8

Statistical Measures of Dependence for Financial Data

David S. Matteson, Nicholas A. James and William B. Nicholson
Cornell University, USA

8.1 Introduction

The analysis of financial and econometric data is typified by non-Gaussian multivariate observations that exhibit complex dependencies: heavy-tailed and skewed marginal distributions are commonly encountered; serial dependence, such as autocorrelation and conditional heteroscedasticity, appear in time-ordered sequences; and nonlinear, higher-order, and tail dependence are widespread. Illustrations of serial dependence, nonnormality, and nonlinear dependence are shown in Figure 8.1.

When data are assumed to be jointly Gaussian, all dependence is linear, and therefore only pairwise among the variables. In this setting, Pearson's product-moment correlation coefficient uniquely characterizes the sign and strength of any such dependence.

Definition 8.1 *For random variables X and Y with joint density f_{XY},* **Pearson's correlation coefficient** *is defined as*

$$\rho_P(X,Y) = E\left[\frac{(X-\mu_X)}{\sqrt{\sigma_X^2}}\frac{(Y-\mu_Y)}{\sqrt{\sigma_Y^2}}\right] = \int\int \frac{(x-\mu_X)}{\sqrt{\sigma_X^2}}\frac{(y-\mu_Y)}{\sqrt{\sigma_Y^2}} f_{XY}(x,y)\,dx\,dy,$$

where $\mu_X = E(X) = \int\int x f_{XY}(x,y)\,dx\,dy$, and $\sigma_X^2 = E[(X-\mu_X)^2] = \int\int (x-\mu_X)^2 f_{XY}(x,y)\,dx\,dy$, are the mean and variance of X, respectively, and μ_Y and σ_Y^2 are defined similarly.

This conventional measure of pairwise linear association is well-defined provided σ_X^2 and σ_Y^2 are positive and finite, in which case $\rho_P(X,Y) \in [-1,+1]$. A value of -1 or $+1$ indicates

Financial Signal Processing and Machine Learning, First Edition.
Edited by Ali N. Akansu, Sanjeev R. Kulkarni and Dmitry Malioutov.

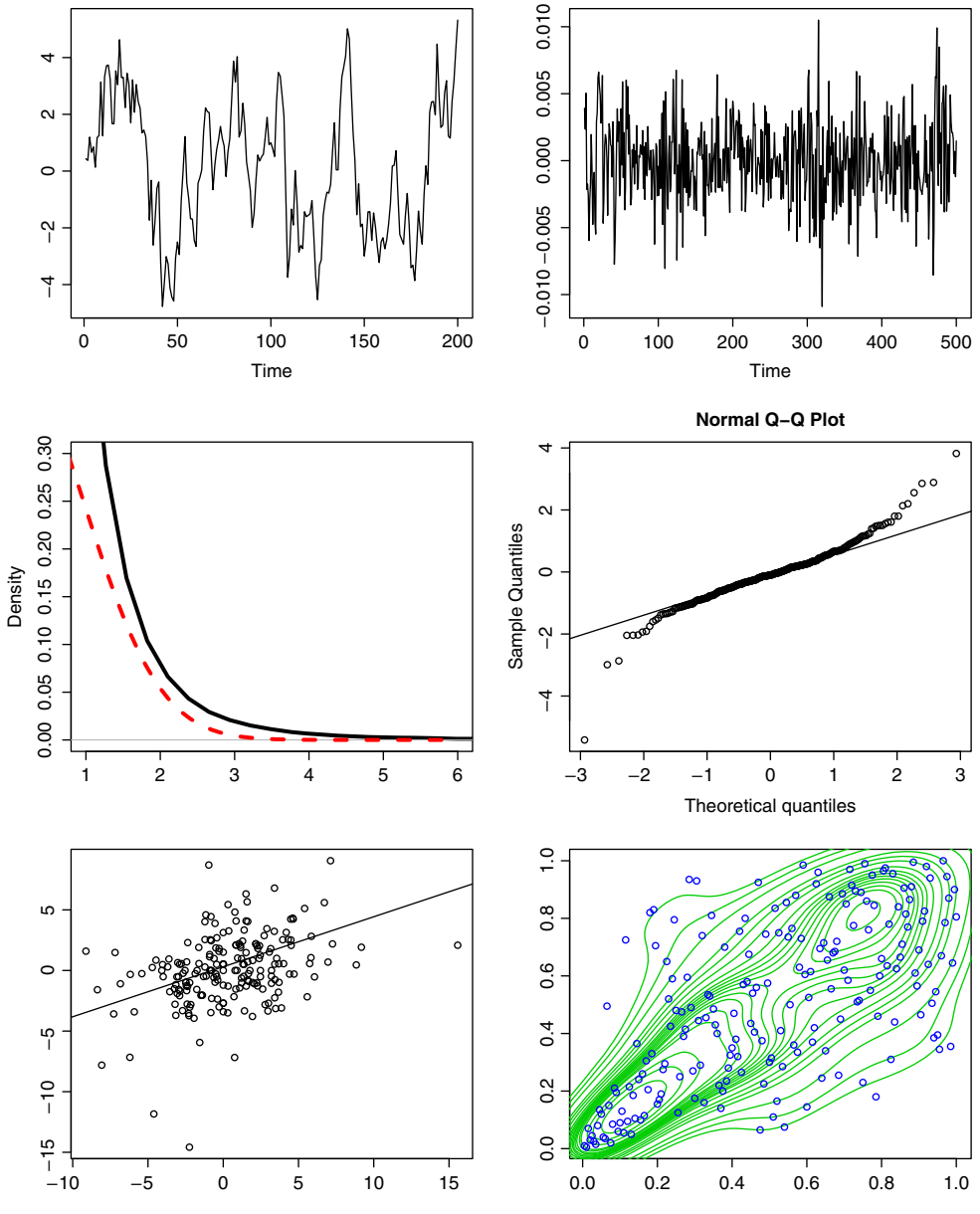

Figure 8.1 Top left: Strong and persistent positive autocorrelation, that is, persistence in local level; top right: moderate volatility clustering, that is, i.e., persistence in local variation. Middle left: Right tail density estimates of Gaussian versus heavy- or thick-tailed data; middle right: sample quantiles of heavy-tailed data versus the corresponding quantiles of the Gaussian distribution. Bottom left: Linear regression line fit to non-Gaussian data; right: corresponding estimated density contours of the normalized sample ranks, which show a positive association that is stronger in the lower left quadrant compared to the upper right.

perfect negative or positive linear dependence, respectively, and in either case X and Y have an exact linear relationship. Negative and positive values indicate negative and positive linear associations, respectively, while a value of 0 indicates no linear dependence.

A sample estimator $\hat{\rho}_P$ is typically defined by replacing expectations with empirical expectations in the above definition. For a paired random sample of n observations $(\boldsymbol{X}_{1:n}, \boldsymbol{Y}_{1:n}) = \{(x_i, y_i)\}_{i=1}^n$, define

$$\hat{\rho}_P(\boldsymbol{X}_{1:n}, \boldsymbol{Y}_{1:n}) = \frac{1}{n-1} \sum_{i=1}^{n} \frac{(x_i - \hat{\mu}_X)}{\sqrt{\hat{\sigma}_X^2}} \frac{(y_i - \hat{\mu}_Y)}{\sqrt{\hat{\sigma}_Y^2}} \; : \quad \hat{\mu}_X = \frac{1}{n} \sum_{i=1}^{n} x_i; \;\; \hat{\sigma}_X^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \hat{\mu}_X)^2,$$

where $\hat{\mu}_Y$ and $\hat{\sigma}_Y^2$ are defined similarly. For jointly Gaussian variables (X, Y), zero correlation is equivalent to independence. For an independent and identically distributed (i.i.d.) sample from the bivariate normal distribution, inference regarding ρ_P can be conducted using the following asymptotic approximation:

$$\sqrt{n}(\hat{\rho}_P - \rho_P) \xrightarrow{\mathcal{D}} \mathcal{N}[0, (1 - \rho_P^2)^2],$$

in which $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution. Under the null hypothesis of zero correlation, this expression simplifies to $\sqrt{n}\hat{\rho}_P \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)$. More generally, for an i.i.d sample from an arbitrary distribution with finite fourth moments, $E(X^4)$ and $E(Y^4)$, a variance-stabilizing transformation may be applied (Fisher's transformation) to obtain the alternative asymptotic approximation (cf. Ferguson, 1996)

$$\frac{\sqrt{n}}{2} \left(\log \; \frac{1 + \hat{\rho}_P}{1 - \hat{\rho}_P} - \log \; \frac{1 + \rho_P}{1 - \rho_P} \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

Although the previous approximation is quite general, assuming finite fourth moments may be unreasonable in numerous financial applications where extreme events are common. Furthermore, Pearson's correlation coefficient measures the strength of linear relationships only. There are many situations in which correlations are zero but a strong nonlinear relationship exists, such that variables are highly dependent. In Section 8.2, we discuss several robust measures of correlation and pairwise association, and illustrate their application in measuring serial dependence in time-ordered data. In Section 8.3, we consider multivariate extensions and Granger causality, and introduce measures of mutual independence. Finally, in Section 8.4 we explore copulas and their financial applications.

8.2 Robust Measures of Correlation and Autocorrelation

Financial data are often time-ordered, and intertemporal dependence is commonplace. The autocovariance and autocorrelation functions are extensions of covariance and Pearson's correlation to a time-dependent setting, respectively.

Definition 8.2 *For a univariate ordered sequence of random variables $\{X_t\}$, the* **autocovariance** γ *and* **autocorrelation** ρ *at indices q and r (Shumway and Stoffer, 2011) are defined as*

$$\gamma(X_q, X_r) = \gamma(q, r) = E[(X_q - \mu_q)(X_r - \mu_r)]$$

and

$$\rho_P(X_q, X_r) = \rho_P(q, r) = \frac{\gamma(q, r)}{\sqrt{\gamma(q, q)}\sqrt{\gamma(r, r)}},$$

respectively, in which $\mu_q = E(X_q)$ and $\mu_r = E(X_r)$. The above quantities are well-defined provided $E(X_t^2)$ is finite for all t; however, estimating these quantities from an observed sequence $\boldsymbol{X}_{1:n} = \{x_t\}_{t=1}^n$ requires either multiple i.i.d. realizations of the entire sequence (uncommon in finance), or some additional assumptions. The first basic assumption is that the observations are equally spaced, and t denotes their discrete-time index. We will refer to such sequences generically as *time-series*. In finance, this assumption may only hold approximately. For example, a sequence of daily market closing asset prices may only be available for weekdays, with additional gaps on holidays, or intraday asset transaction prices may be reported every hundredth of a second, but there may be no transaction at many of these times. In either case, the consecutive observations are commonly regarded as equally spaced, for simplicity.

The next basic assumption is some form of distributional invariance over time, such as stationarity.

Definition 8.3 *A univariate sequence of random variables* $\{X_t\}$ *is* **weakly (***or* **covariance) stationary** *if and only if*

$$E(X_t) = E(X_{t-h}) = \mu \quad \textit{and} \quad \gamma(t, t-h) = \gamma(|h|) = \gamma_h \quad \forall t, h, \textit{ and} \quad \gamma_0 < \infty.$$

This implies that the means and variances are finite and constant, and the autocovariance is constant with respect to t, and only depends on the relative time lag h between observations.

For any k-tuple of indices $t_1, t_2, \ldots, t_k$*, let* $F_{t_1,t_2,\ldots,t_k}(\cdot)$ *denote the joint distribution function of* $(X_{t_1}, X_{t_2}, \ldots, X_{t_k})$*. Then, the sequence is* **strictly stationary** *if and only if*

$$F_{t_1,t_2,\ldots,t_k}(\cdot) = F_{t_1-h,t_2-h,\ldots,t_k-h}(\cdot) \quad \forall h, k, \textit{ and } \forall t_1, t_2, \ldots, t_k.$$

This implies that the joint distributions of all k-tuples are invariant to a common time shift h such that their relative time lags remain constant.

Strict stationary implies weak stationarity provided the variance is also finite.

Now, under the weak stationarity assumption, the parameters $\gamma_h = \gamma(h)$ for $h = 0, 1, 2, \ldots$ denote the autocovariance function of $\{X_t\}$ with respect to the time lag h, and the corresponding autocorrelation function (ACF) is defined as $\rho_P(h) = \gamma_h/\gamma_0$. Under the weak stationarity assumption, the joint distribution of the random variables $(X_1, \ldots, X_n)$ has mean vector $\boldsymbol{\mu_X} = \mathbf{1}\mu$, where $\mathbf{1}$ denotes a length n vector of ones, and a symmetric Toeplitz covariance matrix $\Sigma_{\boldsymbol{X}}$, with $[\Sigma_{\boldsymbol{X}}]_{i,j} = \gamma(|i-j|)$. Furthermore, both $\Sigma_{\boldsymbol{X}}$ and the corresponding correlation matrix $\Omega_{\boldsymbol{X}}$ are positive definite for any stationary sequence. For an observed stationary time-series $\boldsymbol{X}_{1:n} = \{x_t\}_{t=1}^n$, the mean is estimated as before ($\hat{\mu}_X$), while autocovariances and autocorrelations are commonly estimated as $\hat{\gamma}(\boldsymbol{X}_{1:n}; h) = \hat{\gamma}(h) = \frac{1}{n}\sum_{t=h+1}^n (x_t - \hat{\mu}_X)(x_{t-h} - \hat{\mu}_X)$ and $\hat{\rho}_P(\boldsymbol{X}_{1:n}; h) = \hat{\rho}_P(h) = \hat{\gamma}(h)/\hat{\gamma}(0)$, respectively. Using the scaling $\frac{1}{n}$ as opposed to $\frac{1}{n-h}$ assures that the corresponding estimated covariance and correlation matrices $[\hat{\Sigma}_{\boldsymbol{X}}]_{i,j} = \hat{\gamma}(|i-j|)$ and $[\hat{\Omega}_{\boldsymbol{X}}]_{i,j} = \hat{\rho}_P(|i-j|)$ are both positive definite (McLeod and Jimenéz, 1984).

8.2.1 Transformations and Rank-Based Methods

Pearson's correlation and the autocorrelation function are commonly interpreted under an implicit joint normality assumption on the data. The alternative measures discussed in this subsection offer robustness to outlying and extreme observations and consider nonlinear dependencies, all of which are common in financial data.

8.2.1.1 Huber-type Correlations

Transformations may be applied to define a robust correlation measure between pairs of random variables (X, Y). For example, let μ_R and σ_R denote robust location and scale parameters, such as the trimmed mean and trimmed standard deviation. And let ψ denote a bounded monotone function, such as $\psi(x; k) = xI_{|x|\leq k} + \text{sgn}(x)kI_{|x|>k}$ (cf. Huber, 1981), where k is some positive constant and I_A is the indicator function of an event A. Then, a robust covariance and correlation may be defined as

$$\gamma_R(X, Y) = \sigma_R(X)\sigma_R(Y)E\left[\psi\left(\frac{X - \mu_R(X)}{\sigma_R(X)}\right)\psi\left(\frac{Y - \mu_R(Y)}{\sigma_R(Y)}\right)\right]$$

and

$$\rho_R(X, Y) = \frac{\gamma_R(X, Y)}{\sqrt{\gamma_R(X, X)}\sqrt{\gamma_R(Y, Y)}},$$

respectively (cf. Maronna *et al.*, 2006). Although these measures are robust to outlying and extreme observations, they depend on the choice of transformation ψ. They provide an intuitive measure of association, however, for an arbitrary joint distribution on (X, Y), $\rho_R(X, Y) \neq \rho_P(X, Y)$, in general. Sample versions are obtained by replacing the expectation, μ_R and σ_R, by their sample estimates in the above expressions. Asymptotic sampling distributions can be derived, but the setting is more complicated than above. Finally, robust pairwise covariances and correlations such as these may be used to define corresponding covariance and correlation matrices for random vectors, but the result is not positive definite (or affine equivariant), in general.

8.2.1.2 Kendall's Tau

Rank-based methods measure all monotonic relationships and are resistant to outliers. Kendall's tau (Kendall, 1938) is a nonparametric measure of dependence, for which the sample estimate considers the pairwise agreement between two ranked lists.

Definition 8.4 *For random variables X and Y,* **Kendall's tau** *is defined as*

$$\begin{aligned}\tau(X, Y) &= P[(X - X^*)(Y - Y^*) > 0] - P[(X - X^*)(Y - Y^*) < 0] \\ &= E\{sgn[(X - X^*)(Y - Y^*)]\},\end{aligned} \tag{8.1}$$

in which (X^*, Y^*) *denotes an i.i.d. copy of* (X, Y).

A sample estimator $\hat{\tau}$ considers the pairwise agreement between two ranked lists. For a paired random sample of n observations $(\boldsymbol{X}_{1:n}, \boldsymbol{Y}_{1:n}) = \{(x_i, y_i)\}_{i=1}^{n}$, the sample analog of (8.1) is defined as

$$\hat{\tau}(\boldsymbol{X}_{1:n}, \boldsymbol{Y}_{1:n}) = \binom{n}{2}^{-1} \sum_{j=2}^{n} \sum_{i=1}^{j-1} \operatorname{sgn}(x_j - x_i)\operatorname{sgn}(y_j - y_i),$$

and, in the absence of ties, is also equal to

$$\frac{\text{number of concordant pairs} - \text{number of discordant pairs}}{\text{number of concordant pairs} + \text{number of discordant pairs}},$$

in which a pair of observations $(x_j, y_j), (x_i, y_i)$ are concordant if $\operatorname{sgn}(x_j - x_i) = \operatorname{sgn}(y_j - y_i)$, and discordant otherwise. These definitions may also be extended to define analogous autocorrelation functions.

Definition 8.5 *The* **ACF based on Kendall's tau** *(Bingham and Schmidt, 2006) for a stationary sequence of random variables $\{X_t\}$ is defined as*

$$\begin{aligned}\tau(X_t; h) = \tau_h &= P[(X_t - X_t^*)(X_{t-h} - X_{t-h}^*) > 0] - P[(X_t - X_t^*)(X_{t-h} - X_{t-h}^*) < 0] \\ &= E\{sgn[(X_t - X_t^*)(X_{t-h} - X_{t-h}^*)]\},\end{aligned} \tag{8.2}$$

in which (X_t^*, X_{t-h}^*) denotes an i.d.d. copy of (X_t, X_{t-h}).

For an observed time-series $\boldsymbol{X}_{1:n} = \{x_t\}_{t=1}^{n}$, the sample analog of (8.2) is defined as

$$\hat{\tau}(\boldsymbol{X}_{1:n}; h) = \hat{\tau}_h = \binom{n-h}{2}^{-1} \sum_{t=2}^{n-h} \sum_{i=1}^{t-1} \operatorname{sgn}(x_t - x_i)\operatorname{sgn}(x_{t+h} - x_{i+h}),$$

which is not assured to have a corresponding estimated correlation matrix that is positive definite.

In certain applications, Kendall's tau has a distinct advantage over Pearson's correlation in that it is invariant to increasing monotonic transformations of one or both variables. For example, $\tau(X, Y) = \tau(X, \log\ Y) = \tau(\log\ X, \log\ Y)$, for positive random variables, and similarly for $\hat{\tau}$. Bingham and Schmidt (2006) consider an example based on a time-series of squared asset returns, which exhibits abnormally high dependence at a specific lag. This is exaggerated in a Pearson ACF, as it only measures linear dependence. After applying a data standardization, the magnitude of the Pearson ACF at this lag decreases substantially, whereas the ACF based on Kendall's tau is very similar under both specifications.

Figure 8.2 offers an example of a scenario in which the autocorrelation based on Kendall's tau differs substantially from that produced from a Pearson ACF. It depicts the Bank of America (BOA) stock price and the estimated ACF of the squared daily stock returns from June 2008 to January 2009, a very volatile period that included the failure of Lehman Brothers, Washington Mutual, and several other financial institutions. The ACF based on Kendall's tau detects persistent serial dependence across several lags, while the Pearson ACF detects dependence only at the first and 10th lags. This discrepancy suggests that the underlying serial dependence is highly nonlinear and non-Gaussian in nature.

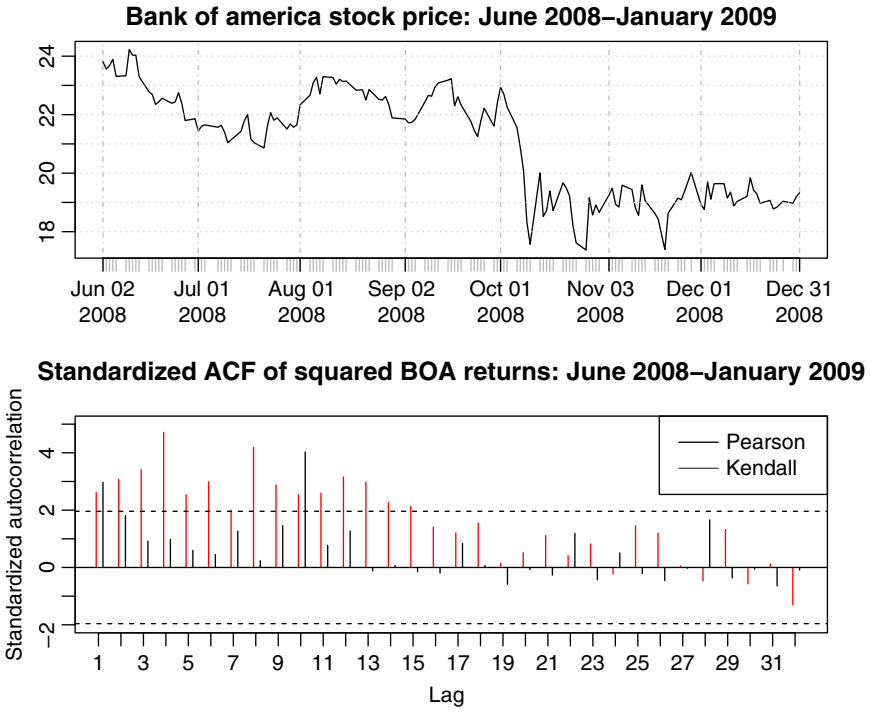

Figure 8.2 Bank of America (BOA) daily closing stock price. Bottom: Standardized (Fisher's transformation) ACF based on Kendall's tau and Pearson's correlation coefficient for the squared daily stock returns.

8.2.1.3 Spearman's Rho

One of the shortcomings of Kendall's tau, as discussed in Helsel and Hirsch (1992), is that since it is purely based on sign, the magnitude of concordance or discordance is not taken into account. Spearman's rho (Spearman, 1904), which computes Pearson's correlation on the normalized ranks in the sample case, does explicitly take magnitude into account. Though Kendall's tau and Spearman's rho are asymptotically equivalent when testing for independence, Spearman's rho can be preferable in certain scenarios. For example, Xu *et al.* (2013) conduct a simulation study with bivariate contaminated normal data and conclude that Spearman's rho is preferable to Kendall's tau in small sample sizes when the population correlation is moderate.

Definition 8.6 *The* **ACF based on Spearman's rho** *(Hollander and Wolfe, 1999) for a stationary sequence of random variables* $\{X_t\}$ *with common marginal distribution* F *is defined as*

$$\rho_S(X_t; h) = \rho_S(h) = \rho_P[F(X_t), F(X_{t-h})],$$

that is, the Pearson correlation of $F(X_t)$ *and* $F(X_{t-h})$.

For an observed time-series $\boldsymbol{X}_{1:n} = \{x_t\}_{t=1}^n$, a sample version may be defined as

$$\hat{\rho}_S(\boldsymbol{X}_{1:n}; h) = \hat{\rho}_S(h) = \frac{12}{n(n^2-1)} \sum_{t=1+h}^{n} \left(r_t - \frac{n+1}{2} \right) \left(r_{t-h} - \frac{n+1}{2} \right),$$

in which $r_t = rank\{x_t : x_t \in \boldsymbol{X}_{1:n}\}$.

8.2.2 Inference

8.2.2.1 Kendall's Tau

We can test the hypothesis of zero correlation, that is,

$$H_0 : \tau_h = 0 \quad vs. \quad H_1 : \tau_h \neq 0,$$

with the decision rule: Reject H_0 if $|\hat{\tau}_h| \leq \kappa_{\alpha/2}$, in which $\kappa_{\alpha/2}$ can be computed from a table of critical values. For larger sample sizes, a normal approximation may be utilized on a standardized version of $\hat{\tau}_h$

$$\hat{\tau}_h^* = \frac{\hat{\tau}_h - E_0(\hat{\tau}_h)}{\sqrt{Var_0(\hat{\tau}_h)}},$$

where $E_0(\hat{\tau}_h)$ and $Var_0(\hat{\tau}_h)$ represent the respective mean and variance of $\hat{\tau}_h$ under the null hypothesis of independence. Under serial independence, Ferguson *et al.* (2000) derive

$$E_0(\hat{\tau}_h) = \begin{cases} \frac{(3n-3h-1)(n-h)}{12} & \text{if } 1 \leq h \leq \frac{n}{2} \\ \frac{(n-h)(n-h-1)}{4} & \text{if } \frac{n}{2} \leq h < n-1. \end{cases}$$

The variance is derived in the case where $h = 1$ as

$$Var_0(\hat{\tau}_h) = \frac{10n^3 - 37n^2 + 27n + 74}{360}.$$

A formula for a general lag h (assuming $n \geq 4h$) is constructed by Šiman (2012) as:

$$Var_0(\hat{\tau}_h) = \frac{2(10n^3 + (7-30h)n^2 + (30h^2 + 46h - 49)n + (-10h^3 - 29h^2 + 114h))}{45(n-h)^2(n-h-1)^2}.$$

It was additionally shown in Ferguson *et al.* (2000) that $\hat{\tau}_h$ is asymptotically standard normal and independent across lags h.

8.2.2.2 Spearman's Rho

When n is large, we consider Fisher's transformation of Spearman's rho

$$\hat{z}_h = \frac{1}{2} \log \left(\frac{1 + \hat{\rho}_S(h)}{1 - \hat{\rho}_S(h)} \right)$$

to test the hypothesis

$$H_0 : \rho_S(h) = 0 \qquad vs. \qquad H_1 : \rho_S(h) \neq 0.$$

The corresponding test statistic is $\hat{z}_h^* = \hat{z}_h / \sqrt{Var(\hat{z}_h)}$. Per Anderson (1954), the standard error of z is approximated as $\sqrt{Var(\hat{z}_h)} = \sqrt{n-3}$, and $\hat{z}_h^*$ can be compared to the standard normal distribution.

Example 8.1 *To illustrate a setting in which rank-based measures are more informative than Pearson's correlation coefficient, we consider data simulated from a cubic moving average model (adapted from Ashley and Patterson, 2001). Let*

$$X_t = a_t + 0.02a_{t-1}^3, \text{ with} \qquad a_t \overset{\text{i.i.d.}}{\sim} F(0, \sigma_a^2).$$

Clearly, there is serial dependence at lag $h = 1$, but it is nonlinear in nature.

We consider three distributions F for the mean zero i.i.d. innovations a_t: standard normal; Student's T distribution with 5 degrees of freedom (hence, heavy tails); and a stable Pareto distribution with shape parameter $\alpha = 1.93$, which has characteristic function $E(e^{i\theta X}) = \exp(-|\theta|^\alpha)$. The choice of $\alpha = 1.93$ is believed by Fama (1965) to adequately capture the leptokurtic nature of US stock returns. From its construction, it is clear that the second moment of this distribution does not exist. One simulation from each process is shown in Figure 8.3. At each iteration,

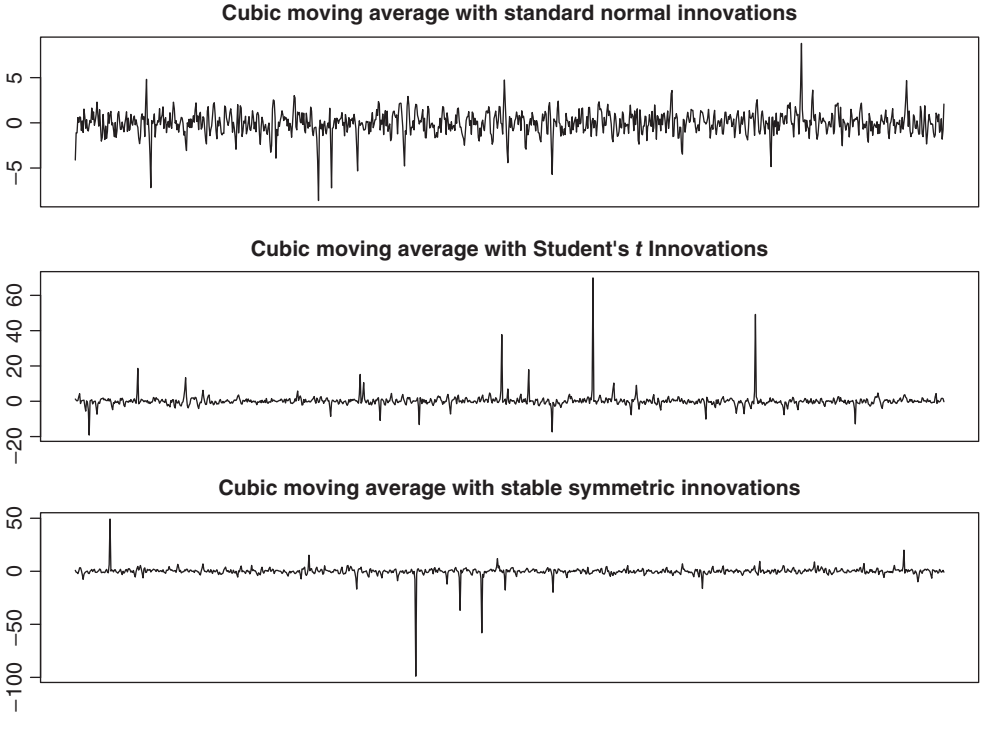

Figure 8.3 Realized time-series simulated from each of the three process models discussed in Example 8.1.

Table 8.1 Percentage of tests failing to reject H_0 of no lag-1 correlation

Error distribution	Pearson	Kendall	Spearman
Standard normal	0.00	0	0
Student's T	0.13	0	0
Stable Pareto	0.53	0	0

1000 observations were drawn from each model and the lag-1 correlations were tested under the hypotheses stated in this chapter. This process was repeated 5000 times, and the results are summarized in Table 8.1.

These simulations provide evidence to support the claim that Pearson's correlation coefficient is robust to nonlinearity, but performs poorly in the presence of leptokurtic distributions and outliers. Hence, one should be judicious in measuring and testing autocorrelation, especially when working with data exhibiting these characteristics.

8.2.3 Misspecification Testing

8.2.3.1 Ljung–Box Test

Assuming a time-series $\boldsymbol{X}_{1:n}$ is i.i.d. with finite variance, it has no autocorrelation at any lag, that is, $\rho_P(h) = 0$ for all $h \geq 1$. Additionally, $\sqrt{n}\hat{\rho}_P(h)$ is asymptotically $N(0,1)$, and $\hat{\rho}_P(1), \ldots, \hat{\rho}_P(m)$ are asymptotically independent. Thus, we can perform a joint test with the hypotheses

$$H_0 : \rho_P(1) = \rho_P(2) = \ldots = \rho_P(m) = 0 \quad vs. \quad H_1 : \rho_P(h) \neq 0 \text{ for some } h = 1, \ldots, m.$$

This hypothesis is often tested on the residuals of a fitted time-series model. For example, consider the residuals $\hat{\boldsymbol{a}}_{1:n}$ from an autoregressive moving average [ARMA(p,q)] model. Either the Ljung–Box statistic $Q^*(m)$ (Ljung and Box, 1978) or the Box–Pierce statistic $Q(m)$ (Box and Pierce, 1970), which are asymptotically equivalent, can be used:

$$Q^*(\hat{\boldsymbol{a}}_{1:n}; m) = n(n+2)\sum_{h=1}^{m} \hat{\rho}(\hat{\boldsymbol{a}}_{1:n}; h)^2/(n-h), Q(\hat{\boldsymbol{a}}_{1:n}; m) = n\sum_{h=1}^{m} \hat{\rho}(\hat{\boldsymbol{a}}_{1:n}; h)^2. \tag{8.3}$$

The null hypothesis is rejected at significance level α if the statistic exceeds the $1-\alpha$ quantile of the χ^2_{m-p-q} distribution. Per Lai and Xing (2008), simulation studies suggest that the Ljung–Box statistic is better approximated by a χ^2_{m-p-q} than the Box–Pierce statistic, hence it is preferred for small to moderate-sized samples. Analogous joint tests are formed for both $\tau(h)$ and $\rho_S(h)$ by leveraging the asymptotic normality and independence across lags by applying similar transformations.

8.2.3.2 Conditional Heteroscedasticity

Traditionally, the Ljung–Box test is performed on the residuals of a linear or conditional mean for a given series. However, in financial applications, conditional variance (also known as

volatility), as opposed to conditional mean, is often a primary parameter of interest. Per Tsay (2010), in many financial applications, a daily asset return series $\boldsymbol{X}_{1:n} = \{x_t\}_{t=1}^n$ generally is either serially uncorrelated or has minor correlations that can be accounted for with a low-order ARMA model. However, substantial autocorrelation may exist in the squared return series $\{x_t^2\}$, and the volatility $\{\sigma_t^2\}$ sequence is the focus of modeling. The autoregressive conditional heteroscedasticity (ARCH) model and its generalization are popular models for forecasting volatility.

Definition 8.7 *Let $\{a_t\}$ denote the innovations from a return series centered by its conditional mean (e.g., ARMA) equation. Then, the* **autoregressive conditional heteroscedasticity model** *of order r (*ARCH_r *model) is defined as*

$$a_t = \sigma_t \varepsilon_t, \quad \varepsilon_t \overset{iid}{\sim} F(0,1),$$
$$\sigma_t^2 = \omega + \alpha_1 a_{t-1}^2 + \ldots + \alpha_r a_{t-r}^2,$$

in which $\{\varepsilon_t\}$ represents a sequence of i.i.d. standardized innovations (Engle, 1982).

Normal random variables are commonly used for F, but other asymmetric and heavy-tailed distributions, such as a standardized skewed Student's T, can be substituted. The volatility equation intercept ω is positive, and the coefficients $\alpha_1, \ldots, \alpha_r$ are constrained to be nonnegative and sum to less than one to ensure positivity of $\{\sigma_t^2\}$ and stationarity of $\{(a_t, \sigma_t^2)\}$. The ARCH model was generalized to allow for a more parsimonious representation when r is large.

Definition 8.8 *The* **generalized ARCH model** *(*$\text{GARCH}_{r,s}$*) for $\{a_t\}$ is defined as above, but with*

$$\sigma_t^2 = \omega + \sum_{i=1}^{r} \alpha_i a_{t-i}^2 + \sum_{j=1}^{s} \beta_j \sigma_{t-j}^2,$$

with $\omega > 0$, $\alpha_i, \beta_j \geq 0$ for positivity, and $\sum_{i=1}^r \alpha_i + \sum_{j=1}^s \beta_j < 1$ for stationarity (Bollerslev, 1986).

8.2.3.3 Parametric Tests for ARCH Effects

Observations from an ARCH or GARCH process are dependent. In particular, the sequence of squared observations $\boldsymbol{a}_{1:n}^2 = \{a_t^2\}_{t=1}^n$ have autocorrelation, which is referred to as ARCH effects. Perhaps the most straightforward test for ARCH effects is the McLeod–Li test, which simply computes the Ljung–Box statistic (8.3) on the squared innovations a_t^2 (or on squared residuals $\hat{a}_t^2$), in place of the innovations a_t. It was shown in McLeod and Li (1983) that under the null hypothesis of independent observations, for fixed m, the scaled correlations $\sqrt{n}[\hat{\rho}_P(\boldsymbol{a}_{1:n}^2; 1), \ldots, \hat{\rho}_P(\boldsymbol{a}_{1:n}^2; m)]$ are asymptotically normally distributed with mean zero and identity covariance matrix. The McLeod–Li statistic is then computed as

$$Q_{MC}^*(\boldsymbol{a}_{1:n}^2; m) = n(n+2) \sum_{h=1}^{m} [\hat{\rho}_P(\boldsymbol{a}_{1:n}^2; h)]^2 / (n-h),$$

which is asymptotically χ_m^2 distributed under H_0, and no adjustment is needed when $\boldsymbol{a}_{1:n}^2$ is replaced by $\hat{\boldsymbol{a}}_{1:n}^2$.

A popular alternative to test for ARCH effects is the Lagrange multiplier test of Engle (1982). Per Tsay (2010), the test is equivalent to using a standard F-statistic to conduct the joint test that all $\alpha_i = 0$ in the regression equation

$$a_t^2 = \omega + \alpha_1 a_{t-1}^2 + \ldots + \alpha_m a_{t-m}^2 + e_t, \tag{8.4}$$

where m is a prespecified lag order. The hypotheses are

$$H_0 : \alpha_1 = \ldots = \alpha_m = 0 \quad vs. \quad H_1 : \alpha_h \neq 0 \text{ for some } h = 1, \ldots, m.$$

To calculate the F-statistic, first define $\text{SSR}_0 = \sum_{t=m+1}^{n}(a_t^2 - \hat{\omega}_0)$, where $\hat{\omega}_0 = \frac{1}{n}\sum_{t=m+1}^{n} a_t^2$, and $\text{SSR}_1 = \sum_{t=m+1}^{n} \hat{e}_t^2$, where $\{\hat{e}_t\}$ represents the residuals from (8.4). The F-statistic is then defined as

$$F(\boldsymbol{a}_{1:n}^2; m) = \frac{(\text{SSR}_0 - \text{SSR}_1)/m}{\text{SSR}_1/(n - 2m - 1)}, \tag{8.5}$$

which is also asymptotically χ_m^2 distributed under H_0 (Luukkonen *et al.*, 1988).

Another test developed by Tsay (1986) specifically looks for quadratic serial dependence by incorporating cross-product terms. Let $\boldsymbol{M}_{t-1} = \text{vech}[(1, x_{t-1}, \ldots, x_{t-p})(1, x_{t-1}, \ldots, x_{t-p})']$. Then, consider the regression

$$x_t = (1, x_{t-1}, \ldots, x_{t-p})'\boldsymbol{\phi} + \boldsymbol{M}_{t-1}\boldsymbol{\alpha} + a_t,$$

where $\boldsymbol{\phi}$ is a $(p+1) \times 1$ coefficients vector and $\boldsymbol{\alpha}$ is a $p(p-1)/2 \times 1$ coefficients vector. If the AR(p) model is adequate, then $\boldsymbol{\alpha}$ should be zero. This can be tested using a standard F-statistic similar to (8.5), but with $[p(p+1)/2, n - p - p(p+1)/2 - 1]$ degrees of freedom.

8.2.3.4 Nonparametric Tests for the ARCH Effect

The BDS test (Brock *et al.*, 1996) originated as a test for the detection of nonrandom chaotic dynamics, but has gained traction as a test for nonlinear serial dependence in financial time-series. The BDS test is based on a correlation integral. To construct a correlation integral for a series $\boldsymbol{X}_{1:n} = \{x_t\}_{t=1}^n$, define a sequence of m-vectors as $\boldsymbol{x}_t^m = (x_t, x_{t-1}, \ldots, x_{t-m+1})'$, in which m is the embedding dimension. The correlation integral measures the fraction of pairs of m-vectors that are close to each other (given some range parameter $\epsilon > 0$). As described in Patterson and Ashley (2000), it counts the number of m-vectors that lie within a hypercube of size ϵ of each other.

Definition 8.9 An empirical correlation integral *of order m and range* $\epsilon > 0$ *is*

$$C_{m,\epsilon}(\boldsymbol{X}_{1:n}) = \frac{2}{n_m(n_m - 1)} \sum_{m \leq s < t \leq n} I(x_t^m, x_s^m; \epsilon), \tag{8.6}$$

in which $n_m = n - m + 1$*, and* $I(x_t^m, x_s^m; \epsilon)$ *is one if* $|x_{t-i} - x_{s-i}| < \epsilon$ *for* $i = 0, \ldots, m-1$*, and zero otherwise.*

Under the null hypothesis of serial independence, the correlation integral factorizes as $C_{m,\epsilon} = (C_{1,\epsilon})^m$. Correlation integrals are U-statistics, so asymptotic theory can readily be applied. The

BDS test statistic is defined in Zivot and Wang (2007) as

$$V_{m,\epsilon} = \sqrt{n}\frac{C_{m,\epsilon} - (C_{1,\epsilon})^m}{s_{m,\epsilon}},$$

where $s_{m,\epsilon}$ is an estimate of the asymptotic standard deviation of $C_{m,\epsilon} - (C_{1,\epsilon})^m$, as derived in Brock *et al.* (1996).

Under mild regularity conditions, the BDS statistic converges in distribution to a standard normal distribution, though convergence can be very slow. As stated in Patterson and Ashley (2000), convergence for large m requires an extremely large sample size. The null hypothesis of i.i.d. data is rejected if $|V_{m,\epsilon}| > Z_{\alpha/2}$. Per Diks (2009), the range parameter ϵ is typically set to 0.5 to 1.5 times the sample standard deviation of the observed series. In practice, the test is constructed for several values of ϵ, which are then compared against one another as a stability check. As an alternative to relying on asymptotic normality, Genest *et al.* (2007) developed a rank-based extension of the BDS test whose finite sample p-values can be constructed by simulation.

The BDS test also functions as a model misspecification test. Brock *et al.* (1996) shows that the asymptotic distribution of the test statistic for residuals is the same as that of true innovations. Similar to the aforementioned parametric tests, it is often performed on the residuals of a fitted model in order to test for any remaining dependence. However, as pointed out by Diks (2009), though this holds for autoregressive models, it does not hold for models in the ARCH family. While it is not free of nuisance parameters for GARCH models, Caporale *et al.* (2005) found the BDS test to perform very well on the logged squared residuals of a GARCH(1,1) model.

8.2.3.5 Comparative Performance of Misspecification Tests

Ashley and Patterson (2001) conducted a simulation study of the BDS, Mcleod–Li, Tsay, Lagrange multiplier, and several other tests over a wide variety of data-generating processes. They found that the BDS test has the highest power against all alternatives, rendering it very useful as a so-called nonlinearity screening test, which can detect any residual nonlinear serial dependence. However, it is not informative as to the type of nonlinear structure that may be present in the observations. The Tsay test was also found to perform well in the detection of self-excited threshold autoregression models.

8.3 Multivariate Extensions

The vector autoregression (VAR) is the multivariate analog of the autoregressive model. It allows simultaneous and dynamic linear dependence across multiple components.

Definition 8.10 *A pth order* **vector autoregression** *is defined as*

$$\mathbf{y}_t = \boldsymbol{\phi} + \boldsymbol{\Phi}_1\mathbf{y}_{t-1} + \ldots + \boldsymbol{\Phi}_p\mathbf{y}_{t-p} + \mathbf{a}_t, \tag{8.7}$$

in which $\boldsymbol{\phi} \in \mathbb{R}^k$, $\mathbf{y}_t, \mathbf{a}_t \in \mathbb{R}^k$ *for all t; each* $\boldsymbol{\Phi}_i$ *is a* $k \times k$ *coefficient matrix; and* $\mathbf{a}_t$ *is a serially uncorrelated weakly stationary sequence (white noise) with mean vector zero and nonsingular* $k \times k$ *covariance matrix* $\boldsymbol{\Sigma}_a$.

8.3.1 Multivariate Volatility

The vector autoregression (8.7) operates under the assumption that the innovations are white noise, but it allows for nonlinear serial dependence. Such dependence is common in financial data, particularly multivariate ARCH effects. In this case, we may decompose the linear and nonlinear dynamics as

$$\mathbf{y}_t = \boldsymbol{\mu}_t + \mathbf{a}_t, \quad \mathbf{a}_t = \boldsymbol{\Sigma}_t^{1/2}\boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \overset{iid}{\sim} F(\mathbf{0}, \boldsymbol{I}),$$

in which $\boldsymbol{\mu}_t$ denotes the conditional mean given the past observations $E(\mathbf{y}_t|\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \ldots)$ (possibly estimated with a VAR); and $\boldsymbol{\epsilon}_t$ represents an i.i.d sequence with mean zero and identity covariance. As with the univariate ARCH case, we want to consider processes in which the conditional covariance matrix $\boldsymbol{\Sigma}_t$ is dependent across time, where $\boldsymbol{\Sigma}_t = Cov(\mathbf{y}_t|\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \ldots) = Cov(\mathbf{a}_t|\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \ldots)$. As outlined in Tsay (2014), there are many challenges in multivariate volatility modeling. The dimension of $\boldsymbol{\Sigma}_t$ increases quadratically with k, and restrictions are needed to ensure $\boldsymbol{\Sigma}_t$ is positive definite for all t.

One way to model multivariate volatility is to extend the conventional GARCH models to a multivariate setting. A simple approach is to model a half-vectorization (vech) of the diagonal and lower triangle elements of $\boldsymbol{\Sigma}_t$.

Definition 8.11 *The* **VECH GARCH model** *(Bollerslev* et al.*, 1992) for a k-dimensional series has volatility defined by the recursion*

$$vech(\boldsymbol{\Sigma}_t) = vech(\boldsymbol{\Omega}) + \sum_{i=1}^{r} \boldsymbol{A}_i \; vech(\boldsymbol{\epsilon}_{t-i}\boldsymbol{\epsilon}_{t-i}^{'}) + \sum_{j=1}^{s} \boldsymbol{B}_j \; vech(\boldsymbol{\Sigma}_{t-j}),$$

in which $\boldsymbol{\Omega}$ is a $k \times k$ symmetric, positive definite matrix; and the coefficient matrices $\boldsymbol{A}_i$ and $\boldsymbol{B}_j$ each have dimension $k(k+1)/2 \times k(k+1)/2$.

As stated in Jondeau *et al.* (2007), this model requires estimating on the order of k^4 parameters, and it is difficult to incorporate restrictions that guarantee the conditional covariances $\boldsymbol{\Sigma}_t$ will be positive definite.

One popular model that addresses some of these concerns is the Baba, Engle, Kraft, and Kroner (BEKK) representation.

Definition 8.12 *The* **BEKK GARCH model** *(Engle and Kroner, 1995) for a k-dimensional series has volatility defined by the recursion*

$$\boldsymbol{\Sigma}_t = \tilde{\boldsymbol{\Omega}} + \sum_{i=1}^{r} \tilde{\boldsymbol{A}}_i^{'} \boldsymbol{\epsilon}_{t-i}\boldsymbol{\epsilon}_{t-i}^{'} \tilde{\boldsymbol{A}}_i + \sum_{j=1}^{s} \tilde{\boldsymbol{B}}_j^{'} \boldsymbol{\Sigma}_{t-j} \tilde{\boldsymbol{B}}_j,$$

in which $\tilde{\boldsymbol{\Omega}}$ is a $k \times k$ symmetric, positive definite matrix; and the coefficient matrices $\tilde{\boldsymbol{A}}_i$ and $\tilde{\boldsymbol{B}}_j$ each have dimension $k \times k$.

The major advantage of this model is that $\boldsymbol{\Sigma}_t$ is positive definite as long as $\tilde{\boldsymbol{\Omega}}$ is positive definite and the sequence is initialized at positive definite values. Estimation includes $k(k+1)/2 + (r+s)k^2$ parameters. In certain cases, to reduce the computational burden, the matrices

$\tilde{\boldsymbol{A}}_i$ and $\tilde{\boldsymbol{B}}_j$ are assumed to be diagonal. However, the BEKK model does not scale well, even in the diagonal case; the number of parameters grows quadratically with the number of variables. For additional model structures, the interested reader can consult Bauwens *et al.* (2006), which conducts a comprehensive survey of multivariate volatility models.

8.3.2 Multivariate Misspecification Testing

Many univariate misspecification tests have multivariate extensions. Referencing Equation (8.7), the lagged cross-covariance matrices of $\{\mathbf{a}_t\}$ can be estimated as

$$\boldsymbol{C}_h = \frac{1}{n}\sum_{t=h+1}^{n} \mathbf{a}_t \mathbf{a}_{t-h}'.$$

Let $\boldsymbol{D}$ denote a diagonal matrix with $[\boldsymbol{D}]_{ii} = [\boldsymbol{C}_0]_{ii}^{1/2}$, then the lagged cross-correlation matrices of $\{\mathbf{a}_t\}$ can be estimated as

$$\boldsymbol{R}_h = \boldsymbol{D}^{-1}\boldsymbol{C}_h\boldsymbol{D}^{-1}.$$

The Ljung–Box test was extended to a multivariate setting by Baillie and Bollerslev (1990), to test the hypotheses

$$H_0 : \boldsymbol{R}_1 = \cdots = \boldsymbol{R}_m = \mathbf{0} \quad vs. \quad H_1 : \mathbf{R}_h \neq \mathbf{0}, \text{ for some } h = 1, \ldots, m.$$

The test statistic recommended by Lütkepohl (2007) is

$$\bar{\boldsymbol{Q}}(\{\mathbf{a}_t\}; m) = n^2 \sum_{h=1}^{m} \text{trace}(\boldsymbol{C}_h' \boldsymbol{C}_0^{-1} \boldsymbol{C}_h \boldsymbol{C}_0^{-1})/(n-h),$$

which asymptotically follows a $\chi^2_{mk^2}$ distribution under H_0. For residuals $\{\hat{\mathbf{a}}_t\}$ from an estimated $VAR_k(p)$ model, the $\boldsymbol{C}_h$ are defined analogously, as is $\bar{\boldsymbol{Q}}(\{\hat{\mathbf{a}}_t\}; m)$, which instead has a $\chi^2_{(m-p)k^2}$ asymptotic distribution under H_0. Furthermore, analogous hypotheses and tests can be conducted for the multivariate version of the ARCH effect based on $\{\mathbf{a}_t^2\}$, $\{\hat{\mathbf{a}}_t^2\}$, or $\{\hat{\boldsymbol{\epsilon}}_t^2\}$, where $\hat{\boldsymbol{\epsilon}}_t = \hat{\boldsymbol{\Sigma}}_t^{-1/2}\hat{\mathbf{a}}_t$, for an estimated volatility sequence $\{\hat{\boldsymbol{\Sigma}}_t\}$.

8.3.3 Granger Causality

In financial applications, it is widely believed (cf. Gallant *et al.*, 1992) that the joint dynamics of stock prices and trading volume can be more informative as to the underlying state of the stock market than stock prices alone. Many studies have explicitly tested for a causal link between stock prices and trading volume. The notion of Granger causality, developed by Granger (1969), can be used to determine whether one time-series helps to predict another. As a simple example, consider a bivariate time-series $\{(x_t, y_t)\}_{t=1}^n$ specified by the equations

$$x_t = \nu_1 + \phi x_{t-1} + \gamma y_{t-1} + a_{1t},$$
$$y_t = \nu_2 + \beta y_{t-1} + \delta x_{t-1} + a_{2t}$$

where $a_{1t}, a_{2t} \text{iid}\sim N(0, \sigma^2)$. The series y_t is said to "Granger-cause" x_t if $\gamma \neq 0$.

More generally, given a prespecified lag length p, consider the marginal model for $\{x_t\}$ defined by

$$x_t = \nu + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + \gamma_1 y_{t-1} + \gamma_2 y_{t-2} + \cdots + \gamma_p y_{t-p} + a_t \tag{8.8}$$

To test for Granger causality, we may test the hypotheses

$$H_0 : \gamma_1 = \cdots = \gamma_p = 0 \quad vs. \quad H_1 : \gamma_h \neq 0 \text{ for some } \quad h = 1, \ldots, p.$$

As a test statistic, one can use the sum of squared residuals from (8.8)

$$RSS_{\text{Full}} = \sum_{t=1}^{n} \hat{a}_t^2,$$

which can then be compared with the sum of squared residuals from a univariate autoregression of $\{x_t\}$, for example,

$$x_t = \nu + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + u_t, \quad RSS_{\text{Null}} = \sum_{i=1}^{n} \hat{u}_i^2.$$

Then a F-statistic is constructed as

$$F(\{(x_t, y_t)\}; p) = \frac{(RSS_{\text{Null}} - RSS_{\text{Full}})/p}{RSS_{\text{Full}}/(n - 2p - 1)} \tag{8.9}$$

which is asymptotically distributed as $F(p, 2p-1)$ under H_0. There are several additional methods to test for Granger causality, but simulation studies by Geweke *et al.* (1983) suggest that (8.9) achieves the best performance. For an overview of additional tests, the interested reader is referred to Chapter 11 of Hamilton (1994).

8.3.4 Nonlinear Granger Causality

The aforementioned test for Granger causality will only elucidate whether $\{y_t\}$ can help to predict $\{x_t\}$ linearly. Potential nonlinear relationships will remain undetected. Consider the following example, procured from Baek and Brock (1992),

$$x_t = \beta y_{t-q} x_{t-p} + a_t$$

There exists an obvious but nonlinear relationship that will not be evident in Granger causality testing. In a similar fashion to the BDS test, Baek and Brock (1992) develop a test for nonlinear Granger causality based on the correlation integral (8.6).

Definition 8.13 (Nonlinear Granger Causality) $\{y_t\}$ *does not nonlinearly Granger-cause* $\{x_t\}$ *if:*

$$Pr(|x_t - x_s| < \epsilon_1 | |x_{t,p} - x_{s,p}| < \epsilon_1, |y_{t,q} - y_{s,q}| < \epsilon_2) = Pr(|x_t - x_s| < \epsilon_1 | |x_{t,p} - x_{s,p}| < \epsilon_1),$$

in which $z_{t,m} = (z_{t-1}, \ldots, z_{t-m})$. *In words: "Given* ϵ_1 *and* ϵ_2, q *lags of Y do not incrementally help to predict next period's value given p lags of X."*

The notion of nonlinear Granger causality has been extensively developed to analyze relationships in finance. In particular, Hiemstra and Jones (1994) examine the relationship between aggregate stock prices and trading volume.

They expand upon Baek and Brock (1992) and give a detailed description of a hypothesis test (henceforth, the H-J test).

Under the null hypothesis of the H-J test, it is assumed that $\{X_t\}$ does not nonlinearly Granger cause $\{Y_t\}$. This is determined by testing conditional independence using finite lags ℓ_x and ℓ_y:

$$Y_{t+1}|X_t^{\ell_x}, Y_t^{\ell_y} \sim Y_{t+1}|Y_t^{\ell_y},$$

where $X_t^{\ell_x} = (X_{t-\ell_x+1}, \ldots, X_t)$ and, $Y_t^{\ell_y} = (Y_{t-\ell_y+1}, \ldots, Y_t)$. Assuming the bivariate time-series $\{X_t, Y_t\}$ is strictly stationary, the H-J test can be interpreted as a statement about the invariant distribution of the $(\ell_x + \ell_y + 1)$ dimensional random vector $W_t = (X_t^{\ell_x}, Y_t^{\ell_y}, Z_t)$, where
$Z_t = Y_{t+1}$. For notational ease, we will drop the time index and refer to this vector as $W = (X, Y, Z)$. Under the null hypothesis, the joint probability density function $f_{X,Y,Z}(x, y, z)$ and its marginals must satisfy:

$$\frac{f_{X,Y,Z}(x,y,z)}{f_{X,Y}(x,y)} = \frac{f_{Y,Z}(y,z)}{f_Y(y)} \iff \frac{f_{X,Y,Z}(x,y,z)}{f_Y(y)} = \frac{f_{X,Y}(x,y)}{f_Y(y)}\frac{f_{Y,Z}(y,z)}{f_Y(y)}, \tag{8.10}$$

for every (x, y, z) in the support of (X, Y, Z). Correlation integrals are used to measure the discrepancies between the left and right sides of (8.10). Given $\epsilon > 0$, the correlation integral for a multivariate random vector is the probability of finding two independent realizations of the vector V: V_1 and V_2, at a distance less than or equal to ϵ, that is,

$$C_v(\epsilon) = P(\|V_1 - V_2\| \leq \epsilon).$$

This leads to the test case:

$$\frac{C_{X,Y,Z}(\epsilon)}{C_{X,Y}(\epsilon)} = \frac{C_{Y,Z}(\epsilon)}{C_Y(\epsilon)}. \tag{8.11}$$

It is then estimated via the sample analog:

$$C_{W,\eta}(\epsilon) = \frac{2}{n(n-1)} \sum \sum_{i<j} I(\|W_i - W_j\| \leq \epsilon). \tag{8.12}$$

However, as shown in Diks and Panchenko (2006), in certain situations this test tends to reject too often under the null hypothesis of no Granger causality. Diks and Panchenko (2006) show that the test statistic used in Hiemstra and Jones (1994) is biased and converges to a nonzero limit, while the variance decreases to zero, which will generate significant values for the test statistic as the length of the series increases. An important instance in which this occurs i between two series with independent ARCH effects.

Instead, Diks and Panchenko (2006) construct a new test statistic that measures dependence between X and Z given $Y = y_i$ locally for each y_i. Define the local density estimator as:

$$\hat{f}_W(W_i) = \frac{(2\epsilon)^{d_W}}{n-1} \sum_{j,j\neq i} I_{ij}^W,$$

where $I_{ij}^W = I(\|W_i - W_j\| < \epsilon)$. The test statistic can then be expressed as:

$$T_n(\epsilon_n) = \frac{(n-1)}{n(n-2)} \sum_i (\hat{f}_{X,Y,Z}(X_i, Y_i, Z_i)\hat{f}_Y(Y_i) - \hat{f}_{X,Y}(X_i, Y_i)\hat{f}_{X,Y}(X_i, Y_i)\hat{f}_{Y,Z}(Y_i, Z_i)).$$

The test is shown to be consistent for $d_x = d_y = d_z = 1$, and bandwidth chosen as:

$$\epsilon_n = Cn^{-\beta},$$

for any $C > 0$ and $\beta \in (\frac{1}{4}, \frac{1}{3})$. Details of derivation for the optimal choice of C are discussed in Diks and Panchenko (2006). The authors then show that under these conditions, the test is asymptotically normally distributed:

$$\sqrt{n}\frac{T_n(\epsilon_n) - q}{S_n} \xrightarrow{d} N(0, 1)$$

where $q = E(f_{X,Y,Z}(X, Y, Z)f_Y(Y) - f_{X,Y}(X, Y)f_{Y,Z}(Y, Z))$; and S_n is an autocorrelation robust estimate of the asymptotic variance. They repeat the empirical study of Hiemstra and Jones (1994) and find weaker evidence of nonlinear Granger causality between S&P 500 returns and volume.

8.4 Copulas

Suppose that we have two random variables X and Y, which we use to create the random vector $\mathbf{Z} = (X, Y)'$. Then the joint distribution of $\mathbf{Z}$ can be decomposed into two parts, one that only depends on the marginal distributions $F_X(x)$ and $F_Y(y)$, and another that is only associated with the dependence between X and Y. The latter is referred to as the copula. More formally, we have the following definition.

Definition 8.14 (Sklar's Theorem (Sklar, 1959)) *Let X and Y be continuous random variables with marginal distributions F_X and F_Y. Furthermore, let U and V be uniform random variables such that $U = F_X(X)$ and $V = F_Y(Y)$. We then define the copula C of (X, Y) as the joint distribution of (U, V).*

One way to model the dependence between two random variables is through the use of a copula. As can be seen from Definition 8.14, a copula is a multivariate distribution for which all marginal distributions are uniform. Thus, copulas are a natural tool for modeling dependence between two random variables since we can ignore the effects of their marginal distributions. For instance, we can express Kendall's tau and Spearman's rho coefficients in terms of the copula.

Theorem 8.1 *Let X and Y be continuous random variables with copula C. Then, if we let $\tau(X, Y)$ and $\rho_S(X, Y)$ represent their corresponding Kendall's tau and Spearman's rho coefficients, we have the following equations:*

$$\tau(X, Y) = 4 \int \int_{[0,1]^2} C(u, v)\, dC(u, v) - 1,$$

$$\rho_S(X, Y) = 12 \int \int_{[0,1]^2} uv\, dC(u, v) - 3 = 12 \int_0^1 \int_0^1 C(u, v)\, du\, dv - 3.$$

Proof. The proof for these equalities can be found in Embrechts *et al.* (2003).

Below are some useful notes regarding copulas.

- $C(0, v) = C(u, 0) = 0$.
- $C(1, v) = v$ and $C(u, 1) = u$.
- For all $0 \leq u_1 < u_2 \leq 1$ and $0 \leq v_1 < v_2 \leq 1$, we have that

$$C(u_2, v_2) - C(u_1, v_2) - C(u_2, v_1) + C(u_1, v_1) > 0,$$

  which implies that C is increasing in both variables.
- Copulas satisfy the following Lipschitz condition:

$$|C(u_2, v_2) - C(u_1, v_1)| \leq |u_2 - u_1| + |v_2 - v_1|.$$

8.4.1 Fitting Copula Models

There are numerous methods for fitting a copula model to data, including parametric, semiparametric, and nonparametric (Choroś *et al.*, 2010; Kim *et al.*, 2007). Suppose we wish to fit bivariate data $\{(x_i, y_i)\}_{i=1}^n$ to a copula C. In many cases, it is more natural to work on the density scale and with the corresponding copula density. The copula density c is obtained through differentiating the copula C, yielding the following relation between the joint density $f(x, y)$ and marginal densities f_X and f_Y:

$$f(x, y) = c[F_X(x), F_Y(y)] \cdot f_X(x) \cdot f_Y(y).$$

First, suppose that we believe the marginal distributions and copula to belong to known families that can be indexed by parameters θ_1, θ_2, and Ω respectively. Since we are selecting a model from a known family, we can select the best model through maximum likelihood. For this, we select $(\hat{\theta}_1, \hat{\theta}_2, \hat{\Omega})$ so as to maximize

$$L(\theta_1, \theta_2, \Omega) = \sum_{i=1}^{n} \log[c\,(F_X(x_i|\theta_1), F_Y(y_i|\theta_2)|\Omega)] + \sum_{i=1}^{n} \log(f_X(x_i|\theta_1)) + \sum_{i=1}^{n} \log(f_Y(y_i|\theta_2)).$$

The typical maximum likelihood approach would attempt to estimate θ_1, θ_2, and Ω all at once. However, the inference function for margins (IFM) method of Joe (1997) estimates the parameters in sequence. First, the parameters for the marginal distributions are estimated to obtain $\hat{\theta}_1$ and $\hat{\theta}_2$. Then, these estimates are used to estimate Ω, which maximizes

$$L(\Omega) = \sum_{i=1}^{n} \log\left[c\left(F_X(x_i|\hat{\theta}_1), F_Y(y_i|\hat{\theta}_2)|\Omega\right)\right].$$

The IFM method has a significant computational advantage to maximum likelihood estimation, and Joe (1997) shows that it is almost as efficient as maximum likelihood estimation in most cases.

If, however, the model is misspecified because of an incorrect marginal model choice, then the maximum likelihood estimates may lose their efficiency, and may in fact not be consistent (Kim *et al.*, 2007). For this reason, semiparametric approaches have also been suggested for

fitting copula models. In this setting, we usually have enough data to make accurate inferences about the marginal distributions, but not enough to easily model the intervariable dependence. Thus, we will assume that the copula belongs to a family that can be indexed by parameter Ω, while no specification about the marginals is made. The approach taken by Genest *et al.* (1995) fits the marginal distributions by using the empirical distribution function to get estimates $\hat{F}_X$ and $\hat{F}_Y$. Then, the fitted copula is selected so as to have parameter $\tilde{\Omega}$, which maximizes

$$L(\Omega) = \sum_{i=1}^{n} \log \left[c \left(\hat{F}_X(x_i), \hat{F}_Y(y_i) | \Omega \right) \right].$$

In Genest *et al.* (1995), it is shown that the estimate $\tilde{\Omega}$ is consistent and asymptotically normal. Furthermore, approaches to estimate the variance of the $\tilde{\Omega}$ estimator are also presented.

Finally, there are nonparametric approaches that can be used to fit copula models. Unlike the marginals, the copula is a hidden model, and thus selecting an appropriate parametric model is difficult. Thus, employing nonparametric methods is a natural choice. In this setting, many approaches fit the copula itself instead of the density because of increased rates of convergence. The natural starting point is to work with the empirical copula process:

$$\hat{C}_n(u, v) = \frac{1}{n} \sum_{i=1}^{n} 1 \left\{ x_i \leq \hat{F}_X^{-1}(u), y_i \leq \hat{F}_Y^{-1}(v) \right\}.$$

For the case where C is the independence copula, Deheuvels (1979) shows consistency and asymptotic normality for the empirical copula; whereas Fermanian *et al.* (2004) show consistency for a more general class of copulas, as well as consistency when using kernel functions to approximate the copula. However, when using kernel approaches, care must be taken when dealing with the boundary, so as to reduce the effect of boundary bias associated with kernels. This issue has been addressed through the use of locally linear kernels in Chen and Huang (2007).

8.4.2 Parametric Copulas

The simplest copula to work with is the product copula $C(u, v) = uv$. This corresponds to the random variables X and Y being independent, and thus is also commonly referred to as the independence copula. Other popular copulas include the Gaussian, t, and Archimedean copulas. The Gaussian copula and some Archimedean copulas include the independence copula as a special case. Additionally, the t copula contains the independence copula as a limiting case.

If the vector $(X, Y)'$ has a bivariate normal distribution, then all of the dependence between X and Y is captured by the correlation matrix Ω. Thus, the copulas for bivariate normal random vectors can be parameterized by their correlation matrix, $C(\cdot\ |\Omega)$. The copulas that are created in such a manner are called Gaussian copulas. If the random vector $(X, Y)'$ has a Gaussian copula, then it is said to have a meta-Gaussian distribution; this is because it is not required that the univariate marginal distributions be normal. For a given correlation matrix, Ω, the Gaussian copula's density, $c(u, v|\Omega)$, is given by

$$\frac{1}{\sqrt{\det(\Omega)}} \exp \left\{ -\frac{1}{2} \begin{pmatrix} \Phi^{-1}(u) \\ \Phi^{-1}(v) \end{pmatrix}' (\Omega^{-1} - I_2) \begin{pmatrix} \Phi^{-1}(u) \\ \Phi^{-1}(v) \end{pmatrix} \right\},$$

where Φ^{-1} is the standard normal quantile function and I_2 is the two-dimensional identity matrix. Fitting this copula to a given dataset can be accomplished by using any of the

techniques of Section 8.4.1. In this case, we can also directly estimate the correlation matrix by using Kendall's tau or Spearman's rho. Using Kendall's tau, we have the following estimate for the correlation matrix:

$$\Omega_{i,j} = \sin\left\{\frac{\pi}{2}\rho_\tau(Z_i, Z_j)\right\} \tag{8.13}$$

where $\mathbf{Z} = (Z_1, Z_2)' = (X, Y)'$. Once this calculation has been carried out, we can either directly use this estimate of Ω, or use the estimate as a starting value for maximum likelihood or pseudo-maximum likelihood estimation. A similar relation exists for Spearman's rho and can be found in Ruppert (2015).

In a similar fashion, we can define a copula that is based upon the bivariate t-distribution. This distribution is completely described by its correlation matrix Ω along with a shape parameter ν. The parameter ν affects both the marginal distributions as well as the dependence, and must thus be included in the copula parameterization, $C(\cdot|\nu, \Omega)$. A random vector that has a t-copula is said to have a t-meta distribution.

Regardless of which approach we choose to employ to fit the model, there is a relatively simple way to obtain an estimate of the correlation matrix, and it closely resembles that used for the Gaussian copula. Using Equation 8.13, we obtain a matrix that we refer to as $\Omega^{(1)}$. There is a chance that the matrix $\Omega^{(1)}$ is not positive definite because it has nonpositive eigenvalues. In this case, we have that $\Omega^{(1)} = O^T D^{(1)} O$, where $D^{(1)}$ is a diagonal matrix of eigenvalue and O is an orthogonal matrix of the corresponding eigenvectors. To obtain positive eigenvalues, we transform $D^{(1)}$ to $D^{(2)}$ by $D^{(2)}_{i,i} = \max(\epsilon, D^{(1)}_{i,i})$ for some small positive constant ϵ. Then $\Omega^{(2)} = O^T D^{(2)} O$ is a positive definite matrix; however, its diagonal entries may not be equal to one. To fix this, we multiply the ith row and column by $(\Omega^{(2)}_{i,i})^{-1/2}$. After performing these steps, we are left with a true correlation matrix Ω. We can then estimate the shape parameter by holding Ω constant and maximizing the following equation:

$$L(\Omega, \nu) = \sum_{i=1}^{n} \log\left[c\left(\hat{F_X}(x_i), \hat{F_Y}(y_i)|\Omega, \nu\right)\right].$$

The final class of copulas we will discuss are Archimedean copulas. This class of copula provides modeling characteristics that are not available with either the Gaussian or t-copulas.

Definition 8.15 *A copula C is said to be Archimedean if it has the following representation:*

$$C(u, v) = \phi^{-1}[\phi(u) + \phi(v)].$$

where $\phi : [0, 1] \mapsto [0, \infty]$ *is a function, called the generator, with the following properties;*

1. ϕ *is a decreasing convex function.*
2. $\phi(0) = \infty$ *and* $\phi(1) = 0$.

Depending on the choice of generator, we are able to model a wide array of dependencies. For instance, the Frank copula can be obtained by selecting

$$\phi(t|\theta) = -\log\left\{\frac{e^{-\theta t} - 1}{e^{-\theta} - 1}\right\}$$

Table 8.2 A table of common Archimedean copulas

Copula	$\phi(t\|\theta)$	$C(u,v)$	Range
Clayton	$\frac{1}{\theta}(t^{-\theta}-1)$	$\max(\{u^{-\theta}+v^{-\theta}\}^{-1/\theta},0)$	$\theta\in[-1,\infty)\backslash\{0\}$
Ali–Mikhail–Haq	$\log\left(\frac{1-\theta(1-t)}{t}\right)$	$\frac{uv}{1-\theta(1-u)(1-v)}$	$\theta\in[-1,1)$
Joe	$-\log(1-(1-t)^{\theta})$	$1-\left((1-u)^{\theta}+(1-v)^{\theta}-(1-u)^{\theta}(1-v)^{\theta}\right)^{1/\theta}$	$\theta\in[1,\infty)$

with $\theta\in\mathbb{R}$. The choice of generator gives the following copula:

$$C(u,v|\theta)=-\frac{1}{\theta}\log\left\{1+\frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1}\right\}.$$

By noting that $\lim_{\theta\to 0}\phi(t|\theta)=-\log(t)$, we have that as θ approaches 0, the Frank copula converges to the independence copula; whereas as $\theta\to-\infty$, we obtain the copula for the random vector $(U,1-U)'$, $C(u,v)=\max(0,u+v-1)$, which is the copula for perfect negative dependence. Similarly, as $\theta\to\infty$, we obtain the copula for the random vector $(U,U)'$, which has perfect positive dependence, $C(u,v)=\min(u,v)$.

Now let us consider the generator $\phi(t|\theta)=[-\log(t)]^{\theta}$ for $\theta\geq 1$. This gives the Gumbel copula

$$C(u,v|\theta)=\exp\ \{-([\log(u)]^{\theta}+[\log(v)]^{\theta})^{1/\theta}\}.$$

Selecting $\theta=1$ results in the independence copula. Also, note that as $\theta\to\infty$, the Gumbel copula converges to the copula for perfect positive dependence. However, unlike the Frank copula, the Gumbel copula is unable to model negative dependence.

Table 8.2 contains some additional examples of Archimedean copulas along with their generators.

8.4.3 Extending beyond Two Random Variables

So far, we have only seen how to model the dependence between two random variables. Naturally, we would also like to model dependence between $m\geq 3$ random variables. In this case, our copula would be a function $C:[0,1]^m\mapsto[0,1]$. However, some difficulties arise when trying to extend the tools for creating bivariate copulas to multivariate settings. For instance, Genest *et al.* (1995) show that if C is a bivariate copula and $F_m(x_1,\ldots,x_m)$, $G_n(y_1,\ldots,y_n)$ are multivariate marginal distributions, then the only copula that satisfies $H(x_1,\ldots,x_m,y_1,\ldots,y_n)=C(F_m(x),G_n(y))$ is the independence copula.

Because of these difficulties, we will only be considering ways to extend Archimedean copulas, as these extensions still remain relatively easy to analyze. More specifically, we will consider the exchangeable Archimedean copula (EAC), nested Archimedean copula (NAC), and pair copula construction (PCC).

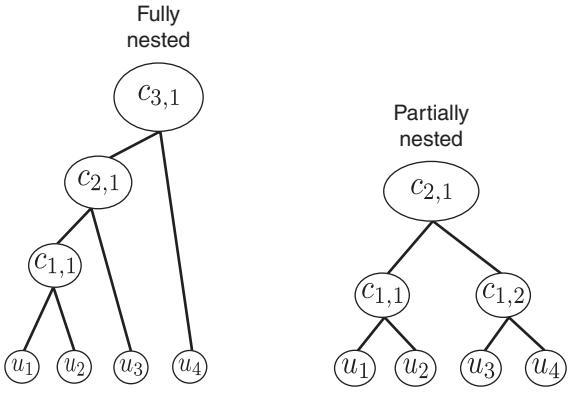

Figure 8.4 Tree representation of the fully nested (left) and partially nested (right) Archimedean copula construction. Leaf nodes represent uniform random variables, while the internal and root nodes represent copulas. Edges indicate which variables or copulas are used in the creation of a new copula.

To start, we consider the simplest extension, EACs. As with the bivariate Archimedean copula, these are tied to a generating function ϕ. In this case, we have that

$$C(u_1, \ldots, u_m) = \phi^{-1}\{\phi(u_1) + \cdots + \phi(u_m)\}.$$

This extension, however, comes at the cost of a greater restriction on the generator ϕ. This restriction is d-monotonic; see McNeil and Nešlehová (2009) for more information.

NACs provide more flexibility than EACs because they allow for the combination of different types of bivariate copulas to obtain higher order copulas. The bivariate copulas are combined hierarchically and thus allow for a convenient graphical representation. Figure 8.4 shows a graphical representation for fully and partially nested Archimedean copulas in the case of four variables. As can be seen from Figure 8.4, NACs only allow one to model $m - 1$ copulas. Thus, all other possible interactions are predefined by the hierarchical structure. For instance, in the partially nested structure, both (u_1, u_3) and (u_1, u_4) are modeled by $c_{2,1}(u_1, u_2, u_3, u_4)$. Fitting a general NAC can be done with maximum likelihood, but in general this must be done recursively and becomes extremely computationally intensive as the number of variables increases. Similarly, sampling from such copulas is also a difficult task (Aas and Berg, 2009), but in the case of fully and partially nested models Hofert (2010) presents algorithms that are efficient for sampling any number of variables.

We finally consider PCCs, which like NACs are created by a hierarchical process. The most popular types of PCCs are canonical vines and drawable vines, commonly referred to as C-vines and D-vines, respectively. For both C- and D-vines, there exist analytic expressions for the densities; see Czado (2010). Thus, both models can be fit by maximum likelihood; however, as with NACs, this must be done in a recursive fashion and can thus be computationally expensive. Unlike NACs, drawing samples from PCCs is much simpler (Aas and Berg, 2009). Furthermore, using PCCs, we are able to model all $m(m-1)/2$ copulas, which is not possible when using NAC extensions. Figure 8.5 shows the graphical representations of C- and D-vines on four variables.

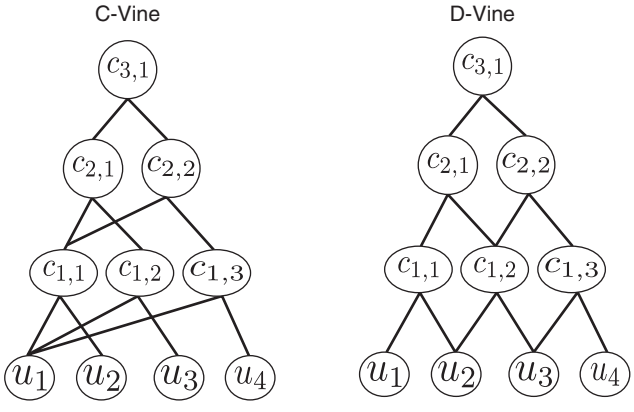

Figure 8.5 Graphical representation of the C-vine (left) and D-vine (right) Archimedean copula construction. Leaf nodes labeled u_i represent uniform random variables, whereas nodes labeled $c_{i,j}$ represent the jth copula at the ith level. Edges indicate which variables or copulas are used in the creation of a new copula.

8.4.4 Software

There are a number of software packages to choose from when working with copulas. Here, we mention a few for working with copulas in R. All of these packages are designed to perform statistical inference of various copula models. They also all have methods for performing model estimation and random sampling. The **copula** package (Hofert *et al.*, 2014) works with elliptical, Archimedean, and a few other copula families. **VineCopula** (Schepsmeier *et al.*, 2012) is designed to perform analysis of vine copulas, while **CDVine** (Brechmann and Schepsmeier, 2013) is specialized for C- and D-vines.

8.5 Types of Dependence

In Section 8.4, we showed that copulas can be used to model the dependence between random variables. We now turn our attention to the dependence structure itself, and when appropriate make connections to copulas. We first describe different types of dependence, and then provide theoretical background.

When dealing with only two random variables, the concept of positive and negative dependence is more straightforward. Suppose that we have random variables X and Y, then positive dependence means that an increase/decrease in the value of X is likely to accompany an increase/decrease in the value of Y. However, if X and Y are negatively dependent, then an increase/decrease in X is likely to be accompanied by a decrease/increase in Y.

8.5.1 Positive and Negative Dependence

We now examine ways to extend the notion of positive and negative dependence to more than two random variables. In this section, we will use the following notation; if $\mathrm{x}, y \in \mathbb{R}^d$, we say that $\mathbf{x} > \mathbf{y}$ if $x_i > y_i$ for all $i = 1, 2, \ldots, d$.

In the bivariate case, the random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$ is positively dependent if all of its components tend to move in the same direction. However, when $d > 2$, there are many different interpretations.

Definition 8.16 *A random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$ is positive upper orthant dependent if for all $\mathbf{x} \in \mathbb{R}^d$*

$$P(\boldsymbol{X} > \mathbf{x}) \geq \prod_{i=1}^{d} P(X_i > x_i). \tag{8.14}$$

Similarly, if

$$P(\boldsymbol{X} \leq \mathbf{x}) \geq \prod_{i=1}^{d} P(X_i \leq x_i), \tag{8.15}$$

we say that $\boldsymbol{X}$ is positive lower orthant dependent.

If we generalize the concept of an orthant, we obtain another form of positive dependence.

Definition 8.17 *A set $\mathcal{U}$ is called an upper set if $\mathbf{x} \in \mathcal{U}$ and $\mathbf{y} > \mathbf{x}$ implies that $\mathbf{y} \in \mathcal{U}$. The complement of an upper set is called a lower set.*

It should be noted that since the lower orthant is not the complement of the upper orthant, it is not a lower set. Then, the concept of positive upper/lower set dependence can be easily expressed using inequalities similar to those in Equations 8.14 and 8.15.

Definition 8.18 *A random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$ is said to be positive upper set dependent, or positive lower set dependent, if*

$$P(\boldsymbol{X} \in \bigcap \mathcal{U}_k) \geq \prod_k P(\boldsymbol{X} \in \mathcal{U}_k)$$

and

$$P(\boldsymbol{X} \in \bigcap \mathcal{L}_k) \geq \prod_k P(\boldsymbol{X} \in \mathcal{L}_k)$$

respectively, where $\mathcal{U}_k$ are upper sets and $\mathcal{L}_k$ are lower sets.

Unlike positive dependence, the concept of negative dependence is not as easily extended to multiple random variables. For instance, it is impossible for more than two random variables to have perfect negative dependence. To see this, suppose that X_1 has perfect negative dependence with X_2 and X_3; this would imply that X_2 and X_3 are positively dependent.

One way to obtain a type of negative dependence is by reversing the inequalities in Equations 8.14 and 8.15. This results in what are called negative upper orthant and negative lower orthant dependence, respectively. However, conditions under which these inequalities hold are difficult to show when $d > 2$. One such condition, based upon the behavior of the multinomial distribution, is given by Block *et al.* (1982).

The versions of positive and negative dependence for multiple random variables presented here are elementary. There are notions of positive/negative association that are related to the covariance structure, as well as other less intuitive concepts such as setwise dependence. Further details about these and other types of dependence can be found in the book by Drouet Mari and Kotz (2001).

8.5.2 Tail Dependence

One type of dependence that we will discuss in more detail has major applications in finance. Tail dependence describes how likely it is that a set of random variables will simultaneously take on extreme values. For instance, when working with a portfolio, tail dependence provides information on how likely it is to observe large, simultaneous losses from a portfolio's assets.

Suppose that we have random variables X and Y with distributions F_X and F_Y, respectively, and let $p \in (0, 1)$. In the most basic case, upper tail dependence is concerned with the following quantity:

$$u(p) = P(X > F_X^{-1}(p) | Y > F_Y^{-1}(p)) = \frac{1 - 2p + C(p, p)}{1 - p} \tag{8.16}$$

where $C(\cdot, \cdot)$ is the copula for X and Y. Similarly, lower tail dependence is concerned with the quantity

$$l(p) = P\left(X \leq F_X^{-1}(p) | Y \leq F_Y^{-1}(p)\right) = \frac{C(p, p)}{p}. \tag{8.17}$$

Let us assume that the variables X and Y measure the amount lost from the components of a two-asset portfolio. We then say that X and Y are asymptotically independent if $u^* = \lim_{p \to 1} u(p)$ exists and is equal to zero. Otherwise, we say that X and Y are asymptotically dependent. Poon *et al.* (2004) provide a nonparametric method for estimating the value of u^*, as well as a way to measure upper tail dependence even if X and Y are asymptotically independent. They also demonstrate the impact of knowing whether variables are asymptotically dependent or independent on the conclusions drawn from statistical analyses.

We now turn our attention to the type of dependence that is under consideration in Equation 8.17. As with Equation 8.16, we will say that X and Y are asymptotically independent if $l^* = \lim_{p \to 0} l(p)$ exists and is equal to zero, and asymptotically dependent if the limit exists and is nonzero. As mentioned by Schmid and Schmidt (2007), one issue with this method is that it only examines values along the diagonal of the copula. Thus, the quantity

$$\frac{C(p, p^2)}{p^2}$$

might have a different limit as $p \to 0$ than l^*. For this reason, Schmid and Schmidt (2007) propose the following measure of tail dependence, which is closely related to Spearman's rho:

$$\rho_L = \lim_{p \to 0} \frac{3}{p^3} \int_0^p \int_0^p C(u, v)\, du\, dv. \tag{8.18}$$

The measures discussed so far have only been for bivariate processes; however, in financial cases, multivariate approaches are typically more applicable. Extensions to higher dimensions are rather simple for Equations 8.16 and 8.17, requiring only a change in the set of variables over which we condition and those for which we wish to find the probability of observing an extreme event. Schmid and Schmidt (2007) provide a general multivariate version of Equation 8.18.

Financial applications are typically multivariate and consider a portfolio of assets rather than individual assets. Such portfolios may exhibit complex dependencies and evolutions. The methods described in this chapter allow robust analysis of non-Gaussian and time-series data, and provide intuitive and tractable tools for multivariate data. They also make relatively weak

assumptions about the underlying dependence structure to maintain broad applicability for financial data analysis.

References

Aas, K. and Berg, D. (2009). Models for construction of multivariate dependence: a comparison study. *The European Journal of Finance*, 15(7–8), 639–659.

Anderson, T.W. (1954). *An introduction to multivariate statistical analysis*. New York: John Wiley & Sons.

Ashley, R.A. and Patterson, D.M. (2001). Nonlinear model specification/diagnostics: insights from a battery of nonlinearity tests. Economics Department Working Paper E99-05. Blacksburg, VA: Virginia Tech.

Baek, E. and Brock, W. (1992). A general test for nonlinear granger causality: Bivariate model. Working Paper. Ames and Madison: Iowa State University and University of Wisconsin at Madison.

Baillie, R.T. and Bollerslev, T. (1990). A multivariate generalized arch approach to modeling risk premia in forward foreign exchange rate markets. *Journal of International Money and Finance*, 9(3), 309–324.

Bauwens, L., Laurent, S. and Rombouts, J.V. (2006). Multivariate GARCH models: a survey. *Journal of Applied Econometrics*, 21(1), 79–109.

Bingham, N.H. and Schmidt R (2006). Interplay between distributional and temporal dependence: an empirical study with high-frequency asset returns. In *From stochastic calculus to mathematical finance*. Berlin: Springer, pp. 69–90.

Block, H.W., Savits, T.H., Shaked, M., *et al.* (1982). Some concepts of negative dependence. *The Annals of Probability*, 10(3), 765–772.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3), 307–327.

Bollerslev, T., Chou, R.Y. and Kroner, K.F. (1992). Arch modeling in finance. *Journal of Econometrics*, 52, 5–59.

Box, G.E. and Pierce, D.A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association*, 65(332), 1509–1526.

Brechmann, E.C. and Schepsmeier, U. (2013). Modeling dependence with c- and d-vine copulas: The R package CDVine. *Journal of Statistical Software*, 52(3), 1–27.

Brock, W., Scheinkman, J.A., Dechert, W.D. and LeBaron, B. (1996). A test for independence based on the correlation dimension. *Econometric Reviews*, 15(3), 197–235.

Caporale, G.M., Ntantamis, C., Pantelidis, T. and Pittis, N. (2005). The BDS test as a test for the adequacy of a GARCH (1, 1) specification: a Monte Carlo study. *Journal of Financial Econometrics*, 3(2), 282–309.

Chen, S.X. and Huang, T.M. (2007). Nonparametric estimation of copula functions for dependence modelling. *Canadian Journal of Statistics*, 35(2), 265–282.

Chorós, B., Ibragimov, R. and Permiakova, E. (2010). Copula estimation. In *Copula theory and its applications* (ed. Jaworski P, Durante F, Hrdle WK and Rychlik T). Lecture Notes in Statistics. Berlin: Springer pp. 77–91.

Czado, C. (2010). Pair-copula constructions of multivariate copulas. In *Copula theory and its applications* (ed. Jaworski P, Durante F, Hrdle WK and Rychlik T). Lecture Notes in Statistics. Berlin: Springer pp. 93–109.

Deheuvels, P. (1979). La fonction de dépendance empirique et ses propriétés: un test non paramétrique d'indépendance. *Academie Royale de Belgique, Bulletin de la Classe des Sciences*, 65(6), 274–292.

Diks, C. (2009). Nonparametric tests for independence. In *Encyclopedia of complexity and systems science*. New York: Springer, pp. 6252–6271.

Diks, C. and Panchenko, V. (2006). A new statistic and practical guidelines for nonparametric Granger causality testing. *Journal of Economic Dynamics and Control*, 30(9), 1647–1669.

Drouet Mari, D. and Kotz, S. (2001). *Correlation and dependence*, vol. 518. Singapore: World Scientific.

Embrechts, P., Lindskog, F. and McNeil, A. (2003). Modelling dependence with copulas and applications to risk management. *Handbook of Heavy Tailed Distributions in Finance*, 8(1), 329–384.

Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50, 987–1007.

Engle, R.F. and Kroner, K.F. (1995). Multivariate simultaneous generalized arch. *Econometric Theory*, 11(01), 122–150.

Fama, E.F. (1965). The behavior of stock-market prices. *Journal of Business*, 38, 34–105.

Ferguson, T.S., Genest, C. and Hallin, M. (2000). Kendall's tau for serial dependence. *Canadian Journal of Statistics*, 28(3), 587–604.

Fermanian, J.D., Radulovic, D., Wegkamp, M., *et al.* (2004). Weak convergence of empirical copula processes. *Bernoulli*, 10(5), 847–860.

Gallant, A.R., Rossi, P.E. and Tauchen, G. (1992). Stock prices and volume. *Review of Financial Studies*, 5(2), 199–242.

Genest, C., Ghoudi, K. and Rémillard, B. (2007). Rank-based extensions of the Brock, Dechert, and Scheinkman test. *Journal of the American Statistical Association*, 102, 1363–1376.

Genest, C., Quesada Molina, J. and Rodríguez Lallena, J. (1995). De l'impossibilité de construire des lois à marges multidimensionnelles données à partir de copules. *Comptes rendus de l'Académie des sciences, Série 1, Mathématique*, 320(6), 723–726.

Geweke, J., Meese, R. and Dent, W. (1983). Comparing alternative tests of causality in temporal systems: analytic results and experimental evidence. *Journal of Econometrics*, 21(2), 161–194.

Granger, C.W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37, 424–438.

Hamilton, J.D. (1994). *Time series analysis*, vol. 2. Princeton, NJ: Princeton University Press.

Helsel, D.R. and Hirsch, R.M. (1992). *Statistical methods in water resources*, vol. 49. Amsterdam: Elsevier.

Hiemstra, C. and Jones, J.D. (1994). Testing for linear and nonlinear granger causality in the stock price-volume relation. *The Journal of Finance*, 49(5), 1639–1664.

Hofert, M. (2010). Construction and sampling of nested Archimedean copulas. In *Copula theory and its applications* (ed. Jaworski P, Durante F, Hrdle WK and Rychlik T). Lecture Notes in Statistics. Berlin: Springer pp. 147–160.

Hofert, M., Kojadinovic, I., Maechler, M. and Yan, J. (2014). *Copula: multivariate dependence with copulas*. R package version 0.999-8.

Hollander, M. and Wolfe, D.A. (1999). *Nonparametric statistical methods*. New York: John Wiley & Sons.

Joe, H. (1997). *Multivariate models and multivariate dependence concepts*, vol. 73. Boca Raton, FL: CRC Press.

Jondeau, E., Poon, S.H. and Rockinger, M. (2007). *Financial modeling under non-Gaussian distributions*. Berlin: Springer.

Kendall, M.G. (1938). A new measure of rank correlation. *Biometrika*, 30(1–2): 81–93.

Kim, G., Silvapulle, M.J. and Silvapulle, P. (2007). Comparison of semiparametric and parametric methods for estimating copulas. *Computational Statistics & Data Analysis*, 51(6), 2836–2850.

Lai, T.L. and Xing, H. (2008). *Statistical models and methods for financial markets*. New York: Springer.

Ljung, G.M. and Box, G.E. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2), 297–303.

Lütkepohl, H. (2007). *New introduction to multiple time series analysis*. New York: Springer.

Luukkonen, R., Saikkonen, P. and Teräsvirta, T. (1988). Testing linearity against smooth transition autoregressive models. *Biometrika*, 75(3), 491–499.

Maronna, R.A., Martin, D.R. and Yohai, V.J. (2006). *Robust statistics: theory and methods*. Hoboken, NJ: John Wiley & Sons.

McLeod, A.I. and Jimenéz, C. (1984). Nonnegative definiteness of the sample autocovariance function. *The American Statistician*, 38(4), 297–298.

McLeod, A.I. and Li, W.K. (1983). Diagnostic checking arma time series models using squared-residual autocorrelations. *Journal of Time Series Analysis*, 4(4), 269–273.

McNeil, A.J. and Nešlehovà, J. (2009). Multivariate Archimedean copulas, d-monotone functions and 1-norm symmetric distributions. *The Annals of Statistics*, 37, 3059–3097.

Patterson, D.M. and Ashley, R.A. (2000). *A nonlinear time series workshop: a toolkit for detecting and identifying nonlinear serial dependence*, vol. 2. New York: Springer.

Poon, S.H., Rockinger, M. and Tawn, J. (2004). Extreme value dependence in financial markets: diagnostics, models, and financial implications. *Review of Financial Studies*, 17(2), 581–610.

Ruppert, D., and Matteson, D.S. (2015). *Statistics and data analysis for financial engineering*.

Schepsmeier, U., Stoeber, J., Brechmann, E. and Gräler, B. (2012). *Vinecopula: statistical inference of vine copulas*. R package version.

Schmid, F. and Schmidt, R. (2007). Multivariate conditional versions of Spearman's rho and related measures of tail dependence. *Journal of Multivariate Analysis*, 98(6), 1123–1140.

Shumway, R.H., and Stoffer, D.S. (2011). *Time Series Analysis and Its Applications: With R Examples*. Springer Texts in Statistics.

Ŝiman, M. (2012). On Kendall's autocorrelations. *Communications in Statistics: Theory and Methods*, 41(10), 1733–1738.

Sklar, M. (1959). Fonctions de Répartition à n Dimensions et Leurs Marges. *Publications de l'Institut de Statistique de L'Université de Paris*, 8, 229–231.

Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, 15(1), 72–101.

Tsay, R. (2014). *Multivariate time series analysis with R and financial applications*. Berlin: Springer.

Tsay, R.S. (1986). Nonlinearity tests for time series. *Biometrika*, 73(2), 461–466.

Tsay, R.S. (2010). *Analysis of Financial Time Series*. John Wiley & Sons.

Xu, W., Hou, Y., Hung, Y. and Zou, Y. (2013). A comparative analysis of Spearman's rho and Kendall's tau in normal and contaminated normal models. *Signal Processing*, 93(1), 261–276.

Zivot, E. and Wang, J. (2007). *Modeling financial time series with S-PLUSR*, vol. 191. New York: Springer.