Chapter 9

Multivariate Volatility, Dependence and Copulas

Modeling the conditional covariance of the assets in a portfolio is more challenging than modeling the variance of the portfolio. There are two challenges unique to the multivariate problem: ensuring that the conditional covariance is positive definite and finding a parsimonious specification that limits the number of model parameters in applications to large portfolios. This chapter covers standard moving-average covariance models, multivariate ARCH and Realized Covariance. While correlations are a key component of portfolio optimization, these measures are insufficient to fully characterize the joint behavior of asset returns, especially when markets are turbulent. This chapter introduces leading alternative measures of cross-asset dependence and then concludes with an introduction to a general framework for modeling multivariate returns using copulas.

9.1 Introduction

Multivariate volatility or covariance modeling is a crucial ingredient in modern portfolio management. It is applied to many 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 finds the portfolio with the minimum variance subject to achieving a required expected return. 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 objectives other than those in the Markowitz optimization problem. For example, fund managers may be selecting investment opportunities based on beliefs about fundamental imbalances between a firm and its competitors. Accurate measurement of asset return covariance is essential when assessing the portfolio's sensitivity to new positions. The sensitivity to the existing

portfolio may be the deciding factor when evaluating multiple investment opportunities that have similar risk-return characteristics.

- Value-at-Risk - Naive $\alpha - VaR$ estimators scale the standard deviation of a portfolio by a constant value that depends on the quantile α. The conditional covariance allows the VaR sensitivity of the positions in the portfolio to be examined.

- Credit Pricing - Many credit products are written on a basket of bonds, and the correlation between the defaults of the underlying bonds is essential when determining the value of the derivative.

- Correlation Trading - Recent financial innovations allow correlation to be directly traded. The traded correlation is formally an equicorrelation (See 9.3.5). These products allow accurate correlation predictions to be used as the basis of a profitable trading strategy.

This chapter begins with an overview of simple, static estimators of covariance which are widely used. Attention then turns to dynamic models of conditional covariance based on the ARCH framework. Realized covariance, which exploits ultra-high frequency data in the same manner as realized variance, is then introduced as an improved estimator of the covariance. This chapter concludes with an examination of non-linear dependence measures and copulas, a recent introduction to financial econometrics that enables complex multivariate models to be flexibly constructed.

9.2 Preliminaries

Most volatility models are built using either returns, which is appropriate if the time horizon is small and 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 demeaned returns are $\boldsymbol{\epsilon}_t = \mathbf{r}_t - \boldsymbol{\mu}_t$ where $\boldsymbol{\mu}_t \equiv \mathrm{E}_{t-1}[\mathbf{r}_t]$ is the conditional mean.

The conditional covariance, $\boldsymbol{\Sigma}_t \equiv \mathrm{E}_{t-1}\left[\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t'\right]$, is assumed to be a k by k positive definite matrix. Some models make 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) and $\boldsymbol{\sigma}_t$ is a k by vector of conditional standard deviations. 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

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

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

Some models use 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 $\mathbf{0}$, 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 on models capable of predicting the time-t covariance using information in $\mathcal{F}_{t-1}$. Multi-step forecasting is possible from many models in this chapter by direct recursion, simulation or bootstrapping. Alternatively, direct forecasting techniques can be used to mix higher frequency data (e.g., daily) with longer forecast horizons (e.g., 2-week or one month).

9.2.1 Synchronization

Synchronization is a significant concern when measuring and modeling covariance, and non-synchronous returns can occur for a variety of reasons:

- Market opening and closing time differences – Most assets trade in liquid markets for only a fraction of the day. Differences in market hours frequently occur when modeling the return of assets that trade in different venues. The NYSE closes at either 20:00 or 21:00 GMT, depending on whether the U.S. east coast is using Eastern Standard or Daylight Time (EDT or EST). The London Stock Exchange closes at 15:30 or 16:30 GMT, depending on whether the U.K. is on British Summer Time (BST). Changes in U.S. equity prices that occur after the LSE closes are not reflected in U.K. equity prices 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. Closing prices, which are computed at the end of the trading day, do not reflect the same information in these three markets.

- Market closures due to public holidays – Markets are closed for public holidays which differ across geographies. Closures can even differ across markets, especially across asset class, within a country due to historical conventions.

- Delays in opening or closing – Assets that trade on the same exchange may be subject to opening or closing delays. For example, the gap between the first-to-open and the last-to-open stock in the S&P 500 can be as long as 15 minutes. While the range of closing times of the constituents is narrower, these are also not perfectly synchronized. These seemingly small differences lead to challenges when measuring the covariance using intra-daily (high-frequency) 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]

[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) was acquired by another company, and so did not trade.

There are three solutions to address biases that arise when modeling non-synchronous data. The first is to use relatively low-frequency returns. When using daily data, the NYSE and LSE are typically simultaneously open for 2 hours of 6½ hour U.S. trading day (30%). Using multi-day returns partially mitigates the lack of standardized opening hours since developments in U.S. equities on one day affect prices 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 weekly returns (5-day), 28 out of 32.5 hours are synchronized (86%). The downside of aggregating returns is the loss of data: parameter estimators are less efficient when and low-frequency return measurement makes it difficult to adjust portfolios for change in risk due to recent news.

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, returns constructed using prices sampled 1 hour before the LSE closes, which typically corresponds to 10:30 Eastern time, are synchronized. Daily returns constructed from these prices should capture all of the covariance between these assets. This approach is only a partial solution since many markets overlap in their trading hours, and so it is not applicable when measuring the covariance of a broad internationally diversified portfolio.

The third solution is to synchronize the non-synchronous returns using a vector moving average (Burns, Engle, and Mezrich, 1998). Suppose returns are ordered so that the first to close is in position 1, the second to close is in position 2, and so on until the last to close is 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 modeling the leading equity index of 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. News from the ASX on day t appears in the LSE, NYSE, and TSE on the same day. 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 affect $t+1$ returns in the ASX and LSE. Finally, news which comes out when the TSE is open shows 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 simplifies estimation since $r_t^{TSE} = \epsilon_t^{TSE}$, and so the model for r_t^{NYSE} is a MA(1)-X. Given estimates of ϵ_t^{NYSE}, the model for r_t^{LSE} is also a MA(1)-X. This recursive MA(1)-X structure applies to the remaining assets in the model.

In vector form, this adjustment model is

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

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

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

$\mathbf{\Theta}$ captures any components in asset return j correlated with the return to asset return i when market i closes later than the where j. In essence this procedure "brings forward" the fraction of the return which has not yet occurred when asset j closes. 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 that rely on closed-form parameter estimators are widely used as benchmarks. These models are localized using rolling-windows, and so have a limited ability to adapt to changing market conditions.

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, standard choices for n are 22 (monthly), 66 (quarterly), or 252 (annual). When returns are measured monthly, standard choices for n are 12 (annual) or 60. When variance and correlations are time-varying, moving average covariances are imprecise measures; they simultaneously give too little weight to recent observations and place too much on observations in the distant past.

9.3.2 Exponentially Weighted Moving Average Covariance

Exponentially weighted moving averages (EWMA) are an alternative to moving average covariance estimators and allow for more weight on recent information. EWMAs have been popularized in the volatility literature by RiskMetrics, which is introduced as a standard VaR model 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

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

An EWMA covariance estimator depends on an initial value for $\mathbf{\Sigma}_1$, which is usually set to the average covariance over the first m days for some $m > k$ or to the full-sample covariance. The single remaining parameter, λ, is usually to a value close to 1.

Definition 9.3 (RiskMetrics 1994 Covariance)**.** The RiskMetrics 1994 Covariance is computed as an EWMA with $\lambda = .94$ for daily data or $\lambda = .99$ for monthly (J.P.Morgan/Reuters, 1996).

The RiskMetrics EWMA estimator, formally known as RM1994, has been updated to RM2006. The improved covariance estimator uses a model with a longer memory than RM1996. Long memory processes have weights that 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). This structure simplifies estimation since incorporating a new observation into the conditional covariance only requires updating the values of a small number of EWMAs.

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

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

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

The 2006 update is a 3-parameter model that includes a logarithmic decay factor, τ_0 (1560), a lower cut-off, τ_1 (4), and an upper cutoff $\tau_{\max}$ (512), where the suggested values are in parentheses. One additional parameter, ρ, is required to operationalize the model, and RiskMetrics suggests $\sqrt{2}$ (Zumbach, 2007). [2]

Both RiskMetrics covariance estimators can be expressed as weighted averages of the outer-products of shocks, $\mathbf{\Sigma}_t = \sum_{i=1}^{\infty} \gamma_i \boldsymbol{\epsilon}_{t-i}\boldsymbol{\epsilon}_{t-1}'$, for a set of weights $\{\gamma_i\}$. Figure 9.1 contains a plot of the weights on the 120 most recent observations from both estimators. The updated methodology places both more weight on recent data and more weight on values in the distant past relative to the RM1996 model. Computing the number of periods before 99% of the weight is accumulated, or $\min_n \sum_{i=0}^{n} \gamma_i \geq .99$, is a simple method to compare the two methodologies. In RM1994, 99% of the weight accumulates in 75 observations when $\lambda = 0.94$ – the RM2006 methodology takes 619 days. The first 75 weights in the RM2006 model contain 83% of the weight, and so 1/6 of the total weight is assigned to returns more than 2 months in the past.

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

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

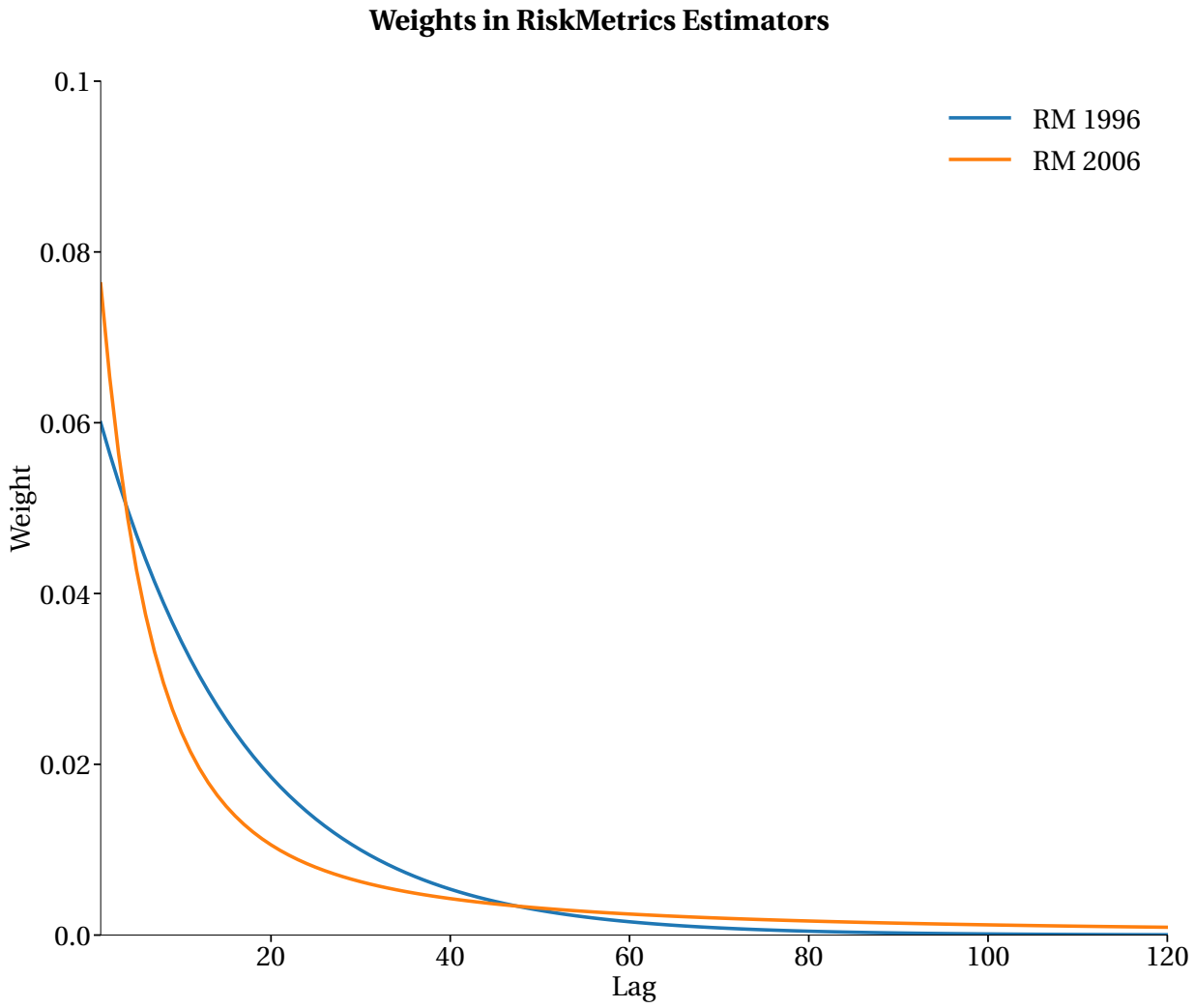

Figure 9.1: These two lines show the weights assigned to the lagged outer-product of returns $\left(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t'\right)$ in the 1994 and 2006 versions of the RiskMetrics methodology. The 2006 version places more weight on recent shocks and more weight on shocks in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

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 (Sharpe, 1964; Lintner, 1965; Black, 1972), the intertemporal CAP-M (Merton, 1973) and Arbitrage Pricing Theory (Roll, 1977). Moving average factor covariance estimators are restricted moving average covariance estimators where the covariance between assets 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 is 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 regression residuals.

Imposing a factor structure on the covariance has one key advantage: factor covariance estimators are positive definite when the number of periods used to estimate the factor covariance is larger than the number of factors ($n > p$). The standard moving average covariance estimator is only positive definite when the number of observations is larger than the number of assets ($n > k$). This feature facilitates application of factor covariance estimators in very large portfolios.

Structure can be imposed on the factor loadings estimator to improve covariance estimates in heterogeneous portfolios. Loadings on unrelated factors can be restricted to zero. For example, suppose a portfolio hold of equity and credit instruments, and that a total of 5 factors are used to model the covariance – on common to all assets, two specific to equities and two specific to bonds. The factor covariance is a 5 by 5 matrix, and the factor loadings for all assets have only three non-zero coefficients: the common factor and two asset-class specific factors. Zero restrictions on the factor loadings allow for application to large, complex portfolios, even in cases where many factors are needed to capture the systematic risk components in the portfolio.

9.3.4 Principal Component Covariance

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

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

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 principal component objective function can be computed from an eigenvalue decomposition of the outer product of $\mathbf{Y}$, $\boldsymbol{\Upsilon} = \mathbf{Y}'\mathbf{Y} = \sum_{t=1}^{T}\mathbf{y}_t\mathbf{y}_t'$.

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 (Eigenvalues)**.** 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 function.

Definition 9.8 (Eigenvectors)**.** A k by 1 vector $\mathbf{u}$ is an eigenvector corresponding to an eigenvalue λ 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* $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}'$ *where* $\mathbf{U}$ *is an orthonormal matrix* ($\mathbf{U}' = \mathbf{U}^{-1}$) *containing the eigenvectors of* $\mathbf{A}$ *in its columns and* $\mathbf{\Lambda}$ *is a diagonal matrix with the eigenvalues* $\lambda_1, \lambda_2, \ldots, \lambda_k$ *of* $\mathbf{A}$ *along its diagonal.*

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

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

$\mathbf{F} = \mathbf{Y}\mathbf{U}$ 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]

The construction of the factor returns is the only difference between PCA-based covariance estimators and factor estimators. Factor estimators use observable portfolio returns to measure common exposure. In PCA-based covariance models, the factors are estimated from the returns, and so additional assets are not needed to measure common exposures.

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. $\mathbf{\Omega}_t$ is a diagonal matrix with diagonal elements $\omega_{j,t+1}^2 = n^{-1}\sum_{i=1}^{n} \eta_{j,t-1}^2$ where $\eta_{i,t} = r_{i,t} - \mathbf{f}_t'\boldsymbol{\beta}_{i,t}$ are the residuals from a p-factor principal component analysis.

The number of factors, p, is the only parameter used to implement a PCA covariance estimator. The simple approach is to use a fixed number of factors based on experience or empirical regularities, e.g., selecting three factors when modeling with equity returns. The leading data-based approach is to select the number of factors by minimizing an information criterion such as those proposed in Bai and Ng (2002),

[3] The factors and factor loadings are only identified up to ± 1.

Principal Component Analysis of the S&P 500

$k = 378$	1	2	3	4	5	6	7	8	9	10
Partial R^2	0.327	0.038	0.035	0.025	0.023	0.018	0.015	0.010	0.010	0.008
Cumulative R^2	0.327	0.366	0.401	0.426	0.449	0.467	0.482	0.492	0.502	0.510

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 are included if the asset is in the S&P 500 for 50% of the sample (k reports the number of firms that satisfy this criterion). The second line contains the cumulative R^2 of a p-factor model for the first 10 factors.

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

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. The Bai and Ng information criterion is similar to other information criteria such as the HQIC or BIC. The first term, $\ln(V(p, \hat{\mathbf{f}}^p))$, measures the fit of a p-component model. Increasing p always improves the fit, and $p = \max(k, T)$ always perfectly explains the observed data. The second term, $p\frac{k+T}{kT}\ln\left(\frac{kT}{k+T}\right)$, is a penalty that increases in p. Trading off these two leads to a consistent choice of p in data sets that are both long (large T) and wide (large k).

9.3.4.1 Interpreting the components

Factors extracted using PCA 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 are 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 are 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 30-50% of the total variation in a large panel of equity returns.

9.3.4.2 Alternative methods

Principal components are often 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 using low-frequency returns (e.g., a 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 place more weight on the high variance series – fitting these high variance series produces the largest decrease in overall residual variance and the largest in R^2 for a fixed p. Using the correlation focuses the PCA estimator on the common (or systemic) variation rather than the variation of a small number of high variance asset returns.

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 σ_i and σ_j are the volatilities of assets i and j, respectively. The correlation parameter is *not* indexed by i or j, and it is common to all assets. This estimator is misspecified whenever $k > 2$, and is generally only appropriate for assets where the majority of the pairwise correlations are homogeneous and positive.[4]

[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.

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 ρ_t is estimated using one of the estimators below.

The equicorrelation can be estimated using a moment-based estimator or a maximum-likelihood estimator. Define $\epsilon_{p,t}$ as the equally weighted portfolio return. It is 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 are identical. The moment-based estimator replaces population values with estimates,

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

and the equicorrelation is estimated using

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

Maximum likelihood, assuming returns are multivariate Gaussian, can alternatively be used to estimate the equicorrelation using standardized residuals $u_{j,t} = \epsilon_{j,t}/\sigma_{j,t}$. The estimator for ρ 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 ρ. This likelihood is computationally similar to univariate likelihood for any k and so maximization is very fast even when k is large.[5]

[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

$k = 378$	Equicorrelation	1-Factor R^2 (S&P 500)	3-Factor R^2 (Fama-French)
	0.291	0.282	0.313

Table 9.2: Full sample correlation measures of the S&P 500. Returns on an asset are included if the asset is in the S&P 500 for more than 50% of the sample (k reports the number of firms that satisfy this criterion). The 1-factor R^2 is from a model using the return on the S&P 500, and the 3-factor R^2 is from a model that uses the returns on the 3 Fama-French portfolios.

9.3.6 Application: S&P 500

The S&P 500 is used to illustrate the moving-average covariance estimators. The CRSP database provides daily return data for all constituents of the S&P 500. The sample runs from January 1, 1984, until December 31, 2018. The returns on firms are included in the data set is available for at least 50% of the sample.[6]

Table 9.1 contains the number of assets which meet this criterion (k) and both the partial and cumulative R^2 for the first 10 principal components. The first explains a substantial amount of the data (32.7%) and the next four combine to explain 42.6% of the cross-sectional variation. If returns did not follow a factor structure, then each principal component is expected to explain approximately 0.25% of the variation. Table 9.2 contains the full-sample equicorrelation, 1-factor R^2 using the S&P 500 index as the observable factor and the 3-factor R^2 using the 3 Fama-French portfolios as factors.[7] The average correlation and the 1-factor fit is similar to that in the 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. The difference between the 1- and 3-factor observable and PCA models is due to the lack of cross-sectional variation in firm size among the components of the S&P 500 when compared to all assets in CRSP.

Figure 9.2 contains a plot of the 1-year moving average equicorrelation and 1- and 3-factor PCA R^2. Each component asset is included in the calculation if all returns are present in the 1-year window. Periods of high volatility, such as the end of the dot-com bubble and late 2008, also have a 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, and only the levels differ. PCA selects the factors to maximizes the fit in the cross-section, and so must produce a higher R^2 than the observable models for a given number of factors.

[6] The Expectations-Maximization algorithm allows PCA to be applied in data sets containing missing values. The algorithm begins with a guess for the missing values, usually the mean of the non-missing values for each variable. The augmented data set is then used to estimate a p factor model (the *maximization* step). The missing values are then replaced with the fitted p components (the *expectations* step). These two steps are repeated until the process converges in the sense that the change in the fitted values for the missing coefficients is small.

[7] These estimators are computed using missing values. The observable factor models are estimated using only the common sample where the factor and the individual asset are present. The equicorrelation is estimated by standardizing each series to have mean 0 and unit variance, and then computing the MLE of the correlation of these values treated as if they are a single series.

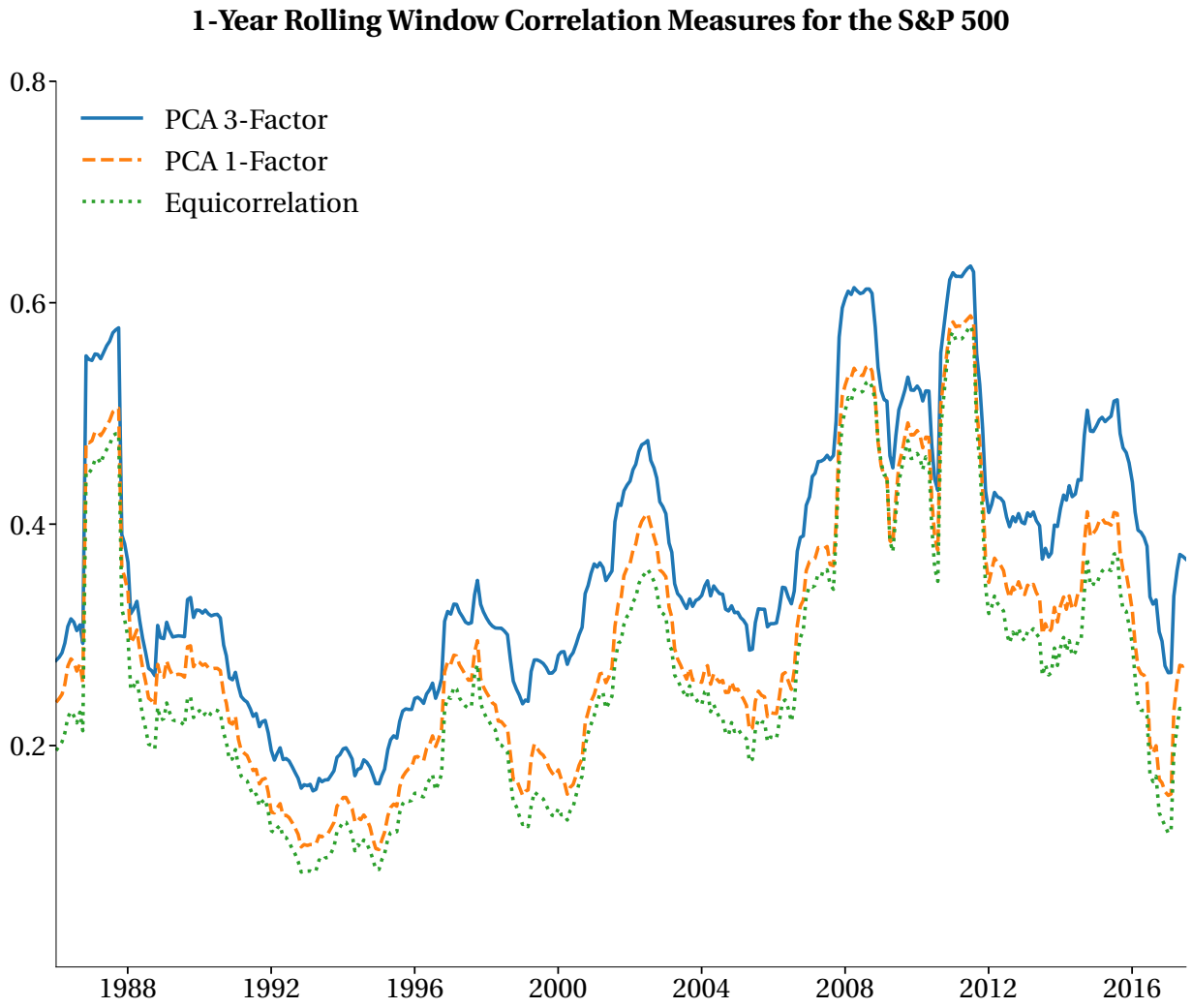

Figure 9.2: Three views of the average correlation of the S&P 500. The PCA measures are the R^2 of models with 1 and 3 factors. Each estimate is computed using a 1-year rolling window and is plotted against the center of the rolling window. All three measures roughly agree about the changes in the average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

9.4 Multivariate ARCH Models

9.4.1 Vector GARCH (*vec*)

The Vector GARCH model uses a specification that naturally extends the univariate GARCH model to a model of the conditional covariance (Bollerslev, Engle, and Wooldridge, 1988). The model is defined using the vec of the conditional covariance, which stacks the elements of the covariance into a vector.

Definition 9.11 (Vector GARCH)**.** The covariance in a vector GARCH(1,1) model (*vec*) evolves

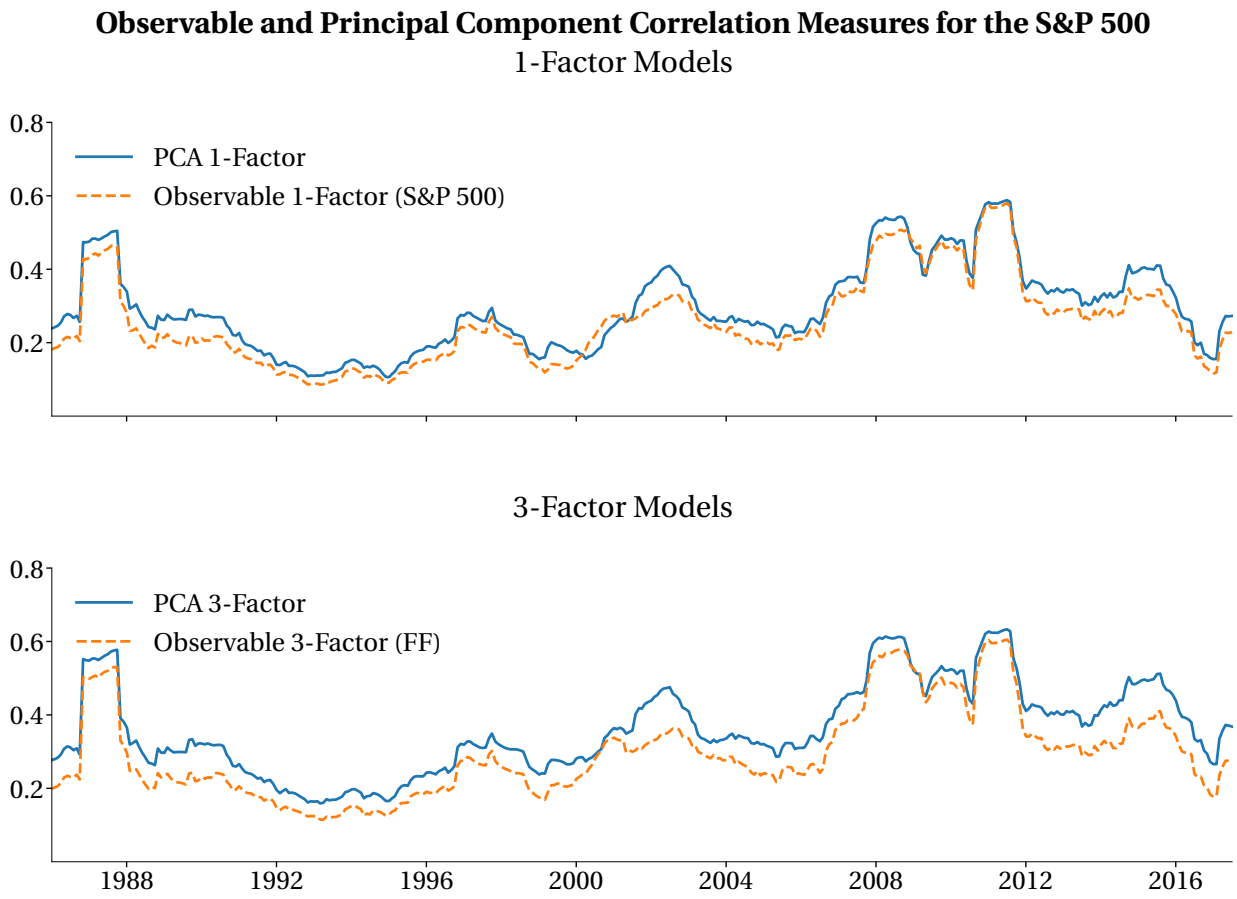

Figure 9.3: The top panel plots the R^2 for 1-factor PCA and an observable factor model which uses the return on the S&P 500 as the observable factor. The bottom contains the same for 3-factor PCA and the Fama-French 3-factor model. Each estimate is computed using a 1-year rolling window and is plotted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

according to

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

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. $\boldsymbol{\Sigma}_{t-1}^{1/2}$ 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 square or cross-product to influence each term in the conditional covariance. To understand the richness of the specification, consider the evolution of the conditional covariance in a bivariate model,

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

The repeated elements are needed to ensure that the conditional covariance is symmetric.

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 both past squared returns and the cross-product. In practice, it is difficult to use the vector GARCH model since it is challenging to determine the restrictions on $\mathbf{A}$ and $\mathbf{B}$ necessary to guarantee that $\mathbf{\Sigma}_t$ is positive definite.

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 be diagonal matrices so that the elements of $\mathbf{\Sigma}_t$ evolve according to

$$\mathbf{\Sigma}_t = \mathbf{C} + \tilde{\mathbf{A}} \odot \boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}'_{t-1} + \tilde{\mathbf{B}} \odot \mathbf{\Sigma}_{t-1} \tag{9.21}$$

where $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{B}}$ are symmetric parameter matrices and $\odot$ the is Hadamard product operator.[8] All elements of $\mathbf{\Sigma}_t$ evolve using GARCH(1,1)-like dynamics, 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 on the parameters to ensure that the conditional covariance is positive definite. Ding and Engle (2001) develop one set of sufficient constraints on the parameters in the Matrix GARCH model (see section 9.4.3).

9.4.2 BEKK GARCH

The BEKK (Baba, Engle, Kraft, and Kroner) GARCH model directly addresses the difficulties in determining constraints on the parameters in a vec specification (Engle and Kroner, 1995). BEKK models rely on two results from linear algebra to ensure that the conditional covariance is positive definite: quadratic forms are positive semi-definite, and the sum of a positive semi-definite matrix and a positive definite matrix is positive definite.

8

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}.

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

$$\boldsymbol{\Sigma}_t = \mathbf{C}\mathbf{C}' + \mathbf{A}\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1}'\mathbf{A}' + \mathbf{B}\boldsymbol{\Sigma}_{t-1}\mathbf{B}' \tag{9.22}$$

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

The BEKK is 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,

$$\text{vec}(\boldsymbol{\Sigma}_t) = \text{vec}\left(\mathbf{C}\mathbf{C}'\right) + \mathbf{A} \otimes \mathbf{A}\text{vec}\left(\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1}'\right) + \mathbf{B} \otimes \mathbf{B}\text{vec}(\boldsymbol{\Sigma}_{t-1}). \tag{9.23}$$

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

$$\begin{aligned}\begin{bmatrix} \sigma_{11,t} & \sigma_{12,t} \\ \sigma_{12,t} & \sigma_{22,t} \end{bmatrix} &= \begin{bmatrix} c_{11} & 0 \\ c_{12} & c_{22} \end{bmatrix}\begin{bmatrix} c_{11} & 0 \\ c_{21} & c_{22} \end{bmatrix}' \\ &+ \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\begin{bmatrix} \epsilon_{1,t-1}^2 & \epsilon_{1,t-1}\epsilon_{2,t-1} \\ \epsilon_{1,t-1}\epsilon_{2,t-1} & \epsilon_{2,t-1}^2 \end{bmatrix}\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}' \\ &+ \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}\begin{bmatrix} \sigma_{11,t-1} & \sigma_{12,t-1} \\ \sigma_{12,t-1} & \sigma_{22,t-1} \end{bmatrix}\begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}'.\end{aligned} \tag{9.24}$$

The conditional variance of the first asset is

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

The other conditional variance and the conditional covariance have similar forms that depend on both squared returns and the cross-product. Estimation of full BEKK models is difficult in portfolios with only a moderate number of assets since 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 the growth rate in 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

$$\boldsymbol{\Sigma}_t = \mathbf{C}\mathbf{C}' + \tilde{\mathbf{A}}\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1}'\tilde{\mathbf{A}}' + \tilde{\mathbf{B}}\boldsymbol{\Sigma}_{t-1}\tilde{\mathbf{B}}'. \tag{9.25}$$

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

$$\sigma_{ij,t} = \tilde{c}_{ij} + a_i a_j \epsilon_{i,t-1}\epsilon_{j,t-1} + b_i b_j \sigma_{ij,t-1} \tag{9.26}$$

where $\tilde{c}_{ij}$ is the ij$^{\text{th}}$element of $\mathbf{C}\mathbf{C}'$. This specification is similar to the diagonal vec except that the parameters are shared across 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

$$\boldsymbol{\Sigma}_t = \mathbf{C}\mathbf{C}' + a^2\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1} + b^2\boldsymbol{\Sigma}_{t-1} \tag{9.27}$$

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 be covariance targeted, which simplifies parameter estimation. Covariance targeting replaces the intercept ($\mathbf{C}\mathbf{C}'$) with a consistent estimator, $(1 - a^2 - b^2)\overline{\boldsymbol{\Sigma}}$, where $\overline{\boldsymbol{\Sigma}}$ is the long-run covariance, $\mathrm{E}[\boldsymbol{\Sigma}_t]$. $\overline{\boldsymbol{\Sigma}}$ is estimated using the outer product of returns, $\widehat{\overline{\boldsymbol{\Sigma}}} = T^{-1}\sum_{t=1}^{T}\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t'$. The two remaining parameters, a and b, are then estimated conditioning on the estimate of the unconditional covariance of returns,

$$\boldsymbol{\Sigma}_t = (1 - a^2 - b^2)\widehat{\overline{\boldsymbol{\Sigma}}} + a^2\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1} + b^2\boldsymbol{\Sigma}_{t-1}. \tag{9.28}$$

This 2-step estimator reduces the number of parameters that need to be simultaneously estimated using numerical methods from $2 + k(k+1)/2$ to 2. The reduction in the parameter space allows covariance-targeted scalar BEKK models to be applied in large portfolios ($k > 50$).

9.4.3 Matrix GARCH (M-GARCH)

Matrix GARCH imposes structure on the parameters of a diagonal vec that ensure that the estimated conditional covariances are positive definite (Ding and Engle, 2001).

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

$$\boldsymbol{\Sigma}_t = \mathbf{C}\mathbf{C}' + \mathbf{A}\mathbf{A}' \odot \boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1}' + \mathbf{B}\mathbf{B}' \odot \boldsymbol{\Sigma}_{t-1} \tag{9.29}$$

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

Ding and Engle (2001) show that if $\mathbf{U}$ and $\mathbf{V}$ are positive semi-definite matrices, then $\mathbf{U} \odot \mathbf{V}$ is also. Combining this result with the result that quadratic forms are positive semi-definite ensures that $\boldsymbol{\Sigma}_t$ is positive definite if $\mathbf{C}$ has full rank. The diagonal Matrix GARCH, which restricts $\mathbf{A}$ and $\mathbf{B}$ to be vectors, is equivalent to the diagonal BEKK model.

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

$$\boldsymbol{\Sigma}_t = \mathbf{C}\mathbf{C}' + \mathbf{a}\mathbf{a}' \odot \boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1}' + \mathbf{b}\mathbf{b}' \odot \boldsymbol{\Sigma}_{t-1} \tag{9.30}$$

where $\mathbf{C}$ is a lower triangular matrix 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 modeling the conditional covariance. CCC GARCH decomposes the conditional covariance into k conditional variances and the conditional correlation, which is assumed to be constant,

$$\mathbf{\Sigma}_t = \mathbf{D}_t \mathbf{R} \mathbf{D}_t. \tag{9.31}$$

$\mathbf{D}_t$ is a diagonal matrix composed of the conditional standard deviations,

$$\mathbf{D}_t = \begin{bmatrix} \sigma_{1,t} & 0 & 0 & \dots & 0 \\ 0 & \sigma_{2,t} & 0 & \dots & 0 \\ 0 & 0 & \sigma_{3,t} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \sigma_{k,t} \end{bmatrix} \tag{9.32}$$

where $\sigma_{i,t} = \sqrt{\sigma_{ii,t}}$ is the standard deviation the i$^{\text{th}}$asset return. The conditional variances are typically modeled using GARCH(1,1) models,

$$\sigma_{ii,t} = \omega_i + \alpha_i \epsilon_{i,t-1}^2 + \beta_i \sigma_{ii,t-1} \tag{9.33}$$

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

$$\mathbf{R} = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} & \dots & \rho_{1k} \\ \rho_{12} & 1 & \rho_{23} & \dots & \rho_{2k} \\ \rho_{13} & \rho_{23} & 1 & \dots & \rho_{3k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{1k} & \rho_{2k} & \rho_{3k} & \dots & 1 \end{bmatrix}. \tag{9.34}$$

The conditional covariance matrix is constructed from 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.

$$\mathbf{\Sigma}_t = \begin{bmatrix} \sigma_{11,t} & \rho_{12}\sigma_{1,t}\sigma_{2,t} & \rho_{13}\sigma_{1,t}\sigma_{3,t} & \dots & \rho_{1k}\sigma_{1,t}\sigma_{k,t} \\ \rho_{12}\sigma_{1,t}\sigma_{2,t} & \sigma_{22,t} & \rho_{23}\sigma_{2,t}\sigma_{3,t} & \dots & \rho_{2k}\sigma_{2,t}\sigma_{k,t} \\ \rho_{13}\sigma_{1,t}\sigma_{3,t} & \rho_{23}\sigma_{2,t}\sigma_{3,t} & \sigma_{33,t} & \dots & \rho_{3k}\sigma_{3,t}\sigma_{k,t} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{1k}\sigma_{1,t}\sigma_{k,t} & \rho_{2k}\sigma_{2,t}\sigma_{k,t} & \rho_{3k}\sigma_{3,t}\sigma_{k,t} & \dots & \sigma_{kk,t} \end{bmatrix}. \tag{9.35}$$

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 applied to the standardized residuals.

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

$$\mathbf{\Sigma}_t = \begin{bmatrix} \sigma_{11,t} & \rho_{12}\sigma_{1,t}\sigma_{2,t} & \rho_{13}\sigma_{1,t}\sigma_{3,t} & \dots & \rho_{1k}\sigma_{1,t}\sigma_{k,t} \\ \rho_{12}\sigma_{1,t}\sigma_{2,t} & \sigma_{22,t} & \rho_{23}\sigma_{2,t}\sigma_{3,t} & \dots & \rho_{2k}\sigma_{2,t}\sigma_{k,t} \\ \rho_{13}\sigma_{1,t}\sigma_{3,t} & \rho_{23}\sigma_{2,t}\sigma_{3,t} & \sigma_{33,t} & \dots & \rho_{3k}\sigma_{3,t}\sigma_{k,t} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{1k}\sigma_{1,t}\sigma_{k,t} & \rho_{2k}\sigma_{2,t}\sigma_{k,t} & \rho_{3k}\sigma_{3,t}\sigma_{k,t} & \dots & \sigma_{kk,t} \end{bmatrix} \tag{9.36}$$

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

9.4.5 Dynamic Conditional Correlation (DCC)

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

Definition 9.20 (Dynamic Conditional Correlation GARCH)**.** The covariance in a Dynamic Conditional Correlation (DCC)-GARCH model evolves according to

$$\mathbf{\Sigma}_t = \mathbf{D}_t\mathbf{R}_t\mathbf{D}_t. \tag{9.37}$$

where

$$\mathbf{R}_t = \mathbf{Q}_t^*\mathbf{Q}_t\mathbf{Q}_t^*, \tag{9.38}$$

$$\begin{aligned} \mathbf{Q}_t &= (1-a-b)\overline{\mathbf{R}} + a\mathbf{u}_{t-1}\mathbf{u}_{t-1}' + b\mathbf{Q}_{t-1}, & (9.39)\\ &= (1-a-b)\overline{\mathbf{R}} + a\left(\mathbf{R}_{t-1}^{\frac{1}{2}}\mathbf{e}_{t-1}\right)\left(\mathbf{R}_{t-1}^{\frac{1}{2}}\mathbf{e}_{t-1}\right)' + b\mathbf{Q}_{t-1}, & (9.40) \end{aligned}$$

$$\mathbf{Q}_t^* = (\mathbf{Q}_t \odot \mathbf{I}_k)^{-\frac{1}{2}} \tag{9.41}$$

$\mathbf{u}_t$ 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 $\mathbf{I}_k$. $\mathbf{D}_t$ is a diagonal matrix with the conditional standard deviation of asset i on the i$^{\text{th}}$diagonal position. The conditional variance, $\sigma_{ii,t}$, $i = 1, 2, \dots, k$, evolve according to some univariate GARCH process for asset i, usually a GARCH(1,1) and are identical to eq. 9.33.

The $\mathbf{Q}_t$ processresembles a covariance targeting BEKK (eq. 9.28). Eqs. 9.38 and 9.41 are needed to ensure that $\mathbf{R}_t$ is a correlation matrix with diagonal elements equal to 1. This structure allows for three-step estimation. The first two steps are identical to those in the CCC GARCH model. The third step conditions on the estimate of the long-run correlation when estimating the parameters of the dynamics, a and b.[9]

[9]The three-step estimator is biased, even in large samples. Only two-step estimation – where the variances are first estimated, and then all correlation parameters are jointly estimated – produces consistent parameter estimates in DCC models.

9.4.6 Orthogonal GARCH (OGARCH)

The principal 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 principal components to model the conditional covariance by assuming that the factors are conditionally uncorrelated.[10]

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

$$\boldsymbol{\Sigma}_t = \boldsymbol{\beta}\boldsymbol{\Sigma}_t^f\boldsymbol{\beta}' + \boldsymbol{\Omega} \tag{9.42}$$

where $\boldsymbol{\beta}$ is the k by p matrix of factor loadings corresponding to the p factors with the highest total R^2. The conditional covariance of the factors is assumed diagonal,

$$\boldsymbol{\Sigma}_t^f = \begin{bmatrix} \psi_{1,t}^2 & 0 & 0 & \dots & 0 \\ 0 & \psi_{2,t}^2 & 0 & \dots & 0 \\ 0 & 0 & \psi_{3,t}^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \psi_{l,t}^2 \end{bmatrix}, \tag{9.43}$$

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

$$\begin{aligned} \psi_{i,t}^2 &= \varphi_i + \alpha_i f_{i,t-1}^2 + \beta_i \psi_{i,t-1}^2 && (9.44) \\ &= \varphi_i + \alpha_i \psi_{i,t-1}^2 e_{t,t-1}^2 + \beta_i \psi_{i,t-1}^2 && (9.45) \end{aligned}$$

where $\{\mathbf{e}_t\}$ is a sequence of i.i.d. innovations with mean $\mathbf{0}$ and covariance $\mathbf{I}_k$.

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

$$\boldsymbol{\Omega} = \begin{bmatrix} \omega_1^2 & 0 & 0 & \dots & 0 \\ 0 & \omega_2^2 & 0 & \dots & 0 \\ 0 & 0 & \omega_3^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \omega_l^2 \end{bmatrix}, \tag{9.46}$$

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

$$\omega_i^2 = \sum_{t=1}^{T} \eta_{i,t}^2 = \sum_{t=1}^{T} \left(\epsilon_{i,t} - \mathbf{f}_t\boldsymbol{\beta}_i\right)^2. \tag{9.47}$$

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

[10] Principal components are estimated using the outer-product or the unconditional covariance of returns, and so only guarantee that the factors are unconditionally uncorrelated.

9.4.7 Conditional Asymmetries

Standard multivariate ARCH models are symmetric since they only depend on the outer product of returns, and so have news impact curves that are identical for $\boldsymbol{\epsilon}_t$ and $-\boldsymbol{\epsilon}_t$. Most models can be modified to allow for conditional asymmetries in covariance, a feature that may be important when modeling returns in some asset classes, e.g., equities. 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

$$\boldsymbol{\Sigma}_t = \mathbf{C}\mathbf{C}' + \mathbf{A}\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1}'\mathbf{A}' + \mathbf{G}\boldsymbol{\zeta}_{t-1}\boldsymbol{\zeta}_{t-1}'\mathbf{G}' + \mathbf{B}\boldsymbol{\Sigma}_{t-1}\mathbf{B}' \tag{9.48}$$

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

An asymmetric version of Matrix GARCH can be similarly constructed so that

$$\boldsymbol{\Sigma}_t = \mathbf{C}\mathbf{C}' + \mathbf{A}\mathbf{A}' \odot \boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1}' + \mathbf{G}\mathbf{G}' \odot \boldsymbol{\zeta}_{t-1}\boldsymbol{\zeta}_{t-1}' + \mathbf{B}\mathbf{B}' \odot \boldsymbol{\Sigma}_{t-1} \tag{9.49}$$

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

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

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 driven by cross products of returns. The asymmetry only has an effect on the conditional covariance between two assets if both markets experience "bad" news (negative returns). Cappiello, Engle, and Sheppard (2006) propose an asymmetric extension to the DCC model.

9.4.8 Fitting Multivariate GARCH Models

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

$$f(\boldsymbol{\epsilon}_t; \boldsymbol{\theta}) = (2\pi)^{-\frac{k}{2}} |\boldsymbol{\Sigma}_t|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\boldsymbol{\epsilon}_t'\boldsymbol{\Sigma}_t^{-1}\boldsymbol{\epsilon}_t\right) \tag{9.50}$$

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 is relatively small (less than 10) or when the model is tightly parameterized (e.g., scalar BEKK). The log-likelihood in larger, more complex models is difficult to optimize 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 evaluating the log-likelihood is increasing in the number of unknown parameters and grows at rate k^3 in most multivariate ARCH models.

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

CCC GARCH Correlation

	Large Cap	Small Cap	Bond
Large Cap	1	0.718	−0.258
Small Cap	0.718	1	−0.259
Bond	−0.258	−0.259	1

Unconditional Correlation

	Large Cap	Small Cap	Bond
Large Cap	1	0.803	−0.306
Small Cap	0.803	1	−0.305
Bond	−0.306	−0.305	1

Table 9.3: The top panel reports the estimates of the conditional correlation from a CCC GARCH model for three mutual funds spanning large-cap stocks (OAKMX), small-cap stocks (FSLCX), and long government bond returns (WHOSX). The bottom panel contains the estimtes of the unconditional correlation computed from the unfiltered returns.

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

9.4.9 Application: Mutual Fund Returns

Three mutual funds are used to illustrate the differences (and similarities) of multivariate ARCH models. The three funds are:

- Oakmark I (OAKMX), a large-cap fund;
- Fidelity Small Cap Stock (FSLCX), a small-cap fund which seeks to invest in firms with capitalizations similar to those in the Russell 2000 or S&P 600; and
- Wasatch-Hoisington US Treasury (WHOSX), a fund which invests at least 90% of AUM in U.S. Treasury securities.

All data comes from the CRSP database, and data between January 1, 1998, and December 31, 2018, is used to estimate model parameters. Table 9.3 contains the estimated correlation from

the CCC-GARCH model where each volatility series is modeled using a GARCH(1,1). The correlations between these assets are large and positive for the equity funds and negative, on average, between the equity funds and the bond fund. The bottom panel reports the unconditional correlation of the returns. These values are all larger in magnitude than the conditional correlations. The conditional volatilities of the three series tend to comove, and the periods with high volatility have a disproportionate impact on the estimated covariance.

Table 9.4 contains the parameters of the dynamics of six models: the DCC, scalar BEKK, an asymmetric scalar BEKK, Matrix GARCH and the asymmetric extension of the Matrix GARCH model. The estimates of the parameters in the DCC are typical – the two parameters sum to nearly 1 and α is smaller in magnitude than the values typically found in volatility models. These estimates indicate that correlation is very persistent but less dynamic than volatility. The parameters in the scalar BEKK and asymmetric scalar BEKK are similar to what one typically finds in a volatility model, although the asymmetry is weak. The Matrix GARCH parameters are fairly homogeneous although the treasury fund is less responsive to news (i.e., has smaller coefficient in $\mathbf{AA}'$). In the asymmetric Matrix GARCH model, the response to "bad" news is not homogeneous. The equities have large asymmetries while the bond fund does not. This heterogeneity explains the small asymmetry parameter in the asymmetric scalar BEKK.

Figure 9.4 plots the annualized volatility for these series from four models: the CCC (standard GARCH(1,1)), the two RiskMetrics methodologies, and the asymmetric scalar BEKK. All volatilities are similar which is 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 correlation estimated in the CCC GARCH (which is constant), the estimated correlations are also substantially similar.

9.5 Realized Covariance

Realized Covariance uses ultra-high-frequency data (trade data) to estimate the integrated covariance over some period, usually a day. Suppose prices followed a k-variate continuous time diffusion,

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

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

$$\int_0^1 \boldsymbol{\Sigma}_s ds$$

where the bounds 0 and 1 represent the (arbitrary) interval over which the covariance is estimated. The integrated covariance is the multivariate analog of the integrated variance introduced in Chapter 7.[11]

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

[11] In the presence of jumps, Realized Covariance estimates the quadratic covariation, which is the integrated covariance plus the outer product of the jumps

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

Multivariate GARCH Model Estimates

	α	γ	β
DCC	0.009 (3.4)	–	0.990 (4.9)
Scalar BEKK	0.062 (143.0)	–	0.918 (89.6)
Asym. Scalar BEKK	0.056 (158.9)	0.021 (84.7)	0.911 (65.8)

	$\mathbf{AA}'$			$\mathbf{GG}'$			$\mathbf{BB}'$		
Matrix GARCH	0.092 (5.37)	0.090 (6.03)	0.048 (2.34)	–	–	–	0.875 (36.74)	0.885 (32.35)	0.910 (14.62)
	0.090 (6.03)	0.087 (5.06)	0.047 (2.85)	–	–	–	0.885 (32.35)	0.895 (30.99)	0.921 (20.78)
	0.048 (2.34)	0.047 (2.85)	0.043 (6.20)	–	–	–	0.910 (14.62)	0.921 (20.78)	0.947 (111.86)
Asymmetric Matrix GARCH	0.073 (2.86)	0.068 (4.33)	0.050 (2.51)	0.038 (0.99)	0.042 (1.35)	−0.007 (−0.31)	0.872 (30.58)	0.883 (23.59)	0.908 (15.76)
	0.068 (4.33)	0.063 (1.44)	0.048 (2.68)	0.042 (1.35)	0.047 (1.50)	−0.008 (−0.44)	0.883 (23.59)	0.893 (17.55)	0.919 (19.18)
	0.050 (2.51)	0.048 (2.68)	0.043 (6.38)	−0.007 (−0.31)	−0.008 (−0.44)	0.001 (0.08)	0.908 (15.76)	0.919 (19.18)	0.946 (69.10)

Table 9.4: Parameter estimates (t-stats in parenthesis) from multivariate ARCH models for three mutual funds representing distinct investment styles: small-cap stocks (FSLCX), large-cap stocks (OAKMX), and long government bond returns (WHOSX). The top panel contains results for DCC, scalar BEKK and asymmetric scalar BEKK. The bottom panel contains estimation results for Matrix GARCH and the asymmetric extension to the Matrix GARCH model.

Definition 9.22 (Realized Covariance)**.** The m-sample Realized Covariance is defined

$$RC_t^{(m)} = \sum_{i=1}^{m} \mathbf{r}_{i,t}\mathbf{r}_{i,t}' = (\mathbf{p}_{i,t} - \mathbf{p}_{i-1,t})(\mathbf{p}_{i,t} - \mathbf{p}_{i-1,t})', \tag{9.51}$$

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

Prices should, in principle, be sampled as frequently as possible to maximize the precision of the Realized Covariance estimator. In practice, frequent sampling 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, if asset i trades at 10:00:00 and the last trade of asset j occurs at 9:59:50, then the estimated covariance is biased towards 0.

The conventional method to address these two concerns is to sample relatively infrequently, for example, every 5 minutes.

where Δp_s are the jumps.

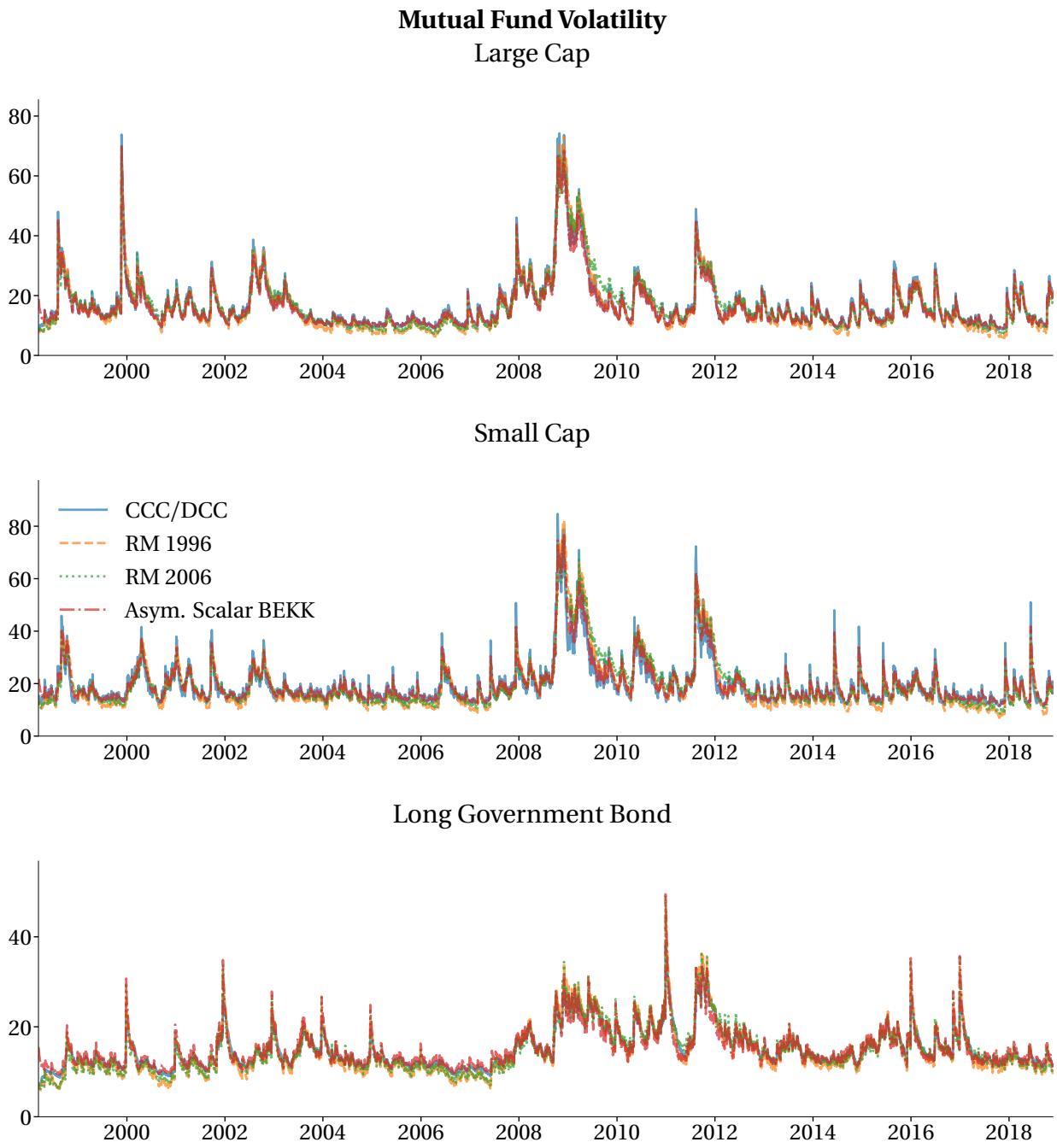

Figure 9.4: The three panels plot the estimated annualized volatility of the three mutual funds.

The standard Realized Covariance estimator can be improved using *subsampling*. For example, suppose prices are available every minute, but that microstructure concerns (noise and synchronization) limit sampling to 10-minute returns. 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 esti-

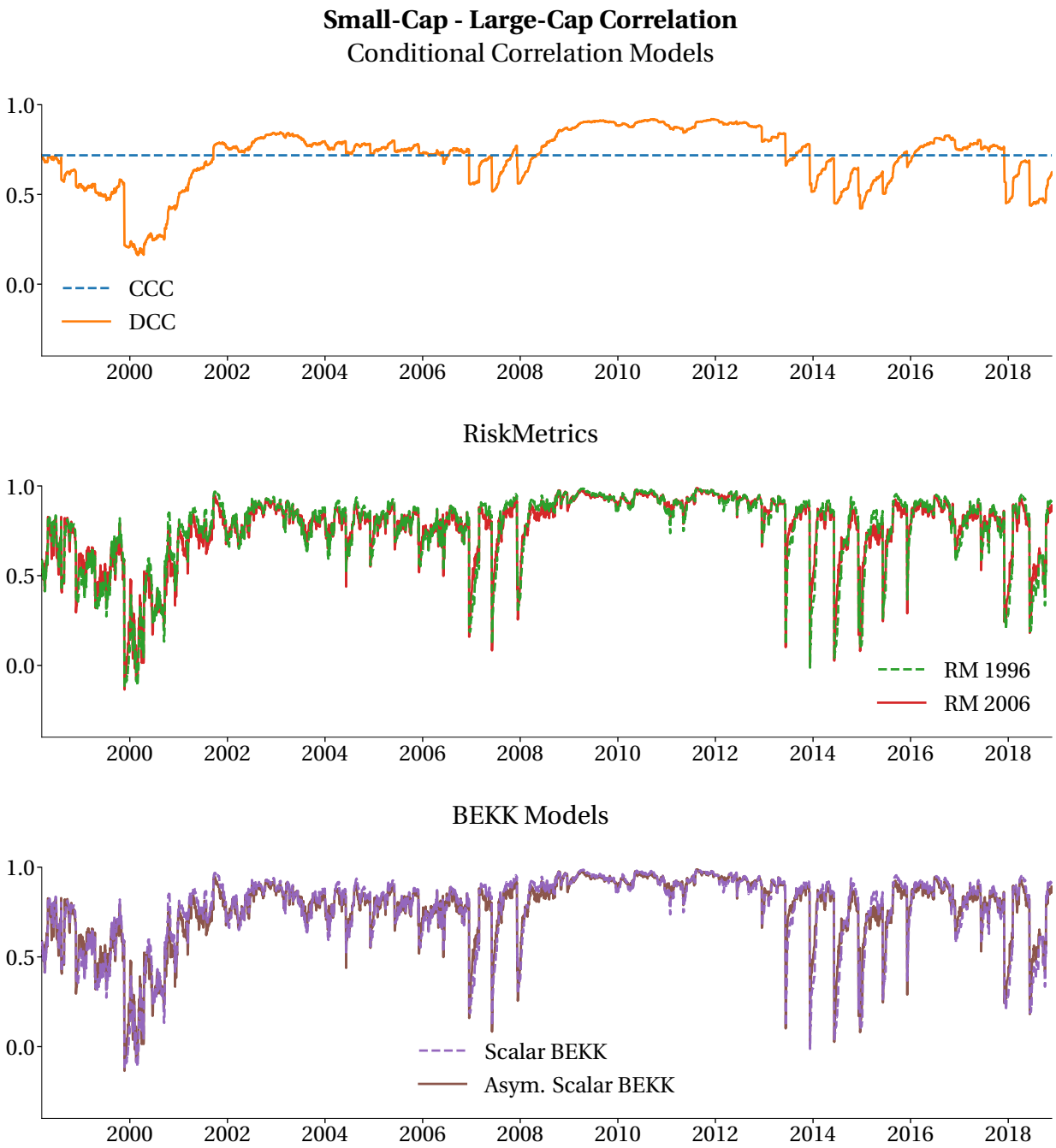

Figure 9.5: The three panels show the estimated conditional correlation between the large-cap fund and the small-cap fund from 6 models.

mator is defined

$$RC_{t,SS}^{(m,n)} = \frac{m}{n(m-n+1)} \sum_{i=1}^{m-n+1} \sum_{j=1}^{n} \mathbf{r}_{i+j-1,t} \mathbf{r}'_{i+j-1,t} \tag{9.52}$$

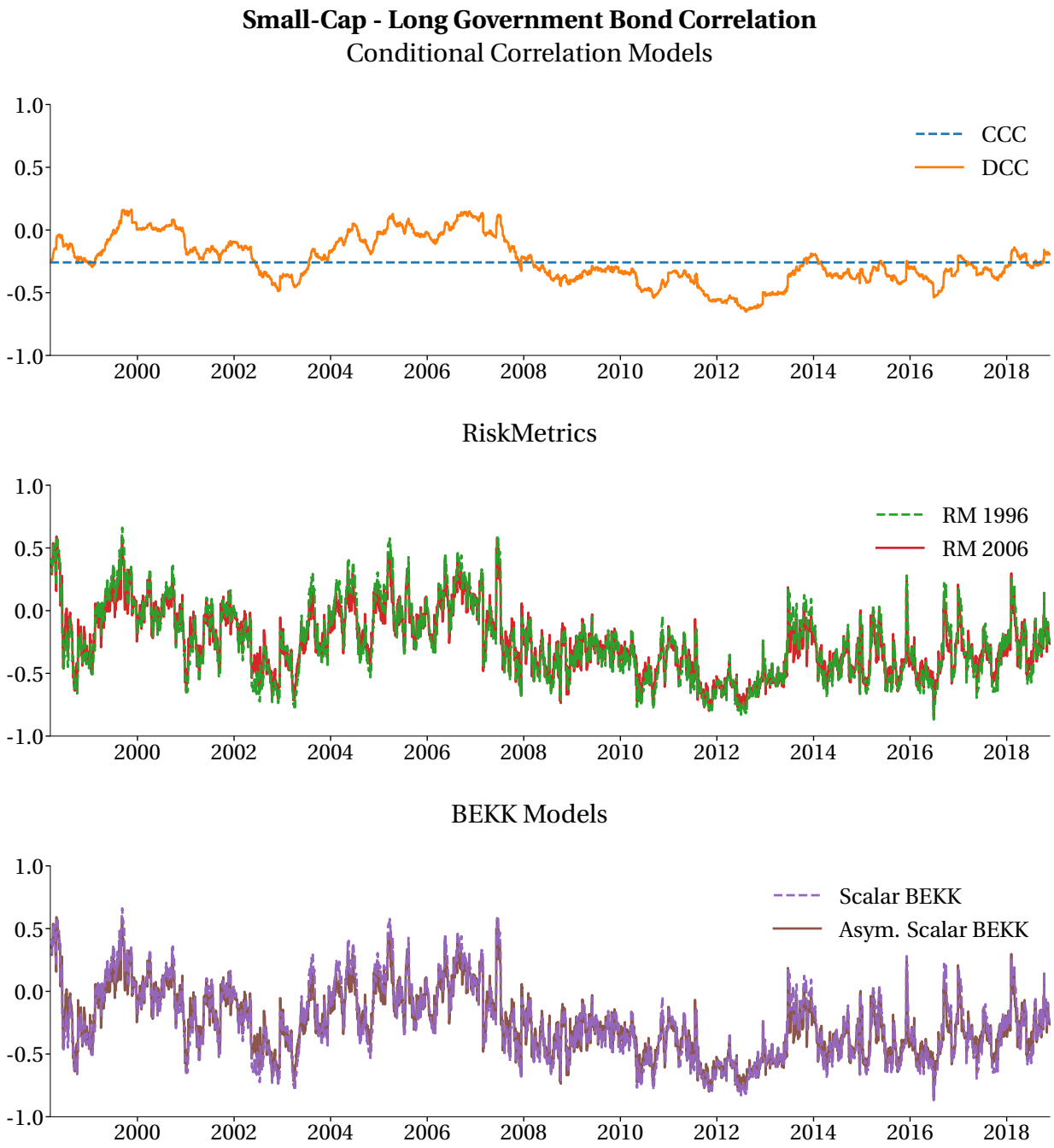

Figure 9.6: The three panels show the estimated conditional correlation between the small-cap fund and the bond fund from 6 models.

$$= \frac{1}{n}\sum_{j=1}^{n}\frac{m}{(m-n+1)}\sum_{i=1}^{m-n+1}\mathbf{r}_{i+j-1,t}\mathbf{r}'_{i+j-1,t}$$

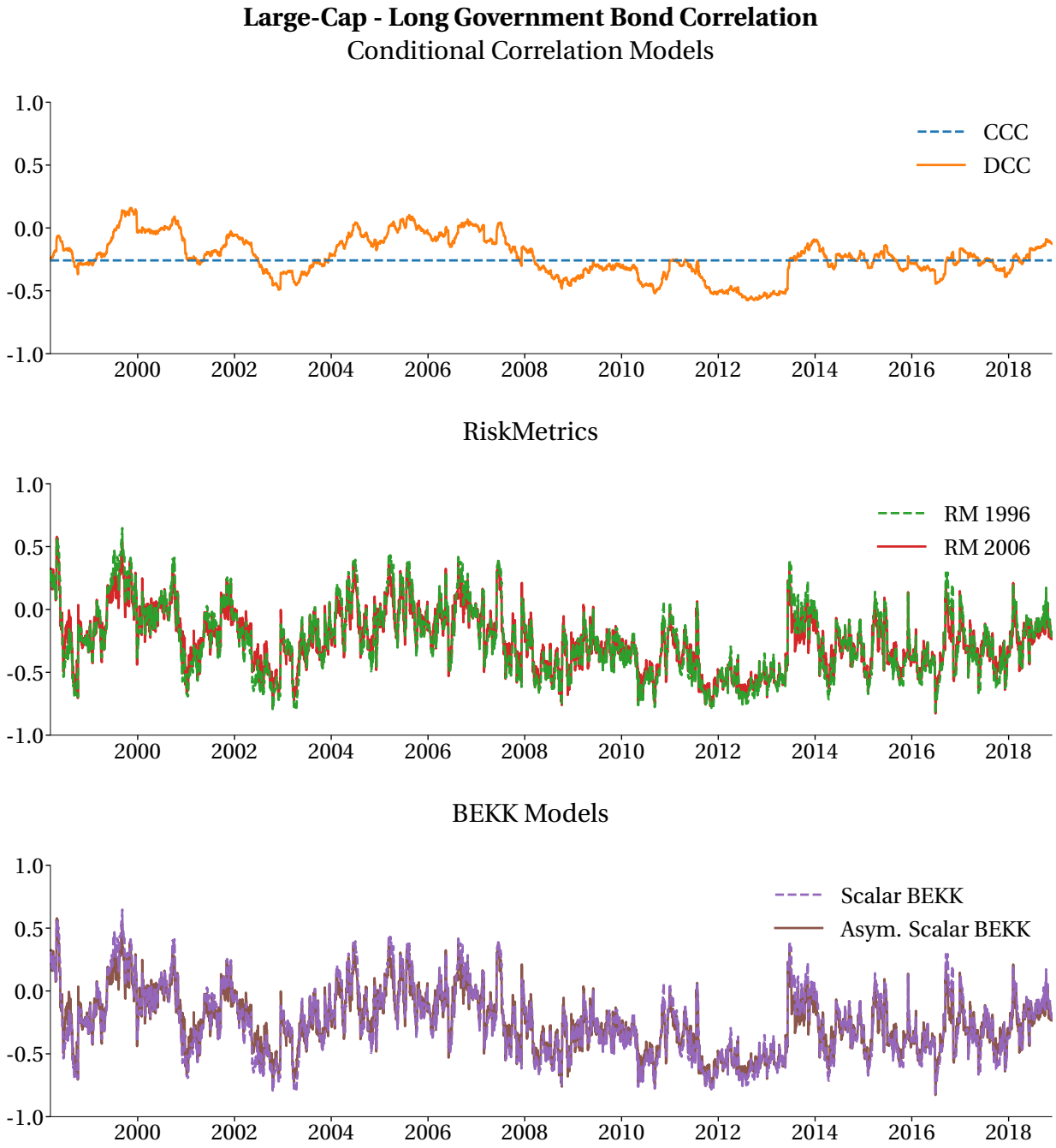

Figure 9.7: The three panels show the estimated conditional correlation between the large-cap fund and the bond fund from 6 models.

$$= \frac{1}{n} \sum_{j=1}^{n} \widetilde{RC}_{j,t},$$

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

For example, suppose data is available from 9:30:00 to 16:00:00, and that prices are sampled every minute. The standard Realized Covariance uses returns constructed from prices sampled at 9:30:00, 9:40:00, 9:50:00, ... The subsampled Realized Covariance uses returns computed from all 10-minute windows, i.e., 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 is the number of 1-minute returns available over a 6.5 hour day (390), and n is the number of 1-minute returns in the desired sampling frequency of 10-minutes (10).

Barndorff-Nielsen, Hansen, Lunde, and Shephard (2011) propose an alternative method to estimate the integrated 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 the Newey and West (1987) covariance estimator.

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

$$RK_t = \mathbf{\Gamma}_0 + \sum_{i=1}^{h} K\left(\frac{i}{H+1}\right)\left(\mathbf{\Gamma}_i + \mathbf{\Gamma}_i'\right) \tag{9.53}$$

$$\mathbf{\Gamma}_j = \sum_{i=j+1}^{\tilde{m}} \tilde{\mathbf{r}}_{i,t}\tilde{\mathbf{r}}_{i-j,t}'$$

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 constructed by sampling prices using last-price interpolation only after all assets have traded. For example, Table 9.5 contains a set of simulated trade times for SPY, a leading ETF that tracks the S&P 500, and GLD, an ETF that tracks the price of gold. A tick (✓) indicates that a trade occurs at the timestamp in the first column. A tick in the refresh column indicates that this timestamp is a refresh time. The final two columns contain the timestamps of the prices used to compute the refresh-time returns.

The recommended kernel is Parzen's kernel,

$$K(x) = \begin{cases} 1 - 6x^2 + 6x^3 & 0 > x \geq \frac{1}{2} \\ 2(1-x)^3 & \frac{1}{2} > x \geq 1 \\ 0 & x > 1 \end{cases} \tag{9.54}$$

The bandwidth parameter, H, plays a crucial role in the accuracy of Realized Kernels. A discussion of the estimation of the bandwidth is beyond the scope of these notes. See Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008) and Barndorff-Nielsen, Hansen, Lunde, and Shephard (2011) for detailed discussions.

9.5.1 Realized Correlation and Beta

Realized Correlation is the realized analog of the usual correlation estimator, and is defined using the Realized Covariance.

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

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

Trade Time	SPY	GLD	Refresh	SPY Time	GLD 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 illustrates refresh-time price construction. Prices are sampled after all assets have traded using last-price interpolation. Refresh-time sampling usually eliminated some of the data, e.g., the 9:30:03 trade of SPY and prices are not perfectly synchronized, e.g., the 9:30:08 refresh-time price which uses the SPY price from 9:30:08 and the GLD price from 9:30:07.

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 the definition of a regression β (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

$$RC = \begin{bmatrix} RC_{ii} & RC_{fi}' \\ RC_{fi} & RC_{ff} \end{bmatrix}$$

where RC_{ii} is the Realized Variance of the asset, 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 Realized 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, are accurate measures of the exposure to changes in the market.

9.5.2 Modeling Realized Covariance

Modified multivariate ARCH models can be used to modeling and forecast Realized Covariance and Realized Kernels . The basic assumption is that the *mean* of the Realized Covariance, conditional on the time $t-1$ information, is $\mathbf{\Sigma}_t$,

$$RC_t|\mathcal{F}_{t-1} \sim F(\mathbf{\Sigma}_t, \boldsymbol{\upsilon}) \tag{9.55}$$

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-valued shock which has conditional expectation $\mathbf{I}_k$,

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

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 but different from those in eq. 9.55. 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 $\xi_t \overset{\text{i.i.d.}}{\sim} F(1, \boldsymbol{\upsilon})$.

Most multivariate ARCH models can be adapted by replacing the outer product of the shocks with the Realized Covariance. For example, 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 no more difficult than the estimation of the parameters in a multivariate ARCH model, and the parameters can be estimated by maximizing the Wishart log-likelihood. See Noureldin, Shephard, and Sheppard (2012) for details.

9.5.3 Application: ETF Realized Covariance

Exchange-traded funds have emerged as popular instruments that facilitate investing in assets that are often difficult to access for retail investors. They trade like stocks but are backed by other assets or derivative securities. Three ETFs are used to highlight some of the issues unique to Realized Covariance that are not important when modeling a single asset. The funds used are the SPDR S&P 500 ETF (SPY), which tracks the S&P 500, SPDR Gold Trust (GLD), which aims to track to the spot price of gold, and iShares 7-10 Year Treasury Bond ETF (IEF), which tracks the return on intermediate maturity U.S. government debt. The data used in this application run from the start of 2008 until the end of 2018. The estimators are implemented using only transaction data (trades) that are available during the normal trading hours of 9:30 (Eastern/US) to 16:00.

Figure 9.8 shows the average number of transactions per day for the three ETFs. There are substantial differences in the liquidity of the three funds. IEF trades about 800 times per day, on average, over the sample. In some months, the average number of transaction is as low as 250, while in periods of higher liquidity the fund is traded over 1,000 times per day. The S&P 500-tracking ETF consistently trades over 80,000 times per day. The U.S. trading day last 6.5 hours and so the time between trades ranges between 30 and 90 seconds for IEF and is less than half a second for SPY. The liquidity of the least liquid asset always serves an upper bound for the

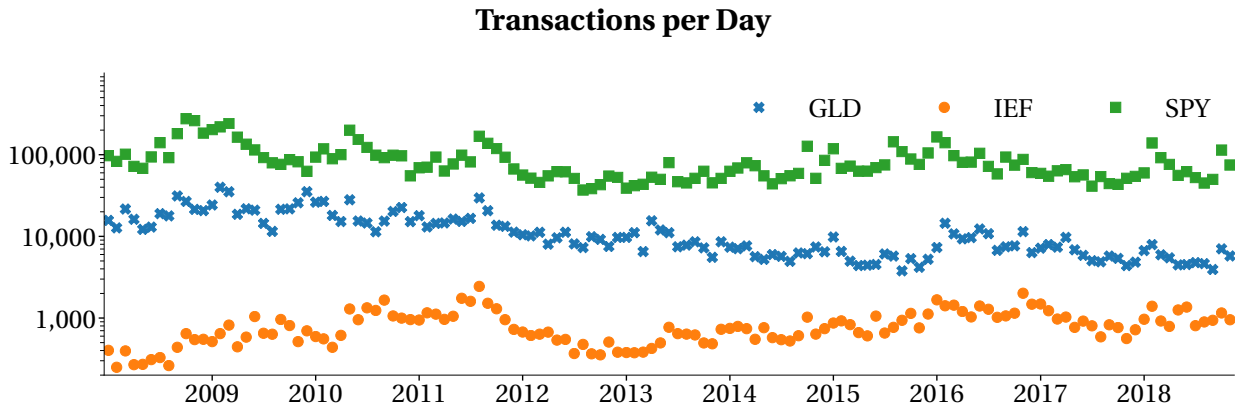

Figure 9.8: The average number of daily transactions in each month of the sample for the three ETFs: SPDR S&P 500 ETF, SPDR Gold Trust (GLD), and iShares 7-10 Year Treasury Bond ETF.

sampling frequency used when estimating RC. In this application, sampling more frequently than 30 seconds is likely to produce a sharp reduction in covariance and correlation for pairs involving IEF. GLD's liquidity is consistently between IEF and SPY and trades typically occur every 3 seconds.

Figure 9.9 contains two signature plots. The top is known as the pseudo-correlation signature and plots the time-averaged Realized Covariance standardized by the average cross-product of realized volatilities sampled at a single (conservative) frequency.

Definition 9.27 (Pseudo-Correlation Signature Plot)**.** The pseudo-correlation signature plot displays the time-series average of Realized Covariance

$$\overline{RCorr}^{(m)}_{ij,t} = \frac{T^{-1}\sum_{t=1}^{T} RC^{(m)}_{ij,t}}{\overline{RVol}_i \times \overline{RVol}_j}$$

where m is the number of samples and $\overline{RVol}_\bullet = \sqrt{T^{-1}\sum_{t=1}^{T} RC^{(q)}_{\bullet\bullet,t}}$ is a the square root of the average Realized Variance using q-samples. q is chosen to produce an accurate RV that is free from microstructure effects. An equivalent representation displays the amount of time, either in calendar time or tick time (number of trades between observations) along the x-axis.

The pseudo-correlations all diverge from 0 as the sampling interval grows. The pseudo-correlation between the Gold and the S&P 500 ETFs appears to reach its long-run level when sampling prices every 2 minutes. This sampling interval is surprisingly long considering that the slower of these two assets, GLD, trades about every 3 seconds on average. The pseudo-correlations involving the U.S. bond ETF continue to move away from 0 until the sample interval is 10 minutes, which reflects the lower liquidity in this ETF.

The slow convergence of both series is known as the Epps Effect (Epps, 1979). Epps first documented that correlations converge to 0 as the sampling frequency increases. There are two reasons why the correlations converge to zero as the sampling frequency increases: the numerator (covariance) reducing in magnitude or the denominator increasing due to bid-ask bounce. The bottom panel of Figure 9.9 plots the annualized cross-volatility signature of the two series.

The cross-volatility signatures are remarkably flat, and so the changes in the pseudo-correlation signature are due to changes in the covariances.

Definition 9.28 (Cross-volatility Signature Plot)**.** The cross-volatility signature plot displays the square-root of the time-series average of of the product of two Realized Variances,

$$\overline{XVol}_{ij,t}^{(m)} = \sqrt{T^{-1}\sum_{t=1}^{T} RV_{i,t}^{(m)}} \times \sqrt{T^{-1}\sum_{t=1}^{T} RV_{j,t}^{(m)}}$$

where m is the number of samples and $RV_{\bullet,t}^{(m)} = RC_{\bullet\bullet,t}^{(m)}$ are the diagonal elements of the Realized Covariance matrix. An equivalent representation displays the amount of time, whether in calendar time or tick time (number of trades between observations) along the X-axis. It is often presented in annualized terms,

$$\text{Ann.}\overline{XVol}_{ij,t}^{(m)} = \sqrt{252 \times \overline{XVol}_{ij,t}^{(m)}}.$$

The pseudo-correlation signature plot can be misleading if covariance does not consistently have the same sign. For example, suppose two assets have a long-run correlation near 0 but have persistent deviations where their correlation is either positive or negative for long periods. The pseudo-correlation signature may appear flat for all sampling times even though the correlation is not well estimated. An alternative is to use a R^2-signature plot which is defined by transforming the Realized Covariances into the Realized β and an idiosyncratic variance. In the model $Y_i = \alpha + \beta X_i + \epsilon_i$, the variance of the idiosyncratic residual is $V[Y] - \beta^2 V[X]$. Scaling this variance by the variance of Y produces $\frac{V[Y]-\beta^2 V[X]}{V[Y]} = 1 - R^2$.

The R^2 signature plot displays the scaled average residual variance,

$$\overline{R^2}_{ij,t}^{(m)} = 1 - \frac{T^{-1}\sum_{t=1}^{T} RC_{ii,t}^{(m)} - \left(RC_{ij,t}^{(m)}\right)^2 / RC_{jj,t}^{(m)}}{T^{-1}\sum_{t=1}^{T} RC_{ii,t}^{(m)}} = 1 - \frac{T^{-1}\sum_{t=1}^{T} RC_{ii,t}^{(m)} - \left(R\beta_{ij,t}^{(m)}\right)^2 RC_{jj,t}^{(m)}}{T^{-1}\sum_{t=1}^{T} RC_{ii,t}^{(m)}}$$

where $RC_t^{(m)}$ is the m-sample Realized Covariance in month j.

The R^2 could be low if the variance components were large, which may happen if the returns are contaminated by market microstructure noise, or if the β is not accurately measured.

Finally, Figure 9.10 shows the estimated correlations of these ETFs. The dashed line shows the average correlation computed by transforming the time-averaged Realized Covariance to a correlation. The S&P 500 tracking ETF is negativity correlated, on average, with both Gold and U.S. Treasuries. The correlation is unusually low near the start of 2016 and is also near the bottom of its range at the end of 2018. Gold and Treasuries, on the other hand, were highly correlated in 2017 and only returned to their long-run level towards the end of the sample.

9.6 Measuring Dependence

Covariance does not completely characterize the dependence between asset returns. It only measures the linear dependence between the returns and so may be misleading if assets have nonlinear relationships.

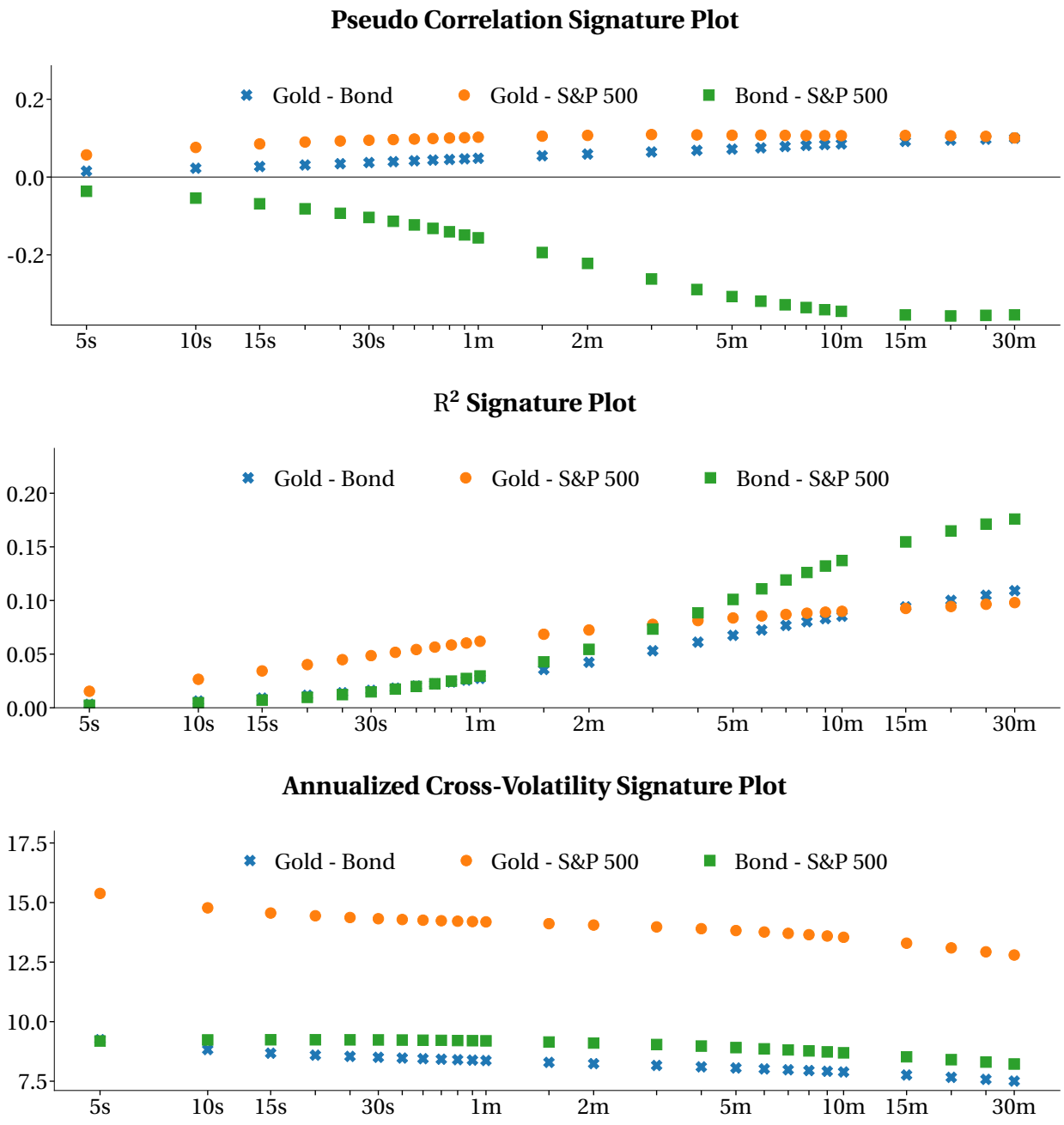

Figure 9.9: The top panel contains the pseudo-correlation signature for the three assets pairs defined as the ratio of the average covariance sampled at different frequencies standardized by a single, fixed-sampling interval cross-product of volatilities. The middle panel plots an alternative sign-free signature plot constructed by squaring the realized correlations. The bottom plot shows the average annualized cross-volatility for average Realized Variances sampled at different frequencies.

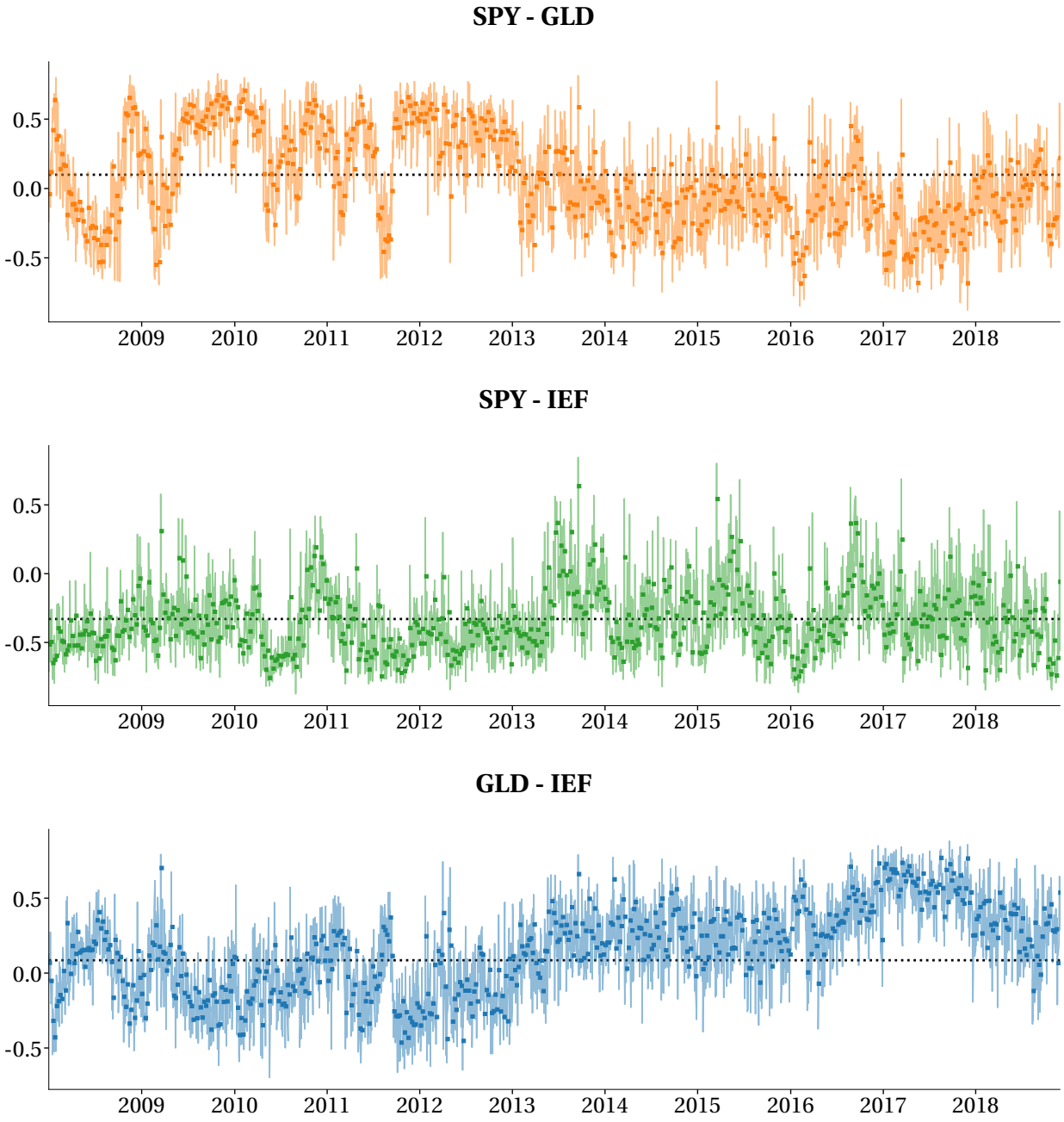

Figure 9.10: Plot of the Realized Correlations between the three ETFs: SPDR S&P 500, SPDR Gold Trust, and iShares 7-10 Year Treasury Bond ETF. All realized correlations are estimated using RC^{SS} based on 15-minute returns subsampled from prices sampled every 5 seconds ($m = 4,680$, $n = 26$). The markers show the weekly realized correlation computed from weekly-averaged Realized Covariances.

9.6.1 Linear Dependence

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

Definition 9.29 (Linear (Pearson) Correlation)**.** The linear (Pearson) correlation between two random variables X and Y is

$$\rho = \frac{\text{Cov}[X, Y]}{\sqrt{\text{V}[X]\text{V}[Y]}}. \tag{9.56}$$

The sample correlation estimator is

$$\hat{\rho} = \frac{\sum_{t=1}^{T}(X_t - \hat{\mu}_x)\left(Y_t - \hat{\mu}_y\right)}{\sqrt{\sum_{t=1}^{T}(Y_s - \hat{\mu}_x)^2 \sum_{s=1}^{T}\left(Y_t - \hat{\mu}_y\right)^2}}. \tag{9.57}$$

where $\hat{\mu}_x$ and $\hat{\mu}_y$ are the sample means of X_t and Y_t, respectively.

Linear correlation measures the strength of the linear relationship between standardized versions of X and Y. Correlation is invariant to affine increasing transformations of X or Y (i.e., $a + b\,Y$,$b > 0$). 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 when X and Y follow a bivariate normal distribution. Moreover, two distributions can have the same correlation yet have different behavior during extreme events.

9.6.2 Non-linear Dependence

Many measures have been designed to overcome the shortcomings of linear 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 λ 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 λ increases, the correlation becomes arbitrarily small despite the perfect dependence between X and Y. Rank correlation is robust to increasing non-linear transformations, and the rank correlation between X and Y is 1 for any odd power λ.

Definition 9.30 (Rank (Spearman) Correlation)**.** The rank (Spearman) correlation between two random variables X and Y is

$$\rho_s(X, Y) = \text{Corr}\left(F_X(X), F_Y(Y)\right) = \frac{\text{Cov}\left[F_X(X), F_Y(Y)\right]}{\sqrt{\text{V}\left[F_X(X)\right]\text{V}\left[F_Y(Y)\right]}} = 12\text{Cov}\left[F_X(X), F_Y(Y)\right] \tag{9.58}$$

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

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 analog 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,

$$\hat{\rho}_s = \frac{\sum_{i=1}^n \left(\frac{R_{X,i}}{n+1} - \frac{1}{2}\right)\left(\frac{R_{Y,i}}{n+1} - \frac{1}{2}\right)}{\sqrt{\sum_{i=1}^n \left(\frac{R_{X,i}}{n+1} - \frac{1}{2}\right)^2}\sqrt{\sum_{j=1}^n \left(\frac{R_{Y,i}}{n+1} - \frac{1}{2}\right)^2}} = \frac{\sum_{i=1}^n \left(R_{X,i} - \frac{n+1}{2}\right)\left(R_{Y,i} - \frac{n+1}{2}\right)}{\sqrt{\sum_{i=1}^n \left(R_{X,i} - \frac{n+1}{2}\right)^2}\sqrt{\sum_{j=1}^n \left(R_{Y,j} - \frac{n+1}{2}\right)^2}}$$

where $\frac{R_{X,i}}{n+1}$ is the empirical quantile of X_i.

9.6.2.2 Kendall's τ

Kendall's τ is an alternative measure of non-linear dependence which is based on the idea of concordance. Concordance is defined using the signs of pairs of random variables.

Definition 9.31 (Concordant Pair)**.** The pairs of random variables (X_i, Y_i) and (X_j, Y_j) are concordant if $\text{sgn}\left(X_i - X_j\right) = \text{sgn}\left(Y_i - Y_j\right)$ where $\text{sgn}(\cdot)$ is the sign function which returns -1 for negative values, 0 for zero, and +1 for positive values (equivalently defined as $\text{sgn}\left(\left(X_i - X_j\right)\left(Y_i - Y_j\right)\right)$).

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

Definition 9.32 (Kendall's τ)**.** Kendall τ is defined

$$\tau = \Pr\left(\text{sgn}\left(X_i - X_j\right) = \text{sgn}\left(Y_i - Y_j\right)\right) - \Pr\left(\text{sgn}\left(X_i - X_j\right) \neq \text{sgn}\left(Y_i - Y_j\right)\right) \tag{9.59}$$

The estimator of Kendall's τ uses the sample analogs 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 τ is

$$\begin{aligned}
\tau &= \frac{n_c - n_d}{\frac{1}{2}n(n-1)} && (9.60)\\
&= \frac{n_c}{\frac{1}{2}n(n-1)} - \frac{n_d}{\frac{1}{2}n(n-1)} && (9.61)\\
&= \widehat{\Pr}\left(\text{sgn}\left(X_i - X_j\right) = \text{sgn}\left(Y_i - Y_j\right)\right) - \widehat{\Pr}\left(\text{sgn}\left(X_i - X_j\right) \neq \text{sgn}\left(Y_i - Y_j\right)\right) && (9.62)
\end{aligned}$$

where $\widehat{\Pr}$ denotes the empirical probability. Kendall's τ 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 τ 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 τ is increasing as the

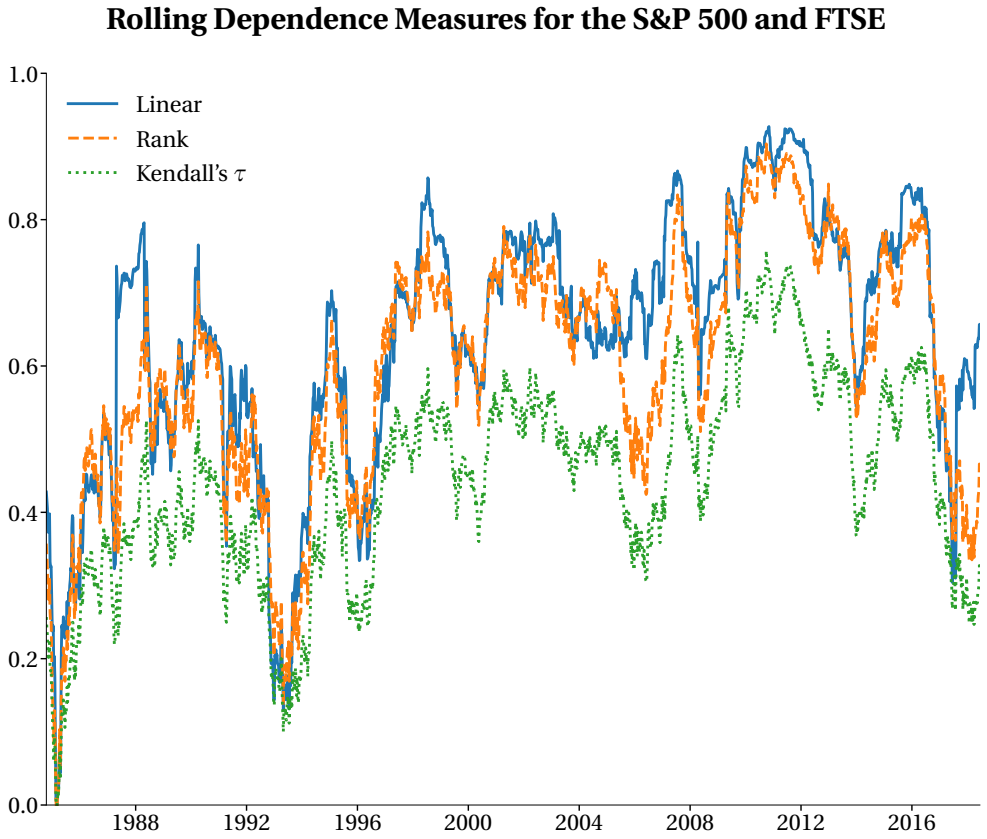

Figure 9.11: Plot of rolling linear correlation, rank correlation and Kendall's τ between weekly returns on the S&P 500 and the FTSE estimated using 1-year moving windows. The measures broadly agree about the changes in dependence but not the level.

Dependence Measures for Weekly FTSE and S&P 500 Returns

Linear (Pearson)	0.678 (0.027)	Rank (Spearman)	0.613 (0.031)	Kendall's τ	0.446 (0.027)

Table 9.6: Linear and rank correlation and Kendall's τ (bootstrap std. error in parenthesis) for weekly returns for the S&P 500 and FTSE 100.

concordance between the pairs increases. Like rank correlation, Kendall's τ is also invariant to increasing transformation since a pair that is concordant before the transformation (i.e., $X_i > X_j$ and $Y_i > Y_j$) is also concordant after a strictly increasing transformation (i.e., $T_1(X_i) > T_1(X_j)$ and $T_2(Y_i) > T_2(Y_j)$).

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. Ex-

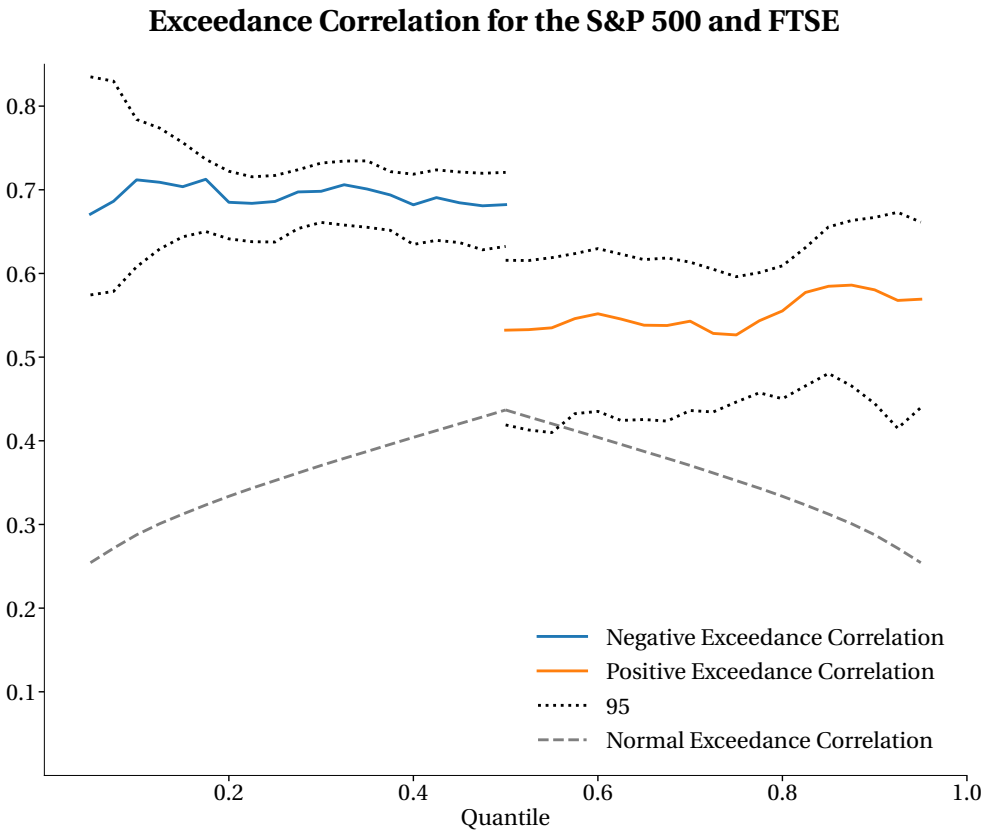

Figure 9.12: Plot of the exceedance correlations with 95% bootstrap confidence intervals for weekly returns on the S&P 500 and FTSE (each series is divided by its sample standard deviation). There is a substantial asymmetry between the positive and negative exceedance correlations.

ceedance correlation measures the correlation between the variables *conditional* on both variables taking values in their upper or lower tail.

Definition 9.33 (Exceedance Correlation)**.** The exceedance correlation at level κ is defined as

$$\rho^{+}(\kappa) = \text{Corr}\left[X, Y | X > \kappa, Y > \kappa\right] \tag{9.63}$$

$$\rho^{-}(\kappa) = \text{Corr}\left[X, Y | X < -\kappa, Y < -\kappa\right] \tag{9.64}$$

Exceedance correlation is 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 κ_X and κ_Y, which are often used if the series do not have the same variance. Series-specific thresholds 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

$$\hat{\rho}^{+}(\kappa) = \frac{\hat{\sigma}^{+}_{xy}(\kappa)}{\hat{\sigma}^{+}_{x}(\kappa)\hat{\sigma}^{+}_{y}(\kappa)}, \quad \hat{\rho}^{-}(\kappa) = \frac{\hat{\sigma}^{-}_{xy}(\kappa)}{\hat{\sigma}^{-}_{x}(\kappa)\hat{\sigma}^{-}_{y}(\kappa)} \tag{9.65}$$

where

$$\begin{aligned}
\hat{\sigma}^{+}_{XY}(\kappa) &= \frac{\sum_{t=1}^{T}\left(X_t - \mu_X^{+}(\kappa)\right)\left(Y_t - \mu_Y^{+}(\kappa)\right) I_{[X_t>\kappa \cap Y_t>\kappa]}}{T_\kappa^{+}} \\
\hat{\sigma}^{-}_{XY}(\kappa) &= \frac{\sum_{t=1}^{T}\left(X_t - \mu_X^{-}(\kappa)\right)\left(Y_t - \mu_Y^{-}(\kappa)\right) I_{[X_t<-\kappa \cap Y_t<-\kappa]}}{T_\kappa^{-}} \\
\hat{\mu}_X^{+}(\kappa) &= \frac{\sum_{t=1}^{t} X_t I_{[X_t>\kappa \cap Y_t>\kappa]}}{T_\kappa^{+}}, \quad \hat{\sigma}_X^{2+}(\kappa) = \frac{\sum_{t=1}^{t}\left(X_t - \hat{\mu}_X^{+}(\kappa)\right)^2 I_{[X_t>\kappa \cap Y_t>\kappa]}}{T_\kappa^{+}} \\
\hat{\mu}_X^{-}(\kappa) &= \frac{\sum_{t=1}^{t} X_t I_{[X_t<-\kappa \cap Y_t<-\kappa]}}{T_\kappa^{-}}, \quad \hat{\sigma}_X^{2-}(\kappa) = \frac{\sum_{t=1}^{t}\left(X_t - \hat{\mu}_X^{-}(\kappa)\right)^2 I_{[X_t<-\kappa \cap Y_t<-\kappa]}}{T_\kappa^{-}} \\
T_\kappa^{+} &= \sum_{t=1}^{T} I_{[X_t>\kappa \cap Y_t>\kappa]}, \quad T_\kappa^{-} = \sum_{t=1}^{T} I_{[X_t<-\kappa \cap Y_t<-\kappa]}
\end{aligned}$$

where the quantities for Y are similarly defined. Exceedance correlation can only be estimated if the region where $X < \kappa$ and $Y < \kappa$ is populated with data, and it is possible for some assets that this region is empty. Empty regions may occur when measuring the exceedance correlation between assets that have strong negative dependence (e.g., 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,

$$\begin{aligned}
\tilde{X}_t^{+}(\kappa) &= \frac{X_t - \mu_x^{+}(\kappa)}{\sigma_X^{+}(\kappa)}, \quad \tilde{X}_t^{-}(\kappa) = \frac{X_t - \mu_X^{-}(\kappa)}{\sigma_X^{-}(\kappa)}, \\
\tilde{Y}_t^{+}(\kappa) &= \frac{Y_t - \mu_Y^{+}(\kappa)}{\sigma_Y^{+}(\kappa)}, \quad \tilde{Y}_t^{-}(\kappa) = \frac{Y_t - \mu_Y^{-}(\kappa)}{\sigma_Y^{-}(\kappa)}.
\end{aligned}$$

These form the basis of the moment conditions,

$$\begin{aligned}
&\frac{T}{T_\kappa^{+}}\left(\tilde{X}^{+}(\kappa)\tilde{Y}^{+}(\kappa) - \rho^{+}(\kappa)\right) I_{[X_t>\kappa \cap Y_t>\kappa]} \\
&\frac{T}{T_\kappa^{-}}\left(\tilde{X}^{-}(\kappa)\tilde{Y}^{-}(\kappa) - \rho^{-}(\kappa)\right) I_{[X_t<-\kappa \cap Y_t<-\kappa]}.
\end{aligned} \tag{9.66}$$

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 $\boldsymbol{\kappa}$ is a vector of thresholds $\kappa_1, \kappa_2, \ldots, \kappa_n$, then

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

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

$$\hat{\boldsymbol{\Omega}} = \hat{\boldsymbol{\Gamma}}_0 + \sum_{l=1}^{L} w_j\left(\hat{\boldsymbol{\Gamma}}_l + \hat{\boldsymbol{\Gamma}}_l'\right) \tag{9.67}$$

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 9.34 (Exceedance Beta)**.** The exceedance beta at level κ is defined as

$$\begin{aligned} \beta^+(\kappa) &= \frac{\text{Cov}\left(X, Y | X > \kappa, Y > \kappa\right)}{\text{V}\left(X | X > \kappa, Y > \kappa\right)} = \frac{\sigma_Y^+(\kappa)}{\sigma_X^+(\kappa)}\rho^+(\kappa) \\ \beta^-(\kappa) &= \frac{\text{Cov}\left(X, Y | X < -\kappa, Y < -\kappa\right)}{\text{V}\left(X | X < -\kappa, Y < -\kappa\right)} = \frac{\sigma_Y^-(\kappa)}{\sigma_X^-(\kappa)}\rho^-(\kappa) \end{aligned} \tag{9.68}$$

Sample exceedance betas are computed using the sample analogs,

$$\hat{\beta}^+(\kappa) = \frac{\hat{\sigma}_{XY}^+(\kappa)}{\hat{\sigma}_X^{2+}(\kappa)}, \text{ and } \hat{\beta}^-(\kappa) = \frac{\hat{\sigma}_{XY}^-(\kappa)}{\hat{\sigma}_X^{2-}(\kappa)}, \tag{9.69}$$

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

$$\begin{aligned} &\frac{T}{T_\kappa^+}\left(\frac{\sigma_Y^+(\kappa)}{\sigma_X^+(\kappa)}\tilde{X}^+(\kappa)\tilde{Y}^+(\kappa) - \beta^+(\kappa)\right) I_{[X_t>\kappa \cap Y_t>\kappa]} \\ &\frac{T}{T_\kappa^-}\left(\frac{\sigma_Y^-(\kappa)}{\sigma_X^-(\kappa)}\tilde{X}^-(\kappa)\tilde{Y}^-(\kappa) - \beta^-(\kappa)\right) I_{[X_t<-\kappa \cap Y_t<-\kappa]}. \end{aligned} \tag{9.70}$$

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

Daily data for the entire history of both the S&P 500 and the FTSE 100 is provided by Yahoo! Finance. Table 9.6 contains the three correlations and standard errors computed using the bootstrap where weekly returns are used to bias due to nonsynchronous returns (all overlapping 5-day returns are used to estimate all estimators). The linear correlation is the largest, followed by the rank and Kendall's τ. Figure 9.11 plots these same three measures only using 252-day moving averages. The three measures broadly agree about changes in the level of dependence.

Figure 9.12 plots the negative and positive exceedance correlation along with 95% confidence intervals computed using the bootstrap. The exceedance thresholds are chosen using quantiles of each series. The negative exceedance correlation is computed for thresholds less than or equal to 50%, and positive is computed for 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 of figure 9.13 is from a standard CAP-M calibrated to match a typical S&P 500 stock. The market return is simulated from a standardized t_6 with the same variance as the S&P 500 in the past ten years. The idiosyncratic variance is similarly calibrated to the cross-section of idiosyncratic variances.

The simulated data in the top panel is computed from

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

where $\epsilon_{i,t}$ is i.i.d. normally distributed and has the same variance as the average idiosyncratic variance in the cross-section of S&P 500 constituents. The simulated data shown in the bottom panel is generated according to

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

where $z_{i,t} = \exp\left(-10 r_{m,t} I_{[r_{m,t}<0]}\right)$ introduce heteroskedasticity so that the idiosyncratic variance is smaller 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 that have applications in risk management and credit and derivative pricing. Copulas allow a distribution to be decomposed where the dependence between assets is separated from the marginal distribution of each asset. Recall that a k-variate random variable $\mathbf{X}$ has a cumulative 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 rigid. For example, the multivariate Student's t requires all margins to have the same degree-of-freedom parameter, and so the chance of seeing extreme returns – more then 3σ away from the mean – must be the same for all assets. While this assumption may be reasonable when modeling equity index returns, extremely heavy tails are not plausible in other asset classes, e.g., bond or foreign exchange returns. Copulas provide a flexible mechanism to model the marginal distributions *separately* from the dependence structure, and so provide a richer framework for specifying multivariate distributions than the standard set of multivariate distribution functions.

Recall the definition of the marginal distribution of X_1.

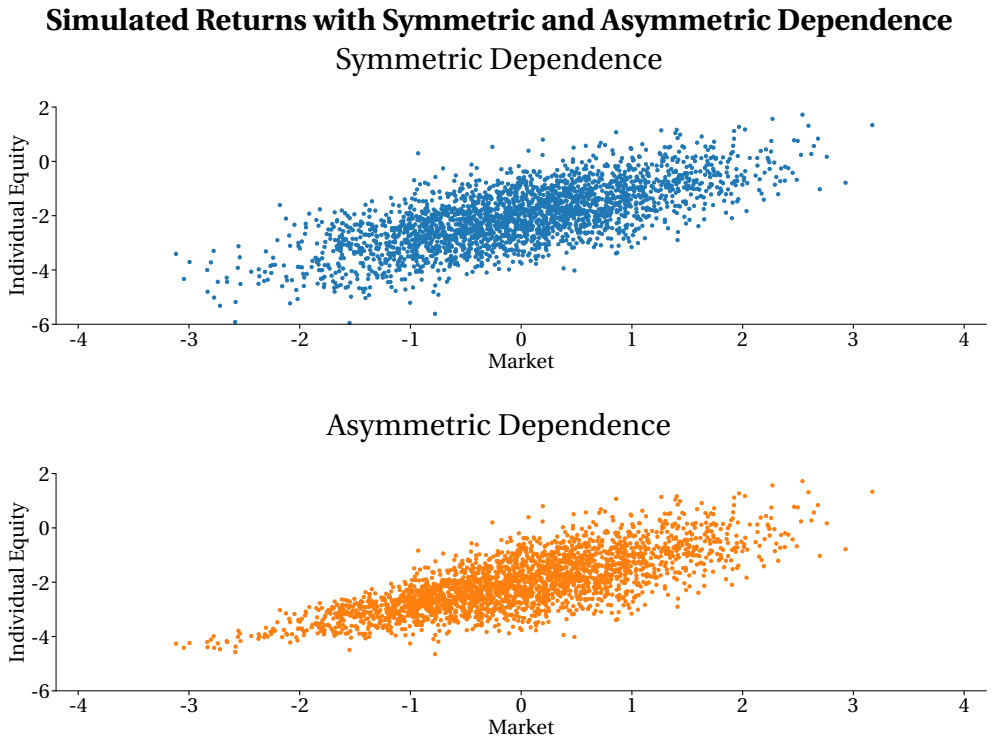

Figure 9.13: These graphs show simulated returns from a CAP-M where the market has a t_6 distribution with the same variance as the S&P 500. The idiosyncratic shock is normally distributed with mean 0, and its variance matches the variance of the idiosyncratic errors of the S&P 500 constituents. The asymmetric dependence is introduced through idiosyncratic error heteroskedasticity where the error variance is $\sigma_\epsilon \exp\left(1/2 r_m I_{[r_m<0]}\right)$. The idiosyncratic component has a smaller variance when the market return is negative than when the market return is positive.

Definition 9.35 (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

$$f_i(x_i) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} f_X(x) dx_1 \ldots dx_{i-1} dx_{i+1} \ldots dx_k.$$

The marginal density contains only information about the probability of observing values of X_i. For example, if X is a bivariate random variable with continuous support, then the marginal density of X_1 is

$$f_1(x_1) = \int_{-\infty}^{\infty} f_X(x_1, x_2) dx_2.$$

The marginal distribution,

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

contains all of the information about the probability of observing values of X_1, and importantly $F_{X_1}(X_1) \sim U(0,1)$. The transformation removes the information contained in the marginal distribution about the probability of observing different values of X_1.

This probability integral transformation applies to both X_1 and X_2, and so $U_1 = F_{X_1}(x_1)$ and $U_2 = F_{X_2}(x_2)$ only information about the dependence between the two random variables. The distribution that 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* the marginal distributions and the dependence.

9.7.1 Basic Theory

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

Definition 9.36 (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 four fundamental properties:

- $C(u_1, u_2, \ldots, u_k)$ is increasing in each component u_i;
- $C(0, \ldots, u_j, \ldots, 0) = 0$;
- $C(1, \ldots, u_j, \ldots, 1) = u_j$; and
- 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.

Sklar's theorem provides the critical insight that explains how a joint distribution is related to its marginal distributions and the copula that link them. (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 has two important implications. First, it ensures that the copula is unique whenever the margins are continuous, which is usually the case in financial applications. Second, it shows that a copula can be constructed from any distribution function that has known marginal distributions. 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 many standard distribution functions 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. A copula is a k-variate distribution, and so

when the copula density exists it can be derived by differentiating the distribution with respect to each component random variable,

$$c(u_1, u_2, \ldots, u_k) = \frac{\partial^k C(u_1, u_2, \ldots, u_k)}{\partial u_1 \partial u_2 \ldots \partial u_k}. \tag{9.71}$$

9.7.2 Tail Dependence

One final measure of dependence, tail dependence, is useful in understanding risks in portfolios and for comparing copulas. Tail dependence is more of a theoretical construction than a measure that is directly estimated (although it is possible to estimate tail dependence).

Definition 9.37 (Tail Dependence)**.** The upper and lower tail dependence, τ^U and τ^L respectively, are defined as the conditional probability of an extreme event,

$$\tau^U = \lim_{u \to 1-} \Pr\left[X > F_X^{-1}(u) \middle| Y > F_Y^{-1}(u)\right] \tag{9.72}$$

$$\tau^L = \lim_{u \to 0+} \Pr\left[X < F_X(u) \middle| Y < F_Y(u)\right] \tag{9.73}$$

where the limits are taken from above for τ^U and below for τ^L.

Tail dependence measures the probability X takes an extreme value given Y takes an extreme value. The dependence between a portfolio and assets used as hedges is particularly important when the portfolio suffers a loss day, and so has a return in its lower tail.

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

$$\tau^L = \lim_{u \to 0+} \frac{C(u, u)}{u} \tag{9.74}$$

$$\tau^U = \lim_{u \to 1-} \frac{1 - 2u + C(u, u)}{1 - u} \tag{9.75}$$

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

9.7.3 Copulas

A large number of copulas have been developed. Some, such as the Gaussian, are implicitly defined from standard distributions. Others have been designed 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 which depends only on the product of the input values.

Definition 9.38 (Independence Copula)**.** The independence copula is

$$C(u_1, u_2, \ldots, u_k) = \prod_{i=1}^{k} u_i \tag{9.76}$$

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 9.39 (Comonotonicity Copula)**.** The comonotonicity copula is

$$C(u_1, u_2, \ldots, u_k) = \min(u_1, u_2, \ldots, u_k) \tag{9.77}$$

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 using the k-variate Gaussian distribution, $\Phi_k(\cdot)$, and the univariate Gaussian distribution, $\Phi(\cdot)$.

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

$$C(u_1, u_2, \ldots, u_k) = \Phi_k\left(\Phi^{-1}(u_1), \Phi^{-1}(u_2), \ldots, \Phi^{-1}(u_k)\right) \tag{9.78}$$

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

Recall that if U is a uniform random variable then $X = \Phi^{-1}(U)$ is distributed standard normal. This transformation allows the Gaussian copula density to be implicitly defined using the inverse distribution function. The Gaussian copula density is

$$c(u_1, u_2, \ldots, u_k) = \frac{(2\pi)^{-\frac{k}{2}} |\mathbf{R}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\boldsymbol{\eta}'\mathbf{R}^{-1}\boldsymbol{\eta}\right)}{\phi(\Phi^{-1}(u_1)) \ldots \phi(\Phi^{-1}(u_k))} \tag{9.79}$$

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 using the multivariate Student's t distribution.

Definition 9.41 (Student's Copula)**.** The Student's t copula is

$$C(u_1, u_2, \ldots, u_k) = t_{k,\nu}\left(t_\nu^{-1}(u_1), t_\nu^{-1}(u_2), \ldots, t_\nu^{-1}(u_k)\right) \tag{9.80}$$

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

Note that while the Student's t distribution is superficially similar to a normal distribution, variables that are distributed multivariate t_ν are substantially more dependent if ν is small (3 – 8). A multivariate Student's t is defined as a multivariate normal divided by a single, common, independent χ_ν^2 standardized to have mean 1. When ν is small, the chance of seeing a small value in the denominator is large, and since this divisor is common, all series 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 dependent than the upper tail, and so it may be appropriate for modeling the returns of some financial assets, e.g., equities.

Definition 9.42 (Clayton Copula)**.** The Clayton copula is

$$C(u_1, u_2) = \left(u_1^{-\theta} + u_2^{-\theta} - 1\right)^{-1/\theta}, \ \theta > 0 \tag{9.81}$$

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 9.43 (Gumbel Copula)**.** The Gumbel copula is

$$C(u_1, u_2) = \exp\left[-\left((-\ln u_1)^\theta + (-\ln u_2)^\theta\right)^{1/\theta}\right], \ \theta \geq 1 \tag{9.82}$$

The Gumbel copula exhibits upper tail dependence that is increasing in θ. It approaches to the independence copula as $\theta \to 1$. Because upper tail dependence is relatively rare among financial assets, a "rotated" version of the Gumbel is more useful when modeling financial asset returns.

Let $C(u_1, u_2)$ be a bivariate copula. The rotated version[12] of the copula is given by

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

[12]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 capture lower tail dependence rather than upper tail dependence.

Definition 9.44 (Rotated Gumbel Copula)**.** The rotated Gumbel copula is

$$C^R(u_1, u_2) = u_1 + u_2 - 1 + \exp\left[-\left((-\ln(1-u_1))^\theta + (-\ln(1-u_2))^\theta\right)^{1/\theta}\right], \ \theta \geq 1 \tag{9.83}$$

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

The rotated Gumbel copula density is

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

This copula density is identical to the Gumbel copula density only using $1-u_1$ and $1-u_2$ as its arguments. The rotation moves values near zero, where the dependence is low, to be near one, where the dependence is higher.

9.7.3.7 Joe-Clayton Copula

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

Definition 9.45 (Joe-Clayton Copula)**.** The Joe-Clayton copula is

$$C(u_1, u_2) = 1 - \left(1 - \left[\left(1-(1-u_1)^{\theta_U}\right)^{-\theta_L} + \left(1-(1-u_2)^{\theta_U}\right)^{-\theta_L} - 1\right]^{-1/\theta_L}\right)^{1/\theta_U} \tag{9.84}$$

where the two parameters, θ_L and θ_U are directly related to lower and upper tail dependence through

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

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

Deriving the density of a Joe-Clayton copula is a straightforward, but tedious, calculation. The Joe-Clayton copula is not symmetric, even when the same values for τ^L and τ^U are used. This asymmetry may be acceptable, but if symmetry is preferred a symmetrized copula can be constructed by averaging a copula with its rotated counterpart.

Definition 9.46 (Symmetrized Copula)**.** Let $C(u_1, u_2)$ be an asymmetric bivariate copula. The symmetrized version of the copula is given by

$$C^S(u_1, u_2) = \frac{1}{2}\left(C(u_1, u_2) + C^R(1-u_1, 1-u_2)\right) \tag{9.85}$$

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

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

Copula	τ^L	τ^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 τ^L and τ^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.

9.7.4 Tail Dependence in Copulas

The copulas presented in the previous section all have different functional forms, and so produce 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 $|\rho| = 1$, 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 hypercube (or unit square in a bivariate copula), 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 which occurs in the small squares of in $[0, 0.05] \times [0, 0.05]$ and $[.95, 1] \times [.95, 1]$, lower and upper 5% of each margin. Transforming the marginal distribution of each series to be standard normal is a superior method to visualize the dependence in the copula. This visualization ensures that any differences are attributable to the copula while distributing the interesting aspects over a wider range of values. It also projects the dependence structure into a familiar space that more closely resembles two financial asset returns.

Figure 9.14 contains plots of 4 copulas. The top two panels show the independence copula and the comonotonicity copula as distributions on $[0, 1] \times [0, 1]$ where curves are isoprobability lines. In distribution space, high dependence appears as an "L" shape and independence appears as a parabola. 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 Gaussian copula is more dependent than the independence copula – a special case of the Gaussian copula when $\rho = 0$ – but less dependent than the comonotonicity copula (except when $\rho = 1$). 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.[13]

[13] Some copulas do not have a copula density, and in these cases, the copula distribution is the only visualization

Figure 9.15 contains plots of 4 copulas depicted as densities with standard normal marginal distributions. The upper left panel contains the Clayton density which has lower-tail dependence ($\theta = 1.5$). The upper right shows the symmetrized Joe-Clayton where $\tau^L = \tau^U = 0.5$, which has both upper and lower tail dependence. The bottom two panels show the rotated Gumbel and symmetrized Gumbel where $\theta = 1.5$. The rotated Gumbel is similar to the Clayton copula although it is not identical.

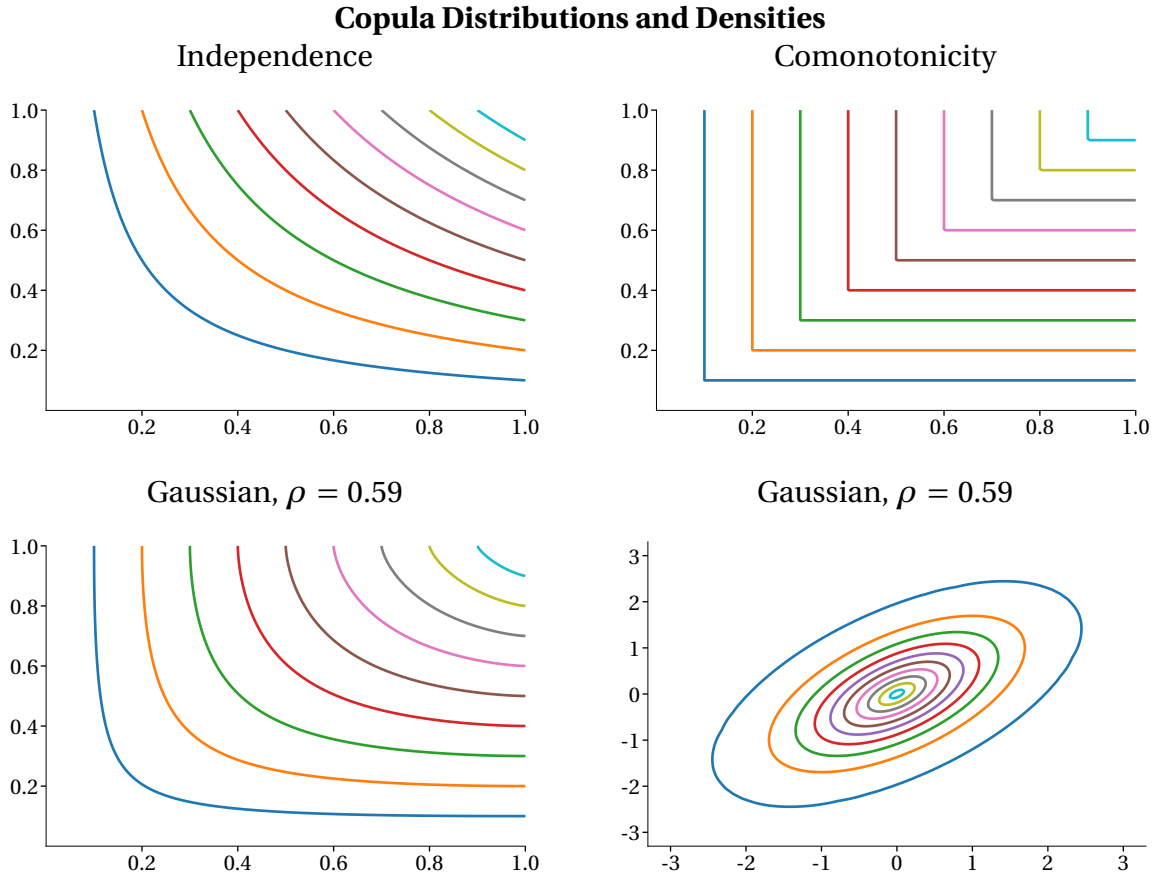

Figure 9.14: The top left panel shows the isoprobability curves of an independence copula. The top right panel shows the isoprobability curves of the comonotonicity copula, which has perfect dependence. The bottom panels contain the Gaussian copula, where the left depicts the copula in distribution space ($[0,1] \times [0,1]$) and the right shows the copula density using standard normal marginal distributions. The correlation of the Gaussian copula is estimated using weekly returns on the S&P 500 and FTSE 100.

9.7.6 Estimation of Copula models

A copula-based model is a joint distribution, and so parameters can be estimated using maximum likelihood. As long as the copula density exists, and the parameters of the margins are

option.

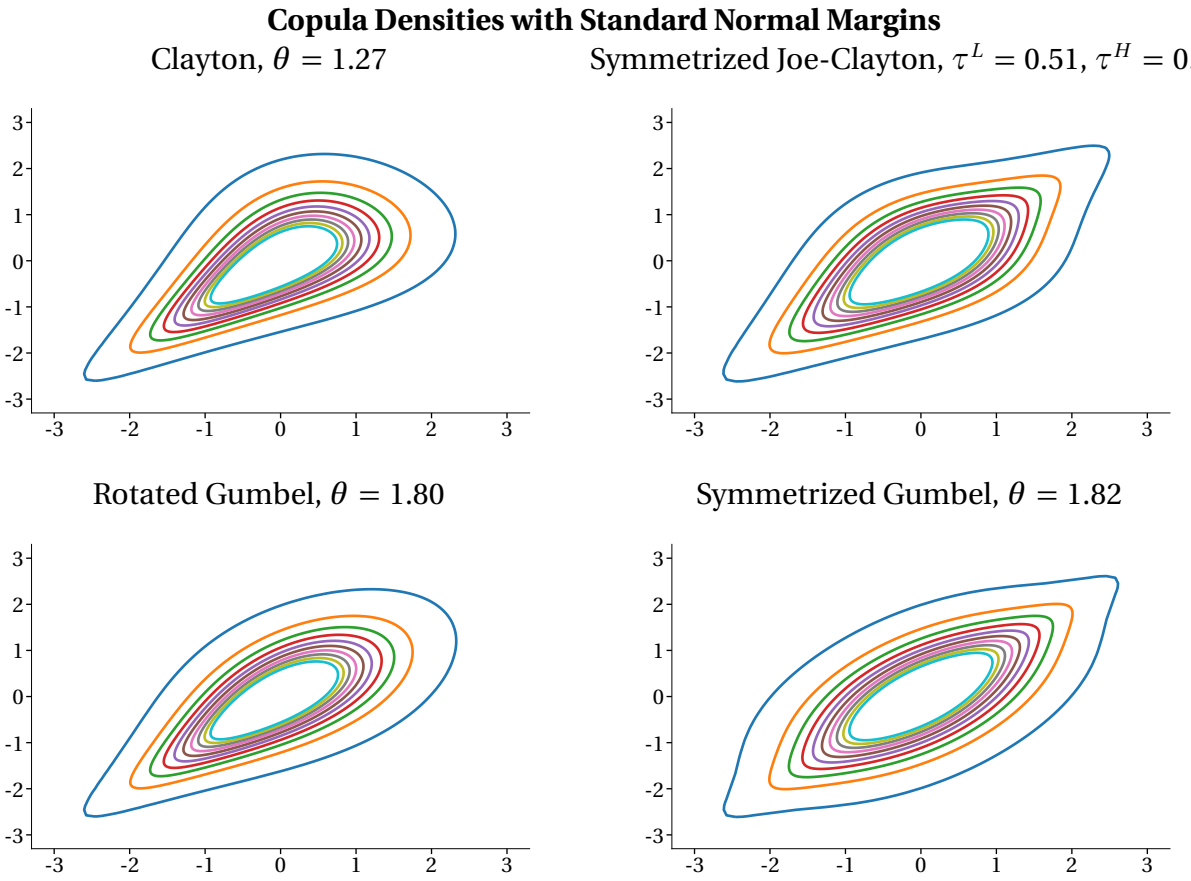

Figure 9.15: These for panels all depict copulas as densities using standard normal margins. All differences in appearance can be attributed to the differences in the copulas. The top left panel contains the Clayton copula density. The top right contains the symmetrized Joe-Clayton. The bottom panels contain the rotated Gumbel which has lower tail dependence and the symmetrized Gumbel. The parameter values are estimated from weekly returns on the S&P 500 and FTSE 100.

distinct from the parameters of the copula (which is almost always the case), the likelihood of a k-variate random variable Y can be written as

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

where $u_{j,t} = F_j^{-1}\left(y_{j,t}; \boldsymbol{\theta}_j\right)$ are the probability integral transformed observations, $\boldsymbol{\theta}_j$ 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,

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

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

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

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

This two-step procedure is not efficient in the sense that the parameter estimates are consistent but have higher variance than if all parameters are simultaneously estimated. In practice, the reduction in precision is typically small. If parameter estimation accuracy is an important consideration, then the two-step estimator can be used as a starting value for an estimator which simultaneously estimates 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 is illustrated using returns of the S&P 500 and the FTSE 100. Weekly returns are used to mitigate issues with non-synchronous closing times. The first example uses the empirical CDF to transform the returns so that the model focuses on the unconditional dependence between the two series. The upper left panel of figure 9.16 contains a scatter plot of the ECDF transformed residuals. The residuals tend to cluster around the 45° line indicating positive dependence (correlation). There are clusters of observations near (0, 0) and (1, 1) that indicating the returns have tail dependence. The normal, students t, Clayton, rotated Gumbel, symmetrized Gumbel and symmetric Joe-Clayton copulas are 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 returns in the lower tails are substantially more dependent.

Copulas can also be used in conditional density models. Combining a constant copula with dynamic models of each margin is similar to using a CCC-GARCH model to estimate the conditional covariance. A conditional joint density model is built using TARCH(1,1,1) volatilities

Dependence Measures for Weekly FTSE and S&P 500 Returns

Copula	θ_1	θ_2	Log. Lik.	τ^L	τ^U
Gaussian	0.645		-486.9	0	0
Clayton	1.275		-460.2	0.581	0
Rot. Gumbel	1.805		-526.2	0.532	0
Sym. Gumbel	1.828		-529.5	0.270	0.270
Sym. Joe-Clayton	0.518	0.417	-539.9	0.518	0.417

Table 9.8: Parameter estimates for the unconditional copula between weekly returns on the S&P 500 and the FTSE 100. Marginal distributions are estimated using empirical CDFs. For the Gaussian copula, θ_1 is the correlation, and in the Joe-Clayton θ_1 is τ^L and θ_2 is τ^U. The third column reports the log likelihood from the copula density. The final two columns report the estimated lower and upper tail dependence.

with skew t errors for each index return series. The copulas are estimated using the conditionally transformed residuals $\hat{u}_{i,t} = F\left(r_{i,t}; \hat{\sigma}_t^2, \hat{\nu}, \hat{\lambda}\right)$ where σ_t^2 is the conditional variance, ν is the degree of freedom, and λ captures the skewness in the standardized residuals. 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 are heavier tailed, and both are negatively skewed.

The parameter estimates using the conditional marginals all indicate less dependence than those estimated using the empirical CDF. This reduction in dependence is due to synchronization between the volatility of the two markets. 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. Unconditional models use data from both high and low volatility periods. Volatility and dependence are linked in financial markets, and so ignoring conditional information tends to higher unconditional dependence than conditional dependence. This phenomenon is similar to the generation of heavy tails in the unconditional distribution of a single asset return – mixing periods of high and low volatility produced heavy tails. The difference in the dependence shows up in the parameter values in the copulas, and in the estimated tail indices, which are uniformly smaller than in their unconditional counterparts. The changes in dependence also appear through the reduction in the improvement in the log-likelihoods of the dependent copulas relative to the Gaussian.

Figure 9.16 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 estimated degree of freedom and skewness parameters. These curves are straight except for the most extreme observations, and so indicate an acceptable fit. The bottom right plot contains the annualized volatility series for the two assets where the coordination in the conditional volatilities is apparent. It also appears the coordination in volatility cycles has strengthened post-2000.

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

Index	α_1	γ_1	β_1	ν	λ
S&P 500	0.026	0.178	0.855	8.924	-0.231
FTSE 100	0.038	0.145	0.861	8.293	-0.138
Copula	θ_1	θ_2	Log. Lik.	τ^L	τ^U
Gaussian	0.621		-439.1	0	0
Clayton	1.126		-399.6	0.540	0
Rot. Gumbel	1.713		-452.4	0.501	0
Sym. Gumbel	1.754		-452.7	0.258	0.258
Sym. Joe-Clayton	0.475	0.357	-448.5	0.475	0.357

Table 9.9: Parameter estimates for the conditional copula between weekly returns on the S&P 500 and the FTSE 100. Marginal distributions are 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, θ_1 is the correlation, and in the Joe-Clayton θ_1 is τ^L and θ_2 is τ^U. The third column reports the log likelihood from the copula density. The final two columns report the estimated lower and upper tail dependence.

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 model dependence using conditional copulas where the copula parameters evolve through time, $C(u_1, u_2; \boldsymbol{\theta}_t)$, using GARCH-like dynamics. Patton (2006) first used this structure 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 $\epsilon_t^2 = (r_t - \mu)^2$ is the natural shock since its conditional expectation is the variance, $\mathrm{E}_{t-1}\left[\epsilon_t^2\right] = \sigma_t^2$. In most copula models there is no obvious equivalent. Creal, Koopman, and Lucas (2013) have recently developed a general framework which can be used to construct a natural shock even in complex models, and have applied their methodology to estimate conditional 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 transformed residuals from the first step are then transformed to be Gaussian, and these are used to estimate the correlation parameters 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 computational tool that has a variety of uses, including estimating standard errors and simulating returns. It is particularly useful when evaluating expressions for asymptotic standard errors that are complex. This appendix provides a *very* brief introduction to bootstrap

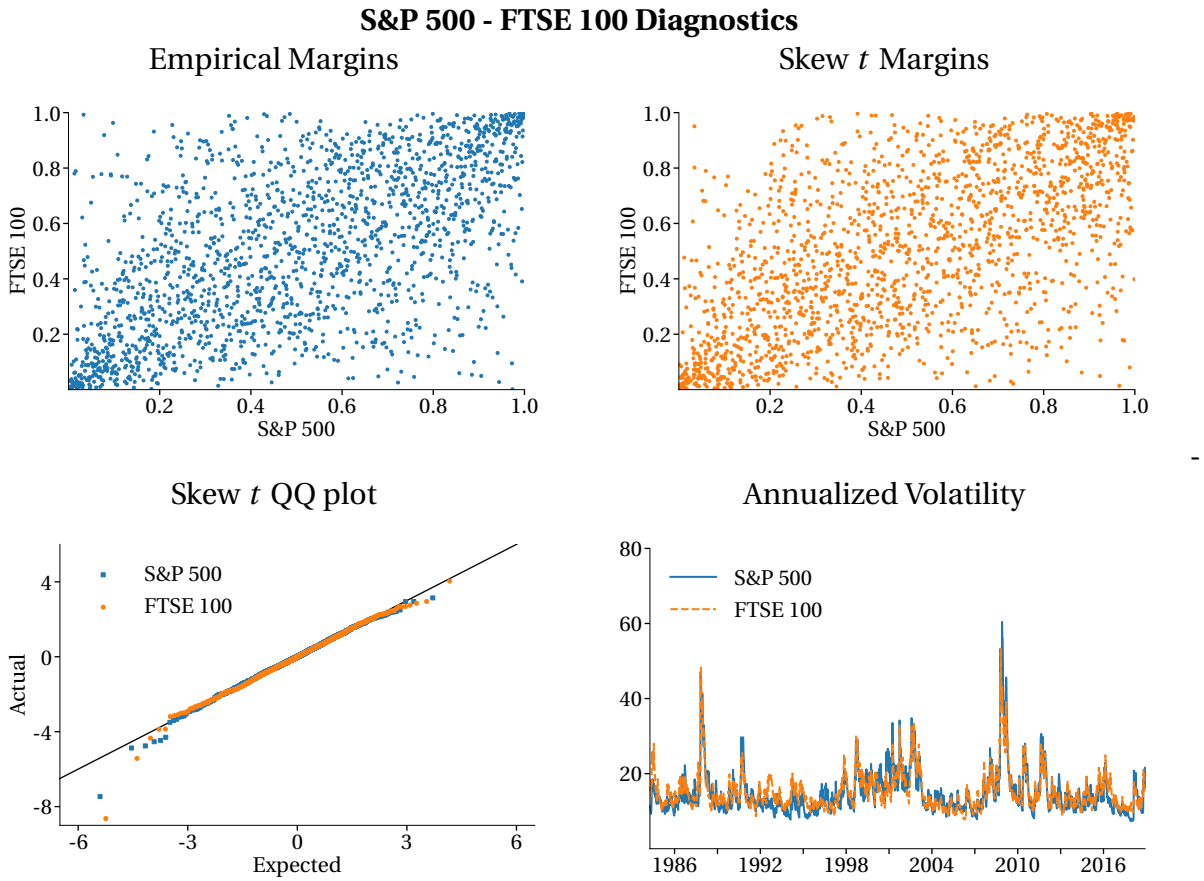

Figure 9.16: These four panels show diagnostics from fitting copulas to weekly returns on the S&P 500 and FTSE 100. The top two panels contain plots of the probability integral transformed residuals. The left panel shows the PITs constructed using the empirical CDF, and so depicts the unconditional dependence. The right contains the PITs from a TARCH(1,1,1) with Skew t errors. The bottom left contains a QQ plot of the data against the typical value from a Skew t. The bottom right plot contains the fit annualized volatility for the two indices.

standard errors. The key intuition that underlies the bootstrap is 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 simulating values from an unknown distribution.

There are many implementations of the bootstrap, and each uses a different sampling scheme when generating bootstrap samples. The assumed data generating process determines which bootstraps are applicable. Bootstrap methods can be classified as parametric or non-parametric. Parametric bootstraps resample model residuals. Nonparametric bootstraps directly resample from the observed data and so do not rely on a model. In many applications both types of bootstraps are valid, and 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 but may require larger

samples 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 Tibshirani (1998) and Chernick (2008).

The i.i.d. bootstrap uses the simplest sampling scheme and is applicable when the data are i.i.d., or more generally when the model errors are not serially correlated.[14]

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

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

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

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

It is implausible to assume that the data are i.i.d. in most applications in finance, and so a bootstrap designed for dependent data is required. The two most common bootstraps for dependent data are the Circular Block Bootstrap and the Stationary Bootstrap (Politis and Romano, 1994). The Circular 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 (Circular Block Bootstrap)**.**

1. *Draw $\tau_1 = \lceil T u \rceil$ where $u \overset{i.i.d.}{\sim} U(0,1)$.*
2. *For $i = 2, \ldots, T$, if $i \mod m \neq 0$, $\tau_i = \tau_{i-1} + 1$ where wrapping is used so that if $\tau_{i-1} = T$ then $\tau_i = 1$. If $i \mod m = 0$ 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 $i = 2, \ldots, T$, 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 $\{\tau_i\}_{i=1}^{T}$.*

[14]The definition of the model error depends on the statistic of interest. For example, when bootstrapping the sample mean, the i.i.d. bootstrap can be used if the data are serially uncorrelated. When bootstrapping a variance estimator, the squared deviations must be uncorrelated. In applications of ML, the scores from the model must be serially uncorrelated. Additionally, when using a nonparametric bootstrap, the i.i.d. bootstrap is only applicable when the model does not impose a time-series structure (i.e., the model is not an ARMA or GARCH).

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

In both the Circular Block Bootstrap and the Stationary Bootstrap, the block length should be chosen to capture most of the dependence in the data. The block size should not be larger than $\sqrt{T}$. Patton, Politis, and White (2009) provide a data-based method to select the block size in these bootstraps.

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 $\tilde{\boldsymbol{\theta}}_j$, $j = 1, 2, \ldots, B$.*
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 $(\hat{\boldsymbol{\theta}} \pm \bar{q}_{\alpha/2})$ 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 $(\hat{\boldsymbol{\theta}} - q_{\alpha/2}, \hat{\boldsymbol{\theta}} + q_{1-\alpha/2})$ using the $\alpha/2$ and $1-\alpha/2$ quantile of η_b, denoted $q_{\alpha/2}$ and $q_{1-\alpha/2}$, respectively. The confidence interval can be equivalently defined as $(\tilde{q}_{\alpha/2}, \tilde{q}_{1-\alpha/2})$ where $\tilde{q}_{\alpha/2}$ is the $\alpha/2$ quantile of the estimators computed from the bootstrap samples, $\left\{\tilde{\boldsymbol{\theta}}_j\right\}_{j=1}^{B}$, and $\tilde{q}_{1-\alpha/2}$ is similarly defined using the $1-\alpha/2$ quantile.*

The bootstrap confidence intervals in this chapter are 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 construction in many situations where estimators have non-standard distributions, e.g., unit roots, and so before computing bootstrap standard errors, it is useful to verify that the bootstrap produces valid inference. In cases where the bootstrap fails, subsampling, a more general statistical technique can 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 models?
3. Other than linear correlation, describe two other measures of dependence.
4. What is Realized Covariance?
5. What are the important considerations when estimating covariance using Realized Covariance?

Exercise 9.2. Answer the following questions.

1. Briefly outline two applications in finance where a multivariate volatility models are useful.
2. Describe two of the main problems faced in multivariate volatility modeling, using two different models to illustrate these problems.
3. Recall that, in a bivariate application, the BEKK model of 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 $\boldsymbol{\epsilon}_t' \equiv [\epsilon_{1,t}, \epsilon_{2,t}]$ is the vector of residuals. Using the result that $\text{vec}(\mathbf{QRS}) = (\mathbf{S}' \otimes \mathbf{Q})\,\text{vec}(\mathbf{R})$, 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 $\text{vec}(\mathbf{\Sigma}_{t-1})$.

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 $\text{vec}(\mathbf{QRS}) = (\mathbf{S}' \otimes \mathbf{Q})\,\text{vec}(\mathbf{R})$, where $\otimes$ is the Kronecker product, rewrite 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 4×1.)

5. Computing "exceedance correlations" on the two-day returns on the S&P 500 index and the FTSE 100 index, we obtain Figure 9.17. 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 measures of dependence.

6. What is Realized Covariance?

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

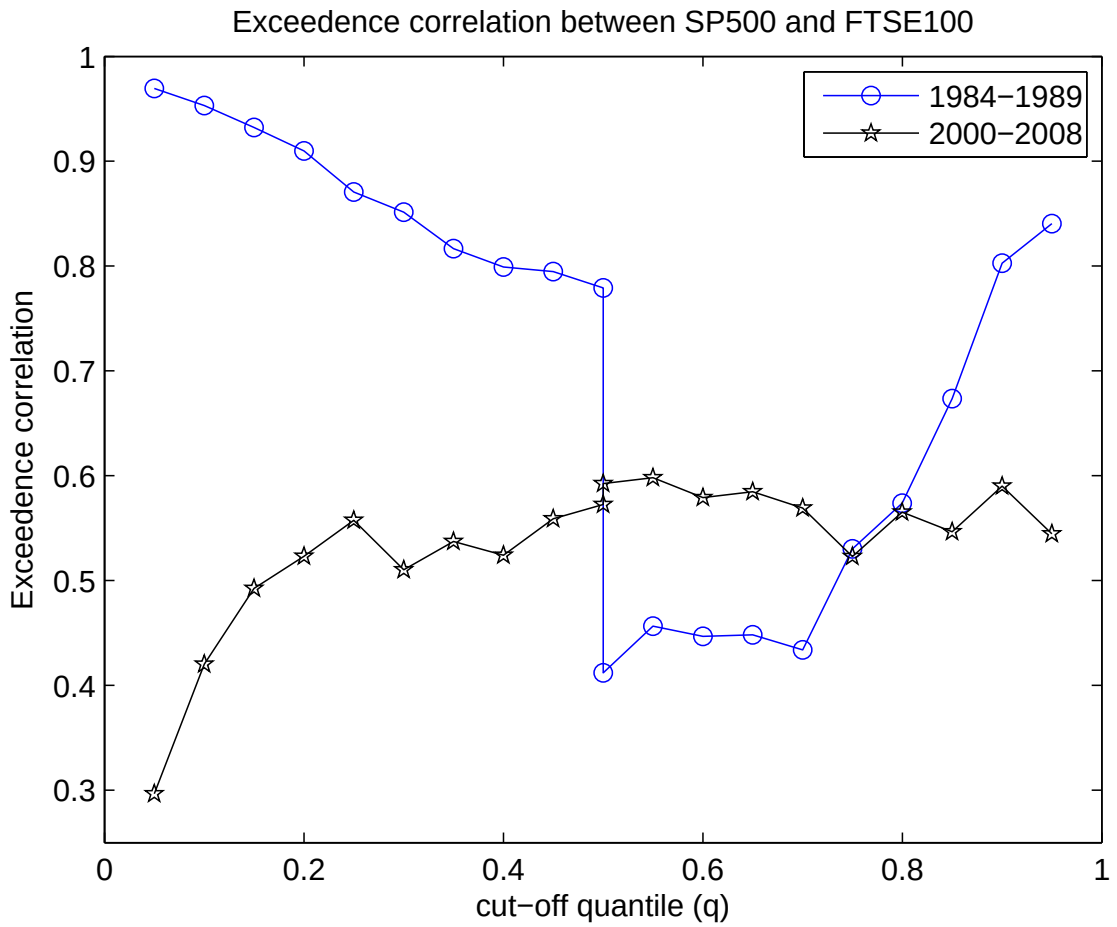

Figure 9.17: Exceedance correlations between two-day returns on the S&P 500 index and the FTSE 100 index. Line with circles uses data from April 1984 to December 1989; line with stars uses data from January 2000 to December 2008.