([ones(n,1) IDP(1:end-1,i)
regress(1:end-1,:),R_eq(2:end,i));

```

Naturally, a diagonal covariance matrix implies that all covariances and hence correlations are constant and identically zero over time. Of course, this is a restriction we have imposed, not an empirical finding suggested by the data, that we have already noticed imply a dynamics that can be usefully captured by a GARCH DCC model in question 2. Interestingly, this model is symmetric to the one in question 2: then, conditional means were constant, but the conditional covariance matrix time-varying; here the opposite occurs and only conditional means are constant. The estimates and their robust standard errors (see MVREGRESS in the Matlab manual for details) are printed on

the Matlab screen in Figure A5.

OLS: US INDEX RETURNS		PARAMETERS	
Constant	8.2368	with robust S.E.	8.6820 t-stat 0.9487
Coeff. on lagged own DY	-2.2145	with robust S.E.	2.0243 t-stat -1.0940
Coeff. on lagged Log-US P	-1.7070	with robust S.E.	2.0050 t-stat -0.8514
Coeff. on lagged Log-UK P	3.7737	with robust S.E.	1.8823 t-stat 2.0048
Coeff. on lagged Log-GER P	-4.3518	with robust S.E.	1.6794 t-stat -2.5913

OLS: UK INDEX RETURNS		PARAMETERS	
Constant	8.4337	with robust S.E.	7.4318 t-stat 1.1348
Coeff. on lagged own DY	1.4662	with robust S.E.	2.1663 t-stat 0.6768
Coeff. on lagged Log-US P	3.1292	with robust S.E.	1.7178 t-stat 1.8216
Coeff. on lagged Log-UK P	-2.5821	with robust S.E.	1.8159 t-stat -1.4219
Coeff. on lagged Log-GER P	-1.4169	with robust S.E.	1.6265 t-stat -0.8711

OLS: German INDEX RETURNS		PARAMETERS	
Constant	6.5938	with robust S.E.	6.0248 t-stat 1.0944
Coeff. on lagged own DY	-0.2172	with robust S.E.	1.1724 t-stat -0.1852
Coeff. on lagged Log-US P	3.0964	with robust S.E.	1.6448 t-stat 1.8826
Coeff. on lagged Log-UK P	0.7813	with robust S.E.	1.8594 t-stat 0.4202
Coeff. on lagged Log-GER P	-5.2690	with robust S.E.	1.5801 t-stat -3.3345

Figure A5: OLS estimates of regression-based conditional mean functions

Interestingly, most coefficients fail to be statistically significant (none of them are in the case of UK excess returns) and when they are, these are marginally significant, with t stats close to the threshold of 2.

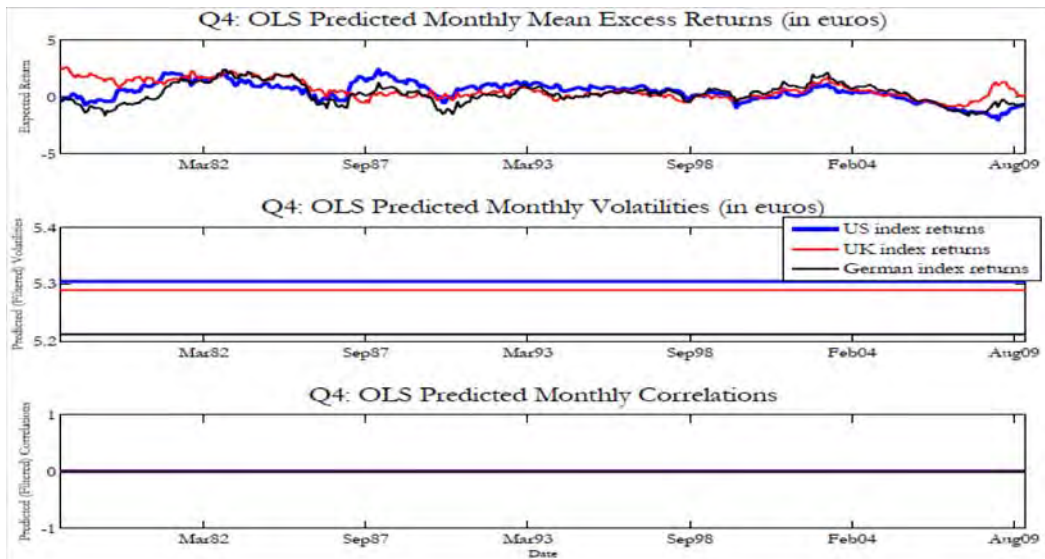


Figure A6: Conditional means, volatilities, and correlations from a homoskedastic contagion model

There is however evidence that past high UK stock prices positively infect US stock returns, while the opposite occurs for past German stock prices. Figure A6 uses the customary three-panel plot to graph the one-month ahead (predicted) expected excess returns, volatilities, and correlations. As it is obvious, all volatilities and correlations are constant (the latter are identically zero, as noted above).

- When we use conditional means, variances, and correlations to compute recursive dynamic Markowitz weights when $\hat{\mu}_t \equiv [E_t(r_{us,t+1}) \ E_t(r_{uk,t+1}) \ E_t(r_{ger,t+1})]'$, we obtain Figure A7,

where we have cut excessively large and small weights at 400% in absolute value.

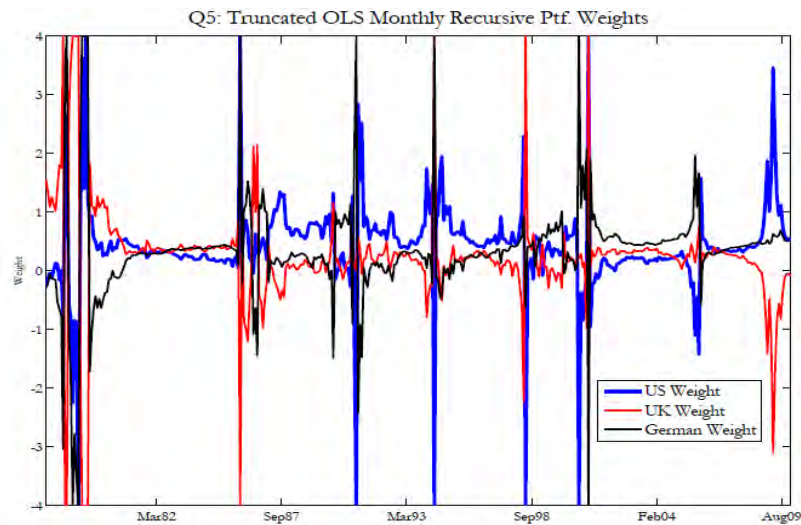


Figure A7: Optimal mean-variance portfolio weights from a homoskedastic, contagion regression model

In fact, the optimal weights occasionally “shoot up” in absolute value to values that are hardly sensible as they exceed 100%. This derives from the fact that in Figure A6—although conditional means are imprecisely estimated—the forecasts are highly unstable, and this carries forward onto optimal MV portfolio shares.

6. We now bring the models of questions 2-3 and 4-5 together, by estimating by iterated MLE (see below for details), the following restricted VAR(1) model that jointly captures the presence of time variation in conditional means, conditional variances, and conditional covariances:⁵³

$$\begin{aligned} r_{us,t+1} &= \beta_0^{us} + \beta_2^{us} r_{us,t} + \epsilon_{us,t+1} \\ r_{uk,t+1} &= \beta_0^{uk} + \beta_3^{uk} r_{uk,t} + \epsilon_{uk,t+1} \\ r_{ger,t+1} &= \beta_0^{ger} + \beta_4^{ger} r_{ger,t} + \epsilon_{ger,t+1} \end{aligned} \quad \begin{bmatrix} \epsilon_{us,t+1} \\ \epsilon_{uk,t+1} \\ \epsilon_{ger,t+1} \end{bmatrix} \sim N\left(\mathbf{0}, \Sigma_{t+1}^{GC/DCC}\right).$$

Notice that by estimating this model you are generalizing the model in question 4 not only because you are allowing variances to vary over time, but also (or especially) because you are no longer constraining the correlations to be always zero and therefore constant over time.⁵⁴ The reference above to “iterated MLE” refers to the fact that Kevin Sheppard’s GARCH/DCC utilities will *not* allow you to jointly estimate the regression coefficients and also the GARCH/DCC dynamic covariance matrix; therefore what you are advised to do is to first estimate the residuals from the regression models, then fit on them your GARCH/DCC model using the code,

for i=1:3

⁵³This VAR(1) is restricted because it boils down to three different AR(1) models stacked on top of each other. However, different national stock markets are still allowed to interact and contagion each other through the dynamic covariance matrix.

⁵⁴Of course, the conditional mean functions are simpler than those in question 4, but you are welcome to propose the same, identical regression to see what happens.

```

        var1=strcat('beta_',country(i,:));
        var2=strcat('SIGMA_',country(i,:));
        var3=strcat('RESID_',country(i,:));
[beta,sigma,resid,cov_b] = mvregress([ones(n,1) R_eq(1:end-1,i)], R_eq(2:end,i));
        assignin('base',var1,beta);
        assignin('base',var2,sigma);
        assignin('base',var3,resid);
        assignin('base',var4,cov_b);
        end
Res = [RESID_US RESID_UK RESID_GER];
[parameters, loglikelihood, SIGMA_3, stderrors, jointscores, diagnostics]=
        dcc(Res,[],1,0,1);

```

and finally proceed to re-estimate the regression model by feasible GLS when the covariance matrix is the one estimated by GARCH/DCC:

```

SIGMA_3_r = [SIGMA_3_11 SIGMA_3_12 SIGMA_3_13;
            SIGMA_3_21 SIGMA_3_22 SIGMA_3_23;
            SIGMA_3_31 SIGMA_3_32 SIGMA_3_33];
X = [ones(n,1) R_eq(1:end-1,1) zeros(n,2) zeros(n,2);
     zeros(n,2) ones(n,1) R_eq(1:end-1,2) zeros(n,2);
     zeros(n,2) zeros(n,2) ones(n,1) R_eq(1:end-1,2)];
Y = [R_eq(2:end,1); R_eq(2:end,2); R_eq(2:end,3)];
BETA = inv(X'*inv(SIGMA_3_r)*X)*X'*inv(SIGMA_3_r)*Y;
OMEGA = inv(X'*inv(SIGMA_3_r)*X);

```

```

=====
----- FIRST STAGE ESTIMATION (GARCH/DCC) -----
=====
GARCH(1,1)-DCC(1)  PARAMETERS
omega US returns   1.2551 with robust S.E. 0.6712 t-stat 1.8699
alpha US returns   0.0565 with robust S.E. 0.0007 t-stat 86.3658
beta US returns    0.9019 with robust S.E. 0.0010 t-stat 869.9187
omega UK returns   0.6142 with robust S.E. 0.6291 t-stat 0.9764
alpha UK returns   0.0788 with robust S.E. 0.0034 t-stat 23.4173
beta UK returns    0.9077 with robust S.E. 0.0012 t-stat 734.5192
omega Germany returns 0.8398 with robust S.E. 0.8844 t-stat 0.9496
alpha Germany returns 0.1195 with robust S.E. 0.0046 t-stat 26.1712
beta Germany returns 0.8615 with robust S.E. 0.0083 t-stat 103.2881
alpha DCC equation  0.0274 with robust S.E. 0.0000 t-stat 845.5459
beta DCC equation   0.9683 with robust S.E. 0.0001 t-stat 17522.0166
MaxLikelihood:     -3339.2544

```

Figure A8: First-stage GARCH/DCC estimates of demeaned excess returns

For a refresher on what FGLS are, how different these are from OLS, etc., please consult Appendix A of this chapter. Figures A8-A9 display the estimated parameters as these are printed on the Matlab screen:

```

=====
===== SECOND STAGE ESTIMATION (MEANS) =====
=====
2SFGLS: US INDEX RETURNS PARAMETERS
Constant          0.3135  with robust S.E. 0.2380  t-stat 1.3172
Coeff. on lagged own return -0.0048  with robust S.E. 0.0423  t-stat -0.1143

2SFGLS: UK INDEX RETURNS PARAMETERS
Constant          0.4299  with robust S.E. 0.2243  t-stat 1.9168
Coeff. on lagged own return 0.0401  with robust S.E. 0.0438  t-stat 0.9147

2SFGLS: German INDEX RETURNS PARAMETERS
Constant          0.3797  with robust S.E. 0.2202  t-stat 1.6929
Coeff. on lagged own return 0.0246  with robust S.E. 0.0423  t-stat 0.5620

```

Figure A9: Second-stage feasible GLS (GARCH/DCC-based) estimates of conditional mean parameters

Also in this case none of the regression coefficients are strongly statistically significant, which should advise caution before basing your mean-variance asset allocation on such parameters. Figure A10 shows the now customary three-panel plot, in which finally conditional means, variances, and correlations are all time-varying.

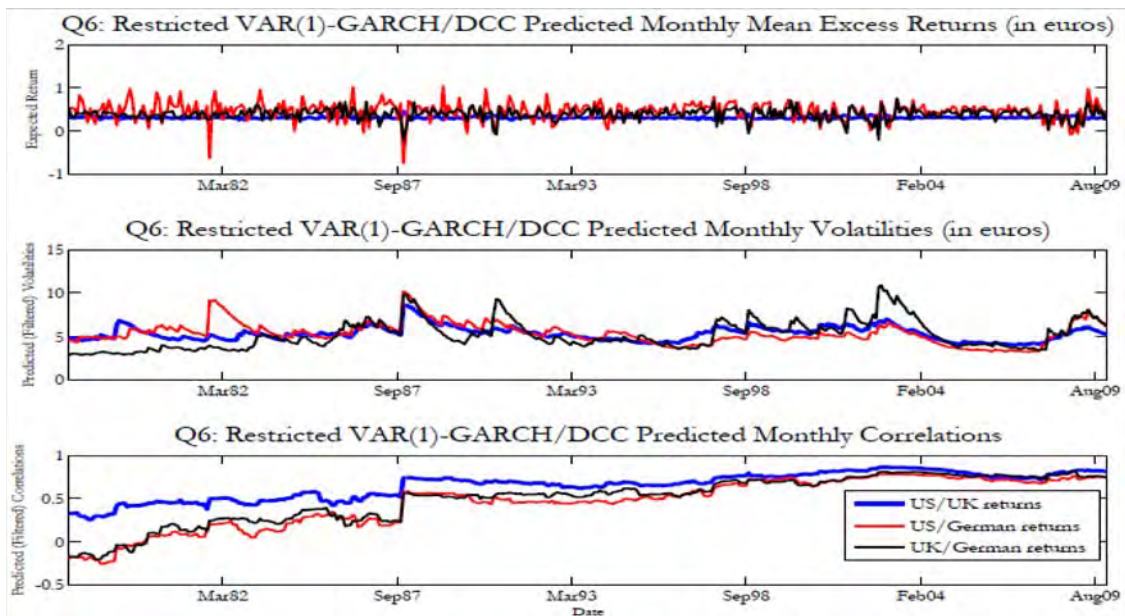


Figure A10: Conditional means, volatilities, and correlations from a heteroskedastic restricted VAR(1) model

- When we use conditional means, variances, and correlations to compute recursive dynamic Markowitz weights when $\hat{\mu}_t$ is derived from the fitted values from the restricted VAR(1) in question 6, as given by the line of command

```

FIT3_vec = [(BETA(1,1)*ones(n,1)+BETA(2,1)*R_eq(1:end-1,1))
            (BETA(3,1)*ones(n,1)+BETA(4,1)*R_eq(1:end-1,2))
            (BETA(5,1)*ones(n,1)+BETA(6,1)*R_eq(1:end-1,3))];

```

we obtain Figure A11, where we have cut excessively large and small weights at 500% in absolute value.

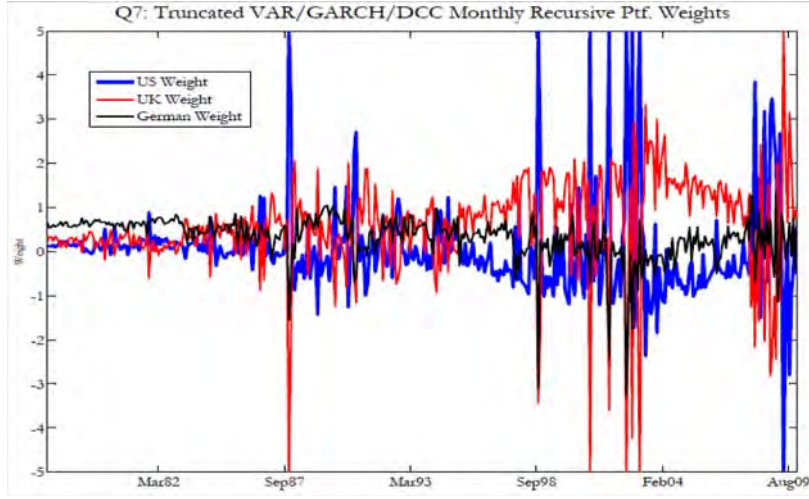


Figure A11: Optimal mean-variance portfolio weights from a heteroskedastic restricted VAR(1) model

8. With reference to the out-of-sample period January 2010 - December 2012 we now proceed to compute optimal Markowitz weights for two models: the unconditional mean GARCH/DCC model of question 2; the VAR/GARCH/DCC of question 6. For both models we have used the same estimated conditional mean (the intercept in the first case, the restricted VAR parameters in the latter) and the GARCH/DCC parameters estimated in questions 2 and 6, but the dynamic covariance matrix has been updated in real time (to produce one-month ahead forecasts) on the basis of the out-of-sample forecast errors obtained in the earlier step. For instance, the GARCH models are used during the out-of-sample period to obtain

$$\hat{\sigma}_{i,\tau+1}^2 = \hat{\omega}_i + \hat{\alpha}_i(\hat{\sigma}_{i,\tau}\hat{z}_{i,\tau})^2 + \hat{\beta}_i\hat{\sigma}_{i,\tau}^2.$$

In the code, this occurs by first computing forecast, then forecast errors, and finally by feeding such errors inside the GARCH DCC model using Kevin Sheppard's *dcc_mvqarch_cov* function:⁵⁵

```

% Computes forecasts
X = [ones(n,1) R_eq(1:end-1,1) zeros(n,2) zeros(n,2);
     zeros(n,2) ones(n,1) R_eq(1:end-1,2) zeros(n,2);
     zeros(n,2) zeros(n,2) ones(n,1) R_eq(1:end-1,2)];
Forc_h = X*BETA;
Forc = [Forc_h(1:n) Forc_h(n+1:2*n) Forc_h(2*n+1:3*n)];
% Computes forecast errors
Pr_err = R_eq(2:end,:) - Forc;

```

⁵⁵The code below refers to the second, restricted VAR(1) model but they are logically equivalent to those applied to the constant mean model.


```
[GAMMA, D]=dcc_mvgarch_cov(par_dcc,Pr_err(:,1),Pr_err(:,2),Pr_err(:,3),[0
1]*corr([Pr_err(:,1) Pr_err(:,2)])*[1;0],[0 1]*corr([Pr_err(:,1) Pr_err(:,3)])*[1;0],[0
1]*corr([Pr_err(:,2) Pr_err(:,3)])*[1;0]);
```

Figure A12 presents the weights obtained in this way.

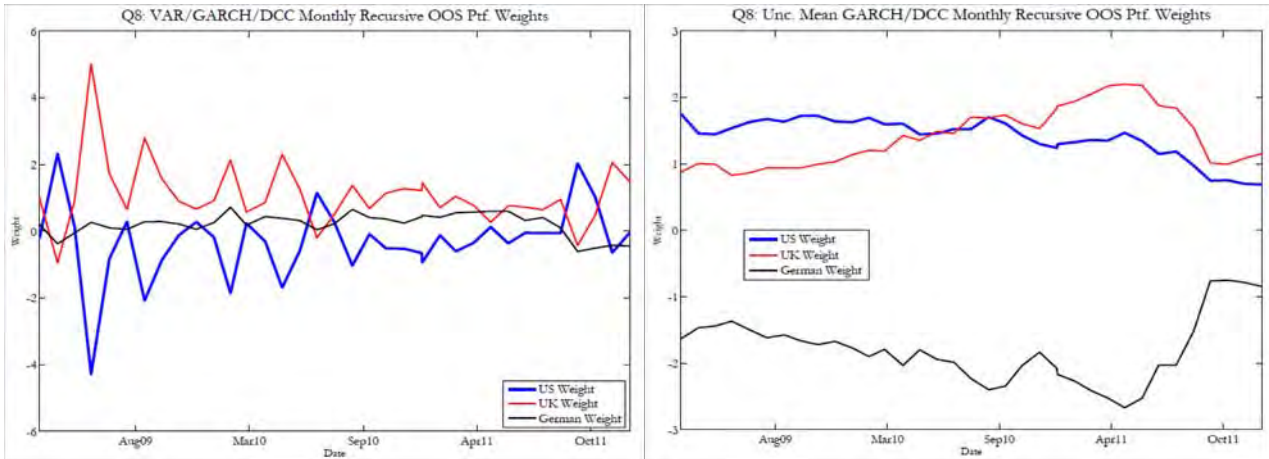


Figure A12: Out-of-sample optimal mean-variance portfolio weights

After obtaining the weights, we compute the realized Sharpe ratios over the out of sample period, which are presented in Figure A13.

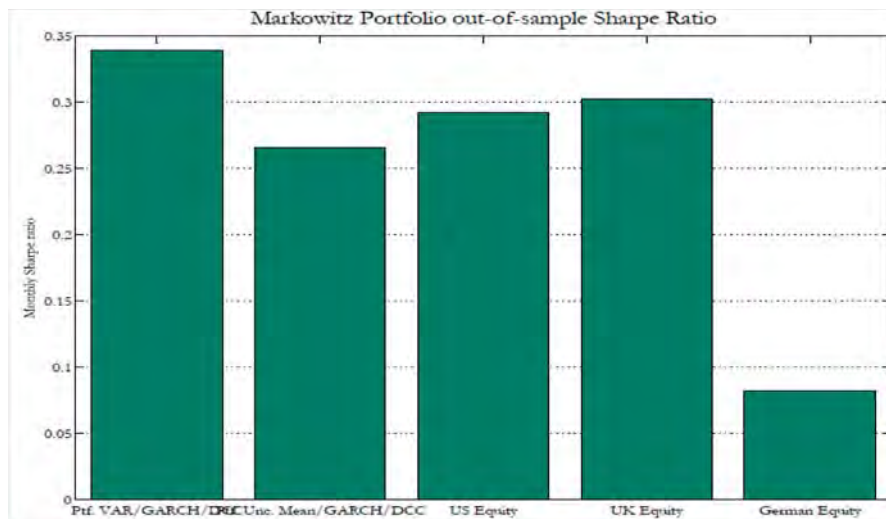


Figure A13: Out-of-sample realized Sharpe ratios from alternative models

Clearly—although these are modest—there can be advantages from adopting a GARCH/DCC model as a base for asset allocation, when the conditional mean model is sufficiently restricted and parsimonious to avoid excessive variation in forecasts of risk premia (expected excess returns).

- At this point, we switch to a different data set and load the daily data set already used on many occasions before. Using daily data, i.e., treating them as high-frequency data (with $m = 22$ per month), we compute monthly non-overlapping cumulative returns and monthly

non-overlapping realized variance (RV) series and realized covariance for US and German stock indices. The following lines of code perform this task:

```

start=1+window;
RV(start,:)=sum(RET(2:2+(window-1),:).^2);
RCOV(start,:)=sum(RET(2:2+(window-1),1).*RET(2:2+(window-1),2));
RET_CUM(start,:)=sum(RET(2:2+(window-1),:));

for t=start+window:(window-1):T
RV(t-(window-1):t,:)=ones(size(t-(window-1):t,2),1)*sum(RET(t-(window-1):t,:).^2);
RCOV(t-(window-1):t,:)=ones(size(t-(window-1):t,2),1)*sum(RET(t-(window-1):t,1).
*RET(t-(window-1):t,2));
RET_CUM(t-(window-1):t,:)=ones(size(t-(window-1):t,2),1)*sum(RET(t-(window-
1):t,:));
end

```

In Figure A14, we plot the resulting monthly realized variances, covariances, and the autocorrelograms of these series.

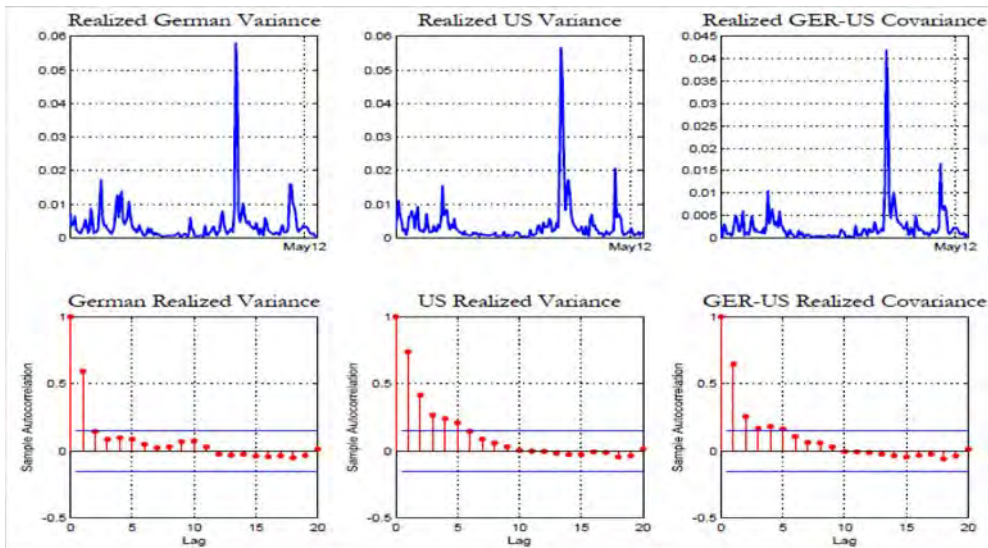


Figure A14: Monthly realized variance and associated autocorrelogram functions

The ACFs emphasize that realized variances and covariances are rather persistent and therefore predictable.

- Finally, we have used the first 100 observations of the monthly realized variance sample that we have obtained in point 9 to estimate two univariate AR(1) models for the US and German RVs and for the US-German RCOV. We use these models to compute one-step ahead forecasts of the US, German and US-German realized variances and covariance from observation 100 to the end of the sample. At this point, we construct an equally weighted portfolio, compute

its returns and the one-step ahead predicted variance and covariances. Figure A15 shows the corresponding Q-Q plots for German, US, and the equally weighted portfolio standardized returns.

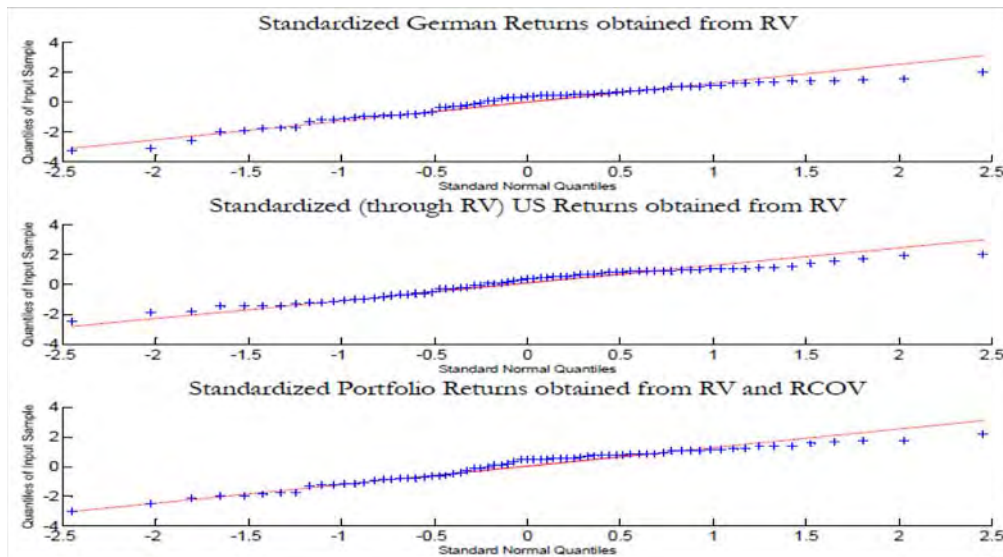


Figure A15: QQ plots for standardized returns obtained from realized variances and covariance

This confirms something mentioned during the lectures: when returns are standardized using realized second moments, the resulting \hat{z}_t^i ($i = US, \text{Germany}$) are approximately normal, even though some problems remain the left tail.

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:

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

where $r_{i,t+1}$ denotes daily returns ($i = \text{EU/Ger, US, UK}$). 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_mvqarch* 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 CTRL+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.

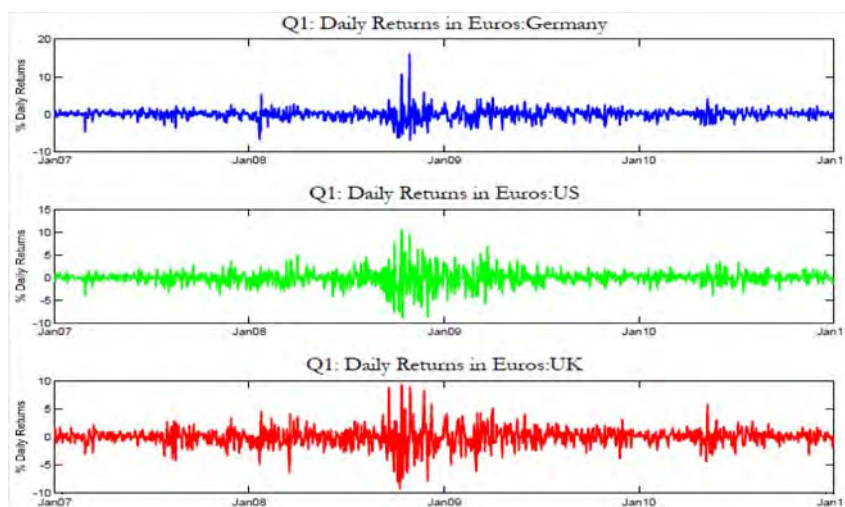


Figure C1: Daily index returns (expressed in euros) for the sample 2007-2010

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;

```

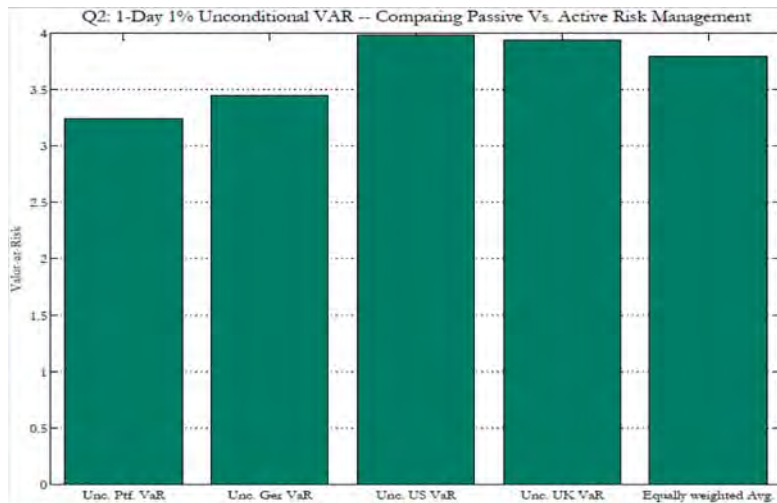


Figure C2: Comparing 1% VaR measures based on unconditional correlation estimates

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 $\mathbf{w}'\Sigma\mathbf{w}$, 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 $\mathbf{w}'\Sigma\mathbf{w}$ 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

⁵⁶This 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.

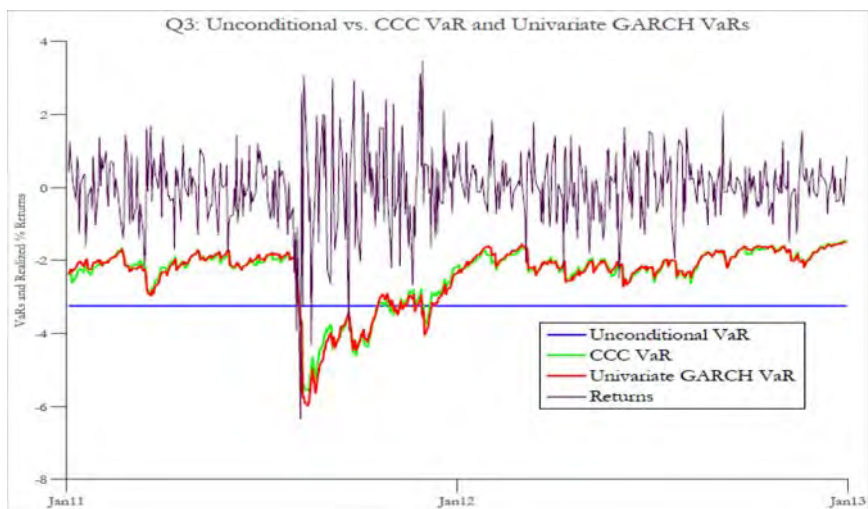


Figure C3: Recursive daily 1% VaR estimates from different alternative models

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 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

⁵⁷A = 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.

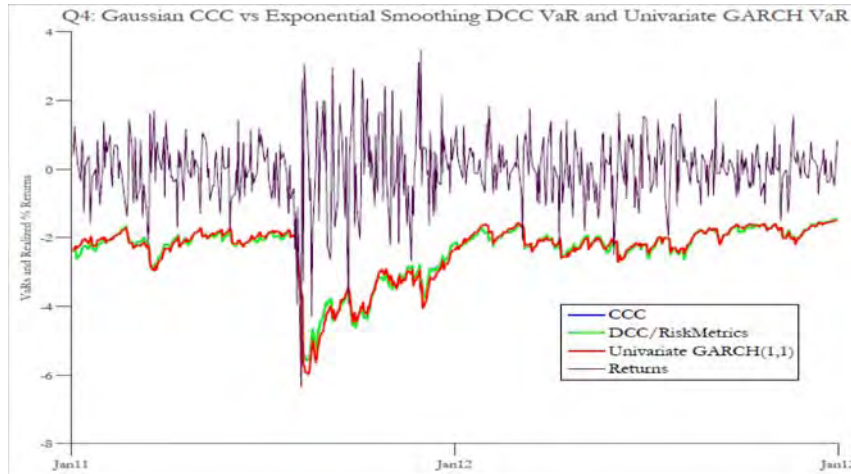


Figure C4: Recursive daily 1% VaR estimates from different alternative models including RiskMetrics-DCC

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.

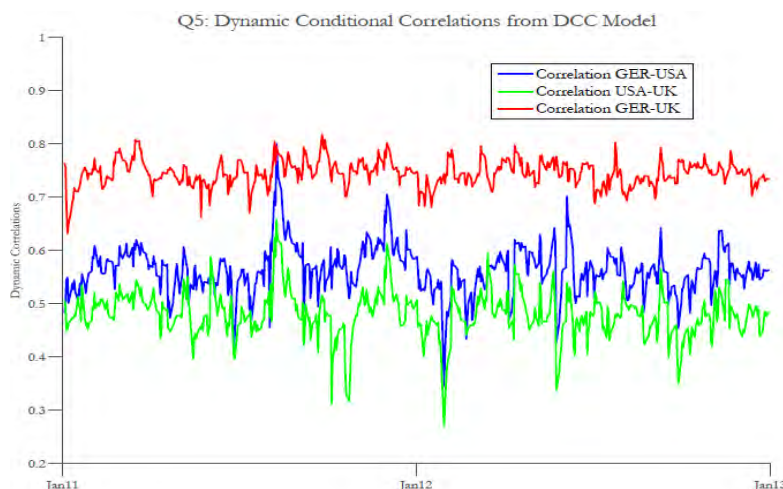


Figure C6: Predicted dynamic conditional correlations from DCC model

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]=
dcm_mvqarch_for(par_dcc,sret_GER,sret_US,sret_UK,Corr_matrix(1,2), ...
...Corr_matrix(1,3),Corr_matrix(2,3));
```

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.

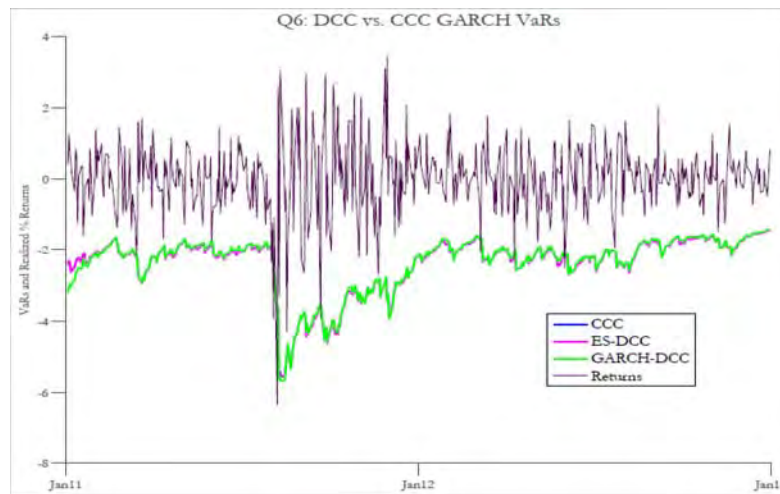


Figure C7: Recursive daily 1% VaR estimates from alternative CCC and DCC models

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_mvlgarch(z_pc,3,1,1);
```

This uses a toolbox function, `o_mvlgarch` 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.

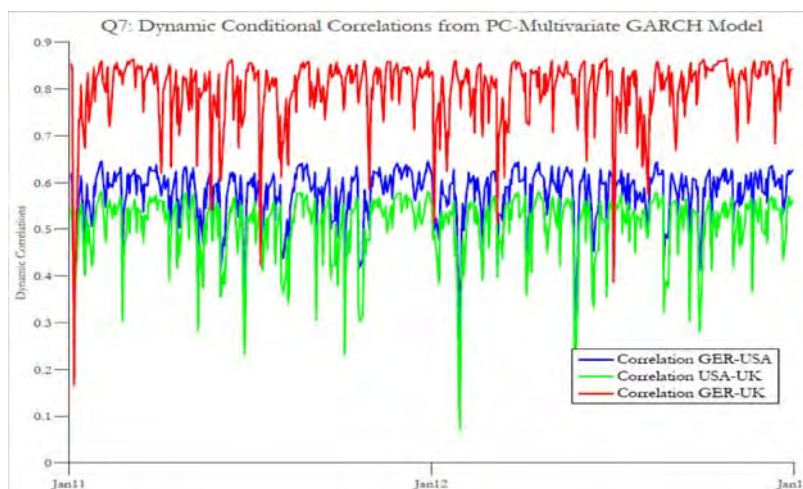


Figure C8: Predicted dynamic conditional correlations from PC/Orthogonal GARCH model

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.

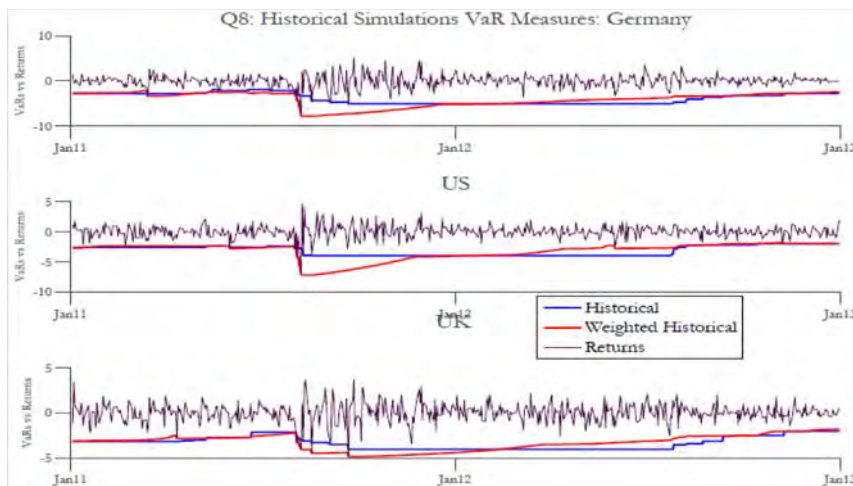


Figure C10: Recursive daily 1% VaR estimates from historical and weighted historical simulations

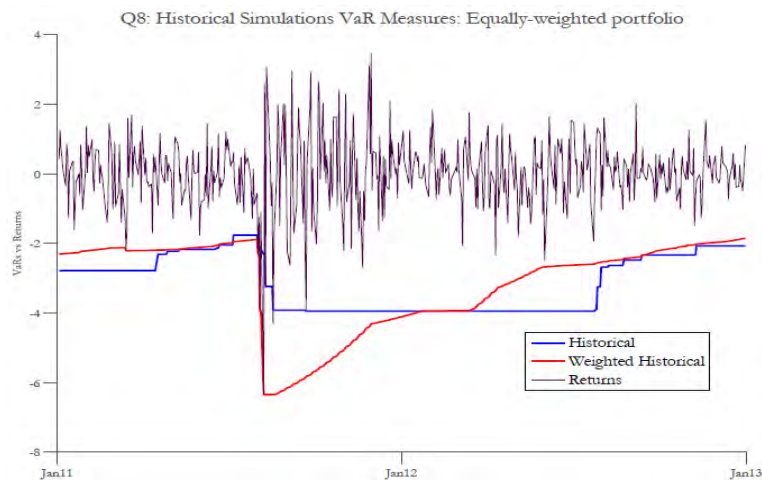


Figure C11: Recursive daily 1% VaR estimates from historical and weighted historical simulations

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_mvgaarch([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_mvgaarch* 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 Finance* 47, 427-465.
- [8] Gouriéroux, 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.

Errata Corrige

(21/05/2013, p. 25) The formula for (matrix) covariance targeting in the DCC is:

$$\mathbf{Q}_{t+1} = (1 - \alpha - \beta)E[\mathbf{z}_t\mathbf{z}'_t] + \alpha\mathbf{z}_t\mathbf{z}'_t + \beta\mathbf{Q}_t.$$

(29/05/2013, p. 12) A term $\sum_{i=1}^N w_i^2 \hat{\sigma}_{i,t+1}^2 [\Phi^{-1}(p)]^2$ had been omitted twice in the formulas appearing in this page. No qualitative effects for the conclusions in this page.