# Chapter 9

# Multivariate Volatility, Dependence and Copulas

Multivariate modeling is in many ways similar to modeling the volatility of a single asset. The primary challenges which arise in the multivariate problem are ensuring that the forecast covariance is positive definite and limiting the number of parameters which need to be estimated as the number of assets becomes large. This chapter covers standard "simple" multivariate models, multivariate ARCH models and realized covariance. Attention then turns to measures of dependence which go beyond simple linear correlation, and the chapter concludes with an introduction to a general framework for modeling multivariate returns which use copulas.

## 9.1 Introduction

Multivariate volatility or covariance modeling is a crucial ingredient in modern portfolio management. If is useful for a number of important tasks including:

- Portfolio Construction - Classic Markowitz (1959) portfolio construction requires an estimate of the covariance of returns, along with the expected returns of the assets, to determine the optimal portfolio weights. The Markowitz problem find the portfolio with the minimum variance subject to achieving a required mean. Alternatively, the Markowitz problem can be formulated as maximizing the expected mean of the portfolio given a constraint on the volatility of the portfolio.

- Portfolio Sensitivity Analysis - Many portfolios are constructed using other objectives than the "pure" Markowitz problem. For example, fund managers may be selecting firms based on beliefs about fundamental imbalances between the firm and its competitors. Return covariance is useful in these portfolios for studying the portfolio sensitivity to

adding additional or liquidating existing positions, especially when multiple investment opportunities exist which have similar risk-return characteristics.

- Value-at-Risk - Portfolio Value-at-Risk often begins with the covariance of the assets held in the portfolio. Naive $VaR$ uses a constant value for the lower $\alpha$-quantile multiplied by the standard deviation of the portfolio. More sophisticated risk measures examine the joint tail behavior – that is, the probability that two or more assets have large, usually negative returns in the same period. Copulas are a natural method for studying the extreme behavior of asset returns.

- Credit Pricing - Many credit products are written on a basket of bonds, and so the correlation between the payouts of the underlying bonds is crucial for determining the value of the portfolio.

- Correlation Trading - Recent financial innovation has produced contracts where correlation can be directly traded. The traded correlation is formally equicorrelation (See 9.3.5) and measuring and predicting correlation is a directly profitable strategy, at least if it can be done well.

This chapter begins with an overview of simple, static estimators for covariance which widely used. The chapter then turns to dynamic models for conditional covariance based on the ARCH framework. The third topic is realized covariance which exploits ultra-high frequency data in the same manner as realized variance. The chapter concludes with an examination of non-linear dependence and copulas, a recent introduction to financial econometrics which allows for the flexible construction of multivariate models.

## 9.2 Preliminaries

Most volatility models are built using either returns, which is appropriate if the time horizon is small and/or the conditional mean is small relative to the conditional volatility, or demeaned returns when using longer time-spans or if working with series with a non-trivial mean (e.g. electricity prices). The $k$ by 1 vector of returns is denoted $\mathbf{r}_t$, and the corresponding demeaned returns are $\boldsymbol{\epsilon}_t = \mathbf{r}_t - \boldsymbol{\mu}_t$ where $\boldsymbol{\mu}_t \equiv \mathrm{E}_{t-1}[\mathbf{r}_t]$.

The conditional covariance, $\mathrm{E}_{t-1}\left[\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t'\right] \equiv \boldsymbol{\Sigma}_t$, is assumed to be a $k$ by $k$ positive definite matrix. Some models will may use of "devolatilized" residuals defined as $u_{i,t} = \epsilon_{i,t}/\sigma_{i,t}$, $i = 1, 2, \ldots, k$, or in matrix notation $\mathbf{u}_t = \boldsymbol{\epsilon}_t \oslash \boldsymbol{\sigma}_t$ where $\oslash$ denoted Hadamard division (element-by-element). Multivariate standardized residuals, which are both "devolatilized" and "decorrelated", are defined $\mathbf{e}_t = \boldsymbol{\Sigma}_t^{-\frac{1}{2}}\boldsymbol{\epsilon}_t$ so that $\mathrm{E}_{t-1}\left[\mathbf{e}_t\mathbf{e}_t'\right] = \mathbf{I}_k$. Some models explicitly parameterize the conditional correlation, $\mathrm{E}_{t-1}\left[\mathbf{u}_t\mathbf{u}_t'\right] \equiv \mathbf{R}_t = \boldsymbol{\Sigma}_t \oslash \left(\boldsymbol{\sigma}_t\boldsymbol{\sigma}_t'\right)$, or equivalently $\mathbf{R}_t = \mathbf{D}_t^{-1}\boldsymbol{\Sigma}_t\mathbf{D}_t^{-1}$ where

$$\mathbf{D}_t = \begin{bmatrix} \sigma_{1,t} & 0 & \dots & 0 \\ 0 & \sigma_{2,t} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \sigma_{k,t} \end{bmatrix}$$

and so $\mathbf{\Sigma}_t = \mathbf{D}_t \mathbf{R}_t \mathbf{D}_t$.

Some models utilize a factor structure to reduce the dimension of the estimation problem. The $p$ by 1 vector of factors is denoted $\mathbf{f}_t$ and the factor returns are assumed to be mean 0 (or demeaned if the assumption of conditional mean 0 is inappropriate). The conditional covariance of the factors is denoted $\mathbf{\Sigma}_t^f \equiv \mathrm{E}_{t-1}\left[\mathbf{f}_t \mathbf{f}_t'\right]$.

This chapter focuses exclusively models capable of predicting the time-$t$ covariance using information in $\mathcal{F}_{t-1}$. Multi-step forecasting is possible from most models in this chapter by direct recursion. Alternatively, direct forecasting techniques can be used to mix higher frequency data (e.g. daily) with longer forecast horizons (e.g. 2-week or monthly).

### 9.2.1 Synchronization

Synchronization is an important concern when working with a cross-section of asset returns, and non-synchronous returns can arise for a number of reasons:

- Closing hour differences – Closing hour differences are more important when working with assets that trade in different markets. The NYSE closes at either 20:00 or 21:00 GMT, depending on whether the U.S. east coast is using EDT or EST. The London Stock Exchange closes at 15:30 or 16:30 GMT, depending whether the U.K. is using GMT or BST. This difference means that changes in U.S. equity prices that occur after the LSE has closed will not appear in U.K. equities until the next trading day.

  Even within the same geographic region markets have different trading hours. Common U.S. equities trade from 9:30 until 16:00 EDT/EST time. U.S. government bond futures are traded using open outcry from 7:20 a.m. to 14:00. Light Sweet Crude futures trade 9:00 - 14:30 in an open outcry session. All of these assets also trade electronically; Light sweet crude trades electronically 18:00 - 17:15 (all but 45 minutes per day) from Sunday to Friday.

- Market closures due to public holidays – Different markets are closed for different holidays which is a particular problem when working with international asset returns.

- Delays in opening/closing – Assets that trade on the same exchange may be subject to opening or closing delays. For example, the gap between the first and last stock to open in the S&P 500 is typically 15 minutes. The closing spread is similar. These small differences lead to problems measuring the covariance when using intra-daily returns.

- Illiquidity/Stale Prices - Some assets trade more than others. The most liquid stock in the S&P 500 has a daily volume that is typically at least 100 times larger than the least liquid.

Illiquidity is problematic when measuring covariance using intra-daily data.[1]

There are three solutions to the problem of non-synchronous data. The first is to use relatively long time-spans. When using daily data, NYSE and LSE data are typically only simultaneously open for 2 hours out of 6.5 (30%). If using multi-day returns, the lack of common opening hours is less problematic since developments in U.S. equities on one day will show up in prices changes in London on the next day. For example, when using 2-day returns, it is as if 8.5 out of the 13 trading hours are synchronous (65%). When using 5 day returns it is as if 28 out of 32.5 hours are synchronized (86%). The downside of using aggregate returns is the loss of data which results in inefficient estimators as well as difficultly in capturing recent changes.

The second solution is to use synchronized prices (also known as pseudo-closing prices). Synchronized prices are collected when all markets are simultaneously open. For example, if using prices of NYSE and LSE listed firms, a natural sampling time would be 1 hour before the LSE closes, which typically corresponds to 10:30 Eastern time. Daily returns constructed from these prices should capture all of the covariance between these assets. The downside of using synchronized prices is that many markets have no common hours, which is an especially acute problem when measuring the covariance of a global portfolio.

The third solution is to synchronize otherwise non-synchronous returns using a vector moving average (Burns, R. F. Engle, and Mezrich, 1998). Suppose returns were ordered in such a way that the first to close was in position 1, the second to close as in position 2, and so on until the last to close was in position $k$. With this ordering, returns on day $t+1$ for asset $i$ may be correlated with the return on day $t$ for asset $j$ whenever $j > i$, and that the return on day $t+1$ should not be correlated with the day $t$ return on asset $j$ when $j \leq i$.

For example, consider returns from assets that trade on the Australian Stock Exchange (UTC 0:00 - 6:10), the London Stock Exchange (UTC 8:00 - 16:30), NYSE (UTC 14:30 - 21:30) and Tokyo Stock Exchange (UTC 18:00 - 0:00 (+1 day)). The ASX closes before any of the others open, and so the contemporaneous correlation with the LSE, NYSE and TSE should pick up all of the correlation between Australian equities and the rest of the world on day $t$. The LSE opens second and so innovations in the LSE on day $t$ may be correlated with changes on the ASX on $t+1$. Similarly innovations in New York after UTC 16:30 will show up in day $t+1$ in the ASX and LSE. Finally news which comes out when the TSE is open will show up in the day $t+1$ return in the 3 other markets. This leads to a triangular structure in a vector moving average,

$$\begin{bmatrix} r_t^{ASX} \\ r_t^{LSE} \\ r_t^{NYSE} \\ r_t^{TSE} \end{bmatrix} = \begin{bmatrix} 0 & \theta_{12} & \theta_{13} & \theta_{14} \\ 0 & 0 & \theta_{23} & \theta_{24} \\ 0 & 0 & 0 & \theta_{34} \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \epsilon_{t-1}^{ASX} \\ \epsilon_{t-1}^{LSE} \\ \epsilon_{t-1}^{NYSE} \\ \epsilon_{t-1}^{TSE} \end{bmatrix} + \begin{bmatrix} \epsilon_t^{ASX} \\ \epsilon_t^{LSE} \\ \epsilon_t^{NYSE} \\ \epsilon_t^{TSE} \end{bmatrix} \tag{9.1}$$

The recursive structure of this system makes estimation simple since $r_t^{TSE} = \epsilon_t^{TSE}$, and so the

[1] On February 26, 2010, the most liquid S&P 500 company was Bank of America (BAC) which had a volume of 96,352,600. The least liquid S&P 500 company was the Washington Post (WPO) which had a volume of 21,000. IMS Healthcare (RX) did not trade since it was acquired by another company.

model for $r_t^{NYSE}$ is a MA(1)-X. Given estimates of $\epsilon_t^{NYSE}$, the model for $r_t^{LSE}$ is also a MA(1)-X.

In vector form this adjustment model is

$$\mathbf{r}_t = \mathbf{\Theta}\boldsymbol{\epsilon}_{t-1} + \boldsymbol{\epsilon}_t$$

where $\mathbf{r}_t$ is the $k$ by 1 vector of nonsynchronous returns. Synchronized returns, $\hat{\mathbf{r}}_t$ are constructed using the VMA parameters as

$$\hat{\mathbf{r}}_t = (\mathbf{I}_k + \mathbf{\Theta})\,\boldsymbol{\epsilon}_t.$$

The role of $\mathbf{\Theta}$ is to capture any components in asset returns $j$ which occur in the returns to asset return $i$ when the market where $i$ trades in closes later than the market where $j$ trades. In essence this procedure "brings forward" the fraction of the return which has not yet occurred by the close of market where asset $j$ trades. Finally, the conditional covariance of $\boldsymbol{\epsilon}_t$ is $\mathbf{\Sigma}_t$, and so the covariance of the synchronized returns is $\mathrm{E}_{t-1}[\hat{\mathbf{r}}_t\hat{\mathbf{r}}_t] = (\mathbf{I}_k + \mathbf{\Theta})\,\mathbf{\Sigma}_t\,(\mathbf{I}_k + \mathbf{\Theta})'$. Implementing this adjustment requires fitting the conditional covariance to the residual from the VMA, $\boldsymbol{\epsilon}_t$, rather than to returns directly.

## 9.3 Simple Models of Multivariate Volatility

Many simple models which do not require complicated parameter estimation are widely used as benchmarks.

### 9.3.1 Moving Average Covariance

The $n$-period moving average is the simplest covariance estimator.

**Definition 9.1** ($n$-period Moving Average Covariance)**.** The $n$-period moving average covariance is defined

$$\mathbf{\Sigma}_t = n^{-1}\sum_{i=1}^{n}\boldsymbol{\epsilon}_{t-i}\boldsymbol{\epsilon}_{t-i}' \tag{9.2}$$

When returns are measured daily, common choices for $n$ include 22 (monthly), 66 (quarterly), and 252 (annual). When returns are measured monthly, common choices for $n$ are 12 (annual) and 60. Moving average covariance are often imprecise measures since they simultaneously give too little weight to recent observations while giving too much to observations in the distant past.

### 9.3.2 Exponentially Weighted Moving Average Covariance

Exponentially weighted moving averages (EWMA) provide an alternative to moving average covariance estimators which allow for more weight on recent information. EWMAs have been popularized in the volatility literature by RiskMetrics, which was introduced in the univariate context in chapter 8.

**Definition 9.2** (Exponentially Weighted Moving Average Covariance)**.** The EWMA covariance is defined recursively as

$$\mathbf{\Sigma}_t = (1-\lambda)\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1}' + \lambda\mathbf{\Sigma}_{t-1} \tag{9.3}$$

for $\lambda \in (0,1)$. EWMA covariance is equivalently defined through the infinite moving average

$$\mathbf{\Sigma}_t = (1-\lambda)\sum_{i=1}^{\infty}\lambda^{i-1}\boldsymbol{\epsilon}_{t-i}\boldsymbol{\epsilon}_{t-i}'. \tag{9.4}$$

Implementation of an EWMA covariance estimator requires an initial value for $\mathbf{\Sigma}_1$, which can be set to the average covariance over the first $m$ days for some $m > k$ or could be set to the full-sample covariance. The single parameter, $\lambda$, is usually set to .94 for daily data and .97 for monthly data based on recommendations from RiskMetrics (J.P.Morgan/Reuters, 1996).

**Definition 9.3** (RiskMetrics 1994 Covariance)**.** The RiskMetrics 1994 Covariance is computed as an EWMA with $\lambda = .94$ for daily data or $\lambda = .97$ for monthly.

The RiskMetrics EWMA estimator is formally known as RM1994, and has been surpassed by RM2006 which uses a *long memory* model for volatility. Long memory requires that the weights on past returns decay hyperbolically ($w \propto i^{-\alpha}$, $\alpha > 0$) rather than exponentially ($w \propto \lambda^i$). The new methodology extends the 1994 methodology by computing the volatility as a weighted sum of EWMAs (eq. 9.5, line 1) rather than a single EWMA (eq. 9.3).

The RM2006 covariance estimator is computed as the average of $m$ EWMA covariances.

**Definition 9.4** (RiskMetrics 2006 Covariance)**.** The RiskMetrics 2006 Covariance is computed as

$$\begin{aligned}
\mathbf{\Sigma}_t &= \sum_{i=1}^{m} w_i \mathbf{\Sigma}_{i,t} \\
\mathbf{\Sigma}_{i,t} &= (1-\lambda_i)\,\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1}' + \lambda_i \mathbf{\Sigma}_{i,t-1} \\
w_i &= \frac{1}{C}\left(1 - \frac{\ln(\tau_i)}{\ln(\tau_0)}\right) \\
\lambda_i &= \exp\left(-\frac{1}{\tau_i}\right) \\
\tau_i &= \tau_1 \rho^{i-1}, \quad i = 1, 2, \ldots, m
\end{aligned} \tag{9.5}$$

where $C$ is a normalization constant which ensures that $\sum_{i=1}^{m} w_i = 1$.

The 2006 methodology uses a 3-parameter model which includes a logarithmic decay factor, $\tau_0$ (1560), a lower cut-off, $\tau_1$ (4), and an upper cutoff $\tau_{\max}$ (512) [suggested values in parentheses], using eq. (9.5) (Zumbach, 2007). One additional parameter, $\rho$, is required to operationalize the

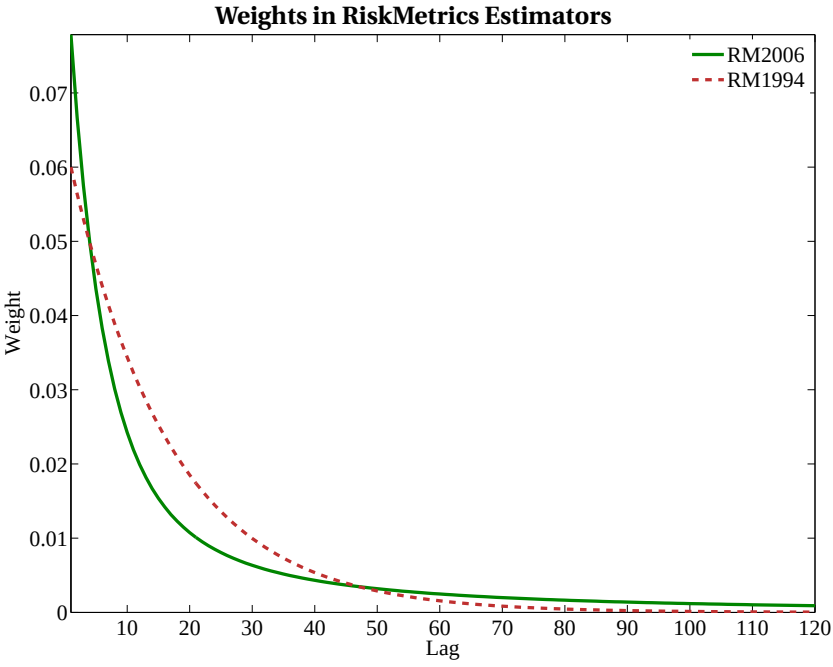


Figure 9.1: These two lines show the weights on lagged outer-product of returns $\left(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t'\right)$ in the 1994 and 2006 versions of the RiskMetrics methodology. The 2006 version features more weight on recent volatility and more weight on volatility in the distant past relative to the 1994 methodology.

---

model, and RiskMetrics suggests $\sqrt{2}$.[2]

Both RiskMetrics covariance estimators can be expressed as $\boldsymbol{\Sigma}_t = \sum_{i=1}^{\infty} \gamma_i \boldsymbol{\epsilon}_{t-i} \boldsymbol{\epsilon}_{t-1}'$ for a set of weights $\{\gamma_i\}$. Figure 9.1 contains a plot of the weights for the 120 most recent observations from both the RM1994 and RM2006 estimators. The new methodology has both higher weight on recent data, and higher weight on data in the distant past. One method to compare the two models is considering how many period it takes for 99% of the weight to have been accumulated, or $\min_n \sum_{i=0}^{n} \gamma_i \geq .99$. For the RM1994 methodology, this happens in 75 days – the RM2006 methodology requires 619 days to achieve the same target. The first 75 weights in the RM2006 estimator contain 83% of the weight, and so 1/6 of the total weight depends on returns more than 6 months in the past.

[2] $\tau_{\max}$ does not directly appear in the equations for the RM2006 framework, but is implicitly included since

$$m = 1 + \frac{\ln\left(\frac{\tau_{\max}}{\tau_1}\right)}{\ln \rho}.$$

### 9.3.3 Observable Factor Covariance

The $n$-period factor model assumes that returns are generated by a strict factor structure and is closely related to the CAP-M (W. Sharpe, 1964; Lintner, 1965; Black, 1972), the intertemporal CAP-M (Merton, 1973) and Arbitrage Pricing Theory (Roll, 1977). Moving average factor covariance estimators can be viewed as restricted versions on the standard moving average covariance estimator where all covariance is attributed to common exposure to a set of factors. The model postulates that the return on the i$^{\text{th}}$asset is generated by a set of $p$ observable factors with returns $\mathbf{f}_t$, an $p$ by 1 set of asset-specific factor loadings, $\boldsymbol{\beta}_i$ and an idiosyncratic shock $\eta_{i,t}$,

$$\epsilon_{i,t} = \mathbf{f}_t'\boldsymbol{\beta}_{i,t} + \eta_{i,t}.$$

The $k$ by 1 vector of returns can be compactly described as

$$\boldsymbol{\epsilon}_t = \boldsymbol{\beta}\mathbf{f}_t + \boldsymbol{\eta}_t$$

where $\boldsymbol{\beta}$ is a $k$ by $p$ matrix of factor loadings and $\boldsymbol{\eta}_t$ is a $k$ by 1 vector of idiosyncratic shocks. The shocks are assumed to be white noise, cross-sectionally uncorrelated ($\mathrm{E}_{t-1}\left[\eta_{i,t}\eta_{j,t}\right] = 0$) and uncorrelated with the factors.

**Definition 9.5** ($n$-period Factor Covariance)**.** The $n$-period factor covariance is defined as

$$\boldsymbol{\Sigma}_t = \boldsymbol{\beta}\boldsymbol{\Sigma}_t^f\boldsymbol{\beta}' + \boldsymbol{\Omega}_t \tag{9.6}$$

where $\boldsymbol{\Sigma}_t^f = n^{-1}\sum_{i=1}^{n}\mathbf{f}_{t-i}\mathbf{f}_{t-i}'$ is the $n$-period moving covariance of the factors,

$$\boldsymbol{\beta}_t = \left(\sum_{i=1}^{n}\mathbf{f}_{t-i}\mathbf{f}_{t-i}'\right)^{-1}\sum_{i=1}^{n}\mathbf{f}_{t-i}\boldsymbol{\epsilon}_{t-i}'$$

is the $p$ by $k$ matrix of factor loadings and $\boldsymbol{\Omega}_t$ is a diagonal matrix with $\omega_{j,t}^2 = n^{-1}\sum_{i=1}^{n}\eta_{j,t-i}^2$ in the j$^{\text{th}}$diagonal position where $\eta_{i,t} = \epsilon_{i,t} - \mathbf{f}_t'\boldsymbol{\beta}_i$ are the regression residuals.

While the moving average factor covariance is a restricted version of the standard moving average covariance estimator, it does have one important advantage: factor covariance estimators are always positive definite as long as the number of periods used to estimate the factor covariance and factor loadings is larger than the number of factors ($n > p$). The standard moving average covariance requires the number of period used in the estimator to be larger than the number of assets to be positive definite ($n > k$). This difference makes the factor structure suitable for large portfolios.

The factor covariance estimator can also be easily extended to allow for different asset classes which may have exposure to different factors by restricting coefficients on unrelated factors to be zero. For example, suppose a portfolio consisted of equity and credit products, and that a total of 5 factors where needed to model the covariance – 1 common to all assets, 2 specific to equities and 2 specific to bonds. The factor covariance would be a 5 by 5 matrix, but the factor

loadings for any asset would only have three non-zero coefficients, the common and equity factors if the asset was an equity or the common and the credit factors if the asset was a bond. Zero restrictions in the factor loadings allows for parsimonious models to be built to model complex portfolios, even in cases where many factors are needed to span the range of assets in a portfolio.

### 9.3.4 Principal Component Covariance

Principal component analysis (PCA) is a statistical technique which can be used to decompose a $T$ by $k$ matrix $\mathbf{Y}$ into a $T$ by $k$ set of orthogonal factors, $\mathbf{F}$,and a $k$ by $k$ set of normalized weights (or factor loadings), $\boldsymbol{\beta}$. Formally the principal component problem can be defined as

$$\min_{\boldsymbol{\beta},\mathbf{F}} (kT)^{-1} \sum_{i=1}^{k} \sum_{t=1}^{T} \left( y_{i,t} - \mathbf{f}_t \boldsymbol{\beta}_i \right)^2 \text{ subject to } \boldsymbol{\beta}'\boldsymbol{\beta} = \mathbf{I}_k \tag{9.7}$$

where $\mathbf{f}_t$ is a 1 by $k$ vector of common factors and $\boldsymbol{\beta}_i$ is a $k$ by 1 vector of factor loadings. The solution to the principle component problem is given by the eigenvalue decomposition of the outer product of $\mathbf{Y}$, $\boldsymbol{\Omega} = \mathbf{Y}'\mathbf{Y} = \sum_{t=1}^{T} \mathbf{y}_t \mathbf{y}_t'$.

**Definition 9.6** (Orthonormal Matrix)**.** A $k$-dimensional orthonormal matrix $\mathbf{U}$ satisfies $\mathbf{U}'\mathbf{U} = \mathbf{I}_k$, and so $\mathbf{U}' = \mathbf{U}^{-1}$.

**Definition 9.7** (Eigenvalue)**.** The eigenvalues of a real, symmetric matrix $k$ by $k$ matrix $\mathbf{A}$ are the $k$ solutions to

$$|\lambda \mathbf{I}_k - \mathbf{A}| = 0 \tag{9.8}$$

where $|\cdot|$ is the determinant.

**Definition 9.8** (Eigenvector)**.** An a $k$ by 1 vector $\mathbf{u}$ is an eigenvector corresponding to an eigenvalue $\lambda$ of a real, symmetric matrix $k$ by $k$ matrix $\mathbf{A}$ if

$$\mathbf{A}\mathbf{u} = \lambda \mathbf{u} \tag{9.9}$$

**Theorem 9.1** (Spectral Decomposition Theorem)**.** *A real, symmetric matrix* $\mathbf{A}$ *can be factored into* $\mathbf{A} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}'$ *where* $\mathbf{U}$ *is an orthonormal matrix (*$\mathbf{U}' = \mathbf{U}^{-1}$*) containing the eigenvectors of* $\mathbf{A}$ *in its columns and* $\boldsymbol{\Lambda}$ *is a diagonal matrix with the eigenvalues* $\lambda_1, \lambda_2, \ldots, \lambda_k$ *of* $\mathbf{A}$ *along its diagonal.*

Since $\mathbf{Y}'\mathbf{Y} = \boldsymbol{\Omega}$ is real and symmetric with eigenvalues $\boldsymbol{\Lambda} = \text{diag}(\lambda_i)_{i=1,\ldots,k}$, the factors the be determined using

$$
\begin{aligned}
\mathbf{Y}'\mathbf{Y} &= \mathbf{U}\mathbf{\Lambda}\mathbf{U}' \\
\mathbf{U}'\mathbf{Y}'\mathbf{Y}\mathbf{U} &= \mathbf{U}'\mathbf{U}\mathbf{\Lambda}\mathbf{U}'\mathbf{U} \\
(\mathbf{Y}\mathbf{U})'(\mathbf{Y}\mathbf{U}) &= \mathbf{\Lambda} && \text{since } \mathbf{U}' = \mathbf{U}^{-1} \\
\mathbf{F}'\mathbf{F} &= \mathbf{\Lambda}.
\end{aligned}
$$

$\mathbf{F} = \mathbf{Y}\mathbf{U}$ is the $T$ by $k$ matrix of factors and $\boldsymbol{\beta} = \mathbf{U}'$ is the $k$ by $k$ matrix of factor loadings. Additionally $\mathbf{F}\boldsymbol{\beta} = \mathbf{F}\mathbf{U}' = \mathbf{Y}\mathbf{U}\mathbf{U}' = \mathbf{Y}$.[3]

Covariance estimation based in Principal Component Analysis (PCA) is virtually identical to observable factor covariance estimation. The sole difference is that the factors are estimated directly from the returns and so covariance estimators using PCA do not require the common factors to be observable.

**Definition 9.9** ($n$-period Principal Component Covariance)**.** The $n$-period principal component covariance is defined as

$$\mathbf{\Sigma}_t = \boldsymbol{\beta}_t' \mathbf{\Sigma}_t^f \boldsymbol{\beta}_t + \mathbf{\Omega}_t \tag{9.10}$$

where $\mathbf{\Sigma}_t^f = n^{-1}\sum_{i=1}^n \mathbf{f}_{t-i}\mathbf{f}_{t-i}'$ is the $n$-period moving covariance of first $p$ principal component factors, $\hat{\boldsymbol{\beta}}_t$ is the $p$ by $k$ matrix of principal component loadings corresponding to the first $p$ factors, and $\mathbf{\Omega}_t$ is a diagonal matrix with $\omega_{j,t+1}^2 = n^{-1}\sum_{i=1}^n \eta_{j,t-1}^2$ on the j$^{\text{th}}$diagonal where $\eta_{i,t} = r_{i,t} - \mathbf{f}_t'\boldsymbol{\beta}_{i,t}$ are the residuals from a $p$-factor principal component analysis.

Selecting the number of factors to use, $p$, is one of the more difficult aspects of implementing a PCA covariance. One method specifies a fixed number of factors based on experience with the data or empirical regularities, for example selecting 3 factors when working with equity returns. An alternative is to select the number of factors by minimizing an information criteria such as those proposed in Bai and Ng (2002),

$$IC(p) = \ln(V(p,\hat{\mathbf{f}}^p)) + p\frac{k+T}{kT}\ln\left(\frac{kT}{k+T}\right)$$

where

$$V(p,\hat{\mathbf{f}}^p) = (kT)^{-1}\sum_{i=1}^k\sum_{t=1}^T \eta_{i,t}^2 \tag{9.11}$$

$$= (kT)^{-1}\sum_{i=1}^k\sum_{t=1}^T \left(r_{i,t} - \boldsymbol{\beta}_i^p \mathbf{f}_t^p\right)^2 \tag{9.12}$$

where $\boldsymbol{\beta}_i^p$ are the $p$ factor loadings for asset $i$, and $\mathbf{f}_t^p$ are the first $p$ factors.

[3] The factors and factor loadings are only identified up a scaling by $\pm 1$.

**Principal Component Analysis of the S&P 500**

| $k = 194$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Partial $R^2$ | 0.263 | 0.039 | 0.031 | 0.023 | 0.019 | 0.016 | 0.014 | 0.013 | 0.012 | 0.011 |
| Cumulative $R^2$ | 0.263 | 0.302 | 0.333 | 0.356 | 0.375 | 0.391 | 0.405 | 0.418 | 0.430 | 0.441 |

Table 9.1: Percentage of variance explained by the first 10 eigenvalues of the outer product matrix of S&P 500 returns. Returns on an asset were only included if an asset was in the S&P 500 for the entire sample ($k$ contains the number which meet this criteria). The second line contains the cumulative $R^2$ of a $p$-factor model for the first 10 factors.

#### 9.3.4.1 Interpreting the components

One nice feature of PCA is that the factors can be easily interpreted in terms of their contribution to total variance using $R^2$. This interpretation is possible since the factors are orthogonal, and so the $R^2$ of a model including $p < k$ factors is the sum of the $R^2$ of the $p$ factors. Suppose the eigenvalues were ordered from largest to smallest and so $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k$ and that the factors associated with eigenvalue $i$ has been ordered such that it appears in column $i$ of $\mathbf{F}$. The $R^2$ associated with factor $i$ is then

$$\frac{\lambda_i}{\lambda_1 + \lambda_2 + \ldots + \lambda_k}$$

and the cumulative $R^2$ of including $p < k$ factors is

$$\frac{\lambda_1 + \lambda_2 + \ldots + \lambda_p}{\lambda_1 + \lambda_2 + \ldots + \lambda_k}.$$

Cumulative $R^2$ is often used to select a subset of the $k$ factors for model building. For example, in equity return data, it is not uncommon for 3–5 factors to explain 30% of the total variation in a large panel of equity returns.

#### 9.3.4.2 Alternative methods

Principal components can also computed on either the covariance matrix of $\mathbf{Y}$ or the correlation matrix of $\mathbf{Y}$. Using the covariance matrix is equivalent to building a model with an intercept,

$$y_{i,t} = \alpha_i + \mathbf{f}_t \boldsymbol{\beta}_i \tag{9.13}$$

which differs from the principal components extracted from the outer product which is equivalent to the model

$$y_{i,t} = \mathbf{f}_t \boldsymbol{\beta}_i. \tag{9.14}$$

When working with asset return data, the difference between principal components extracted from the outer product and the covariance is negligible except in certain markets (e.g. electricity markets) or when the using returns covering long time spans (e.g. one month or more).

Principal components can also be extracted from the sample correlation matrix of $\mathbf{Y}$ which is equivalent to the model

$$\frac{y_{i,t} - \bar{y}_i}{\hat{\sigma}_i} = \mathbf{f}_t \boldsymbol{\beta}_i \tag{9.15}$$

where $\bar{y}_i = T^{-1}\sum_{t=1}^{T} y_{i,t}$ is the mean of $\mathbf{y}_i$ and $\hat{\sigma}_i$ is the sample standard deviation of $\mathbf{y}_i$. PCA is usually run on the correlation matrix when a subset of the series in $\mathbf{Y}$ have variances which are much larger than the others. In cases where the variances differ greatly, principal components extracted from the outer product or covariance will focus on the high variance data series since fitting these produces the largest decrease in residual variance and thus the largest increases in $R^2$.

### 9.3.5 Equicorrelation

Equicorrelation, like factor models, is a restricted covariance estimator. The equicorrelation estimator assumes that the covariance between any two assets can be expressed as $\rho\sigma_i\sigma_j$ where $\sigma_i$ and $\sigma_j$ are the volatilities of assets $i$ and $j$, respectively. The correlation parameter is *not* indexed by $i$ or $j$, and so it is common to all assets. This estimator is clearly mis-specified whenever $k > 2$, and is generally only appropriate for assets where the majority of the pairwise correlations are positive and the correlations are fairly homogeneous.[4]

**Definition 9.10** ($n$-period Moving Average Equicorrelation Covariance)**.** The $n$-period moving average equicorrelation covariance is defined as

$$\mathbf{\Sigma}_t = \begin{bmatrix} \sigma_{1,t}^2 & \rho_t\sigma_{1,t}\sigma_{2,t} & \rho_t\sigma_{1,t}\sigma_{3,t} & \dots & \rho_t\sigma_{1,t}\sigma_{k,t} \\ \rho_t\sigma_{1,t}\sigma_{2,t} & \sigma_{2,t}^2 & \rho_t\sigma_{2,t}\sigma_{3,t} & \dots & \rho_t\sigma_{2,t}\sigma_{k,t} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_t\sigma_{1,t}\sigma_{k,t} & \rho_t\sigma_{2,t}\sigma_{k,t} & \rho_t\sigma_{3,t}\sigma_{k,t} & \dots & \sigma_{k,t}^2 \end{bmatrix} \tag{9.16}$$

where $\sigma_{j,t}^2 = n^{-1}\sum_{i=1}^{n}\epsilon_{j,t}^2$ and $\rho_t$ is estimated using one of the estimators below.

The equicorrelation can be estimated either using a moment-based estimator or using a maximum-likelihood estimator. Defined $\epsilon_{p,t}$ as the equally weighted portfolio return. It is

[4]The positivity constraint is needed to ensure that the covariance is positive definite which requires $\rho \in (-1/(k-1), 1)$, and so for $k$ moderately large, the lower bound is effectively 0.

straightforward to see that

$$\begin{aligned}\mathrm{E}[\epsilon_{p,t}^2] &= k^{-2}\sum_{j=1}^{k}\sigma_{j,t}^2 + 2k^{-2}\sum_{o=1}^{k}\sum_{q=o+1}^{k}\rho\sigma_{o,t}\sigma_{q,t} \\ &= k^{-2}\sum_{j=1}^{k}\sigma_{j,t}^2 + 2\rho k^{-2}\sum_{o=1}^{k}\sum_{q=o+1}^{k}\sigma_{o,t}\sigma_{q,t}\end{aligned} \tag{9.17}$$

if the correlations among all of the pairs of assets were identical. The moment-based estimator replaces population values with estimates,

$$\sigma_{j,t}^2 = n^{-1}\sum_{i=1}^{n}\epsilon_{j,t-i}^2, \quad j = 1,2,\ldots,k,p,$$

and the equicorrelation is estimated using

$$\rho_t = \frac{\sigma_{p,t}^2 - k^{-2}\sum_{j=1}^{k}\sigma_{j,t}^2}{2k^{-2}\sum_{o=1}^{k}\sum_{q=o+1}^{k}\sigma_{o,t}\sigma_{q,t}}.$$

Alternatively maximum likelihood assuming returns are multivariate Gaussian can be used to estimate the equicorrelation using standardized residuals $u_{j,t} = \epsilon_{j,t}/\sigma_{j,t}$. The estimator for $\rho$ can be found by maximizing the likelihood

$$\begin{aligned}L(\rho_t;\mathbf{u}) &= -\frac{1}{2}\sum_{i=1}^{n} k\ln 2\pi + \ln|\mathbf{R}_t| + \mathbf{u}_{t-i}'\mathbf{R}_t^{-1}\mathbf{u}_{t-i} \\ &= \sum_{i=1}^{n} k\ln 2\pi + \ln\left((1-\rho_t)^{k-1}(1+(k-1)\rho_t)\right) \\ &\quad + \frac{1}{(1-\rho_t)}\left[\sum_{j=1}^{k}u_{j,t-i}^2 - \frac{\rho_t}{1+(k-1)\rho_t}\left(\sum_{q=1}^{k}u_{q,t-i}\right)^2\right]\end{aligned} \tag{9.18}$$

where $\mathbf{u}_t$ is a $k$ by 1 vector of standardized residuals and $\mathbf{R}_t$ is a correlation matrix with all non-diagonal elements equal to $\rho$. This likelihood is computationally similar to univariate likelihood for any $k$ and so maximization is very fast even when $k$ is large.[5]

### 9.3.6 Application: S&P 500

The S&P 500 was used to illustrate some of the similarities among the simple covariance estimators. Daily data on all constituents of the S&P 500 was downloaded from CRSP from January

[5]The computation speed of the likelihood can be increased by pre-computing $\sum_{j=1}^{k}u_{j,t-i}^2$ and $\sum_{q=1}^{k}u_{q,t-i}$.

**Correlation Measures for the S&P 500**

| Equicorrelation | 1-Factor $R^2$ (S&P 500) | 3-Factor $R^2$ (Fama-French) |
|---|---|---|
| 0.255 | 0.236 | 0.267 |

Table 9.2: Full sample correlation measures for the S&P 100. Returns on an asset were only included if an asset was in the S&P 500 for the entire sample ($n$ contains the number which meet this criteria). The 1-factor $R^2$ was from a model that used the return on the S&P 500 and the 3-factor $R^2$ was from a model that used the 3 Fama-French factors.

1, 1999 until December 31, 2008. Principal component analysis was conducted on the subset of the returns which was available for each day in the sample. Table 9.1 contains the number of assets which meet this criteria ($k$) and both the partial and cumulative $R^2$ for the first 10 principal components. The first explains a substantial amount of the data – 25%, and the next four combine to explain 35% of the cross-sectional variation. If returns did not follow a factor structure each principal component would be expected to explain approximately 0.5% of the variation. Table 9.2 contains the full-sample equicorrelation, 1-factor $R^2$ using the S&P 500 as the observable factor and the 3-factor $R^2$ using the 3 Fama-French factors. The average correlation and the 1-factor fit is similar to that in 1-factor PCA model, although the 3 Fama-French factors do not appear to work as well as the 3 factors estimated from the data. This is likely due to the lack of cross-sectional variation with respect to size in the S&P 500 when compared to all assets in CRSP.

Figure 9.2 contains a plot of the 252-day moving average equicorrelation and 1- and 3-factor PCA $R^2$. Periods of high volatility, such as the end of the dot-com bubble and late 2008, appear to also have high correlation. The three lines broadly agree about the changes and only differ in level. Figure 9.3 contains plots of the $R^2$ from the 1-factor PCA and the 1-factor model which uses the S&P500 return (top panel) and the 3-factor PCA and the 3 Fama-French factors (bottom panel). The dynamics in all series are similar with the largest differences in the levels, where PCA fits the cross-section of data better than the observable models.

## 9.4 Multivariate ARCH Models

### 9.4.1 Vector GARCH (*vec*)

Vector GARCH was the first multivariate ARCH specification Bollerslev, R. F. Engle, and Wooldridge (1988), and the natural extension of the standard GARCH model. The model is defined using the vec of the conditional covariance.

**Definition 9.11** (Vector GARCH)**.** The covariance in a vector GARCH model (*vec*) evolves accord-

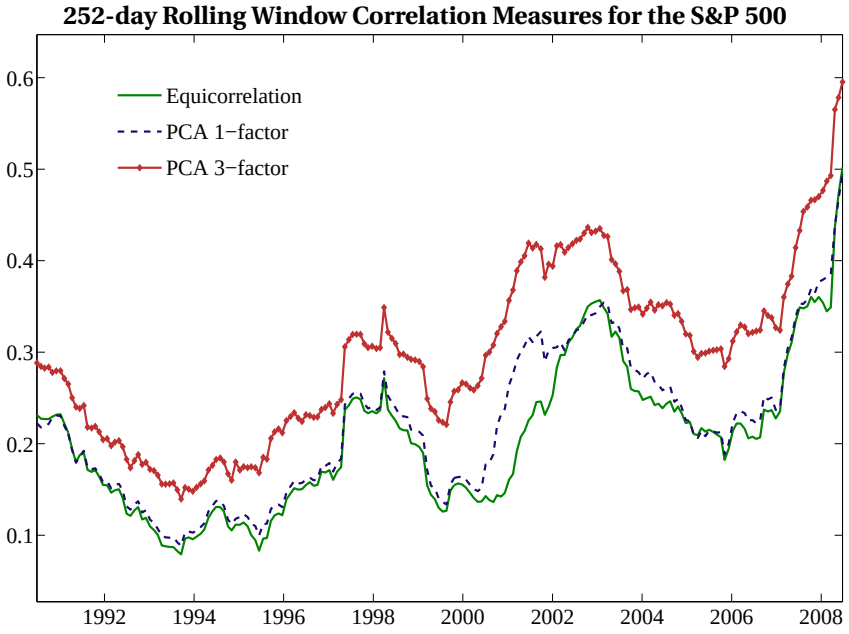


Figure 9.2: Three views into the average correlation in the S&P 500. The PCA measures are the $R^2$ of models with 1 and 3 factors. Each estimate was computed using a 252-day rolling window and is plotted against the center of the rolling window. All three measures roughly agree about the changes in the average correlation.

---

ing to

$$\begin{aligned} \text{vec}(\mathbf{\Sigma}_t) &= \text{vec}(\mathbf{C}) + \mathbf{A}\text{vec}\left(\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1}'\right) + \mathbf{B}\text{vec}(\mathbf{\Sigma}_{t-1}) && (9.19) \\ &= \text{vec}(\mathbf{C}) + \mathbf{A}\text{vec}\left(\mathbf{\Sigma}_{t-1}^{\frac{1}{2}}\mathbf{e}_t\left(\mathbf{\Sigma}_t^{\frac{1}{2}}\mathbf{e}_t\right)'\right) + \mathbf{B}\text{vec}(\mathbf{\Sigma}_{t-1}) && (9.20) \end{aligned}$$

where $\mathbf{C}$ is a $k$ by $k$ positive definite matrix and both $\mathbf{A}$ and $\mathbf{B}$ are $k^2$ by $k^2$ parameter matrices. In the second line, $\mathbf{\Sigma}_{t-1}^{\frac{1}{2}}$ is a matrix square root and $\{\mathbf{e}_t\}$ is a sequence of i.i.d. random variables with mean 0 and covariance $\mathbf{I}_k$, such as a standard multivariate normal.

See eq. 5.9 for the definition of the vec operator. The *vec* allows each cross-product to influence each covariance term. To understand the richness of the specification, consider the evolution of the conditional covariance in a bivariate model,

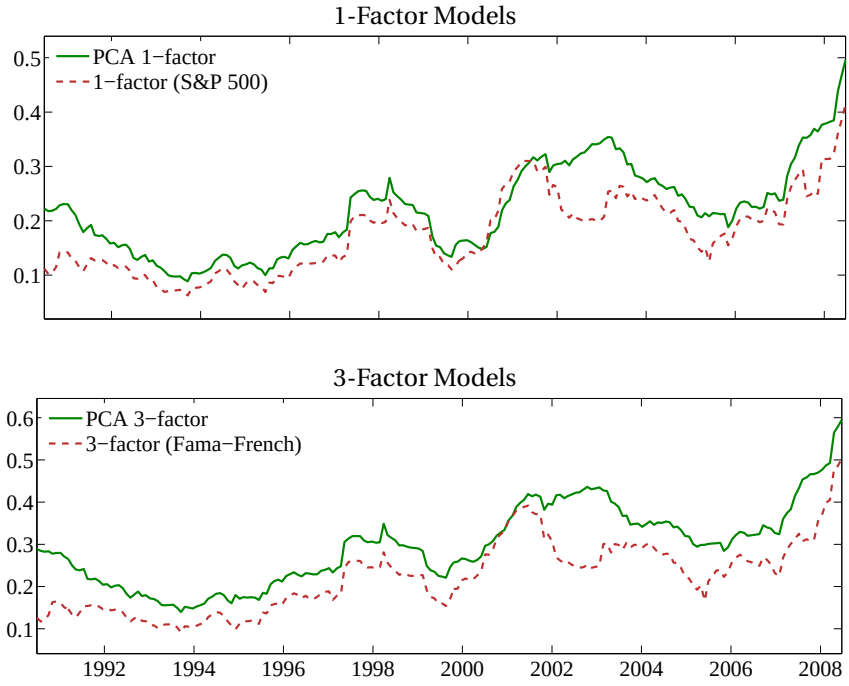


Figure 9.3: The top panel plots the $R^2$ for 1-factor PCA and an observable factor model which uses the return on the S&P 500 as the observable factor. The bottom contains the same for 3-factor PCA and the Fama-French 3-factor model. Each estimate was computed using a 252-day rolling window and is plotted against the center of the rolling window.

---

$$\begin{bmatrix} \sigma_{11,t} \\ \sigma_{12,t} \\ \sigma_{12,t} \\ \sigma_{22,t} \end{bmatrix} = \begin{bmatrix} c_{11} \\ c_{12} \\ c_{12} \\ c_{22} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{42} & a_{44} \end{bmatrix} \begin{bmatrix} \epsilon^2_{1,t-1} \\ \epsilon_{1,t-1}\epsilon_{2,t-1} \\ \epsilon_{1,t-1}\epsilon_{2,t-1} \\ \epsilon^2_{2,t-1} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{32} & b_{33} \\ b_{41} & b_{42} & b_{42} & b_{44} \end{bmatrix} \begin{bmatrix} \sigma_1 \\ \sigma_1 \\ \sigma_1 \\ \sigma_2 \end{bmatrix}$$

The vec operator stacks the elements of the covariance matrix and the outer products of returns. The evolution of the conditional variance of the first asset,

$$\sigma_{11,t} = c_{11} + a_{11}\epsilon^2_{1,t-1} + 2a_{12}\epsilon_{1,t-1}\epsilon_{2,t-1} + a_{13}\epsilon^2_{2,t-1} + b_{11}\sigma_{11,t-1} + 2b_{12}\sigma_{12,t-1} + b_{13}\sigma_{22,t-1},$$

depends on all of the past squared returns and cross-products. In practice it is very difficult to

use the vector GARCH model since finding general conditions on $\mathbf{A}$ and $\mathbf{B}$ which will ensure that $\mathbf{\Sigma}_t$ is positive definite is difficult.

The diagonal vec model has been more successful, primarily because it is relatively straight forward to find conditions which ensure that the conditional covariance is positive semi-definite. The diagonal *vec* model restricts $\mathbf{A}$ and $\mathbf{B}$ to diagonal matrices which means that the elements of $\mathbf{\Sigma}_t$ evolve according to

$$\begin{aligned}\mathbf{\Sigma}_t &= \mathbf{C} + \tilde{\mathbf{A}} \odot \boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1}' + \tilde{\mathbf{B}} \odot \mathbf{\Sigma}_{t-1} && (9.21)\\ &= \mathbf{C} + \tilde{\mathbf{A}} \odot \left(\mathbf{\Sigma}_{t-1}^{\frac{1}{2}}\mathbf{e}_t \left(\mathbf{\Sigma}_t^{\frac{1}{2}}\mathbf{e}_t\right)'\right) + \tilde{\mathbf{B}} \odot \mathbf{\Sigma}_{t-1} && (9.22)\end{aligned}$$

where $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{B}}$ are symmetric parameter matrices and $\odot$ is Hadamard product.[6] All elements of $\mathbf{\Sigma}_t$ evolves according a GARCH(1,1)-like specification, so that

$$\sigma_{ij,t} = c_{ij} + \tilde{a}_{ij}\epsilon_{i,t-1}\epsilon_{j,t-1} + \tilde{b}_{ij}\sigma_{ij,t-1}.$$

The diagonal *vec* still requires restrictions the parameters to ensure that the conditional covariance is positive definite. Ding and R. Engle (2001) develop one set of restrictions in the context of Matrix GARCH (see section 9.4.3).

### 9.4.2 BEKK GARCH

The BEKK (Baba, Engle, Kraft and Kroner) directly addresses the difficulties in finding constraints on the parameters of a vec specification (R. F. Engle and Kroner, 1995). The primary insight of the BEKK is that quadratic forms are positive semi-definite and the sum of a positive semi-definite matrix and a positive definite matrix is positive definite.

**Definition 9.14** (BEKK GARCH)**.** The covariance in a BEKK GARCH(1,1) model evolves according to

$$\begin{aligned}\mathbf{\Sigma}_t &= \mathbf{C}\mathbf{C}' + \mathbf{A}\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1}'\mathbf{A}' + \mathbf{B}\mathbf{\Sigma}_{t-1}\mathbf{B}' && (9.23)\\ &= \mathbf{C}\mathbf{C}' + \mathbf{A}\left(\mathbf{\Sigma}_{t-1}^{\frac{1}{2}}\mathbf{e}_t \left(\mathbf{\Sigma}_t^{\frac{1}{2}}\mathbf{e}_t\right)'\right)\mathbf{A}' + \mathbf{B}\mathbf{\Sigma}_{t-1}\mathbf{B}' && (9.24)\end{aligned}$$

where $\mathbf{C}$ is a $k$ by $k$ lower triangular matrix and $\mathbf{A}$ and $\mathbf{B}$ are $k$ by $k$ parameter matrices.

[6]

**Definition 9.12** (Hadamard Product)**.** Let $\mathbf{A}$ and $\mathbf{B}$ be matrices with the same size. The Hadamard product of $\mathbf{A}$ and $\mathbf{B}$ denoted $\mathbf{A} \odot \mathbf{B}$ is the matrix with ij$^{\text{th}}$element $a_{ij}b_{ij}$.

**Definition 9.13** (Hadamard Quotient)**.** Let $\mathbf{A}$ and $\mathbf{B}$ be matrices with the same size. The Hadamard quotient of $\mathbf{A}$ and $\mathbf{B}$ denoted $\mathbf{A} \oslash \mathbf{B}$ is the matrix with ij$^{\text{th}}$element $a_{ij}/b_{ij}$.

Using the vec operator, the BEKK can be seen as a restricted version of the *vec* specification where $\mathbf{A} \otimes \mathbf{A}$ and $\mathbf{B} \otimes \mathbf{B}$ control the response to recent news and the smoothing, respectively,

$$\begin{aligned} \text{vec}(\mathbf{\Sigma}_t) &= \text{vec}\left(\mathbf{C}\mathbf{C}'\right) + \mathbf{A} \otimes \mathbf{A}\text{vec}\left(\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1}'\right) + \mathbf{B} \otimes \mathbf{B}\text{vec}\left(\mathbf{\Sigma}_{t-1}\right). && (9.25) \\ &= \text{vec}\left(\mathbf{C}\mathbf{C}'\right) + \mathbf{A} \otimes \mathbf{A}\text{vec}\left(\mathbf{\Sigma}_{t-1}^{\frac{1}{2}}\mathbf{e}_t\left(\mathbf{\Sigma}_t^{\frac{1}{2}}\mathbf{e}_t\right)'\right) + \mathbf{B} \otimes \mathbf{B}\text{vec}\left(\mathbf{\Sigma}_{t-1}\right). && (9.26) \end{aligned}$$

The elements of $\mathbf{\Sigma}_t$ generally depend on all cross-products. For example, consider a bivariate BEKK,

$$\begin{aligned} \begin{bmatrix} \sigma_{11,t} & \sigma_{12,t} \\ \sigma_{12,t} & \sigma_{22,t} \end{bmatrix} &= \begin{bmatrix} c_{11} & 0 \\ c_{12} & c_{22} \end{bmatrix} \begin{bmatrix} c_{11} & 0 \\ c_{21} & c_{22} \end{bmatrix}' \\ &+ \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} \epsilon_{1,t-1}^2 & \epsilon_{1,t-1}\epsilon_{2,t-1} \\ \epsilon_{1,t-1}\epsilon_{2,t-1} & \epsilon_{2,t-1}^2 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}' \\ &+ \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} \sigma_{11,t-1} & \sigma_{12,t-1} \\ \sigma_{12,t-1} & \sigma_{22,t-1} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}' \end{aligned} \tag{9.27}$$

The conditional variance of the first asset is given by

$$\sigma_{11,t} = c_{11}^2 + a_{11}^2\epsilon_{1,t-1}^2 + 2a_{11}a_{12}\epsilon_{1,t-1}\epsilon_{2,t-1} + a_{12}^2\epsilon_{2,t-1}^2 + b_{11}^2\sigma_{11,t-1} + 2b_{11}b_{12}\sigma_{12,t-1} + b_{12}^2\sigma_{22,t-1}.$$

The other conditional variance and the conditional covariance have similar forms that depend on both squared returns and the cross-product of returns.

Estimation of full BEKK models rapidly becomes difficult as the number of parameters in the model is $(5k^2 + k)/2$, and so is usually only appropriate for $k \leq 5$. The diagonal BEKK partially addresses some of the number of parameters by restricting $\mathbf{A}$ and $\mathbf{B}$ to be diagonal matrices,

**Definition 9.15** (Diagonal BEKK GARCH)**.** The covariance in a diagonal BEKK-GARCH(1,1) model evolves according to

$$\begin{aligned} \mathbf{\Sigma}_t &= \mathbf{C}\mathbf{C}' + \tilde{\mathbf{A}}\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1}'\tilde{\mathbf{A}}' + \tilde{\mathbf{B}}\mathbf{\Sigma}_{t-1}\tilde{\mathbf{B}}'. && (9.28) \\ &= \mathbf{C}\mathbf{C}' + \tilde{\mathbf{A}}\left(\mathbf{\Sigma}_{t-1}^{\frac{1}{2}}\mathbf{e}_t\left(\mathbf{\Sigma}_t^{\frac{1}{2}}\mathbf{e}_t\right)'\right)\tilde{\mathbf{A}}' + \tilde{\mathbf{B}}\mathbf{\Sigma}_{t-1}\tilde{\mathbf{B}}'. && (9.29) \end{aligned}$$

where $\mathbf{C}$ is a $k$ by $k$ lower triangular matrix and $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{B}}$ are diagonal parameter matrices.

The conditional covariances in a diagonal BEKK evolve according to

$$\sigma_{ij,t} = \tilde{c}_{ij} + a_i a_j \epsilon_{i,t-1}\epsilon_{j,t-1} + b_i b_j \sigma_{ij,t-1} \tag{9.30}$$

where $\tilde{c}_{ij}$ is the ij$^{\text{th}}$element of $\mathbf{CC}'$. This covariance evolution is similar to the diagonal vec specification except that the parameters are shared between different series. The scalar BEKK further restricts the parameter matrices to be common across all assets, and is a particularly simple (and restrictive) model.

**Definition 9.16** (Scalar BEKK GARCH)**.** The covariance in a scalar BEKK-GARCH(1,1) model evolves according to

$$\begin{aligned}\boldsymbol{\Sigma}_t &= \mathbf{CC}' + a^2\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1} + b^2\boldsymbol{\Sigma}_{t-1} && (9.31)\\ &= \mathbf{CC}' + a^2\left(\boldsymbol{\Sigma}_{t-1}^{\frac{1}{2}}\mathbf{e}_t\left(\boldsymbol{\Sigma}_t^{\frac{1}{2}}\mathbf{e}_t\right)'\right) + b^2\boldsymbol{\Sigma}_{t-1} && (9.32)\end{aligned}$$

where $\mathbf{C}$ is a $k$ by $k$ lower triangular matrix and $a$ and $b$ are scalar parameters.

The scalar BEKK has one further advantage: it can easily be covariance targeted. Covariance targeting replaces the intercept ($\mathbf{CC}'$) with a consistent estimator, $(1-a^2-b^2)\overline{\boldsymbol{\Sigma}}$, where $\overline{\boldsymbol{\Sigma}}$ is the long-run variance of the data. $\overline{\boldsymbol{\Sigma}}$ is usually estimated using the outer product of returns, and so $\widehat{\overline{\boldsymbol{\Sigma}}} = T^{-1}\sum_{t=1}^{T}\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t'$. The conditional covariance is then estimated conditioning on the unconditional covariance of returns,

$$\boldsymbol{\Sigma}_t = (1-a^2-b^2)\widehat{\overline{\boldsymbol{\Sigma}}} + a^2\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1} + b^2\boldsymbol{\Sigma}_{t-1} \tag{9.33}$$

and so only $a$ and $b$ reaming to be estimated using maximum likelihood. Scalar BEKK models can be used in large portfolios ($k > 50$), unlike models without targeting.

### 9.4.3 Matrix GARCH (M-GARCH)

Matrix GARCH (Ding and R. Engle, 2001) contains a set of parameterizations which include the diagonal vec and an alternative parameterization of the diagonal BEKK.

**Definition 9.17** (Matrix GARCH)**.** The covariance in a Matrix GARCH(1,1) model evolves according to

$$\begin{aligned}\boldsymbol{\Sigma}_t &= \mathbf{CC}' + \mathbf{AA}' \odot \boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1} + \mathbf{BB}' \odot \boldsymbol{\Sigma}_{t-1} && (9.34)\\ &= \mathbf{CC}' + \mathbf{AA}' \odot \left(\boldsymbol{\Sigma}_{t-1}^{\frac{1}{2}}\mathbf{e}_t\left(\boldsymbol{\Sigma}_t^{\frac{1}{2}}\mathbf{e}_t\right)'\right) + \mathbf{BB}' \odot \boldsymbol{\Sigma}_{t-1} && (9.35)\end{aligned}$$

where $\mathbf{C}$, $\mathbf{A}$ and $\mathbf{B}$ are lower triangular matrices.

Ding and R. Engle (2001) show that if $\mathbf{U}$ and $\mathbf{V}$ are positive semi-definite matrices, then $\mathbf{U} \odot \mathbf{V}$ is also, which, when combined with the quadratic forms in the model, ensures that $\boldsymbol{\Sigma}_t$ will be positive definite as long as $\mathbf{C}$ has full rank. They also propose a diagonal Matrix GARCH specification which is equivalent to the diagonal BEKK.

**Definition 9.18** (Diagonal Matrix GARCH)**.** The covariance in a diagonal Matrix GARCH(1,1) model evolves according to

$$\begin{aligned}\boldsymbol{\Sigma}_t &= \mathbf{C}\mathbf{C}' + \mathbf{a}\mathbf{a}' \odot \boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1}' + \mathbf{b}\mathbf{b}' \odot \boldsymbol{\Sigma}_{t-1} \quad (9.36)\\ &= \mathbf{C}\mathbf{C}' + \mathbf{a}\mathbf{a}' \odot \left(\boldsymbol{\Sigma}_{t-1}^{\frac{1}{2}}\mathbf{e}_t \left(\boldsymbol{\Sigma}_t^{\frac{1}{2}}\mathbf{e}_t\right)'\right) + \mathbf{b}\mathbf{b}' \odot \boldsymbol{\Sigma}_{t-1}, \quad (9.37)\end{aligned}$$

where $\mathbf{C}$ is a lower triangular matrices and $\mathbf{a}$ and $\mathbf{b}$ are $k$ by 1 parameter vectors.

The scalar version of the Matrix GARCH is identical to the scalar BEKK.

### 9.4.4 Constant Conditional Correlation (CCC) GARCH

Constant Conditional Correlation GARCH Bollerslev (1990) uses a different approach to that of the vec, BEKK, and Matrix GARCH. CCC GARCH decomposes the conditional covariance into $k$ conditional variances and the conditional correlation, which is assumed to be constant,

$$\boldsymbol{\Sigma}_t = \mathbf{D}_t\mathbf{R}\mathbf{D}_t. \quad (9.38)$$

$\mathbf{D}_t$ is a diagonal matrix with the conditional standard deviation of the i$^{\text{th}}$asset in its i$^{\text{th}}$diagonal position.

$$\mathbf{D}_t = \begin{bmatrix} \sigma_{1,t} & 0 & 0 & \dots & 0 \\ 0 & \sigma_{2,t} & 0 & \dots & 0 \\ 0 & 0 & \sigma_{3,t} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \sigma_{k,t} \end{bmatrix} \quad (9.39)$$

where $\sigma_{i,t} = \sqrt{\sigma_{ii,t}}$. The conditional variances are typically modeled using standard GARCH(1,1) models,

$$\begin{aligned}\sigma_{ii,t} &= \omega_i + \alpha_i r_{i,t-1}^2 + \beta_i \sigma_{ii,t-1} \quad (9.40)\\ &= \omega_i + \alpha_i \sigma_{ii,t-1} u_{i,t-1}^2 + \beta_i \sigma_{ii,t-1}\end{aligned}$$

where $u_{i,t-1}$ is the i$^{\text{th}}$ element of $\mathbf{u}_t = \mathbf{R}^{\frac{1}{2}}\mathbf{e}_t$ where $\{\mathbf{e}_t\}$ is a sequence of i.i.d. random variables with mean 0 and covariance $\mathbf{I}_k$, such as a standard multivariate normal. Other specifications, such as TARCH or EGARCH, can also be used. It is also possible to model the conditional variances a different models for each asset, a distinct advantage over *vec* and related models. The conditional correlation is constant

$$\mathbf{R} = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} & \dots & \rho_{1k} \\ \rho_{12} & 1 & \rho_{23} & \dots & \rho_{2k} \\ \rho_{13} & \rho_{23} & 1 & \dots & \rho_{3k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{1k} & \rho_{2k} & \rho_{3k} & \dots & 1 \end{bmatrix}. \tag{9.41}$$

The conditional covariance matrix is computed of the conditional standard deviations and the conditional correlation, and so all of the dynamics in the conditional covariance are attributable to changes in the conditional variances.

$$\mathbf{\Sigma}_t = \begin{bmatrix} \sigma_{11,t} & \rho_{12}\sigma_{1,t}\sigma_{2,t} & \rho_{13}\sigma_{1,t}\sigma_{3,t} & \dots & \rho_{1k}\sigma_{1,t}\sigma_{k,t} \\ \rho_{12}\sigma_{1,t}\sigma_{2,t} & \sigma_{22,t} & \rho_{23}\sigma_{2,t}\sigma_{3,t} & \dots & \rho_{2k}\sigma_{2,t}\sigma_{k,t} \\ \rho_{13}\sigma_{1,t}\sigma_{3,t} & \rho_{23}\sigma_{2,t}\sigma_{3,t} & \sigma_{33,t} & \dots & \rho_{3k}\sigma_{3,t}\sigma_{k,t} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{1k}\sigma_{1,t}\sigma_{k,t} & \rho_{2k}\sigma_{2,t}\sigma_{k,t} & \rho_{3k}\sigma_{3,t}\sigma_{k,t} & \dots & \sigma_{kk,t} \end{bmatrix} \tag{9.42}$$

Bollerslev (1990) shows that the CCC GARCH model can be estimated in two steps. The first fits $k$ conditional variance models (e.g. GARCH) and produces the vector of standardized residuals $\mathbf{u}_t$ where $u_{i,t} = \epsilon_{i,t}/\sqrt{\hat{\sigma}_{ii,t}}$. The second step estimates the constant conditional correlation using the standard correlation estimator on the standardized residuals.

**Definition 9.19** (Constant Conditional Correlation GARCH)**.** The covariance in a constant conditional correlation GARCH model evolves according to

$$\mathbf{\Sigma}_t = \begin{bmatrix} \sigma_{11,t} & \rho_{12}\sigma_{1,t}\sigma_{2,t} & \rho_{13}\sigma_{1,t}\sigma_{3,t} & \dots & \rho_{1k}\sigma_{1,t}\sigma_{k,t} \\ \rho_{12}\sigma_{1,t}\sigma_{2,t} & \sigma_{22,t} & \rho_{23}\sigma_{2,t}\sigma_{3,t} & \dots & \rho_{2k}\sigma_{2,t}\sigma_{k,t} \\ \rho_{13}\sigma_{1,t}\sigma_{3,t} & \rho_{23}\sigma_{2,t}\sigma_{3,t} & \sigma_{33,t} & \dots & \rho_{3k}\sigma_{3,t}\sigma_{k,t} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{1k}\sigma_{1,t}\sigma_{k,t} & \rho_{2k}\sigma_{2,t}\sigma_{k,t} & \rho_{3k}\sigma_{3,t}\sigma_{k,t} & \dots & \sigma_{kk,t} \end{bmatrix} \tag{9.43}$$

where $\sigma_{ii,t}$, $i = 1, 2, \dots, k$ evolves according to some univariate GARCH process on asset $i$, usually a GARCH(1,1).

### 9.4.5 Dynamic Conditional Correlation (DCC)

Dynamic Conditional Correlation extends CCC GARCH by introducing simple, scalar BEKK-like dynamics to the conditional correlations, and so $\mathbf{R}$ in the CCC is replaced with $\mathbf{R}_t$ in the DCC (R. F. Engle, 2002b)

**Definition 9.20** (Dynamic Conditional Correlation GARCH)**.** The covariance in a dynamic con-

ditional correlation GARCH model evolves according to

$$\mathbf{\Sigma}_t = \mathbf{D}_t \mathbf{R}_t \mathbf{D}_t. \tag{9.44}$$

where

$$\mathbf{R}_t = \mathbf{Q}_t^* \mathbf{Q}_t \mathbf{Q}_t^*, \tag{9.45}$$

$$\begin{aligned} \mathbf{Q}_t &= (1-a-b)\overline{\mathbf{R}} + a\mathbf{u}_{t-1}\mathbf{u}_{t-1}' + b\mathbf{Q}_{t-1}, & (9.46)\\ &= (1-a-b)\overline{\mathbf{R}} + a\left(\mathbf{R}_{t-1}^{\frac{1}{2}}\mathbf{e}_{t-1}\right)\left(\mathbf{R}_{t-1}^{\frac{1}{2}}\mathbf{e}_{t-1}\right)' + b\mathbf{Q}_{t-1}, & (9.47) \end{aligned}$$

$$\mathbf{Q}_t^* = (\mathbf{Q}_t \odot \mathbf{I}_k)^{-\frac{1}{2}} \tag{9.48}$$

$\mathbf{u}_t$ is the $k$ by 1 vector of standardized returns ($u_{i,t} = \epsilon_{i,t}/\sqrt{\hat{\sigma}_{ii,t}}$) and $\odot$ denotes Hadamard multiplication (element-by-element). $\{\mathbf{e}_t\}$ are a sequence of i.i.d. innovations with mean $\mathbf{0}$ and covariance $\mathbf{I}_k$, such as a standard multivariate normal or possibly a heavy tailed distribution. $\mathbf{D}_t$ is a diagonal matrix with the conditional standard deviation of asset $i$ on the i$^{\text{th}}$diagonal position. The conditional variance, $\sigma_{ii,t}$, $i = 1, 2, \ldots, k$, evolve according to some univariate GARCH process for asset $i$, usually a GARCH(1,1) and are identical to eq. 9.40.

Eqs. 9.45 and 9.48 are needed to ensure that $\mathbf{R}_t$ is a correlation matrix with diagonal elements equal to 1. The $\mathbf{Q}_t$ process is parameterized in a similar manner to the variance targeting BEKK (eq. 9.33) which allows for three step estimation. The first two steps are identical to those of the CCC GARCH model. The third plugs in the estimate of the correlation into eq. 9.46 to estimate the parameters which govern the conditional correlation dynamics, $a$ and $b$.

### 9.4.6 Orthogonal GARCH (OGARCH)

The principle components of a $T$ by $k$ matrix of returns $\boldsymbol{\epsilon}$ are defined as $\mathbf{F} = \boldsymbol{\epsilon}\mathbf{U}$ where $\mathbf{U}$ is the matrix of eigenvectors of the outer product of $\boldsymbol{\epsilon}$. Orthogonal GARCH uses the first $p$ principle components to model the conditional covariance by assuming that the factors are conditionally uncorrelated.[7]

**Definition 9.21** (Orthogonal GARCH)**.** The covariance in an orthogonal GARCH (OGARCH) model evolves according to

$$\mathbf{\Sigma}_t = \boldsymbol{\beta}\mathbf{\Sigma}_t^f\boldsymbol{\beta}' + \mathbf{\Omega} \tag{9.49}$$

where $\boldsymbol{\beta}$ is the $k$ by $p$ matrix of factor loadings corresponding to the $p$ factors with the highest

[7] Principle components uses the unconditional covariance of returns and so only guarantees that the factors are unconditionally uncorrelated.

total $R^2$. The conditional covariance of the factors is assumed diagonal,

$$\mathbf{\Sigma}_t^f = \begin{bmatrix} \psi_{1,t}^2 & 0 & 0 & \dots & 0 \\ 0 & \psi_{2,t}^2 & 0 & \dots & 0 \\ 0 & 0 & \psi_{3,t}^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \psi_{l,t}^2 \end{bmatrix}, \tag{9.50}$$

and the conditional variance of each factor follows a GARCH(1,1) process (other models possible)

$$\begin{aligned} \psi_{i,t}^2 &= \varphi_i + \alpha_i f_{i,t-1}^2 + \beta_i \psi_{i,t-1}^2 & (9.51) \\ &= \varphi_i + \alpha_i \psi_{i,t-1}^2 e_{t,t-1}^2 + \beta_i \psi_{i,t-1}^2 & (9.52) \end{aligned}$$

where $\{\mathbf{e}_t\}$ are a sequence of i.i.d. innovations with mean $\mathbf{0}$ and covariance $\mathbf{I}_k$, such as a standard multivariate normal or possibly a heavy tailed distribution.

The conditional covariance of the residuals is assumed to be constant and diagonal,

$$\mathbf{\Omega} = \begin{bmatrix} \omega_1^2 & 0 & 0 & \dots & 0 \\ 0 & \omega_2^2 & 0 & \dots & 0 \\ 0 & 0 & \omega_3^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \omega_l^2 \end{bmatrix}, \tag{9.53}$$

where each variance is estimated using the residuals from an $p$ factor model,

$$\omega_i^2 = \sum_{t=1}^{T} \eta_{i,t}^2 = \sum_{t=1}^{T} \left(\epsilon_{i,t} - \mathbf{f}_t \boldsymbol{\beta}_i\right)^2. \tag{9.54}$$

Variants of the standard OGARCH model include parameterizations where the number of factors is equal to the number of assets, and so $\mathbf{\Omega} = \mathbf{0}$, and a specification which replaces $\mathbf{\Omega}$ with $\mathbf{\Omega}_t$ where each $\omega_{i,t}^2$ follows a univariate GARCH process.

### 9.4.7 Conditional Asymmetries

Standard multivariate ARCH models are symmetric in the sense that the news impact curves are symmetric for $\boldsymbol{\epsilon}_t$ and $-\boldsymbol{\epsilon}_t$ since they only depend on the outer product of returns. Most models can be modified in a simple way to allow for conditional asymmetries in covariance which may be important when modeling equity returns. Define $\boldsymbol{\zeta}_t = \boldsymbol{\epsilon}_t \odot I_{[\boldsymbol{\epsilon}_t<0]}$ where $I_{[\boldsymbol{\epsilon}_t<0]}$ is a $k$ by 1 vector of indicator variables where the i$^{\text{th}}$position is 1 if $r_{i,t} < 0$. An asymmetric BEKK model can be constructed as

$$\mathbf{\Sigma}_t = \mathbf{C}\mathbf{C}' + \mathbf{A}\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1}'\mathbf{A}' + \mathbf{G}\boldsymbol{\zeta}_{t-1}\boldsymbol{\zeta}_{t-1}'\mathbf{G}' + \mathbf{B}\mathbf{\Sigma}_{t-1}\mathbf{B}' \tag{9.55}$$

where $\mathbf{G}$ is a $k$ by $k$ matrix of parameters which control the covariance response the "bad" news, and when $k = 1$ this model reduces the a GJR-GARCH(1,1,1) model for variance. Diagonal and scalar BEKK models can be similarly adapted.

An asymmetric version of Matrix GARCH can be constructed in a similar manner,

$$\mathbf{\Sigma}_t = \mathbf{C}\mathbf{C}' + \mathbf{A}\mathbf{A}' \odot \boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1}' + \mathbf{G}\mathbf{G}' \odot \boldsymbol{\zeta}_{t-1}\boldsymbol{\zeta}_{t-1}' + \mathbf{B}\mathbf{B}' \odot \mathbf{\Sigma}_{t-1} \tag{9.56}$$

where $\mathbf{G}$ is a lower triangular parameter matrix. The dynamics of the covariances in the asymmetric Matrix GARCH process are given by

$$\sigma_{ij,t} = \tilde{c}_{ij} + \tilde{a}_{ij} r_{i,t-1} r_{j,t-1} + \tilde{g}_{ij} r_{i,t-1} r_{j,t-1} I_{i,t-1} I_{j,t-1} + \tilde{b}_{ij} \sigma_{ij,-t1}$$

where $\tilde{c}_{ij}$ is the ij$^{\text{th}}$element of $\mathbf{C}\mathbf{C}'$ and $\tilde{a}_{ij}$, $\tilde{g}_{ij}$ and $\tilde{b}_{ij}$ are similarly defined. All conditional variances follow GJR-GARCH(1,1,1) models and covariances evolve using similar dynamics only being driven by cross products of returns. The asymmetric term has a slightly different interpretation for covariance since it is only relevant when both indicators are 1 which only occurs if both market experience "bad" news (negative returns). An asymmetric DCC model has been proposed in Cappiello, R. F. Engle, and Sheppard (2006).

### 9.4.8 Fitting Multivariate GARCH Models

Returns are typically assumed to be conditionally multivariate normal, and so model parameters are typically estimated by maximizing the corresponding likelihood function,

$$f(\boldsymbol{\epsilon}_t; \boldsymbol{\theta}) = (2\pi)^{-\frac{k}{2}} \left|\mathbf{\Sigma}_t\right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\boldsymbol{\epsilon}_t'\mathbf{\Sigma}_t^{-1}\boldsymbol{\epsilon}_t\right) \tag{9.57}$$

where $\boldsymbol{\theta}$ contains the collection of parameters in the model. Estimation is, in principle, a simple problem. In practice parameter estimation is only straight-forward when the number of assets, $k$ is relatively small (less than 10) or when the model is tightly parameterized (e.g. scalar BEKK). The problems in estimation arise for two reasons. First, the likelihood is relatively flat and so finding its maximum value is difficult for optimization software. Second, the computational cost of calculating the likelihood is increasing in the number of unknown parameters and typically grows at rate $k^3$ in multivariate ARCH models.

A number of models have been designed to use multi-stage estimation to avoid these problems, including:

- *Covariance Targeting BEKK*: The intercept is concentrated out using the sample covariance of returns, and so only the parameters governing the dynamics of the conditional covariance need to be estimated using numerical methods.

**CCC GARCH Correlation**

| | FSLCX | OAKMX | WHOSX |
|---|---|---|---|
| FSLCX | 1.000 | 0.775 | -0.169 |
| OAKMX | 0.775 | 1.000 | -0.163 |
| WHOSX | -0.169 | -0.163 | 1.000 |

Table 9.3: Conditional correlation form a CCC GARCH model for three mutual funds spanning small cap stocks (FSLCX), large cap stocks (OAKMX), and long government bond returns (WHOSX).

- *Constant Conditional Correlation*: Fitting a CCC GARCH involves fitting $k$ univariate GARCH models and then using a closed form estimator for the constant conditional correlation.
- *Dynamic Conditional Correlation*: Fitting a DCC GARCH combines the first stage of the CCC GARCH with correlation targeting similar to the covariance targeting BEKK.
- *Orthogonal GARCH*: Orthogonal GARCH only involves fitting $p \leq k$ univariate GARCH models and uses a closed form estimator for the idiosyncratic variance.

### 9.4.9 Application: Mutual Fund Returns

Three mutual funds were used to illustrate the differences (and similarities) of the dynamic covariance models. The three funds were:

- Oakmark I (OAKMX) - A broad large cap fund
- Fidelity Small Cap Stock (FSLCX) - A broad small cap fund which seeks to invest in firms with capitalizations similar to those in the Russell 2000 or S&P 600.
- Wasatch-Hoisington US Treasury (WHOSX) - A fund which invests at least 90% of total assets in U.S. Treasury securities and can vary the average duration of assets held from 1 to 25 years, depending on market conditions.

These funds are used to capture interesting aspects of the broad set of investment opportunities. All data was takes from CRSP's database from January 1, 1999 until December 31, 2008. Tables 9.3 contains the estimated correlation from the CCC-GARCH model where each volatility series was assumed to follow a standard GARCH(1,1). This shows that the correlations between these assets are large and positive for the equity fund and negative, on average, between the equity funds and the treasury fund. Table 9.4 contains the parameters of the dynamics for the DCC, scalar BEKK, an asymmetric scalar BEKK, Matrix GARCH and an asymmetric version of the Matrix GARCH. The parameters in the DCC are typical of DCC models – the parameters sum

**Multivariate GARCH Model Estimates**

| | $\alpha$ | $\gamma$ | $\beta$ |
|---|---|---|---|
| DCC | 0.025 (6.84) | – | 0.970 (207) |
| Scalar *vec* | 0.046 (11.6) | – | 0.950 (203) |
| Asymmetric Scalar *vec* | 0.043 (11.6) | 0.009 (2.37) | 0.948 (175) |

| | **AA′** | | | **GG′** | | | **BB′** | | |
|---|---|---|---|---|---|---|---|---|---|
| Matrix GARCH | 0.058 (8.68) | 0.060 (9.33) | 0.033 (8.68) | – | – | – | 0.931 (118) | 0.930 (130) | 0.9 (17 |
| | 0.060 (9.33) | 0.064 (8.75) | 0.035 (8.60) | – | – | – | 0.930 (130) | 0.929 (129) | 0.9 (18 |
| | 0.033 (8.68) | 0.035 (8.60) | 0.035 (5.72) | – | – | – | 0.945 (177) | 0.944 (180) | 0.9 (13 |
| Asymmetric Matrix GARCH | 0.034 (6.01) | 0.032 (6.57) | 0.035 (8.51) | 0.055 (4.89) | 0.060 (5.56) | −0.007 (−4.06) | 0.926 (80.3) | 0.927 (94.6) | 0.9 (13 |
| | 0.032 (6.57) | 0.029 (6.37) | 0.033 (8.44) | 0.060 (5.56) | 0.070 (5.79) | −0.007 (−4.48) | 0.927 (94.6) | 0.929 (107) | 0.9 (15 |
| | 0.035 (8.51) | 0.033 (8.44) | 0.036 (6.30) | −0.007 (−4.06) | −0.007 (−4.48) | 0.001 (2.30) | 0.942 (131) | 0.943 (156) | 0.9 (13 |

Table 9.4: Parameter estimates ($t$-stats in parenthesis) from a selection of multivariate ARCH models for three mutual funds spanning small cap stocks (FSLCX), large cap stocks (OAKMX), and long government bond returns (WHOSX). The top panel contains results for DCC, scalar $vec$ and asymmetric scalar $vec$. The bottom panel contains estimation results for Matrix GARCH and asymmetric version of the Matrix GARCH model, which shows large differences in asymmetries between equity and bond returns.

to nearly 1 and $\alpha$ is smaller than is typically found in univariate models. This indicates that correlation is very persistent but probably moves slower than volatility. The parameters in the scalar BEKK and asymmetric scalar BEKK are similar to what one would typically find in a univariate model, although the asymmetry is weak. The Matrix GARCH parameters are fairly homogeneous although the treasury fund is less responsive to news (smaller coefficient in **AA′**). The most interesting finding in this table is in the asymmetric Matrix GARCH model where the response to "bad" news is very different between the equity funds and the treasury fund. This heterogeneity is the likely the source of the small asymmetry parameter in the asymmetric scalar BEKK.

Figure 9.4 plots the annualized volatility for these series from 4 models: the CCC (standard GARCH(1,1)), the two RiskMetrics methodologies, and the asymmetric scalar BEKK. All volatilities broadly agree which may be surprising given the differences in the models. Figures 9.5, 9.6 and 9.7 plot the correlations as fit from 6 different models. Aside from the CCC GARCH fit (which is constant), all models broadly agree about the correlation dynamics in these series.

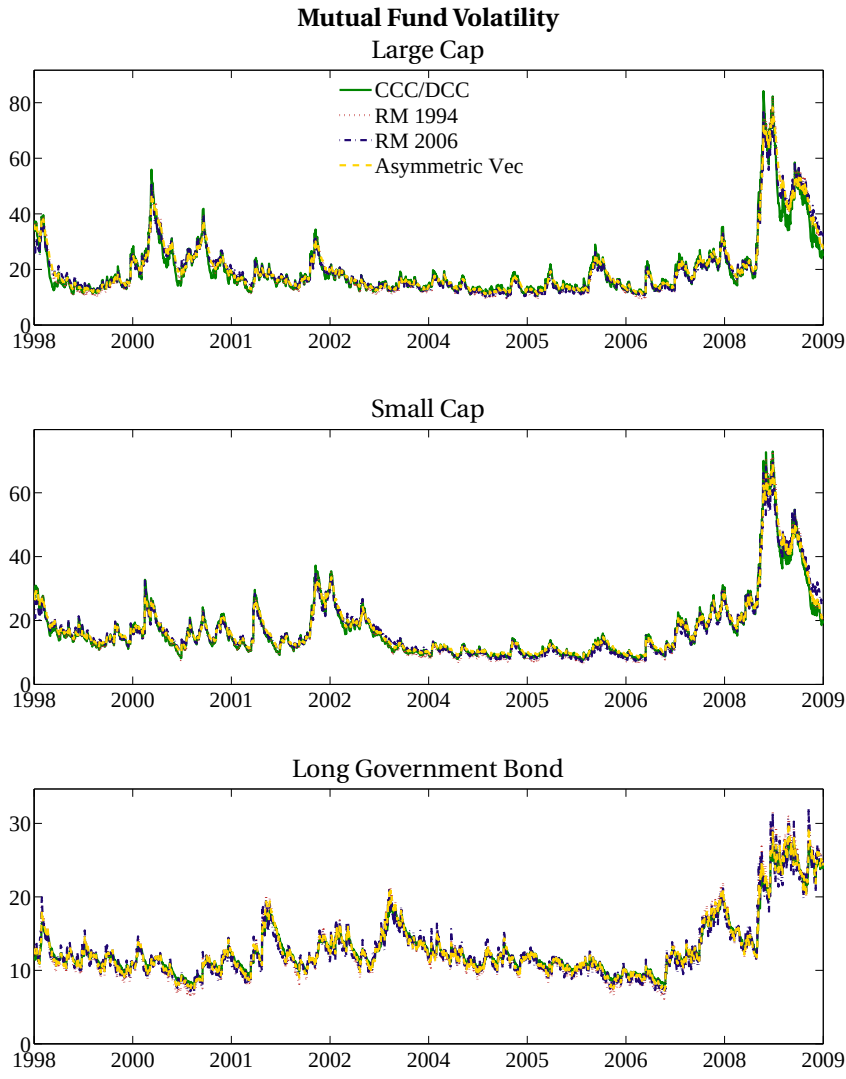


Figure 9.4: The three panels plot the volatility for the three mutual funds spanning small caps, large caps and long government bond returns. The volatilities from all of the models are qualitatively similar and the only visible differences occur when volatility is falling.

## 9.5 Realized Covariance

Realized covariance estimates the integrated covariance over some period, usually a day. Suppose prices followed a $k$-variate continuous time diffusion,

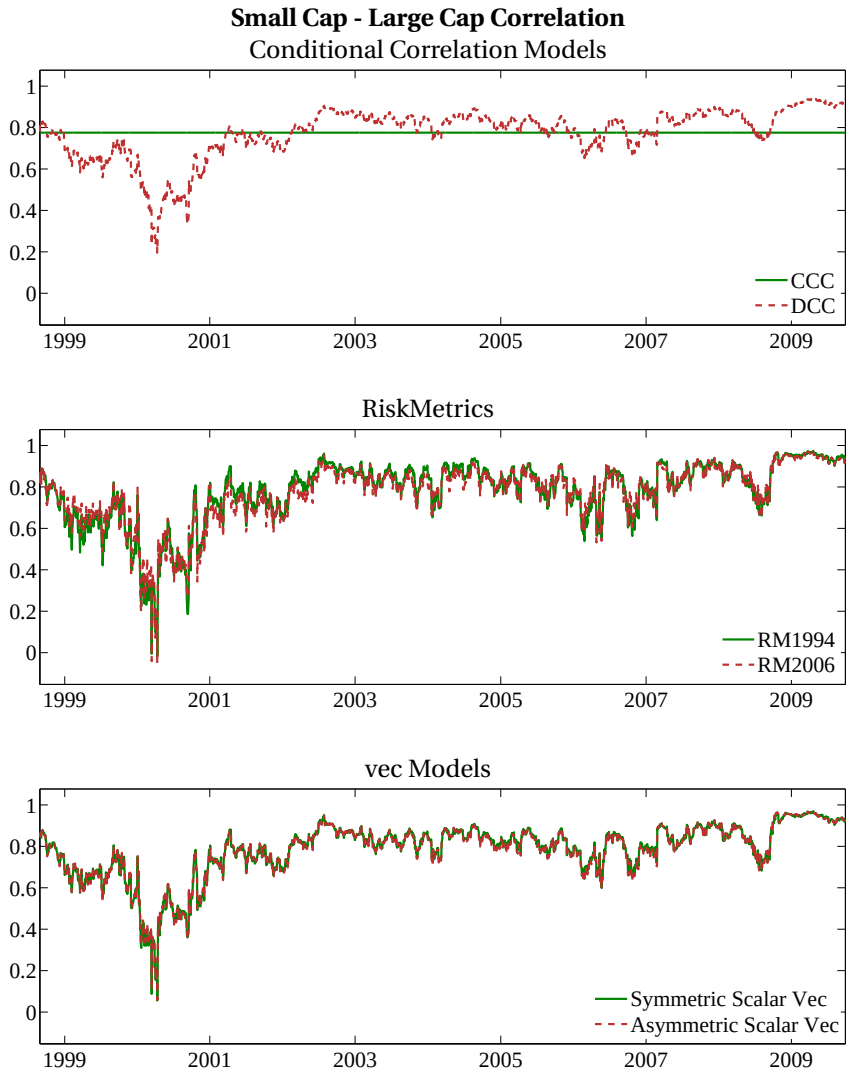


Figure 9.5: The three graphs plot the fit correlation from 6 models. The conditional correlation estimates are broadly similar, aside from the CCC GARCH which assumes that correlation is constant.

---

$$d\mathbf{p}_t = \boldsymbol{\mu}_t dt + \boldsymbol{\Omega}_t d\mathbf{W}_t$$

where $\boldsymbol{\mu}_t$ is the instantaneous drift, $\boldsymbol{\Sigma}_t = \boldsymbol{\Omega}_t \boldsymbol{\Omega}_t'$ is the instantaneous covariance, and $d\mathbf{W}_t$ is a $k$-variate Brownian motion. Realized covariance estimates

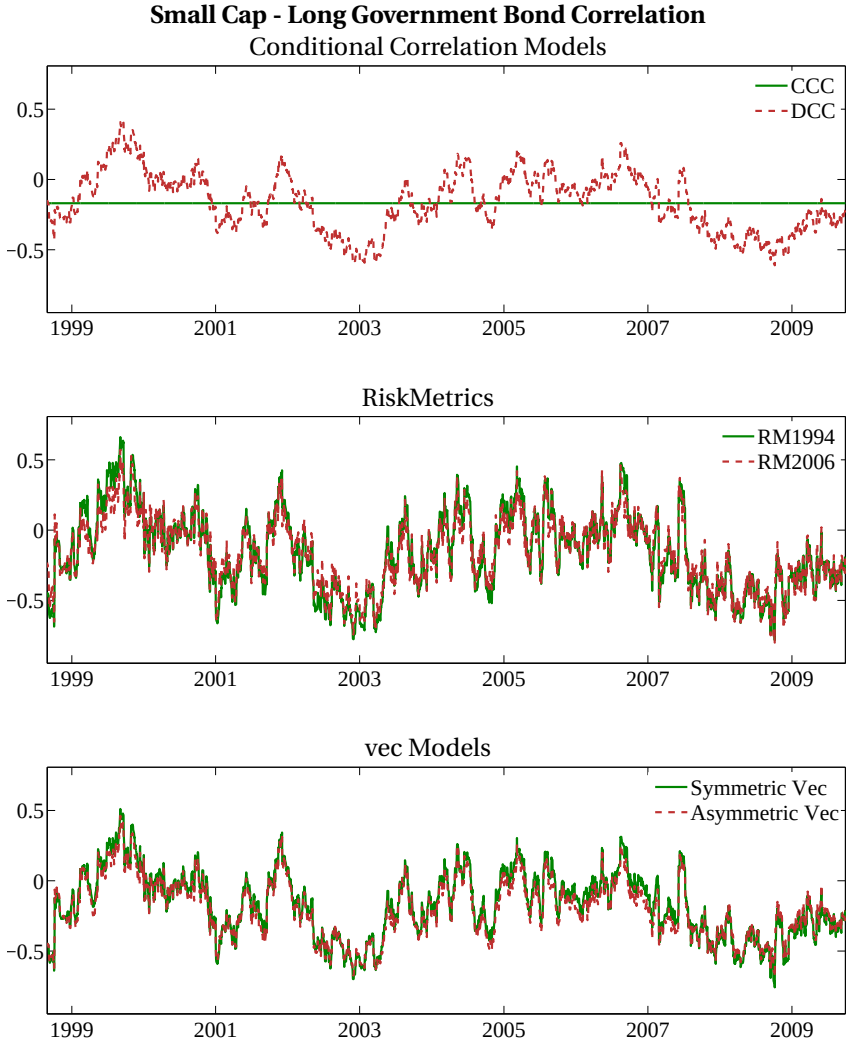


Figure 9.6: The three graphs plot the fit correlation from 6 models. The conditional correlation estimates are broadly similar, aside from the CCC GARCH which assumes that correlation is constant.

---

$$\int_0^1 \Sigma_s \, ds$$

where the bounds 0 and 1 represent the (arbitrary) interval over which the realized covariance is

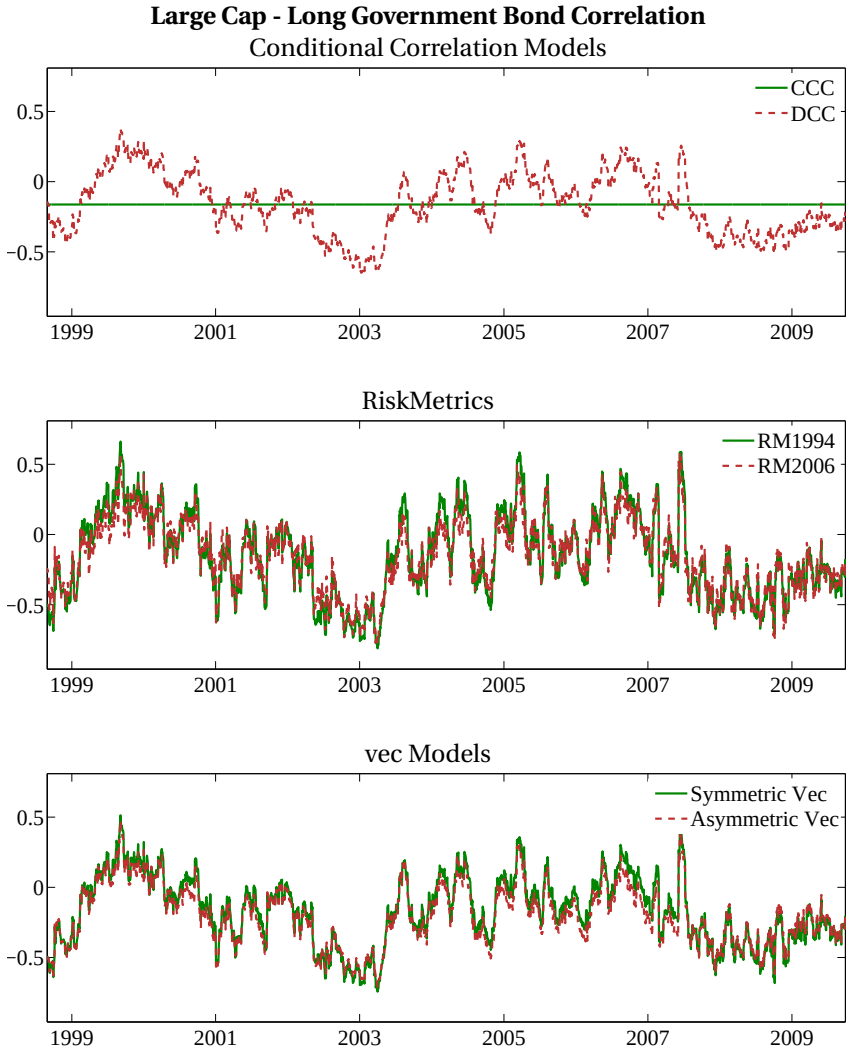


Figure 9.7: The three graphs plot the fit correlation from 6 models. The conditional correlation estimates are broadly similar, aside from the CCC GARCH which assumes that correlation is constant.

computed.[8]

Realized covariance is computed using the outer-product of high-frequency returns.

[8] In the presence of jumps, realized covariance estimates the quadratic covariation, which is the integrated covariance

**Definition 9.22** (Realized Covariance)**.** The realized covariance is defined

$$RC_t = \sum_{i=1}^{m} \mathbf{r}_{i,t}\mathbf{r}_{i,t}' = \left(\mathbf{p}_{i,t} - \mathbf{p}_{i-1,t}\right)\left(\mathbf{p}_{i,t} - \mathbf{p}_{i-1,t}\right)', \tag{9.58}$$

where $\mathbf{r}_{i,t}$ is the i$^{\text{th}}$return on day $t$.

In principle prices should be sampled as frequently as possible to maximize the precision of the realized covariance estimator. In practice this is not possible since:

- Prices, especially transaction prices (trades), are contaminated by noise (e.g. bid-ask bounce).
- Prices are not perfectly synchronized. For example, asset $i$ might trade at 10:00:00 while the most recent trade of asset $j$ might have occurred at 9:59:50. The lack of synchronization will bias the covariance between the two assets toward 0.

The standard method to address these two concerns is to sample relatively infrequently, for example every 5 minutes. An improved method is to use a modified realized covariance estimator which uses *subsampling*. Suppose returns were computed every minute, but that microstructure concerns (noise and synchronization) do not allow for sampling more frequently than every 10 minutes. The subsampled realized covariance uses *all* 10-minute returns, not just non-overlapping ones, to estimate the covariance.

**Definition 9.23** (Subsampled Realized Covariance)**.** The subsampled realized covariance estimator is defined

$$\begin{aligned} RC_t^{SS} &= \frac{m}{n(m-n+1)} \sum_{i=1}^{m-n+1} \sum_{j=1}^{n} \mathbf{r}_{i+j-1,t}\mathbf{r}_{i+j-1,t}' \\ &= \frac{1}{n}\sum_{j=1}^{n} \frac{m}{(m-n+1)} \sum_{i=1}^{m-n+1} \mathbf{r}_{i+j-1,t}\mathbf{r}_{i+j-1,t}' \\ &= \frac{1}{n}\sum_{j=1}^{n} \widetilde{RC}_{j,t}, \end{aligned} \tag{9.59}$$

where there are $m$ high-frequency returns available and the selected sampling time is based on $n$ returns.

For example, suppose data was available from 9:30:00 to 16:00:00, and that prices were sampled every minute. The standard realized covariance would compute returns using prices at 9:30:00,

plus the outer product of the jumps

$$\int_0^1 \mathbf{\Sigma}_s ds + \sum_{0 \le s \le 1} \Delta p_s \Delta p_s',$$

where $\Delta p_s$ are the jumps.

9:40:00, 9:50:00, . . .. The subsampled realized covariance would compute returns using all 10 minute windows, e.g. 9:30:00 and 9:40:00, 9:31:00 and 9:41:00, 9:32:00 and 9:42:00, and so on. In this example $m$, the number of returns available over a 6.5 hour day is 390 and $n$, the number of returns used in the desired sampled window, is 10.

Barndorff-Nielsen, P. R. Hansen, Lunde, and Shephard (2011) recently proposed an alternative method to compute the realized covariance known as a *realized kernel*. It is superficially similar to realized covariance except that realized kernels use a weighting function similar to that in a Newey and West (1987) covariance estimator.

**Definition 9.24** (Realized Kernel)**.** The realized kernel is defined as

$$RK_t = \mathbf{\Gamma}_0 + \sum_{i=1}^{h} K\left(\frac{i}{H+1}\right)\left(\mathbf{\Gamma}_i + \mathbf{\Gamma}_i'\right) \tag{9.60}$$

$$\mathbf{\Gamma}_j = \sum_{i=j+1}^{\tilde{m}} \tilde{\mathbf{r}}_{i,t}\tilde{\mathbf{r}}_{i-j,t}$$

where $\tilde{\mathbf{r}}$ are refresh time returns, $\tilde{m}$ is the number of refresh time returns, $K(\cdot)$ is a kernel weighting function and $H$ is a parameter which controls the bandwidth.

Refresh time returns are needed to ensure that prices are not overly stale, and are computed by sampling all prices using last-price interpolation only when all assets have traded have traded. For example, consider the transactions in table 9.5 which contains a hypothetical series of trade times for MSFT and IBM where ✓ indicates a trade with the time stamp indicated on the left. A ✓ in the refresh column indicates that this time stamp is a refresh time. The final two columns indicate the time stamp of the price which would be used for MSFT and IBM when computing the refresh-time return.

The recommended kernel is Parzen's kernel,

$$K(x) = \begin{cases} 1 - 6x^2 + 6x^3 & 0 > x \geq \frac{1}{2} \\ 2(1-x)^3 & \frac{1}{2} > x \geq 1 \\ 0 & x > 1 \end{cases} \tag{9.61}$$

Selection of the bandwidth parameter, $H$, is an important choice when implementing realized Kernels, although a discussion is the choices needed to correctly determine the bandwidth is beyond the scope of these notes. See Barndorff-Nielsen, P. R. Hansen, et al. (2008)and Barndorff-Nielsen, P. R. Hansen, Lunde, and Shephard (2011) for detailed discussions.

### 9.5.1 Realized Correlation and Beta

Realized Correlation is the realized analogue of the usual correlation estimator, only defined in terms of realized covariance.

| Trade Time | MSFT | IBM | Refresh | MSFT Time | IBM Time |
|---|---|---|---|---|---|
| 9:30:00 | ✓ | ✓ | ✓ | 9:30:00 | 9:30:00 |
| 9:30:01 | ✓ | ✓ | ✓ | 9:30:01 | 9:30:01 |
| 9:30:02 | | | | | |
| 9:30:03 | ✓ | | | | |
| 9:30:04 | ✓ | | | | |
| 9:30:05 | | ✓ | ✓ | 9:30:04 | 9:30:05 |
| 9:30:06 | | ✓ | | | |
| 9:30:07 | | ✓ | | | |
| 9:30:08 | ✓ | | ✓ | 9:30:08 | 9:30:07 |

Table 9.5: This table illustrated the concept of refresh-time sampling. Prices are sampled every time all assets have traded using last-price interpolation. Refresh-time sampling usually eliminated some of the data, as with the 9:30:03 trade of MSFT, and produces some sampling points where the prices are not perfectly synchronized, as with the 9:30:08 refresh-time price which uses a MSFT price from 9:30:08 and an IBM price from 9:30:07.

**Definition 9.25** (Realized Correlation)**.** The realized correlation between two series is defined

$$RCorr = \frac{RC_{ij}}{\sqrt{RC_{ii}RC_{jj}}}$$

where $RC_{ij}$is the realized covariance between assets i and j and $RC_{ii}$ and $RC_{jj}$are the realized variances of assets i and j, respectively.

Realized Betas are similarly defined, only using definition of $\beta$ (which is a function of the covariance).

**Definition 9.26** (Realized Beta)**.** Suppose $RC_t$ is a $k+1$ by $k+1$ realized covariance matrix for an asset and a set of observable factors where the asset is in position 1, so that the realized covariance can be partitioned

$$RC = \begin{bmatrix} RV_i & RC'_{fi} \\ RC_{fi} & RC_{f,f} \end{bmatrix}$$

where $RV_{i,i}$ is the realized variance of the asset being studied, $RC_{if}$ is the $k$ by 1 vector of realized covariance between the asset and the factors, and $RC_{ff}$ is the $k$ by $k$ covariance of the factors. The Realized Beta is defined

$$R\beta = RC_{ff}^{-1}RC_{fi}.$$

In the usual case where there is only one factor, usually the market, the realized beta is the ratio of the realized covariance between the asset and the market to the variance of the market. Realized Betas are similar to other realized measures in that they are model free and, as long as prices can be sampled frequently and have little market microstructure noise, is an accurate

measure of the current exposure to changes in the market.

### 9.5.2 Modeling Realized Covariance

Modeling realized covariance and realized kernels can be accomplished by modifying standard multivariate ARCH models. The basic assumption is that the *mean* of the realized covariance, conditional on the time $t-1$ information, is $\mathbf{\Sigma}_t$,

$$RC_t|\mathcal{F}_{t-1} \sim F(\mathbf{\Sigma}_t, \boldsymbol{\upsilon})$$

where $F(\cdot,\cdot)$ is some distribution with conditional mean $\mathbf{\Sigma}_t$ which may depend on other parameters unrelated to the mean which are contained in $\boldsymbol{\upsilon}$. This assumption implies that the realized covariance is driven by a matrix-values shock which has conditional expectation $\mathbf{I}_k$, and so

$$RC_t = \mathbf{\Sigma}_t^{\frac{1}{2}} \mathbf{\Xi} \mathbf{\Sigma}_t^{\frac{1}{2}}$$

where $\mathbf{\Xi} \overset{\text{i.i.d.}}{\sim} F(\mathbf{I}, \tilde{\boldsymbol{\upsilon}})$ and $\tilde{\boldsymbol{\upsilon}}$ is used to denote that these parameters are related to the original parameters although will generally be different. This assumption is identical to the one made when modeling realized variance as a non-negative process with a multiplicative error (MEM) where it is assumed that $RV_t = \sigma_t^2 \xi_t = \sigma_t \xi_t \sigma_t$ where $\xi_t \overset{\text{i.i.d.}}{\sim} F(1, \boldsymbol{\upsilon})$.

With this assumption most multivariate ARCH models can be used. Consider the standard BEKK model,

$$\mathbf{\Sigma}_t = \mathbf{C}\mathbf{C}' + \mathbf{A}\mathbf{r}_{t-1}\mathbf{r}_{t-1}\mathbf{A}' + \mathbf{B}\mathbf{\Sigma}_{t-1}\mathbf{B}'.$$

The BEKK can be viewed as a multiplicative error model and used for realized covariance by specifying the dynamics as

$$\mathbf{\Sigma}_t = \mathbf{C}\mathbf{C}' + \mathbf{A}RC_{t-1}\mathbf{A}' + \mathbf{B}\mathbf{\Sigma}_{t-1}\mathbf{B}'.$$

Other ARCH models can be similarly adapted by replacing the outer product of returns by the realized covariance or realized kernel. Estimation is not more difficult than the estimation of a multivariate ARCH model since parameters can be estimated using the Wishart likelihood. See Noureldin, Shephard, and Sheppard (2012) for details.

## 9.6 Measuring Dependence

Covariance modeling is the only the first step to understanding risk in portfolios since covariance (and correlation) is only one measure of dependence, and often lacking in many applications.

### 9.6.1 Linear Dependence

Linear or Pearson correlation is the most common measure of dependence.

**Definition 9.27** (Linear (Pearson) Correlation)**.** The linear (Pearson) correlation between two random variables $X$ and $Y$ is

$$\rho = \frac{\text{Cov}[X, Y]}{\sqrt{\text{V}[X]\text{V}[Y]}}. \tag{9.62}$$

The sample correlation is estimated using

$$\hat{\rho} = \frac{\sum_{t=1}^{T}(x_t - \hat{\mu}_x)\left(y_t - \hat{\mu}_y\right)}{\sqrt{\sum_{t=1}^{T}(x_s - \hat{\mu}_x)^2 \sum_{s=1}^{T}\left(y_t - \hat{\mu}_y\right)^2}}. \tag{9.63}$$

where $\hat{\mu}_x$ and $\hat{\mu}_y$ are the sample mean of $x_t$ and $y_t$.

Linear correlation measures the strength of the linear relationship between standardized versions of $X$ and $Y$. Correlation is obviously invariant to affine transformations of $X$ and/or $Y$ (e.g. $a + bX$). It is not, however, invariant to non-linear transformations, even when the non-linear transformation is order preserving (e.g. the log of a non-negative random variable). Linear correlation is also insufficient to characterize the dependence between two random variables, except in the special case where $X$ and $Y$ are follow a bivariate normal distribution. Moreover, two distributions can have the same correlation yet exhibit very different characteristics with respect to the amount of diversification available.

## 9.6.2 Non-linear Dependence

A number of measures have been designed to overcome some of the shortcomings on correlation as a measure of risk. These are broadly classified as measures of non-linear dependence.

### 9.6.2.1 Rank Correlation

Rank correlation, also known as Spearman correlation, is an alternative measure of dependence which can assess the strength of a relationship and is robust to certain non-linear transformations. Suppose $X$ and $Y$ are random variables, $X \overset{\text{i.i.d.}}{\sim} N(0,1)$ and $y \equiv x^\lambda$ where $\lambda$ is odd. If $\lambda = 1$ then $y = x$ and the linear correlation is 1. If $\lambda = 3$ the correlation is .77. If $\lambda = 5$ then the correlation is only .48, despite $y$ being a function of only $x$. As $\lambda$ becomes increasing large the correlation becomes arbitrarily small despite the one-to-one relationship between $X$ and $Y$. Rank correlation is robust to non-linear transformations and so will return a correlation on 1 between $X$ and $Y$ for any power $\lambda$.

**Definition 9.28** (Rank (Spearman) Correlation)**.** The rank (Spearman) correlation between two random variables $X$ and $Y$ is

$$\rho_s(X, Y) = \text{Corr}\left(F_X(X), F_Y(Y)\right) = \frac{\text{Cov}[F_X(X), F_Y(Y)]}{\sqrt{\text{V}[F_X(X)]\text{V}[F_Y(Y)]}} = 12\text{Cov}[F_X(X), F_Y(Y)] \tag{9.64}$$

where the final identity uses the fact that the variance of a uniform (0,1) is $\frac{1}{12}$.

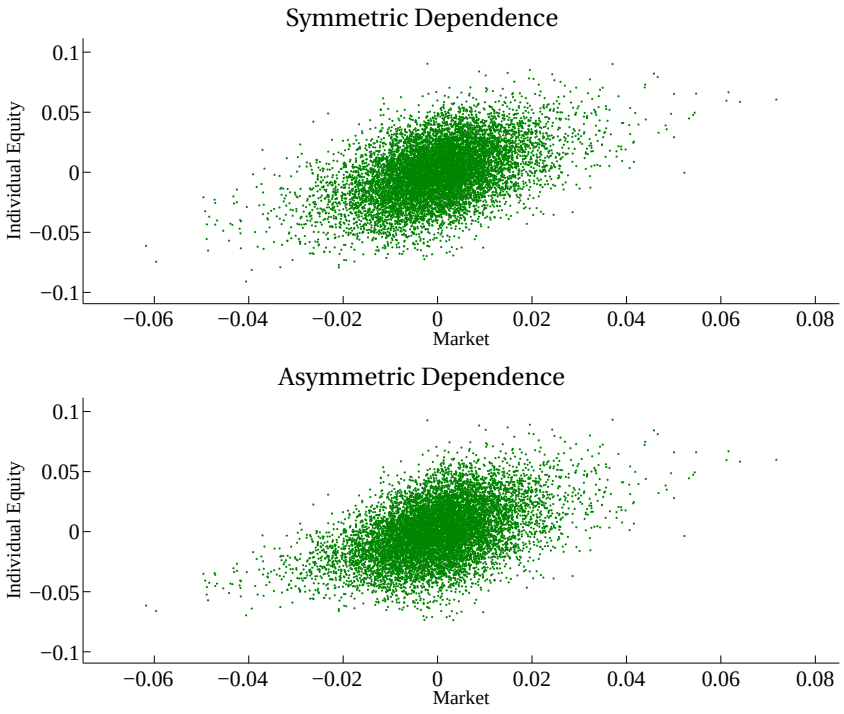


Figure 9.8: These graphs illustrate simulated returns from a CAP-M where the market has a $t_6$ distribution with the same variance as the S&P 500 and the idiosyncratic shock is normal with the same variance as the average idiosyncratic shock in the S&P 500 constituents. The asymmetric dependence was introduced by making the idiosyncratic error heteroskedastic by defining its variance to be $\sigma_\epsilon \exp\left(10 r_m I_{[r_m<0]}\right)$, and so the idiosyncratic component has a smaller variance when the market return is negative then it does when the market return is positive.

---

The rank correlation measures the correlation between the *probability integral transforms* of $X$ and $Y$. The use of the probability integral transform means that rank correlation is preserved under strictly increasing transformations (decreasing monotonic changes the sign), and so $\rho_s(X, Y) = \rho_s(T_1(X), T_2(Y))$ when $T_1$ and $T_2$ are any strictly increasing functions.

The sample analogue of the Spearman correlation makes use of the empirical ranks of the observed data. Define $r_{x,i}$ to be the rank of $x_i$, where a rank of 1 corresponds to the smallest value, a rank of $n$ corresponds to the largest value, where any ties are all assigned the average value of the ranks associated with the values in the tied group. Define $r_{y,i}$ in an identical fashion on $y_i$. The sample rank correlation between $X$ and $Y$ is computed as the sample correlation of the ranks,

$$\rho = \frac{\sum_{i=1}^{n}\left(\frac{r_{x,i}}{n+1}-\frac{1}{2}\right)\left(\frac{r_{y,i}}{n+1}-\frac{1}{2}\right)}{\sqrt{\sum_{i=1}^{n}\left(\frac{r_{x,i}}{n+1}-\frac{1}{2}\right)^2}\sqrt{\sum_{j=1}^{n}\left(\frac{r_{y,i}}{n+1}-\frac{1}{2}\right)^2}} = \frac{\sum_{i=1}^{n}\left(r_{x,i}-\frac{n+1}{2}\right)\left(r_{y,i}-\frac{n+1}{2}\right)}{\sqrt{\sum_{i=1}^{n}\left(r_{x,i}-\frac{n+1}{2}\right)^2}\sqrt{\sum_{j=1}^{n}\left(r_{y,j}-\frac{n+1}{2}\right)^2}}$$

where $\frac{r_{x,i}}{n+1}$ is the empirical quantile of $x_i$. These two formulations are identical since the arguments in the second are linear transformations of the arguments in the first, and linear correlation is robust to linear transformations.

#### 9.6.2.2 Kendall's $\tau$

Kendall's $\tau$ is another measure of non-linear dependence which is based on the idea of concordance. Concordance is defined with respect to differences in the signs of pairs of random variables.

**Definition 9.29** (Concordant Pair)**.** The pairs of random variables $(x_i, y_i)$ and $(x_j, y_j)$ are concordant if $\text{sgn}(x_i - x_j) = \text{sgn}(y_i - y_j)$ where $\text{sgn}(\cdot)$ is the sign function which returns -1 for negative values, 0 for zero, and +1 for positive values (equivalently defined if $(x_i - x_j)(y_i - y_j) > 0$).

If a pair is not concordant then it is *discordant*.

**Definition 9.30** (Kendall's $\tau$)**.** Kendall $\tau$ is defined

$$\tau = \Pr(\text{sgn}(x_i - x_j) = \text{sgn}(y_i - y_j)) - \Pr(\text{sgn}(x_i - x_j) \neq \text{sgn}(y_i - y_j)) \tag{9.65}$$

The estimator of Kendall's $\tau$ uses the obvious sample analogues to the probabilities in the definition. Defined $n_c = \sum_{i=1}^{n}\sum_{j=i+1}^{n} I_{[\text{sgn}(x_i - x_j)=\text{sgn}(y_i - y_j)]}$ as the count of the concordant pairs and $n_d = \frac{1}{2}n(n-1) - n_c$ as the count of discordant pairs. The estimator of $\tau$ is

$$\begin{aligned}
\tau &= \frac{n_c - n_d}{\frac{1}{2}n(n-1)} && (9.66)\\
&= \frac{n_c}{\frac{1}{2}n(n-1)} - \frac{n_d}{\frac{1}{2}n(n-1)} && (9.67)\\
&= \widehat{\Pr}(\text{sgn}(x_i - x_j) = \text{sgn}(y_i - y_j)) - \widehat{\Pr}(\text{sgn}(x_i - x_j) \neq \text{sgn}(y_i - y_j)) && (9.68)
\end{aligned}$$

where $\widehat{\Pr}$ denotes the empirical probability. Kendall's $\tau$ measures the difference between the probability a pair is concordant, $n_c/(\frac{1}{2}n(n-1))$ and the probability a pair is discordant $n_d/(\frac{1}{2}n(n-1))$. Since $\tau$ is the difference between two probabilities it must fall in $[-1, 1]$ where -1 indicates that all pairs are discordant, 1 indicates that all pairs are concordant, and $\tau$ is increasing as the concordance between the pairs increases. Like rank correlation, Kendall's $\tau$ is also invariant to increasing transformation since a pair that was concordant before the transformation (e.g. $x_i > x_j$ and $y_i > y_j$) must also be concordant after a strictly increasing transformation (e.g. $T_1(x_i) > T_1(x_j)$ and $T_2(y_i) > T_2(y_j)$).

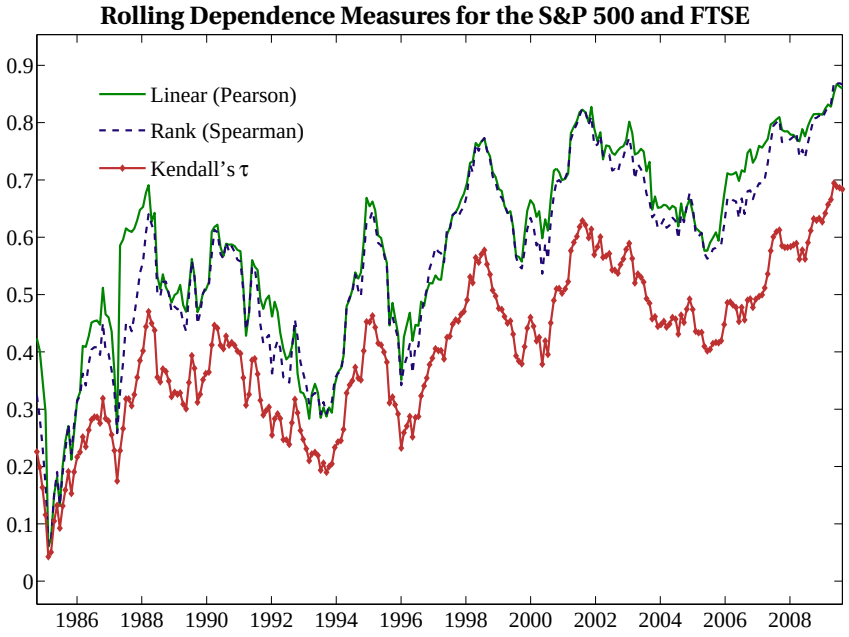


Figure 9.9: Plot of rolling linear correlation, rank correlation and Kendall's $\tau$ between weekly returns on the S&P 500 and the FTSE estimated using 252-day moving windows. The measures broadly agree about the changes in dependence but not the level.

**Dependence Measures for Weekly FTSE and S&P 500 Returns**

| | | | | | |
|---|---|---|---|---|---|
| Linear (Pearson) | 0.660 (0.028) | Rank (Spearman) | 0.593 (0.027) | Kendall's $\tau$ | 0.429 (0.022) |

Table 9.6: Linear and rank correlation and Kendall's $\tau$ (bootstrap std. error in parenthesis) for weekly returns for the S&P 500 and FTSE 100. Weekly returns were used to minimize problems with of non-synchronous closings in these two markets.

#### 9.6.2.3 Exceedance Correlations and Betas

Exceedance correlation, like expected shortfall, is one of many exceedance measures which can be constructed by computing expected values conditional on exceeding some threshold. Exceedance correlation measures the correlation between two variables *conditional* on both variable taking values in either the upper tail or lower tail.

**Definition 9.31** (Exceedance Correlation)**.** The exceedance correlation at level $\kappa$ is defined as

$$\rho^{+}(\kappa) = \text{Corr}\left(x, y | x > \kappa, y > \kappa\right) \tag{9.69}$$

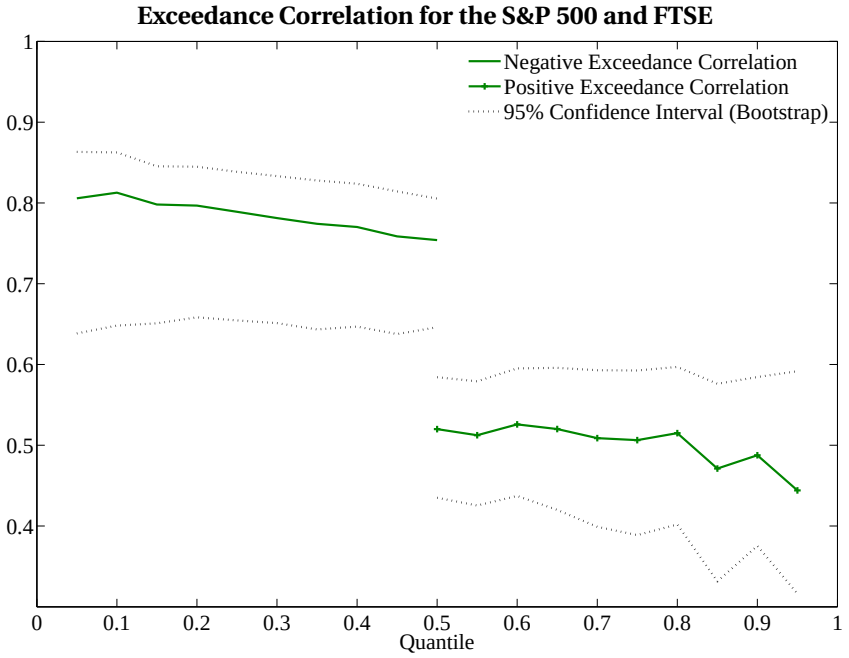


Figure 9.10: Plot of the exceedance correlations with 95% bootstrap confidence intervals for weekly returns on the S&P 500 and FTSE (each series was divided by its sample standard deviation). There is substantial asymmetry between positive and negative exceedance correlation.

---

$$\rho^{-}(\kappa) = \text{Corr}\left(x, y | x < -\kappa, y < -\kappa\right) \tag{9.70}$$

Exceedance correlation are computed using the standard (linear) correlation estimator on the subset of data where both $x > \kappa$ and $y > \kappa$ (positive) or $x < -\kappa$ and $y < -\kappa$. Exceedance correlation can also be defined using series specific cutoff points such as $\kappa_x$ and $\kappa_y$, which are often used if the series doe not have the same variance, and are often set using quantiles of $x$ and $y$ (e.g. the 10% quantile of each). Alternatively exceedance correlations can be computed with data transformed to have unit variance. Sample exceedance correlations are computed as

$$\hat{\rho}^{+}(\kappa) = \frac{\hat{\sigma}^{+}_{xy}(\kappa)}{\hat{\sigma}^{+}_{x}(\kappa)\hat{\sigma}^{+}_{y}(\kappa)}, \quad \hat{\rho}^{-}(\kappa) = \frac{\hat{\sigma}^{-}_{xy}(\kappa)}{\hat{\sigma}^{-}_{x}(\kappa)\hat{\sigma}^{-}_{y}(\kappa)} \tag{9.71}$$

where

$$\hat{\sigma}^{+}_{xy}(\kappa) = \frac{\sum_{t=1}^{T}\left(x_t - \mu^{+}_{x}(\kappa)\right)\left(y_t - \mu^{+}_{y}(\kappa)\right) I_{[x_t > \kappa \cap y_t > \kappa]}}{T^{+}_{\kappa}}$$

$$\hat{\sigma}_{xy}^{-}(\kappa) = \frac{\sum_{t=1}^{T}\left(x_t - \mu_x^{-}(\kappa)\right)\left(y_t - \mu_y^{-}(\kappa)\right) I_{[x_t<-\kappa \cap y_t<-\kappa]}}{T_\kappa^{-}}$$

$$\hat{\mu}_x^{+}(\kappa) = \frac{\sum_{t=1}^{t} x_t I_{[x_t>\kappa \cap y_t>\kappa]}}{T_\kappa^{+}}, \quad \hat{\sigma}_x^{2+}(\kappa) = \frac{\sum_{t=1}^{t}\left(x_t - \hat{\mu}_x^{+}(\kappa)\right)^2 I_{[x_t>\kappa \cap y_t>\kappa]}}{T_\kappa^{+}}$$

$$\hat{\mu}_x^{-}(\kappa) = \frac{\sum_{t=1}^{t} x_t I_{[x_t<-\kappa \cap y_t<-\kappa]}}{T_\kappa^{-}}, \quad \hat{\sigma}_x^{2-}(\kappa) = \frac{\sum_{t=1}^{t}\left(x_t - \hat{\mu}_x^{-}(\kappa)\right)^2 I_{[x_t<-\kappa \cap y_t<-\kappa]}}{T_\kappa^{-}}$$

$$T_\kappa^{+} = \sum_{t=1}^{T} I_{[x_t>\kappa \cap y_t>\kappa]}, \quad T_\kappa^{-} = \sum_{t=1}^{T} I_{[x_t<-\kappa \cap y_t<-\kappa]}$$

where the quantities for $y$ are similarly defined. Exceedance correlation can only be estimated if the where $x < \kappa$ and $y < \kappa$ is populated with data, and it is possible for some assets that this region will empty (e.g. if the assets have strong negative dependence such as with equity and bond returns).

Inference can be conducted using the bootstrap or using analytical methods. Hong, Tu, and Zhou (2007) show that inference on exceedance correlations can be conducted by viewing these estimators as method of moments estimators. Define the standardized exceedance residuals as,

$$\tilde{x}_t^{+}(\kappa) = \frac{x_t - \mu_x^{+}(\kappa)}{\sigma_x^{+}(\kappa)}$$

$$\tilde{x}_t^{-}(\kappa) = \frac{x_t - \mu_x^{-}(\kappa)}{\sigma_x^{-}(\kappa)}$$

$$\tilde{y}_t^{+}(\kappa) = \frac{y_t - \mu_y^{+}(\kappa)}{\sigma_y^{+}(\kappa)}$$

$$\tilde{y}_t^{-}(\kappa) = \frac{y_t - \mu_y^{-}(\kappa)}{\sigma_y^{-}(\kappa)}.$$

These form the basis of the moment conditions,

$$\frac{T}{T_\kappa^{+}}\left(\tilde{x}^{+}(\kappa)\tilde{y}^{+}(\kappa) - \rho^{+}(\kappa)\right) I_{[x_t>\kappa \cap y_t>\kappa]} \tag{9.72}$$

$$\frac{T}{T_\kappa^{-}}\left(\tilde{x}^{-}(\kappa)\tilde{y}^{-}(\kappa) - \rho^{-}(\kappa)\right) I_{[x_t<-\kappa \cap y_t<-\kappa]}. \tag{9.73}$$

Inference on a vector of exceedance correlation can be conducted by stacking the moment conditions and using a HAC covariance estimator such as the Newey and West (1987) estimator. Suppose $\boldsymbol{\kappa}$ was a vector of thresholds $\kappa_1, \kappa_2, \ldots, \kappa_n$, then

$$\sqrt{T}\begin{pmatrix} \hat{\rho}^{+}(\kappa) - \rho^{+}(\kappa) \\ \hat{\rho}^{-}(\kappa) - \rho^{-}(\kappa) \end{pmatrix} \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Omega})$$

$\mathbf{\Omega}$ can be estimated using the moment conditions,

$$\hat{\mathbf{\Omega}} = \hat{\mathbf{\Gamma}}_0 + \sum_{l=1}^{L} w_j \left( \hat{\mathbf{\Gamma}}_l + \hat{\mathbf{\Gamma}}_l' \right) \tag{9.74}$$

where $w_l = 1 - \frac{l}{L+1}$,

$$\hat{\mathbf{\Gamma}}_j = \sum_{t=j+1}^{T} \xi_t \xi_{t-j}$$

and

$$\xi_t = T \begin{bmatrix} \frac{1}{T_{\kappa_1}^+} \left( \tilde{x}^+(\kappa_1)\tilde{y}^+(\kappa_1) - \rho^+(\kappa) \right) I_{[x_t > \kappa_1 \cap y_t > \kappa_1]} \\ \vdots \\ \frac{1}{T_{\kappa_n}^+} \left( \tilde{x}^+(\kappa_n)\tilde{y}^+(\kappa_n) - \rho^+(\kappa_n) \right) I_{[x_t > \kappa_n \cap y_t > \kappa_n]} \\ \frac{1}{T_{\kappa_1}^-} \left( \tilde{x}^-(\kappa_1)\tilde{y}^-(\kappa_1) - \rho^-(\kappa) \right) I_{[x_t > \kappa_1 \cap y_t > \kappa_1]} \\ \vdots \\ \frac{1}{T_{\kappa_n}^-} \left( \tilde{x}^-(\kappa_n)\tilde{y}^-(\kappa_n) - \rho^-(\kappa_n) \right) I_{[x_t > \kappa_n \cap y_t > \kappa_n]} \end{bmatrix}.$$

Exceedance beta is similarly defined, only using the ratio of an exceedance covariance to an exceedance variance.

**Definition 9.32** (Exceedance Beta)**.** The exceedance beta at level $\kappa$ is defined as

$$\beta^+(\kappa) = \frac{\text{Cov}\left(x, y | x > \kappa, y > \kappa\right)}{\text{V}\left(x | x > \kappa, y > \kappa\right)} = \frac{\sigma_y^+(\kappa)}{\sigma_x^+(\kappa)} \rho^+(\kappa) \tag{9.75}$$

$$\beta^-(\kappa) = \frac{\text{Cov}\left(x, y | x < -\kappa, y < -\kappa\right)}{\text{V}\left(x | x < -\kappa, y < -\kappa\right)} = \frac{\sigma_y^-(\kappa)}{\sigma_x^-(\kappa)} \rho^-(\kappa) \tag{9.76}$$

Sample exceedance betas are computed using the sample analogues,

$$\hat{\beta}^+(\kappa) = \frac{\hat{\sigma}_{xy}^+(\kappa)}{\hat{\sigma}_x^{2+}(\kappa)}, \text{ and } \hat{\beta}^-(\kappa) = \frac{\hat{\sigma}_{xy}^-(\kappa)}{\hat{\sigma}_x^{2-}(\kappa)}, \tag{9.77}$$

and inference can be conducted in an analogous manner to exceedance correlations using a HAC estimator and the moment conditions

$$\frac{T}{T_\kappa^+} \left( \frac{\sigma_y^+(\kappa)}{\sigma_x^+(\kappa)} \tilde{x}^+(\kappa)\tilde{y}^+(\kappa) - \beta^+(\kappa) \right) I_{[x_t > \kappa \cap y_t > \kappa]} \tag{9.78}$$

$$\frac{T}{T_\kappa^-} \left( \frac{\sigma_y^-(\kappa)}{\sigma_x^-(\kappa)} \tilde{x}^-(\kappa)\tilde{y}^-(\kappa) - \beta^-(\kappa) \right) I_{[x_t < -\kappa \cap y_t < -\kappa]}. \tag{9.79}$$

### 9.6.3 Application: Dependence between the S&P 500 and the FTSE 100

Daily data was downloaded from Yahoo! Finance for the entire history of both the S&P 500 and the FTSE 100. Table 9.6 contains the three correlations and standard errors computed using the bootstrap where weekly returns were used to avoid issues with of return synchronization (overlapping returns were used in all applications). The linear correlation is the largest, followed by the rank and Kendall's $\tau$. Figure 9.9 plots these same three measures only using 252-day moving averages. The three measures broadly agree about the changes in dependence. Figure 9.10 plots the negative and positive exceedance correlation for these two assets and 95% confidence intervals computed using the bootstrap. The exceedance thresholds were chosen using quantiles of each series where negative corresponds to thresholds less than or equal to 50% and positive includes thresholds greater than or equal to 50%. The correlation between these markets differs substantially depending on the sign of the returns.

### 9.6.4 Application: Asymmetric Dependence from Simple Models

Asymmetric dependence can be generated from simple models. The simulated data in both panels in figure 9.8 was from a standard CAP-M calibrated to match a typical S&P 500 stock. The market returns was simulated from a standardized $t_6$ with the same variance as the S&P 500 in the past 10 years and the idiosyncratic variance was similarly calibrated to the cross-section of idiosyncratic variances. The simulated data in the top panel was computed as

$$r_{i,t} = r_{m,t} + \epsilon_{i,t}$$

where $\epsilon_{i,t}$ was i.i.d. normal with same variance as the average variance in the idiosyncratic return from the S&P 500 cross-section. The bottom panel was generated according to

$$r_{i,t} = r_{m,t} + z_{i,t}\epsilon_{i,t}$$

where $z_{i,t} = \exp\left(-10 r_{m,t} I_{[r_{m,t}<0]}\right)$ introduced heteroskedasticity so that the idiosyncratic variance is lower on days where the market is down. This simple change introduces asymmetric dependence between positive and negative returns.

## 9.7 Copulas

Copulas are a relatively new tool in financial econometrics which are useful in risk management, credit pricing and derivatives. Copulas allow the dependence between assets to be separated from the marginal distribution of each asset. Recall that a $k$-variate random variable $\mathbf{X}$ has a distribution function $F(x_1, x_2, \ldots, x_k)$ which maps from the domain of $\mathbf{X}$ to [0,1]. The distribution function contains all of the information about the probability of observing different values of $\mathbf{X}$, and while there are many distribution functions, most are fairly symmetric and in-flexible. For example, the multivariate Student's $t$ requires all margins to have the same degree-of-freedom

parameter which mean that chance of seeing relatively large returns much be the same for all assets. While this assumption may be reasonable for modeling equity index returns, extremely heavy tails are not realistic in a model which contains equity index returns and bond returns or foreign exchange returns, since the latter two are substantially thinner tailed. Copulas provide a flexible mechanism where the marginal distributions can be modeled *separately* from the dependence structure and provide substantially richer framework than working within the set of known (joint) distribution functions.

Recall the definition of the marginal distribution of $X_1$.

**Definition 9.33** (Marginal Density)**.** Let $X = (X_1, X_2, \ldots, X_k)$ be a $k$-variate random variable with joint density $f_X(X)$. The marginal density of $X_i$ is defined

$$f_i(x_i) = \int_{\mathcal{S}(X_1,\ldots X_{i-1},X_{i+1},\ldots,X_k)} f_X(x) d\,x_1 \ldots dx_{i-1} dx_{i+1} \ldots dx_k,$$

where $\mathcal{S}(\cdot)$ is the support of its argument.

The marginal density contains only information about the probability of observing values of $X_i$ and only $X_i$ since all other random variables have been integrated out. For example, if $X$ was a bivariate random variable with continuous support, then the marginal density of $X_1$ is

$$f_1(x_1) = \int_{-\infty}^{\infty} f_X(x_1, x_2)\, d\,x_2.$$

The marginal distribution,

$$F_1(x_1) = \int_{-\infty}^{x_1} f_1(s)\, d\,s,$$

contains all of the information about the probability of observing values of $X_1$, and importantly $F_{X_1}(x_1) \sim U(0,1)$. Since this is true for both $X_1$ and $X_2$, $u_1 = F_{X_1}(x_1)$ and $u_2 = F_{X_2}(x_2)$ must contain only information about the dependence between the two random variables. The distribution which describes the dependence is known as a copula, and so applications built with copulas allow information in marginal distributions to be cleanly separated from the dependence between random variables. This decomposition provides a flexible framework for constructing precise models of *both* marginal distributions and dependence.

### 9.7.1 Basic Theory

A copula is a distribution function for a random variable where each margin is uniform $[0, 1]$.

**Definition 9.34** (Copula)**.** A $k$-dimensional copula is a distribution function on $[0, 1]^k$ with standard uniform marginal distributions, and is denoted $C(u_1, u_2, \ldots, u_k)$.

All copulas all satisfy 4 key properties.

- $C(u_1, u_2, \ldots, u_k)$ is increasing in each component $u_i$.
- $C(0, \ldots, u_j, \ldots, 0) = 0$.
- $C(1, \ldots, u_j, \ldots, 1) = u_j$.
- For all $\mathbf{u} \leq \mathbf{v}$ where inequality holds on a point-by-point basis, the probability of the hypercube bound with corners $\mathbf{u}$ and $\mathbf{v}$ is non-negative.

The key insight which has lead to the popularity of copulas in finance comes from Sklar's theorem (Sklar, 1959).

**Theorem 9.2** (Sklar's Theorem)**.** *Let $F$ be a $k$-variate joint distribution with marginal distributions $F_1$,$F_2$,...,$F_k$. Then there exists a copula $C : [0,1]^k \rightarrow [0,1]$ such that for all $x_1$,$x_2$,...,$x_k$,*

$$\begin{aligned} F(x_1, x_2, \ldots, x_k) &= C(F_1(x_1), F_2(x_2), \ldots, F_k(x_k)) \\ &= C(u_1, u_2, \ldots u_k). \end{aligned}$$

*Additionally, if the margins are continuous then $C$ is unique.*

Sklar's theorem is useful in a number of ways. First, it ensures that the copula is unique whenever the margins are continuous, which is usually the case in financial applications. Second, it shows how any distribution function with known margins can be transformed into a copula. Suppose $F(x_1, x_2, \ldots, x_k)$ is a known distribution function, and that the marginal distribution function of the i$^{\text{th}}$variable is denoted $F_i(\cdot)$. Further assume that the marginal distribution function is invertible, and denote the inverse as $F_i^{-1}(\cdot)$. The copula implicitly defined by $F$ is

$$C(u_1, u_2, \ldots, u_k) = F\left(F_1^{-1}(u_1), F_2^{-1}(u_2), \ldots, F_k^{-1}(u_k)\right).$$

This relationship allows for any distribution function to be used as the basis for a copula, and appears in the definition of the Gaussian and the Student's $t$ copulas.

Copulas are distribution functions for $k$-variate uniforms, and like all distribution functions they may (or may not) have an associated density. When the copula density exists it can be derived by differentiating the distribution with respect to each random variable,

$$c(u_1, u_2, \ldots, u_k) = \frac{\partial^k C(u_1, u_2, \ldots, u_k)}{\partial u_1 \partial u_2 \ldots \partial u_k}. \tag{9.80}$$

This is identical to the relationship between any $k$-variate distribution $F$ and its associated density $f$.

### 9.7.2 Tail Dependence

One final measure of dependence, tail dependence, can be useful in understanding risks in portfolios. Tail dependence of more of a theoretical construction than something that would generally be estimated (although it is possible to estimate tail dependence).

**Definition 9.35** (Tail Dependence)**.** The upper and lower tail dependence, $\tau^U$ and $\tau^L$ respectively, are defined as the probability of an extreme event,

$$\tau^U = \lim_{u \to 1-} \Pr\left[X > F_X^{-1}(u) | Y > F_Y^{-1}(u)\right] \tag{9.81}$$

$$\tau^L = \lim_{u \to 0+} \Pr\left[X < F_X(u) | Y < F_Y(u)\right] \tag{9.82}$$

where the limits are taken from above for $\tau^U$ and below for $\tau^L$.

Tail dependence measures the probability $X$ takes an extreme value given $Y$ takes an extreme value. The performance of or assets used as hedges or portfolio diversification is particularly important when the asset being hedged has had a particularly bad day, characterized by an extreme return in its lower tail.

Lower tail dependence takes a particularly simple form when working in copulas, and is defined

$$\tau^L = \lim_{u \to 0+} \frac{C(u,u)}{u} \tag{9.83}$$

$$\tau^U = \lim_{u \to 1-} \frac{1 - 2u + C(u,u)}{1-u} \tag{9.84}$$

The coefficient of tail dependence is always in $[0,1]$ since it is a probability. When $\tau^U$ $\left(\tau^L\right)$ is 0, then the two series are upper (lower) tail independent. Otherwise the series are tail dependent where higher values indicate more dependence in extreme events.

## 9.7.3 Copulas

A large number of copulas have been produced. Some, such as the Gaussian, are implicitly defined from general distributions. Others have been produced only for uniform random variables. In all expressions for the copulas, $u_i \sim U(0,1)$ are uniform random variables.

#### 9.7.3.1 Independence Copula

The simplest copula is the independence copula, and given by the product of the inputs.

**Definition 9.36** (Independence Copula)**.** The independence copula is

$$C(u_1, u_2, \ldots, u_k) = \prod_{i=1}^{k} u_i \tag{9.85}$$

The independence copula has no parameters.

#### 9.7.3.2 Comonotonicity Copula

The copula with the most dependence is known as the comonotonicity copula.

**Definition 9.37** (Comonotonicity Copula)**.** The comonotonicity copula is

$$C(u_1, u_2, \ldots, u_k) = \min(u_1, u_2, \ldots, u_k) \tag{9.86}$$

The dependence in this copula is *perfect*. The comonotonicity does not have an associated copula density.

#### 9.7.3.3 Gaussian Copula

The Gaussian (normal) copula is implicitly defined in terms of the $k$-variate Gaussian distribution, $\Phi_k(\cdot)$, and the univariate Gaussian distribution, $\Phi(\cdot)$.

**Definition 9.38** (Gaussian Copula)**.** The Gaussian copula is

$$C(u_1, u_2, \ldots, u_k) = \Phi_k\left(\Phi^{-1}(u_1), \Phi^{-1}(u_2), \ldots, \Phi^{-1}(u_k)\right) \tag{9.87}$$

where $\Phi^{-1}(\cdot)$ is the inverse of the Gaussian distribution function.

Recall that if $u$ is a uniform random variable than $\Phi^{-1}(x)$ will have a standard normal distribution. This transformation allows the Gaussian copula density to be implicitly defined using the inverse distribution function. The Gaussian copula density is

$$c(u_1, u_2, \ldots, u_k) = \frac{(2\pi)^{-\frac{k}{2}} |\mathbf{R}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\boldsymbol{\eta}'\mathbf{R}^{-1}\boldsymbol{\eta}\right)}{\phi(\Phi^{-1}(u_1)) \ldots \phi(\Phi^{-1}(u_k))} \tag{9.88}$$

where $\boldsymbol{\eta} = \Phi^{-1}(\mathbf{u})$ is a $k$ by 1 vector where $\eta_i = \Phi^{-1}(u_i)$, $\mathbf{R}$ is a correlation matrix and $\phi(\cdot)$ is the normal pdf. The extra terms in the denominator are present in all implicitly defined copulas since the joint density is the product of the marginal densities and the copula density.

$$\begin{aligned} f_1(x_1) \ldots f_k(x_k)\, c(u_1, \ldots, u_k) &= f(x_1, x_2, \ldots, x_k) \\ c(u_1, \ldots, u_k) &= \frac{f(x_1, x_2, \ldots, x_k)}{f_1(x_1) \ldots f_k(x_k)} \end{aligned}$$

#### 9.7.3.4 Student's $t$ Copula

The Student's $t$ copula is also implicitly defined in an identical manner to the Gaussian copula.

**Definition 9.39** (Student's Copula)**.** The Student's $t$ copula is

$$C(u_1, u_2, \ldots, u_k) = t_{k,\nu}\left(t_\nu^{-1}(u_1), t_\nu^{-1}(u_2), \ldots, t_\nu^{-1}(u_k)\right) \tag{9.89}$$

where $t_{k,\nu}(\cdot)$ is the $k$-variate Student's $t$ distribution function with $\nu$ degrees of freedom and $t_\nu^{-1}$ is the inverse of the univariate Student's $t$ distribution function with $\nu$ degrees of freedom.

Note that while a Student's $t$ is superficially similar to a normal, variables which have a multivariate $t_\nu$ distributed are often substantially more dependant, at least when $\nu$ is small (3 – 8). A multivariate Student's $t$ which is a multivariate normal divided by a single, common, independent $\chi^2_\nu$ standardized to have mean 1. When $\nu$ is small, the chance of seeing a small value in the denominator is large, and since this divisor is common, all series will tend to take relatively large values simultaneously.

#### 9.7.3.5 Clayton Copula

The Clayton copula exhibits asymmetric dependence for most parameter values. The lower tail is more dependant than the upper tail and so it may be appropriate for financial asset such as equity returns.

**Definition 9.40** (Clayton Copula)**.** The Clayton copula is

$$C(u_1, u_2) = \left(u_1^{-\theta} + u_2^{-\theta} - 1\right)^{-1/\theta}, \ \theta > 0 \tag{9.90}$$

The Clayton copula limits to the independence copula when $\theta \to 0$. The copula density can be found by differentiating the Copula with respect to $u_1$ and $u_2$, and so is

$$c(u_1, u_2) = (\theta + 1)\, u_1^{-\theta-1} u_2^{-\theta-1} \left(u_1^{-\theta} + u_2^{-\theta} - 1\right)^{-1/\theta-2}.$$

#### 9.7.3.6 Gumbel and Rotated Gumbel Copula

The Gumbel copula exhibits asymmetric dependence in the upper tail rather than the lower tail.

**Definition 9.41** (Gumbel Copula)**.** The Gumbel copula is

$$C(u_1, u_2) = \exp\left[-\left((-\ln u_1)^\theta + (-\ln u_2)^\theta\right)^{1/\theta}\right], \ \theta \geq 1 \tag{9.91}$$

The Gumbel copula exhibits upper tail dependence which increases as $\theta$ grows larger, and limits to the independence copula when $\theta \to 1$. Because upper tail dependence is relatively rare among financial assets, a "rotated" version of the Gumbel may be more appropriate.

**Definition 9.42** (Rotated (Survival) Copula)**.** Let $C(u_1, u_2)$ be a bivariate copula. The rotated version[9] of the copula is given by

$$C^R(u_1, u_2) = u_1 + u_2 - 1 + C(1 - u_1, 1 - u_2).$$

[9]The rotated copula is commonly known as the survival copula, since rather than computing the probability of observing values smaller than $(u_1, u_2)$, it computes the probability of seeing values larger than $(u_1, u_2)$.

Using this definition allows the rotated Gumbel copula to be constructed which will have lower tail dependence rather than the upper tail dependence found in the usual Gumbel copula.

**Definition 9.43** (Rotated Gumbel Copula)**.** The rotated Gumbel copula is

$$C^R(u_1, u_2) = u_1 + u_2 - 1 + \exp\left[-\left((-\ln(1-u_1))^\theta + (-\ln(1-u_2))^\theta\right)^{1/\theta}\right],\ \theta \geq 1 \quad (9.92)$$

The rotated Gumbel is the Gumbel copula using $1-u_1$ and $1-u_2$ as its arguments. The extra terms are needed to satisfy the 4 properties of a copula. The rotated Gumbel copula density is tedious to compute, but is presented here.

The rotated Gumbel copula density is

$$\begin{aligned} c(u_1, u_2) &= \frac{\exp\left[-\left((-\ln(1-u_1))^\theta + (-\ln(1-u_2))^\theta\right)^{1/\theta}\right]((-\ln(1-u_1))(-\ln(1-u_2)))^{\theta-1}}{(1-u_1)(1-u_2)((-\ln(1-u_1)) + (-\ln(1-u_2)))^{2-1/\theta}} \\ &\times \left(\left((-\ln(1-u_1))^\theta + (-\ln(1-u_2))^\theta\right)^{1/\theta} + \theta - 1\right). \end{aligned}$$

This copula density is identical to the Gumbel copula density only using $1-u_1$ and $1-u_2$ as its arguments. This is the "rotation" since values of the original Gumbel near 0, where the dependence is low, are near 1 after the rotation, where the dependence is higher.

#### 9.7.3.7 Joe-Clayton Copula

The Joe-Clayton copula allows for asymmetric dependence is both tails.

**Definition 9.44** (Joe-Clayton Copula)**.** The Joe-Clayton copula is

$$C(u_1, u_2) = 1 - \left(\left[\left(1-(1-u_1)^{\theta_U}\right)^{-\theta_L} + \left(1-(1-u_2)^{\theta_U}\right)^{-\theta_L} - 1\right]^{-1/\theta_L}\right)^{1/\theta_U} \quad (9.93)$$

where the two parameters, $\theta_L$ and $\theta_U$ are directly related to lower and upper tail dependence through

$$\theta_L = -\frac{1}{\log_2\left(\tau^L\right)},\ \ \theta_U = \frac{1}{\log_2\left(2-\tau^U\right)}$$

where both coefficients of tail dependence satisfy $0 < \tau^i < 1$, $i = L, U$.

The copula density is a straightforward, although tedious, derivation. The Joe-Clayton copula is not symmetric, even when the same values for $\tau^L$ and $\tau^U$ are used. This may be acceptable, but if symmetry is preferred a symmetric copula can be constructed by averaging a copula with its rotated counterpart.

| Copula | $\tau^L$ | $\tau^U$ | Notes |
|---|---|---|---|
| Gaussian | 0 | 0 | $\lvert\rho\rvert < 1$ |
| Students $t$ | $2t_{\nu+1}(w)$ | $2t_{\nu+1}(w)$ | $w = -\sqrt{\nu+1}\sqrt{1-\rho}/\sqrt{1+\rho}$ |
| Clayton | $2^{-\frac{1}{\theta}}$ | 0 | |
| Gumbel | 0 | $2-2^{\frac{1}{\theta}}$ | Rotated Swaps $\tau^L$ and $\tau^U$ |
| Symmetrized Gumbel | $1-2^{\frac{1-\theta}{\theta}}$ | $1-2^{\frac{1-\theta}{\theta}}$ | |
| Joe-Clayton | $2^{-\frac{1}{\theta_L}}$ | $2-2^{\frac{1}{\theta_U}}$ | Also Symmetrized JC |

Table 9.7: The relationship between parameter values and tail dependence for the copulas in section 9.7.3. $t_{\nu+1}(\cdot)$ is the CDF of a univariate Student's $t$ distribution with $\nu+1$ degree of freedom.

**Definition 9.45** (Symmetrized Copula)**.** Let $C(u_1, u_2)$ be an asymmetric bivariate copula. The symmetrized version of the copula is given by

$$C^S(u_1, u_2) = \frac{1}{2}\left(C(u_1, u_2) + C^R(1-u_1, 1-u_2)\right) \tag{9.94}$$

If $C(u_1, u_2)$ is already symmetric, then $C(u_1, u_2) = C^R(1-u_1, 1-u_2)$ and so the $C^S(u_1, u_2)$ must also be symmetric. The copula density, assuming it exists, is simple for symmetrized copulas since

$$c^S(u_1, u_2) = \frac{1}{2}(c(u_1, u_2) + c(1-u_1, 1-u_2))$$

which follows since the derivative of the rotated copula with respect to $u_1$ and $u_2$ only depends on the term involving $C(1-u_1, 1-u_2)$.

### 9.7.4 Tail Dependence in Copulas

The copulas presented in the previous section all have different functional forms, and so will lead to different distributions. One simple method to compare the different forms is through the tail dependence. Table 9.7 show the relationship between the tail dependence in the different copulas and their parameters. The Gaussian has no tail dependence except in the extreme case when $\lvert\rho\rvert = 1$, in which case tail dependence is 1 in both tails. Other copulas, such as the Clayton and Gumbel, have asymmetric tail dependence.

### 9.7.5 Visualizing Copulas

Copulas are defined on the unit hyper-cube (or unit square when there are two variables) and so one obvious method to inspect the difference between two is to plot the distribution function or the density on its default domain. This visualization method does not facilitate inspecting the tail dependence since the interesting dependence occurs in the small squares of in $[0, 0.05] \times$

$[0, 0.05]$ and $[.95, 1] \times [.95, 1]$ which correspond to the lower and upper 5% of each margin. A superior method to visually inspect copulas is to compare the joint densities where the marginal distribution of each series is a standard normal. This visualization technique ensures that all differences are attributable to the copula which spreading the interesting aspects of a copula over a wider range and so allowing for more detail to be seen about the dependence.

Figure 9.11 contains plots of 4 copulas. The top tow panels contains the independence copula and the comonotonicity copula as distributions on $[0, 1] \times [0, 1]$ where curves are isoprobability lines. In distribution space, increasing dependence appears as "L" shapes, which independence appears as a semi-circle. The bottom two figures contain the normal copula distribution and the Gaussian copula density using normal margins, where in both cases the correlation is $\rho = 0.5$. The dependence in the Gaussian copula is heavier than in the independence copula – a special case of the Gaussian copula when $\rho = 0$ – but lighter than in the comonotonicity copula. The density has both a Gaussian copula and Gaussian margins and so depicts a bivariate normal. The density function shows the dependence between the two series in a more transparent manner and so is usually preferable.[10]

Figure 9.12 contains plots of 4 copulas depicted as densities where the margin of each series is standard normal. The upper left panel contains the Clayton density which has strong tail lower dependence ($\theta = 1.5$). The upper right contains the symmetrized Joe-Clayton where $\tau^L = \tau^U = 0.5$ which has both upper and lower tail dependence. The bottom two panels contain the Gumbel and symmetrized Gumbel where $\theta = 1.5$. The rotated Gumbel is similar to the Clayton copula although it is not identical.

### 9.7.6 Estimation of Copula models

Copula models are estimated using maximum likelihood which is natural since the specify a complete distribution for the data. As long as the copula density exists, and the parameters of the margins are distinct from the parameters of the copula (which is almost always the case), the likelihood or a $k$-variate random variable $Y$ can be written as

$$f(\mathbf{y}_t; \boldsymbol{\theta}, \boldsymbol{\psi}) = f_1(y_{1,t}; \boldsymbol{\theta}_1) f_2(y_{2,t}; \boldsymbol{\theta}_2) \dots f_k(y_{k,t}; \boldsymbol{\theta}_k) c\left(u_{1,t}, u_{2,t}, \dots, u_{k,t}; \boldsymbol{\psi}\right)$$

where $u_{j,t} = F_j^{-1}\left(y_{j,t}; \boldsymbol{\theta}_j\right)$ are the probability integral transformed data,$\boldsymbol{\theta}_j$ are the parameters specific to marginal model $j$ and $\boldsymbol{\psi}$ are the parameters of the copula. The log likelihood is then the sum of the marginal log likelihoods and the copula log likelihood,

$$l(\boldsymbol{\theta}, \boldsymbol{\psi}; \mathbf{y}) = \ln f_1(y_1; \boldsymbol{\theta}_1) + \ln f_2(y_2; \boldsymbol{\theta}_2) + \dots + \ln f_k(y_k; \boldsymbol{\theta}_k) + \ln c\left(u_1, u_2, \dots, u_k; \boldsymbol{\psi}\right).$$

This decomposition allows for consistent estimation of the parameters in two steps:

[10] Some copulas do not have a copula density and in the cases the copula distribution is the only method to visually compare copulas.

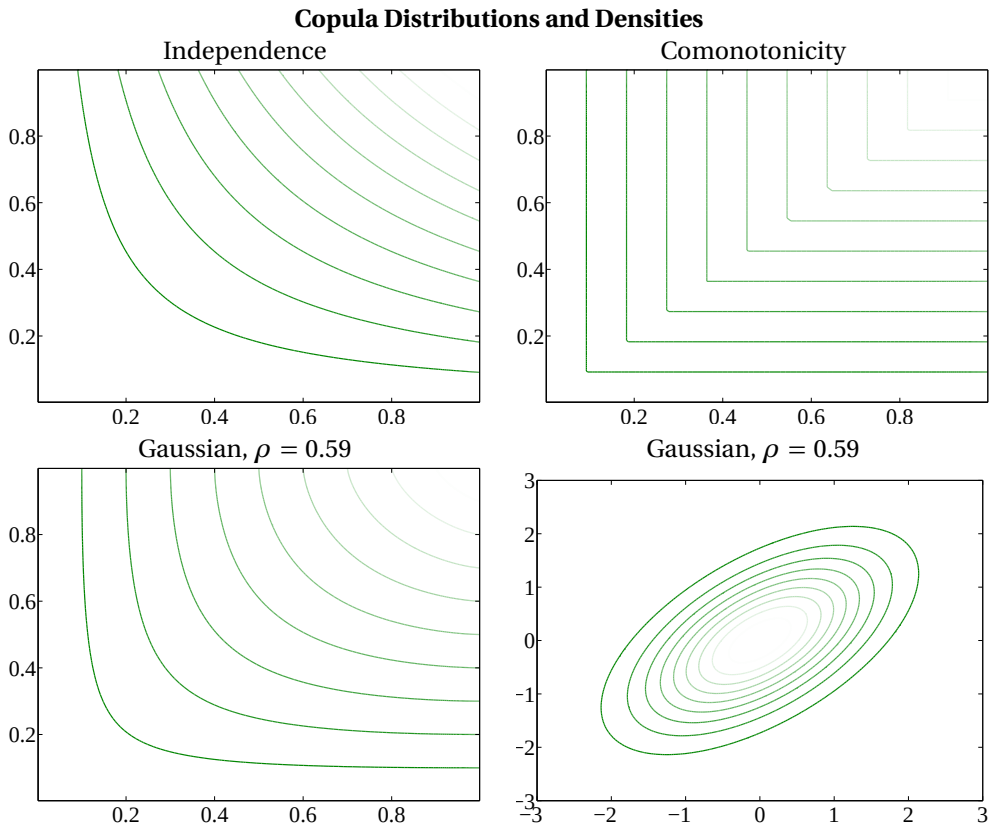


Figure 9.11: The top left panel contains the independence copula. The top right panel contains the comonotonicity copula which as perfect dependence. The bottom panels contain the Gaussian copula, where the left depicts the copula in distribution space ($[0,1] \times [0,1]$) and the right shows the copula a as density with standard normal margins. The parameter values were estimated in the application to the S&P 500 and FTSE 100 returns.

---

1. For each margin $j$, estimate $\boldsymbol{\theta}_j$ using quasi maximum likelihood as the solution to

$$\arg\max_{\boldsymbol{\theta}_j} \sum_{t=1}^{T} \ln f_j\left(y_{j,t}; \boldsymbol{\theta}_j\right)$$

   When fitting models using copulas it is also important to verify that the marginal models are adequate using the diagnostics for univariate densities described in chapter 8.

2. Using the probability integral transformed residuals evaluated at the estimated values,

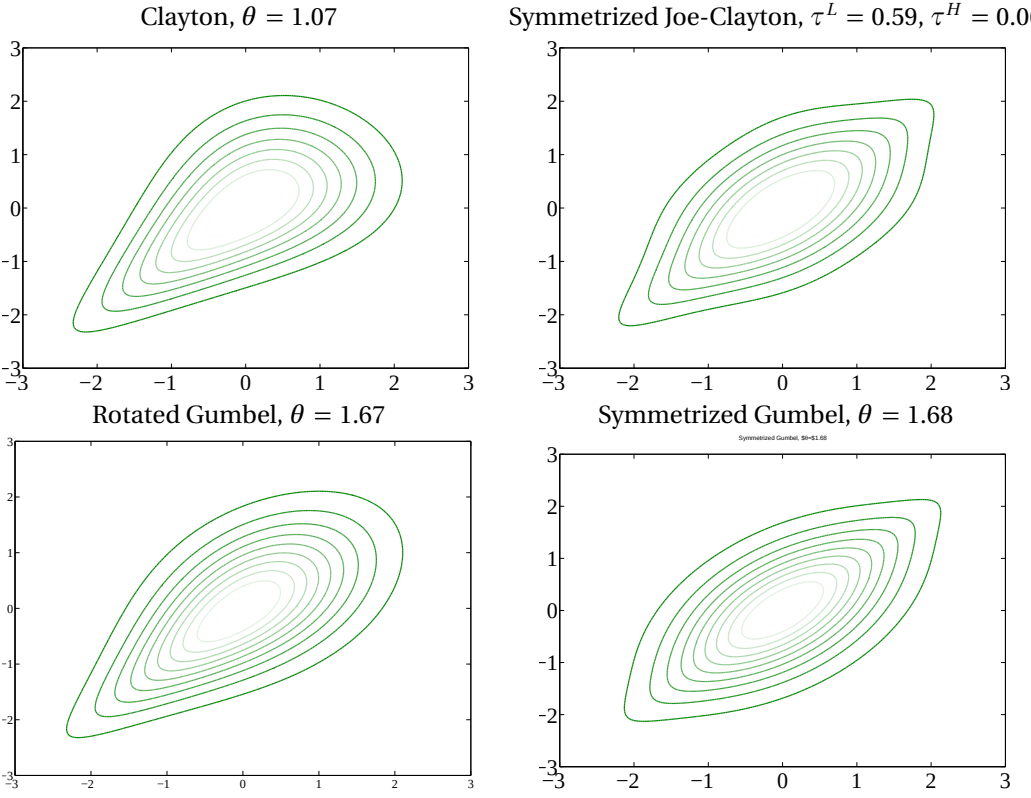


Figure 9.12: These for panels all depict copulas as densities using standard normal margins. All differences in appearance can be attributed to the differences in the copulas. The top left panel contains the Clayton copula density. The top right contains the symmetrized Joe-Clayton. The bottom panels contain the rotated Gumbel which has lower tail dependence and the symmetrized Gumbel. The parameter values were estimated in the application to the S&P 500 and FTSE 100 returns.

---

$\hat{u}_{j,t} = F^{-1}\left(y_{j,t}; \hat{\boldsymbol{\theta}}_j\right)$, estimate the parameters of the copula as

$$\arg\max_{\boldsymbol{\psi}} \sum_{t=1}^{T} \ln c\left(\hat{u}_{1,t}, \hat{u}_{2,t}, \ldots, \hat{u}_{k,t}; \boldsymbol{\psi}\right)$$

This two step procedure is not efficient in the sense that the parameter estimates are consistent but have higher variance than would be the case if all parameters were simultaneously estimated. In practice the reduction in precision is typically small, and one alternative is to use the two-step estimator as a starting value for an estimator which simultaneously estimates the all parameters.

Standard errors can be computed from the two-step estimation procedure by treating it as a two-step GMM problem where the scores of the marginal log likelihoods and the copula are the moment conditions (See section 6.10 for a discussion)

An alternative estimation procedure uses *nonparametric* models for the margins. Nonparametric margins are typically employed when characterizing the distribution of the margins are not particularly important, and so the first step can be replaced through the use of the empirical CDF. The empirical CDF estimates the $\hat{u}_{j,t} = \text{rank}(y_{j,t})/(T+1)$ where rank is the ordinal rank of observation $t$ among the $T$ observations. The empirical CDF is uniform by construction. Using the empirical CDF is not generally appropriate when the data have time-series dependence (e.g. volatility clustering) or when forecasting is an important consideration.

### 9.7.7 Application: Dependence between the S&P 500 and the FTSE 100

The use of copulas will be illustrated using the S&P 500 and FTSE 100 data. Returns were computed using 5-trading days to mitigate issues with non-synchronous trading. The first example will use the empirical CDF to estimate the probability integral transformed residuals so that the model will focus on the unconditional dependence between the two series. The upper left panel of figure 9.13 contains a scatter plot of the ECDF transformed residuals. The residuals tend to cluster around the 45º line indicating positive dependence (correlation). There are also obvious clusters near (0, 0) and (1,1) indicating dependence in the extremes. The normal, students $t$, Clayton, rotated Gumbel, symmetrized Gumbel and symmetric Joe-Clayton were all estimated. Parameter estimates and copula log-likelihoods are reported in table 9.8. The Joe-Clayton fits the data the best, followed by the symmetrized Gumbel and then the rotated Gumbel. The Clayton and the Gaussian both appear to fit the data substantially worse than the others. In the Joe-Clayton, both tails appear to have some dependence, although unsurprisingly the lower tail is substantially more dependent.

Copulas can also be used with dynamic models. Using a constant copula with dynamic models for the margins is similar to using a CCC-GARCH model for modeling conditional covariance. A dynamic distribution model was built using TARCH(1,1,1) distributions for each index return where the innovations are modeled using Hansen's Skew $t$. The same set of copulas were estimated using the conditionally transformed residuals $u_{i,t} = F\left(y_{i,t}; \sigma_t^2, \nu, \lambda\right)$ where $\sigma_t^2$ is the conditional variance, $\nu$ is the degree of freedom, and $\lambda$ which controls the skewness. Parameter estimates are reported in table 9.9. The top panel reports the parameter estimates from the TARCH model. Both series have persistent volatility although the leverage effect is stronger in the S&P 500 than it is in the FTSE 100. Standardized residuals in the S&P 500 were also heavier tailed with a degree of freedom of 8 versus 12 for the FTSE 100, and both were negatively skewed. The parameter estimates from the copulas all indicate less dependence than in the model build using the empirical CDF. This is a common finding and is due to synchronization between the two markets of volatility. Coordinated periods of high volatility leads to large returns in both series at the same time, even when the standardized shock is only moderately large. Mixing periods of high and low volatility across markets tends to increase unconditional dependence

**Dependence Measures for Weekly FTSE and S&P 500 Returns**

| Copula | $\theta_1$ | $\theta_2$ | Log Lik. |
|---|---|---|---|
| Gaussian | 0.619 | | 305.9 |
| Clayton | 1.165 | | 285.3 |
| Rotated Gumbel | 1.741 | | 331.7 |
| Symmetrized Gumbel | 1.775 | | 342.1 |
| Symmetrized Joe-Clayton | 0.606 | 0.177 | 346.2 |

Table 9.8: Parameter estimates for the unconditional copula between weekly returns on the S&P 500 and the FTSE 100. Marginal distributions were estimated using empirical CDFs. For the Gaussian copula, $\theta_1$ is the correlation, and in the Joe-Clayton $\theta_1$ is $\tau^L$ and $\theta_2$ is $\tau^U$. The final column reports the log likelihood from the copula density.

in the same way that missing periods of high and low volatility leads to heavy tails in the same market. The difference in the dependence shows up in the parameter values in the copulas where are uniformly lower than in their unconditional counterparts, and through the reduction in the range of log-likelihoods relative to the Gaussian.

Figure 9.13 contains some diagnostic plots related to fitting the conditional copula. The top right panel contains the scatter plot of the probability integral transformed residuals using from the TARCH. While these appear similar to the plot from the empirical CDF, the amount of clustering near $(0, 0)$ and $(1, 1)$ is slightly lower. The bottom left panel contains a QQ plot of the actual returns against the expected returns using the degree of freedom and skewness parameters estimated on the two indices. These curves are straight except for the most extreme observations, indicating an acceptable fit. The bottom right plot contains the annualized volatility series for the two assets where the coordination in volatility cycles is apparent. It also appears the coordination in volatility cycles has strengthened post-2000.

### 9.7.8 Dynamic Copulas

This chapter has focused on static copulas of the form $C(u_1, u_2; \boldsymbol{\theta})$. It is possible to construct dynamics copulas where dependence parameters change through time which leads to a conditional copula of the form $C(u_1, u_2; \boldsymbol{\theta}_t)$. This was first done in Patton (2006) in an application to exchange rates. The primary difficulty in specifying dynamic copula models is in determining the form of the "shock". In ARCH-type volatility models $r_t^2$ is the natural shock since it has the correct conditional expectation. In most copula models there isn't a single, unique equivalent. Creal, Koopman, and A. Lucas (2013) have recently developed a general framework which can be used to construct a natural shock even in complex models, and have applied this in the context of copulas.

DCC can also be used as a dynamic Gaussian copula where the first step is modified from fitting the conditional variance to fitting the conditional distribution. Probability integral trans-

**Conditional Copula Estimates for Weekly FTSE and S&P 500 Returns**

| Index | $\alpha_1$ | $\gamma_1$ | $\beta_1$ | $\nu$ | $\lambda$ |
|---|---|---|---|---|---|
| S&P 500 | 0.003 | 0.259 | 0.843 | 8.247 | -0.182 |
| FTSE 100 | 0.059 | 0.129 | 0.846 | 12.328 | -0.152 |

| Copula | $\theta_1$ | $\theta_2$ | Log Lik. |
|---|---|---|---|
| Gaussian | 0.586 | | 267.2 |
| Clayton | 1.068 | | 239.9 |
| Rotated Gumbel | 1.667 | | 279.7 |
| Symmetrized Gumbel | 1.679 | | 284.4 |
| Symmetrized Joe-Clayton | 0.586 | 0.057 | 284.3 |

Table 9.9: Parameter estimates for the conditional copula between weekly returns on the S&P 500 and the FTSE 100. Marginal distributions were estimated using a TARCH(1,1,1) with Hansen's Skew $t$ error. Parameter estimates from the marginal models are reported in the top panel. The bottom panel contains parameter estimates from copulas fit using the conditionally probability integral transformed residuals. For the Gaussian copula, $\theta_1$ is the correlation, and in the Joe-Clayton $\theta_1$ is $\tau^L$ and $\theta_2$ is $\tau^U$. The final column reports the log likelihood from the copula density.

formed residuals from the modified first step can then be transformed to be Gaussian, which in turn can be used in the second step of the DCC estimator. The combined model has flexible marginal distributions and a Gaussian copula.

## 9.A Bootstrap Standard Errors

The Bootstrap is a powerful tool which has a variety of uses, although it is primarily used for computing standard errors as an alternative to "plug-in" estimators used in most inference. Moreover, in some applications, expressions for asymptotic standard errors cannot be directly computed and so the bootstrap is the only viable method to make inference. This appendix provides a *very* brief introduction to computing bootstrap standard errors. The idea behind the bootstrap is very simple. If $\{\mathbf{r}_t\}$ is a sample of $T$ data points from some unknown joint distribution $F$, then $\{\mathbf{r}_t\}$ can be used to simulate (via re-sampling) from the unknown distribution $F$. The name bootstrap comes from the expression "To pull yourself up by your bootstraps", a seemingly impossible task, much like simulation from an unknown distribution

There are many bootstraps available and different bootstraps are appropriate for different types of data. Bootstrap methods can be classified as parametric or non-parametric. Parametric bootstraps make use of residuals as the basis of the re-sampling. Nonparametric bootstraps make use of the raw data. In many applications both types of bootstraps are available and

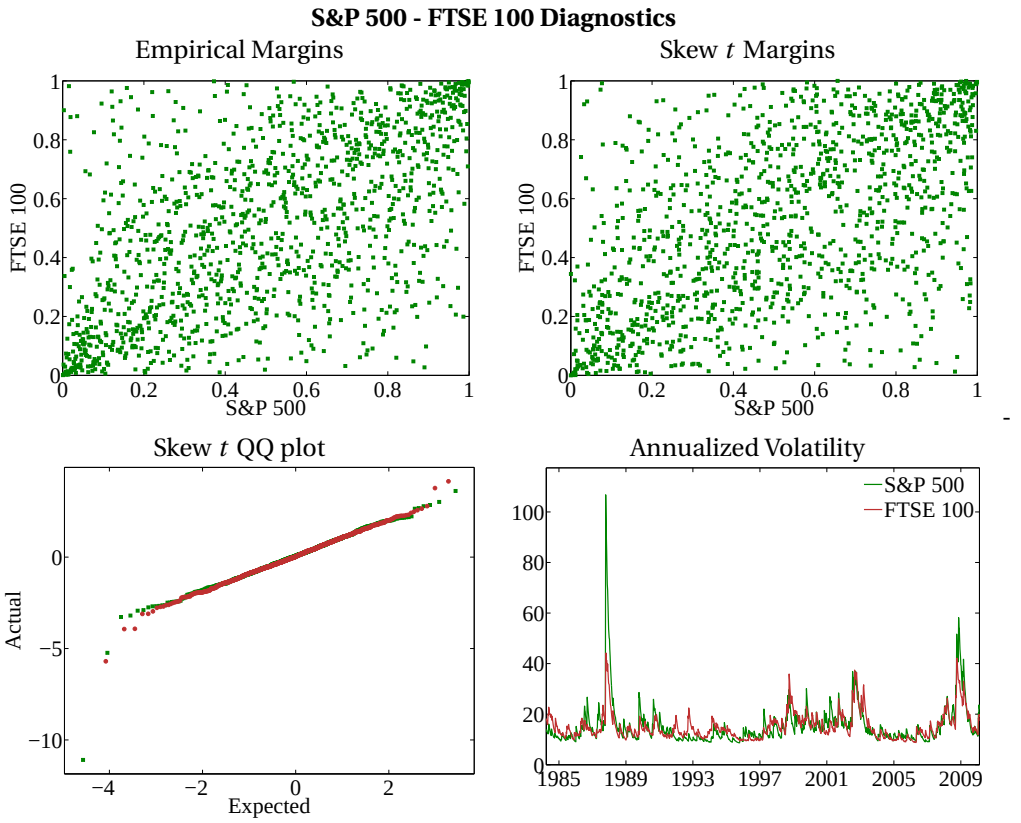


Figure 9.13: These four panels contain diagnostics for fitting copulas to weekly returns on the S&P 500 and FTSE 100. The top two panels are scatter plots of the probability integral transformed residuals. The left contain the PIT from using the empirical CDF and so is a depiction of the unconditional dependence. The right contains the PIT from the TARCH(1,1,1) with Skew $t$ errors. The bottom left contains a QQ plot of the data against the expected value from a Skew $t$. The bottom right plot contains the fit annualized volatility for the two indices.

---

the choice between the two is similar to the choice between parametric and non-parametric estimators: parametric estimators are precise but may be misleading if the model is misspecified while non-parametric estimators are consistent although may require large amounts of data to be reliable. This appendix describes three bootstraps and one method to compute standard errors using a nonparametric bootstrap method. Comprehensive treatments of the bootstrap can be found in Efron and R. J. Tibshirani (1998) and Chernick (2008).

The simplest form of the bootstrap is the i.i.d. bootstrap, which is applicable when the data are i.i.d.

**Algorithm 9.1** (IID Bootstrap)**.** *1. Draw $T$ indices $\tau_i = \lceil T u_i \rceil$ where $u_i \overset{i.i.d.}{\sim} U(0,1)$*

2. *Construct an artificial time series using the indices $\{\tau_i\}_{i=1}^{T}$,*

$$y_{\tau_1}\ y_{\tau_2}\ \dots\ y_{\tau_T}.$$

3. *Repeat steps 1–2 a total of $B$ times.*

In most applications with financial data an assumption of i.i.d. errors is not plausible and a bootstrap appropriate for dependant data is necessary. The two most common bootstraps for dependant data are the block bootstrap and the stationary bootstrap (Politis and Romano, 1994). The block bootstrap is based on the idea of drawing blocks of data which are sufficiently long so that the blocks are approximately i.i.d.

**Algorithm 9.2** (Block Bootstrap)**.**

1. *Draw $\tau_1 = \lceil T u \rceil$ where $u \overset{i.i.d.}{\sim} U(0,1)$.*
2. *For $i = 2, \dots, T$, if $i \bmod m \neq 0$, $\tau_i = \tau_{i-1} + 1$ where wrapping is used so that if $\tau_{i-1} = T$ then $\tau_i = 1$. If $i \bmod m = 0$ when $\tau_i = \lceil T u \rceil$ where $u \overset{i.i.d.}{\sim} U(0,1)$.*
3. *Construct an artificial time series using the indices $\{\tau_i\}_{i=1}^{T}$.*
4. *Repeat steps 1 – 3 a total of $B$ times.*

The stationary bootstrap is closely related to the block bootstrap. The only difference is that it uses blocks with lengths that are exponentially distributed with an *average length* of $m$.

**Algorithm 9.3** (Stationary Bootstrap)**.**

1. *Draw $\tau_1 = \lceil T u \rceil$ where $u \overset{i.i.d.}{\sim} U(0,1)$.*
2. *For $i = 2, \dots, T$, draw a standard uniform $v \overset{i.i.d.}{\sim} U(0,1)$. If $v > 1/m$, $\tau_i = \tau_{i-1} + 1$, where wrapping is used so that if $\tau_{i-1} = T$ then $\tau_i = 1$ . If $v \leq 1/m$, $\tau_i \lceil T u \rceil$ where $u \overset{i.i.d.}{\sim} U(0,1)$*
3. *Construct an artificial time series using the indices $\{\tau_i\}_{i=1}^{T}$.*
4. *Repeat steps 1 – 3 a total of $B$ times.*

These bootstraps dictate how to re-sample the data. The re-sampled data are then used to make inference on statistics of interest.

**Algorithm 9.4** (Bootstrap Parameter Covariance Estimation)**.**

1. *Begin by computing the statistic of interest $\hat{\boldsymbol{\theta}}$ using the original sample.*
2. *Using a bootstrap appropriate for the dependence in the data, estimate the statistic of interest on the $B$ artificial samples, and denote these estimates as $\tilde{\boldsymbol{\theta}}_j$, $j = 1, 2, \dots, B$*

3. *Construct confidence intervals using:*

   (a) *(Inference using standard deviation) Estimate the variance of* $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0$ *as*

   $$B^{-1}\sum_{b=1}^{B}\left(\tilde{\boldsymbol{\theta}}_b - \hat{\boldsymbol{\theta}}\right)^2$$

   (b) *(Inference using symmetric quantiles) Construct bootstrap errors as* $\eta_b = \tilde{\boldsymbol{\theta}}_b - \hat{\boldsymbol{\theta}}$*, and construct the* $1-\alpha$ *confidence interval* $(\boldsymbol{\theta} \pm \bar{q}_{\alpha/2})$ *using the* $1-\alpha/2$ *quantile of* $|\eta_b|$*, denoted* $\bar{q}_{1-\alpha/2}$*.*

   (c) *(Inference using asymmetric quantiles) Construct bootstrap errors as* $\eta_b = \tilde{\boldsymbol{\theta}}_b - \hat{\boldsymbol{\theta}}$*, and construct the* $1-\alpha$ *confidence interval* $(\boldsymbol{\theta} - q_{\alpha/2}, \boldsymbol{\theta} + q_{1-\alpha/2})$ *using the* $\alpha/2$ *and* $1-\alpha/2$ *quantile of* $\eta_b$*, denoted* $q_{\alpha/2}$ *and* $q_{1-\alpha/2}$*, respectively.*

The bootstrap confidence intervals in this chapter were all computed using this algorithm and a stationary bootstrap with $m \propto \sqrt{T}$.

*Warning*: The bootstrap is broadly applicable in cases where parameters are asymptotically normal such as in regression with stationary data. They are either not appropriate or require special attention in many situations, e.g. unit roots, and so before computing bootstrap standard errors, it is useful to verify that the bootstrap will lead to correct inference. In cases where the bootstrap fails, a more general statistical technique *subsampling* can usually be used to make correct inference.

## Shorter Problems

**Problem 9.1.** Describe the observable factor covariance model and the exponentially weighted moving average covariance model. Discuss the relative strengths and weaknesses of these two models.

**Problem 9.2.** Describe one multivariate GARCH model and one multivariate volatility model which is not a GARCH specification. Describe the relative strengths and weaknesses of these two models.

**Problem 9.3.** Discuss three alternative models for conditional *co*variance.

**Problem 9.4.** What is Exceedance Correlation?

**Problem 9.5.** Compare and contrast linear and rank correlation.

## Longer Questions

**Exercise 9.1.** Answer the following questions about covariance modeling

1. Describe the similarities between the RiskMetrics 1994 and RiskMetrics 2006 methodologies.

2. Describe two multivariate GARCH models. What are the strengths and weaknesses of these model?

3. Other than linear correlation, describe two other measures of dependence.

4. What is Realized Covariance?

5. What the important considerations when using Realized Covariance?

**Exercise 9.2.** Answer the following questions.

1. Briefly outline two applications in finance where a multivariate volatility model would be useful.

2. Describe two of the main problems faced in multivariate volatility modeling, using two different models to illustrate these problems.

3. Recall that, for a bivariate application, the BEKK model for a time-varying conditional covariance matrix is:

$$\begin{bmatrix} \sigma_{11,t} & \sigma_{12,t} \\ \sigma_{12,t} & \sigma_{22,t} \end{bmatrix} \equiv \mathbf{\Sigma}_t = \mathbf{C}\mathbf{C}' + \mathbf{A}\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1}'\mathbf{A}' + \mathbf{B}\mathbf{\Sigma}_{t-1}\mathbf{B}'$$

where $\mathbf{C}$ is a lower triangular matrix, and $\boldsymbol{\epsilon}_t' \equiv \left[\epsilon_{1,t}, \epsilon_{2,t}\right]$ is the vector of residuals. Using the result that $\text{vec}(\mathbf{QRS}) = \left(\mathbf{S}' \otimes \mathbf{Q}\right)\text{vec}(\mathbf{R})$, where $\otimes$ is the Kronecker product, re-write the BEKK model for $\text{vec}(\mathbf{\Sigma}_t)$ rather than $\mathbf{\Sigma}_t$.

4. Estimating this model on two-day returns on the S&P 500 index and the FTSE 100 index over the period 4 April 1984 to 30 December 2008, we find:

$$\hat{\mathbf{C}} = \begin{bmatrix} 0.15 & 0 \\ 0.19 & 0.20 \end{bmatrix}, \quad \hat{\mathbf{B}} = \begin{bmatrix} 0.97 & -0.01 \\ -0.01 & 0.92 \end{bmatrix}, \quad \hat{\mathbf{A}} = \begin{bmatrix} 0.25 & 0.03 \\ 0.05 & 0.32 \end{bmatrix}$$

Using your answer from (c), compute the $(1,1)$ element of the coefficient matrix on $\text{vec}\left(\mathbf{\Sigma}_{t-1}\right)$.

**Exercise 9.3.** Answer the following questions.

1. For a set of two asset returns, recall that the BEKK model for a time-varying conditional covariance matrix is:

$$\begin{bmatrix} h_{11t} & h_{12t} \\ h_{12t} & h_{22t} \end{bmatrix} \equiv \mathbf{\Sigma}_t = \mathbf{C}\mathbf{C}' + \mathbf{B}\mathbf{\Sigma}_{t-1}\mathbf{B}' + \mathbf{A}\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1}'\mathbf{A}'$$

where $\mathbf{C}$ is a lower triangular matrix, and $\boldsymbol{\epsilon}_t' \equiv [\boldsymbol{\epsilon}_{1t}, \boldsymbol{\epsilon}_{2t}]$ is the vector of residuals.

2. Describe two of the main problems faced in multivariate volatility modeling, and how the BEKK model overcomes or does not overcome these problems.

3. Using the result that $\text{vec}(\mathbf{QRS}) = \left(\mathbf{S}' \otimes \mathbf{Q}\right) \text{vec}(\mathbf{R})$, where $\otimes$ is the Kronecker product, re-write the BEKK model for $\text{vec}(\mathbf{\Sigma}_t)$ rather than $\mathbf{\Sigma}_t$.

4. Estimating this model on two-day returns on the S&P 500 index and the FTSE 100 index over the period 4 April 1984 to 30 December 2008, we find:

$$\hat{\mathbf{C}} = \begin{bmatrix} 0.15 & 0 \\ 0.19 & 0.20 \end{bmatrix}, \quad \hat{\mathbf{B}} = \begin{bmatrix} 0.97 & -0.01 \\ -0.01 & 0.92 \end{bmatrix}, \quad \hat{\mathbf{A}} = \begin{bmatrix} 0.25 & 0.03 \\ 0.05 & 0.32 \end{bmatrix}$$

   Using your answer from (b), compute the estimated intercept vector in the $\text{vec}(\mathbf{\Sigma}_t)$ representation of the BEKK model. (Hint: this vector is $4 \times 1$.)

5. Computing "exceedance correlations" on the two-day returns on the S&P 500 index and the FTSE 100 index, we obtain Figure 9.14. Describe what exceedance correlations are, and what feature(s) of the data they are designed to measure.

6. What does the figure tell us about the dependence between returns on the S&P 500 index and returns on the FTSE 100 index?

**Exercise 9.4.** Answer the following questions about covariance modeling:

1. Describe the RiskMetrics 1994 methodology for modeling the conditional covariance.

2. How does the RiskMetrics 2006 methodology differ from the 1994 methodology for modeling the conditional covariance?

3. Describe one multivariate GARCH model. What are the strengths and weaknesses of the model?

4. How is the 5% portfolio $VaR$ computed when using the RiskMetrics 1994 methodology?

5. Other than linear correlation, describe two other measures of dependence.

6. What is Realized Covariance?

7. What are the important considerations when using Realized Covariance?

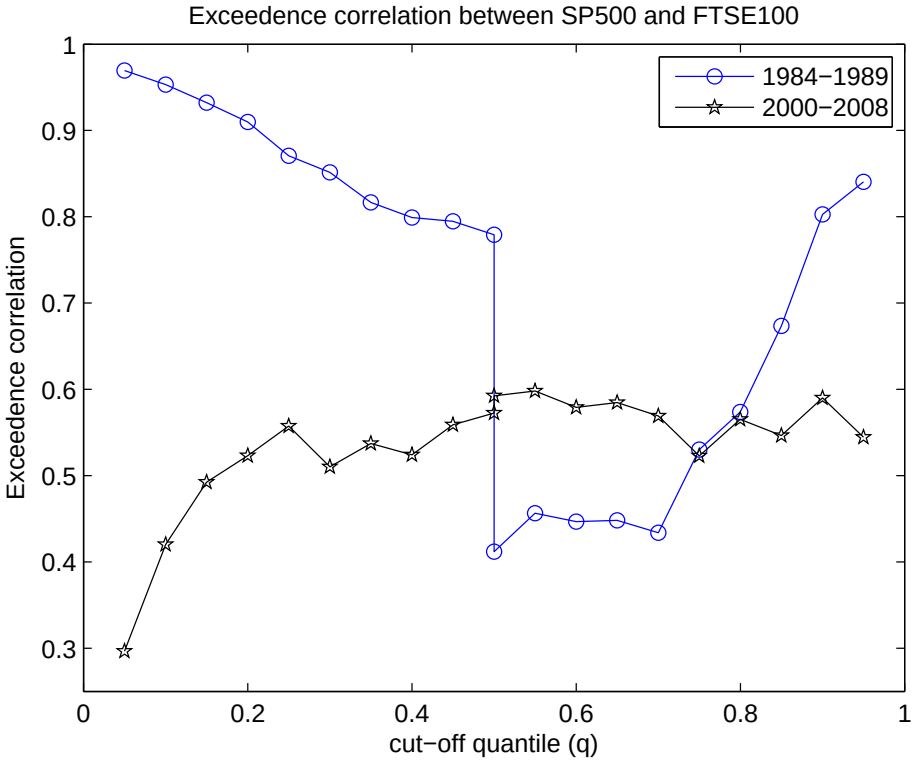


Figure 9.14: Exceedance correlations between two-day returns on the S&P 500 index and the FTSE 100 index. Line with circles uses data from April 1984 to December 1989; line with stars uses data from January 2000 to December 2008.