6

Modeling Correlation

In this chapter, we turn to the modeling of the multivariate distribution of asset returns. The most important characteristic of the multivariate distribution relies on the dependency parameter, i.e., the parameter that measures the strength of the link between two series. For a number of standard distributions (namely, those belonging to the elliptical family, which includes the normal and the Student t distributions), the dependency is simply measured by the Pearson's correlation coefficient. We often refer to correlation when meaning dependency. In practice, however, asset returns do not belong to this elliptical family and the dependency must be modeled with specific techniques. The modeling of the dependency parameter between asset returns has very important consequences in a number of financial applications.

An abundant literature has addressed the issue of how correlation between stock market returns varies when markets become agitated. The time-variability of correlation between returns is crucial from an asset management point of view. Asset allocation is often based on the use of a correlation matrix computed over a given sample period. If correlation increases during turbulent periods, the benefits of diversification would disappear when they are the most needed, i.e., during crashes. Implications for risk management are also obvious because the time-variability of the dependency parameter complicates Value at Risk (VaR) computation significantly.

As described in Chapter 2, the first tests of constancy of the dependency parameter were based on testing the equality of linear correlation coefficients computed before and after a crash. This approach has been found to be misleading, however, because conditioning the estimation of the correlation coefficient on the sample period induces an estimator bias. Subsequently, most tests of the constant correlation hypothesis have been based on the following approach: First, estimate the joint dynamics of stock-market returns and then describe how conditional correlations vary over time. There are several ways to model the joint dynamics of a number of series. The most widely used approach is the multivariate GARCH model.[1] The central idea of the first generation of such models was that covariances had to be modeled using the same type of specification of variances in the univariate GARCH model (Kraft and Engle, 1982, and Bollerslev, Chou, and Kroner, 1992). These models, however, raise a problem of dimensionality because the number of parameters to be estimated increases dramatically with the number of series. Afterwards, most extensions have tried to reduce the computational burden.

Another difficulty of multivariate GARCH models relies on the choice of the conditional distribution of returns. When the marginal distributions are Gaussian, the extension to the multivariate case is trivial, because the joint distribution is simply a multivariate Gaussian distribution. However, for more general distributions, the multivariate extension is far from trivial. Often, it simply does not exist. In fact, an explicit multivariate extension exists only in very few cases. This is the case in particular for the Student t distribution. In other cases, a solution consists in constructing an implicit multivariate distribution by using copula functions. Due to the non-normality found in most financial return series, copula functions have had a great success, because they relate in an easy way very complicated marginal distributions. However, there is no free lunch. When copula functions are used, computing moments or, more generally, dealing with the integration of the joint distribution, becomes analytically intractable. Therefore, only numerical algorithms can be implemented. Unfortunately, in many financial applications, even with a few number of assets, such a numerical integration is too demanding to be performed in a reasonable length of time.

In this chapter, we describe how the dependency parameter may be modeled in the context of a multivariate GARCH model (Section 6.1). We present tests of constancy of the conditional dependency parameter. Notice that we will not provide a full description of the multivariate GARCH models. Rather, this section should be viewed as a brief introduction to this methodology, before applying multivariate models with non-normal distributions.[2] Then, we consider two aspects of the multivariate extension of non-Gaussian distributions. In Section 6.2, we consider the use of explicit multivariate distributions,

---

[1] An alternative approach is the multivariate Markov-switching model. For instance, Ramchand and Susmel (1998) and Ang and Bekaert (2002) test within this framework the hypothesis of a constant international conditional correlation between stock markets. Some papers also consider how correlation varies when stock-market indices are simultaneously affected by very large (positive or negative) fluctuations. Engle and Manganelli (2004) focus on the modeling of large realizations using quantile regressions. Longin and Solnik (2001), using extreme value theory, find that dependency increases more during downside movements than during upside movements. Poon, Rockinger, and Tawn (2004) provide an alternative statistical framework to test conditional dependency between extreme returns.

[2] In addition, it is worth emphasizing that the modeling of dependency in the context of extreme events will be detailed in Chapter 7.

which are generally difficult to estimate but allow a more efficient computation of moments. Then, in Section 6.3, we present the approach based on copula, whose estimation is much easier, although applications are more restricted.

6.1 Multivariate GARCH models

We now consider a random vector $x_t = (x_{1,t}, \cdots, x_{n,t})'$ whose joint dynamics is given by

$$x_t = \mu_t(\theta) + \varepsilon_t, \tag{6.1}$$

$$\varepsilon_t = \Sigma_t^{1/2}(\theta)\, z_t, \tag{6.2}$$

where $\mu_t(\theta)$ denotes the $(n, 1)$ vector of conditional means, $\Sigma_t(\theta)$ denotes the (n, n) conditional covariance matrix of the error term ε_t, and θ is the vector of unknown parameters. The standardized innovation vector z_t is *iid* with mean $E[z_t] = 0$ and covariance matrix $V[z_t] = I_n$. $\Sigma_t^{1/2}(\theta)$ denotes the Cholesky decomposition of $\Sigma_t(\theta)$. In this chapter, we assume that z_t is drawn from the multivariate normal $\mathcal{N}(0, I_n)$ distribution.

Several parameterizations have been proposed for Σ_t. The main issue to be addressed is the dimensionality of the parameter vector when the number of variables n increases. Obviously, it is desirable that most statistical features highlighted in the univariate context be incorporated in the multivariate framework, in particular in terms of asymmetry and tail behavior. Some additional specific issues related to the multivariate framework have to be addressed. First, we have to deal with the conditions guaranteeing that the covariance matrix is positive definite at each date t. A second issue is whether conditional correlations have to be modeled instead of conditional covariance, and whether they have to be time-varying.

It should be noticed that we do not discuss the positivity and stationarity conditions, which have been widely studied in the literature. Very complete and comprehensive surveys of multivariate GARCH models may be found in Bollerslev, Engle, and Nelson (1994) and Bauwens, Laurent, and Rombouts (2005).

For further use, we define D_t the (n, n) diagonal matrix with the conditional variances σ_i^2 along the diagonal, so that $\{D_t\}_{ii} = \{\Sigma_t\}_{ii}$ and $\{D_t\}_{ij} = 0$, $\forall i \neq j$, for $i, j = 1, \cdots, n$. We also define R_t, the (n, n) matrix of conditional correlations of ε_t, as $R_t = D_t^{-1/2} \Sigma_t D_t^{-1/2} = \{\rho_t\}_{ij}$. We deduce the $(n, 1)$ vector of normalized innovations $u_t = D_t^{-1/2} \varepsilon_t$. Notice that u_t differs from standardized innovations $z_t = \Sigma_t^{-1/2} \varepsilon_t$, because they are not orthogonalized.

6.1.1 Vectorial and diagonal GARCH models

Vech model

The first multivariate GARCH model, proposed by Kraft and Engle (1982) and Bollerslev, Chou, and Kroner (1992), assumes that each element of the covariance matrix is a linear function of the most recent past cross-products of errors and conditional variances and covariances. The Vech GARCH(p, q) model is defined as

$$\operatorname{vech}(\varSigma_t) = \operatorname{vech}(\varOmega) + \sum_{i=1}^{p} A_i \operatorname{vech}\left(\varepsilon_{t-i}\varepsilon'_{t-i}\right) + \sum_{j=1}^{q} B_j \operatorname{vech}(\varSigma_{t-j}), \tag{6.3}$$

where $\varOmega$ is an (n, n) positive definite and symmetric matrix, A_i and B_j are $(n(n+1)/2, n(n+1)/2)$ matrices; vech$(\cdot)$ is the operator that stacks the lower triangular elements of an (n, n) matrix as an $(n(n+1)/2, 1)$ vector. The number of parameters is $[n(n+1)/2][1 + (p+q)n(n+1)/2]$. Although this specification is very flexible, the large number of parameters (proportional to n^4) renders this model very difficult to handle. In addition, conditions that ensure that the conditional covariance matrices are positive definite are difficult to verify and impose.

Example: In the case $p = q = 1$ and $n = 2$, this model reduces to

$$\begin{pmatrix} \sigma^2_{1,t} \\ \sigma_{12,t} \\ \sigma^2_{2,t} \end{pmatrix} = \begin{pmatrix} \omega_{11} \\ \omega_{12} \\ \omega_{22} \end{pmatrix} + \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} \varepsilon^2_{1,t-1} \\ \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon^2_{2,t-1} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix} \begin{pmatrix} \sigma^2_{1,t-1} \\ \sigma_{12,t-1} \\ \sigma^2_{2,t-1} \end{pmatrix},$$

with 21 unknown parameters.

Diagonal vec model

In order to reduce the number of unknown parameters, Bollerslev, Engle, and Wooldridge (1988) have proposed the diagonal vec model, in which the matrices A_i^* and B_j^* are all taken to be diagonal (n, n) matrices: each element of the covariance matrix $(\sigma_{ij,t})$ only depends on the corresponding past elements $\sigma_{ij,t-1}$ and $\varepsilon_{i,t-1}\varepsilon_{j,t-1}$

$$\varSigma_t = \varOmega^* + \sum_{i=1}^{p} A_i^* \odot \left(\varepsilon_{t-i}\varepsilon'_{t-i}\right) + \sum_{j=1}^{q} B_j^* \odot \varSigma_{t-j},$$

where $\varOmega^*$ is an (n, n) positive definite and symmetric matrix, A_i^* and B_j^* are (n, n) symmetric matrices and $\odot$ denotes the Hadamard product.[3] This model

[3] The Hadamard product defines the element-wise product of two matrices. So we have $\{A \odot B\}_{ij} = A_{ij}B_{ij}$.

has a natural interpretation, because covariances as well as variances have a GARCH-type specification. In addition, it reduces the number of unknown parameters considerably to $[n(n+1)/2](p+q+1)$, so that it is proportional to n^2 only.

Example: In the case $p = q = 1$ and $n = 2$, this model reduces to

$$\begin{aligned}\Sigma_t &= \begin{pmatrix} \omega_{11}^* & \omega_{12}^* \\ \omega_{12}^* & \omega_{22}^* \end{pmatrix} \\ &+ \begin{pmatrix} a_{11}^* & a_{12}^* \\ a_{12}^* & a_{22}^* \end{pmatrix} \odot \begin{pmatrix} \varepsilon_{1,t-1}^2 & \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon_{1,t-1}\varepsilon_{2,t-1} & \varepsilon_{2,t-1}^2 \end{pmatrix} \\ &+ \begin{pmatrix} b_{11}^* & b_{12}^* \\ b_{12}^* & b_{22}^* \end{pmatrix} \odot \begin{pmatrix} \sigma_{1,t-1}^2 & \sigma_{12,t-1} \\ \sigma_{12,t-1} & \sigma_{2,t-1}^2 \end{pmatrix}, \\ \Sigma_t &= \begin{pmatrix} \omega_{11}^* & \omega_{12}^* \\ \omega_{12}^* & \omega_{22}^* \end{pmatrix} + \begin{pmatrix} a_{11}^*\varepsilon_{1,t-1}^2 & a_{12}^*\varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ a_{12}^*\varepsilon_{1,t-1}\varepsilon_{2,t-1} & a_{22}^*\varepsilon_{2,t-1}^2 \end{pmatrix} \\ &+ \begin{pmatrix} b_{11}^*\sigma_{1,t-1}^2 & b_{12}^*\sigma_{12,t-1} \\ b_{12}^*\sigma_{12,t-1} & b_{22}^*\sigma_{2,t-1}^2 \end{pmatrix},\end{aligned}$$

where $\Sigma_t = \begin{pmatrix} \sigma_{1,t}^2 & \sigma_{12,t} \\ \sigma_{12,t} & \sigma_{2,t}^2 \end{pmatrix}$. In such a case, there are 9 unknown parameters.

BEKK model

An alternative representation is the BEKK representation described by Engle and Kroner (1995)[4]

$$\Sigma_t = \tilde{\Omega} + \sum_{i=1}^{p} \tilde{A}_i' \varepsilon_{t-i}\varepsilon_{t-i}' \tilde{A}_i + \sum_{j=1}^{q} \tilde{B}_j' \Sigma_{t-j} \tilde{B}_j,$$

where $\tilde{\Omega}$ is an (n,n) positive definite and symmetric matrix, and $\tilde{A}_i$ and $\tilde{B}_j$ are (n,n) matrices. This specification involves $[n(n+1)/2] + (p+q)n^2$ unknown parameters. The main advantage of this specification is that the conditional covariance matrix is positive definite as long as $\tilde{\Omega}$ also is.

Example: In the case $p = q = 1$ and $n = 2$, this model reduces to

$$\begin{aligned}\Sigma_t &= \begin{pmatrix} \tilde{\omega}_{11} & \tilde{\omega}_{12} \\ \tilde{\omega}_{12} & \tilde{\omega}_{22} \end{pmatrix} \\ &+ \begin{pmatrix} \tilde{a}_{11} & \tilde{a}_{12} \\ \tilde{a}_{21} & \tilde{a}_{22} \end{pmatrix}' \begin{pmatrix} \varepsilon_{1,t-1}^2 & \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon_{1,t-1}\varepsilon_{2,t-1} & \varepsilon_{2,t-1}^2 \end{pmatrix} \begin{pmatrix} \tilde{a}_{11} & \tilde{a}_{21} \\ \tilde{a}_{12} & \tilde{a}_{22} \end{pmatrix} \\ &+ \begin{pmatrix} \tilde{b}_{11} & \tilde{b}_{12} \\ \tilde{b}_{21} & \tilde{b}_{22} \end{pmatrix}' \begin{pmatrix} \sigma_{1,t-1}^2 & \sigma_{12,t-1} \\ \sigma_{12,t-1} & \sigma_{2,t-1}^2 \end{pmatrix} \begin{pmatrix} \tilde{b}_{11} & \tilde{b}_{21} \\ \tilde{b}_{12} & \tilde{b}_{22} \end{pmatrix},\end{aligned}$$

[4] The acronym BEKK stands for Baba, Engle, Kraft, and Kroner.

with 11 unknown parameters. To reduce the computational burden, this model may be further constrained. The diagonal BEKK model is written as

$$\begin{aligned}\Sigma_t &= \begin{pmatrix} \tilde{\omega}_{11} & \tilde{\omega}_{12} \\ \tilde{\omega}_{12} & \tilde{\omega}_{22} \end{pmatrix} \\ &+ \begin{pmatrix} \tilde{a}_{11} & 0 \\ 0 & \tilde{a}_{22} \end{pmatrix} \begin{pmatrix} \varepsilon_{1,t-1}^2 & \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon_{1,t-1}\varepsilon_{2,t-1} & \varepsilon_{2,t-1}^2 \end{pmatrix} \begin{pmatrix} \tilde{a}_{11} & 0 \\ 0 & \tilde{a}_{22} \end{pmatrix} \\ &+ \begin{pmatrix} \tilde{b}_{11} & 0 \\ 0 & \tilde{b}_{22} \end{pmatrix} \begin{pmatrix} \sigma_{1,t-1}^2 & \sigma_{12,t-1} \\ \sigma_{12,t-1} & \sigma_{2,t-1}^2 \end{pmatrix} \begin{pmatrix} \tilde{b}_{11} & 0 \\ 0 & \tilde{b}_{22} \end{pmatrix},\end{aligned}$$

while the scalar BEKK model is

$$\begin{aligned}\Sigma_t = \begin{pmatrix} \tilde{\omega}_{11} & \tilde{\omega}_{12} \\ \tilde{\omega}_{12} & \tilde{\omega}_{22} \end{pmatrix} + \tilde{a}^2 \begin{pmatrix} \varepsilon_{1,t-1}^2 & \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon_{1,t-1}\varepsilon_{2,t-1} & \varepsilon_{2,t-1}^2 \end{pmatrix} \\ + \tilde{b}^2 \begin{pmatrix} \sigma_{1,t-1}^2 & \sigma_{12,t-1} \\ \sigma_{12,t-1} & \sigma_{2,t-1}^2 \end{pmatrix}.\end{aligned}$$

Dealing with the constant term

As discussed in Engle and Mezrich (1996), these models can be estimated with the additional constraint that the long-run covariance matrix is equal to the sample covariance matrix. This approach is often called *variance targeting.* It reduces the number of parameters dramatically and often gives improved performance in finite sample. For instance, in the case of the Vech model, we have the following parameterization

$$\text{vech}\,(\Omega) = \left(I_n - \sum_{i=1}^{p} A_i - \sum_{j=1}^{q} B_j \right) \text{vech}\,(S)\,,$$

where $S = \frac{1}{T}\sum_{t=1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_t'$ is the sample covariance matrix of residuals.

6.1.2 Dealing with large-dimensional systems

One general problem with multivariate GARCH models is the problem of dimensionality. The specifications described above would be very difficult to implement for large-dimensional portfolios because of the large number of parameters to be estimated. The Factor GARCH approach (and its extensions) reduces the number of variables (the factors) that have to be modeled with a multivariate GARCH dynamics. The Flexible GARCH model is based on a decentralized estimation of the covariance matrix.

Factor GARCH model

The Factor GARCH (or F-GARCH) has been proposed by Engle, Ng, and Rothschild (1990) to further reduce the number of parameters to be estimated. Engle, Ng, and Rothschild (1992) have proposed an extension of the F-GARCH model that allows distinguishing between the dynamic and static structure of asset returns. Other work based on F-GARCH models is by King, Sentana, and Wadhwani (1994) or Sentana and Fiorentini (2001).

The idea of this parameterization is that the joint dynamics of the $(n, 1)$ vector of returns x_t can be correctly described using a small number of factors K $(K < n)$. This model has been used by Bollerslev and Engle (1993) to model common persistence in stock market volatilities. The description of x_t is given by the $(K, 1)$ vector of factors f_t and the (n, K) matrix B of time-invariant factor loadings

$$x_t = Bf_t + \varepsilon_t.$$

Assume that the error term ε_t has constant (n, n) conditional covariance matrix Ω, that the K factors f_t have conditional covariance matrix Λ_t and that ε_t and f_t are uncorrelated. Then, the conditional covariance matrix of x_t is equal to[5]

$$V_{t-1}\left[x_t\right] = \Sigma_t = \Omega + B\Lambda_t B'. \tag{6.4}$$

If we assume now that the conditional covariance matrix of factors Λ_t is diagonal with elements $\lambda_{k,t}$ or if off-diagonal elements are constant and combined into Ω, then the model can be simplified as

$$\Sigma_t = \Omega + \sum_{k=1}^{K} \beta_k \beta_k' \lambda_{k,t},$$

where β_k denotes the kth column in B.

F-GARCH models have several interesting implications (see Engle, Ng, and Rothschild, 1990). First, the conditional covariance matrix Σ_t is guaranteed to be positive semi-definite. Second, we can always construct portfolios of assets that have the same conditional variance $\lambda_{k,t}$ as factors (up to a constant term). To see this, consider the portfolio $r_{k,t} = \phi_k' x_t$ where $\phi_k' \beta_j = 1$ if $j = k$ and 0 otherwise. Then, the conditional variance of $r_{k,t}$ is given by

$$V_{t-1}\left[r_{k,t}\right] = \phi_k' \Sigma_t \phi_k = \psi_k + \lambda_{k,t},$$

with $\psi_k = \phi_k' \Omega \phi_k$. Therefore, the portfolio $r_{k,t}$ has exactly the same time variation as the kth factor, so that it can be called factor-representing portfolio. This property indicates that the information in the factor-representing portfolios is sufficient for predicting the variances and covariances of individual asset returns. If we assume that the dynamics of each component $\lambda_{k,t}$ is

[5] In the case where ε_t and f_t are correlated with constant correlation matrix, we also obtain (6.4), with the constant matrix Ω regrouping terms of the covariance matrix of ε_t and the conditional covariance matrix of ε_t and f_t.

given by univariate GARCH(1, 1) models, we obtain the following conditional variances

$$V_{t-1}[r_{k,t}] = \omega_k + \alpha_k \left(\phi_k' \varepsilon_{t-1}\right)^2 + \gamma_k V_{t-2}[r_{k,t-1}]'.$$

Finally, if $K < n$, we can always construct $n - K$ portfolios of assets, i.e., linear combinations of x_t, which have constant variance.

Orthogonal models

The Orthogonal GARCH model (or O-GARCH) has been proposed by Alexander and Chibumba (1997) and Alexander (2001). It assumes that the observed data can be obtained by a linear transformation of a set of uncorrelated components and the matrix of the linear transformation is an orthogonal matrix. The (n, n) covariance matrix Σ_t is generated by m univariate GARCH models, where $m \leq n$ is determined using principal component analysis. The main interest of this approach is that it avoids estimating off-diagonal components of the multivariate GARCH parameter matrices, because the model is estimated not with original data but with its principal components that are by construction unconditionally uncorrelated.

The first step is to compute the sample (n, n) correlation matrix $\bar{R}$ of the n normalized innovations $u_t = D^{-1/2}\varepsilon_t$, where $D = \operatorname{diag}\left(\sigma_1^2, \cdots, \sigma_n^2\right)$ with σ_i^2 the sample variance of ε_t. Then, we assume that the data are generated by an orthogonal transformation of a small number K ($K \leq n$) of factors f_t. The matrix of this transformation is given by the eigenvectors of $\bar{R}$. The O-GARCH model is therefore defined as

$$\begin{aligned} u_t &= V_K f_t, \\ V_K &= P_K \Lambda_K^{1/2} = P_K \times \operatorname{diag}\left(\lambda_1^{1/2}, \cdots, \lambda_K^{1/2}\right), \end{aligned}$$

where Λ_K is the (K, K) diagonal matrix of eigenvalues of $\bar{R}$ (ranked from the largest to the smallest, $\lambda_1 \geq \cdots \geq \lambda_K$), P_K the associated (n, K) orthogonal matrix of eigenvectors, and $V_K = P_K \Lambda_K^{1/2}$. The unconditional covariance of f_t is the identity matrix of order K, by construction. The conditional covariance matrix of f_t is assumed to be a diagonal matrix (denoted Q_t) where each diagonal element is specified as a univariate GARCH model. The vector of factors $f_t = (f_{1,t}, \cdots, f_{K,t})'$ is thus characterized by

$$\begin{aligned} E_{t-1}[f_t] &= 0, \\ V_{t-1}[f_t] &= Q_t = \operatorname{diag}\left(\sigma_{f_1,t}^2, \cdots, \sigma_{f_K,t}^2\right), \\ \sigma_{f_i,t}^2 &= \left(1 - \sum_{h=1}^{p} \alpha_{ih} - \sum_{h=1}^{q} \beta_{ih}\right) + \sum_{h=1}^{p} \alpha_{ih} f_{i,t-h}^2 + \sum_{h=1}^{q} \beta_{ih} \sigma_{f_i,t-h}^2. \end{aligned}$$

Finally, the conditional covariance matrix of ε_t is simply given by

$$\Sigma_t = V_{t-1}[\varepsilon_t] = D^{1/2} R_t D^{1/2},$$

where

$$R_t = V_{t-1}[u_t] = V_m Q_t V_m'.$$

The parameters to be estimated are the following: The individual variances σ_i^2 of ε_t (included in D), the sample correlation matrix of u_t (the decomposition of which provides us with V_m), and the parameters (α_{ih} and β_{ih}, for $i = 1, \cdots, K$) of the univariate GARCH models. We notice that the individual variances σ_i^2 can be estimated by the sample analogue of D. In addition, once the number of factors K is chosen, V_K can be computed directly from the sample counterpart of $\bar{R}$. Consequently, the estimation burden reduces itself to the estimation of the individual GARCH models for the K factors.

The O-GARCH model implicitly assumes that the observed data can be linearly transformed into a set of uncorrelated components by means of an orthogonal matrix. These unobserved components can be interpreted as a set of uncorrelated factors that drive the market, similar to that in the F-GARCH model. The orthogonality assumption appears to be rather restrictive, however. Van der Weide (2002) has proposed the Generalized Orthogonal GARCH model (or GO-GARCH), in which innovations are linked by any possible invertible matrix. For this purpose, he argues that, when V is the map that links the uncorrelated components f_t with the observed process u_t, then there exists an orthogonal matrix U such that $V = P\Lambda^{1/2}U$, with $|U| = 1$. The O-GARCH model implicitly assumes $U = I_K$.

The matrix U can be represented as a product of $K(K-1)/2$ rotation matrices $U = \prod_{i<j} G_{ij}(\theta_{ij})$, with $-\pi \leq \theta_{ij} \leq \pi$, where $G_{ij}(\theta_{ij})$ performs a rotation in the plan spanned by e_i and e_j over an angle θ_{ij} for $i, j = 1, \cdots, n$, and e_i is the ith column of the (n, n) identity matrix. The conditional covariance matrix of ε_t has the same expression as before

$$\Sigma_t = V_{t-1}[\varepsilon_t] = D^{1/2} R_t D^{1/2},$$

where

$$R_t = V_{t-1}[u_t] = V Q_t V',$$

with $V = P\Lambda^{1/2}U$. The GO-GARCH model is stationary, provided the independent GARCH processes are stationary. In the case where $K = n$, the O-GARCH model can be viewed as a GO-GARCH for the particular choice $U = I_n$. Note also that the GO-GARCH model is a special case of the BEKK model, so that its properties can be derived from those of the BEKK model.

As for the O-GARCH model, an estimate of the parameters in D and V can be obtained from the sample analogues of D and $\bar{R}$. Therefore, the remaining parameters to be estimated are the parameters (α_{ih} and β_{ih}, for $i = 1, \cdots, K$) of the univariate GARCH models and the $K(K-1)/2$ rotation angles θ_{ij}. Related approaches have been followed by Vrontos, Dellaportas, and Politis (2003) and by Lanne and Saikkonen (2005).

Flexible GARCH model

Another approach can be taken in order to reduce the computational burden of large-scale multivariate GARCH models. Most models described in Section 6.1.1 are designed to reduce the number of parameters to be estimated. First-generation models impose some additional structure on the general specification. F-GARCH models summarize the information in asset returns through factors. By contrast, the flexible GARCH model proposed by Ledoit, Santa-Clara, and Wolf (2003) does not try to reduce the number of parameters but instead decentralize the estimation problem.

Assume that we have to estimate a diagonal vec model of the form

$$\sigma_{ij,t} = \omega_{ij} + \alpha_{ij}\varepsilon_{i,t-1}\varepsilon_{j,t-1} + \beta_{ij}\sigma_{ij,t-1},$$

where the conditional covariance $\sigma_{ij,t}$ between assets i and j depends on its own lag and on the cross-product between lagged innovations $\varepsilon_{i,t-1}$ and $\varepsilon_{j,t-1}$. In this case, for each (co)variance, we have 3 parameters to estimate, so that for n assets, we would have $3n\left(n+1\right)/2$ parameters. The idea of the flexible GARCH model consists in estimating the parameters $\left\{\omega_{ij}, \alpha_{ij}, \beta_{ij}\right\}$ for each (i, j) separately. Thus, the problem reduces to estimating one-dimensional (when $i = j$) or two-dimensional (when $i \neq j$) models. The difficulty then is to combine the various estimates into matrices $\hat{\Omega} = \{\hat{\omega}_{ij}\}$, $\hat{A} = \{\hat{\alpha}_{ij}\}$ and $\hat{B} = \{\hat{\beta}_{ij}\}$. Since the estimated parameters come from independent estimations, the covariance matrix Σ_t is not guaranteed to be positive semi-definite. Thus, once elements of the Σ_t matrix have been estimated, it will be necessary to use some trick to ensure positive semi-definiteness of this matrix.

The first step of the estimation of the flexible GARCH model corresponds to the (Q)ML estimation of the diagonal and off-diagonal elements of the covariance matrix. Diagonal elements are estimated using the standard univariate GARCH specification

$$\sigma_{ii,t} = \omega_{ii} + \alpha_{ii}\varepsilon_{i,t-1}^2 + \beta_{ii}\sigma_{ii,t-1},$$

with $\omega_{ii} > 0$, $\alpha_{ii} \geq 0$, $\beta_{ii} \geq 0$, and $\alpha_{ii} + \beta_{ii} < 1$. Off-diagonal elements $(\sigma_{ij,t})$ are estimated using the bivariate GARCH model for assets i and j

$$\varepsilon_{(ij),t} = \Sigma_{(ij),t}^{1/2} z_{(ij),t}$$

where

$$\varepsilon_{(ij),t} = \begin{pmatrix} \varepsilon_{i,t} \\ \varepsilon_{j,t} \end{pmatrix}, \qquad z_{(ij),t} = \begin{pmatrix} z_{i,t} \\ z_{j,t} \end{pmatrix}, \qquad \Sigma_{(ij),t} = \begin{pmatrix} \hat{\sigma}_{ii,t} & \sigma_{ij,t} \\ \sigma_{ij,t} & \hat{\sigma}_{jj,t} \end{pmatrix},$$

and

$$\sigma_{ij,t} = \omega_{ij} + \alpha_{ij}\varepsilon_{i,t-1}\varepsilon_{j,t-1} + \beta_{ij}\sigma_{ij,t-1}.$$

The vector $z_{(ij),t}$ is assumed to be normal $\mathcal{N}\left(0, I_2\right)$. Therefore, in the bivariate model, only the parameters pertaining to the covariance $\sigma_{ij,t}$ are estimated, and the conditional variances $\sigma_{ii,t}$ and $\sigma_{jj,t}$ are fixed to their first-stage values $\hat{\sigma}_{ii,t}$ and $\hat{\sigma}_{jj,t}$. In order to ensure that $\Sigma_{(ij),t}$ is positive definite, we impose the following bounds during the estimation: $|\omega_{ij}| \leq \sqrt{\hat{\omega}_{ii}\hat{\omega}_{jj}}$, $0 \leq \alpha_{ij} \leq \sqrt{\hat{\alpha}_{ii}\hat{\alpha}_{jj}}$, and $0 \leq \beta_{ij} \leq \sqrt{\hat{\beta}_{ii}\hat{\beta}_{jj}}$.

Then, Ledoit, Santa-Clara, and Wolf (2003) show how to render the bivariate estimates compatible in the sense that matrix Σ_t is positive definite. Using the matrix notation for the diagonal vec model

$$\Sigma_t = \Omega + A \odot \left(\varepsilon_{t-1}\varepsilon'_{t-1}\right) + B \odot \Sigma_{t-1},$$

and denoting $\div$ the element-wise division, they show that the conditional covariance matrix Σ_t is positive semi-definite if the three matrices $D \equiv \Omega \div (I_n - B)$, A, and B are positive semi-definite and $\alpha_{ii} + \beta_{ii} < 1$, $\forall i = 1, \cdots, n$. Now, if we define $\hat{D} = \hat{\Omega} \div \left(I_n - \hat{B}\right)$, we need to transform the estimated parameter matrices $\hat{A}$, $\hat{B}$, and $\hat{D}$ in order to ensure positive semi-definiteness of the conditional covariance matrix. The new matrices $\tilde{A}$, $\tilde{B}$, and $\tilde{D}$ are chosen to be the closest to $\hat{A}$, $\hat{B}$, and $\hat{D}$ but such that the diagonal parameters obtained from the estimation of the univariate GARCH models remain unchanged. Formally, we have to solve the following problems

$$\min_{\tilde{D}} \left\|\tilde{D} - \hat{D}\right\|$$
$$\text{s.t.} \quad \tilde{D} \text{ is positive semi-definite and } \tilde{d}_{ii} = \hat{d}_{ii}, \forall i = 1, \cdots, n,$$

$$\min_{\tilde{A}} \left\|\tilde{A} - \hat{A}\right\|$$
$$\text{s.t.} \quad \tilde{A} \text{ is positive semi-definite and } \tilde{\alpha}_{ii} = \hat{\alpha}_{ii}, \forall i = 1, \cdots, n,$$

$$\min_{\tilde{B}} \left\|\tilde{B} - \hat{B}\right\|$$
$$\text{s.t.} \quad \tilde{B} \text{ is positive semi-definite and } \tilde{\beta}_{ii} = \hat{\beta}_{ii}, \forall i = 1, \cdots, n.$$

Once these matrices have been obtained, we deduce $\tilde{\Omega} = \tilde{D} \odot \left(I_n - \tilde{B}\right)$. One interesting property of this approach is that we have, by construction, $|\tilde{\alpha}_{ij} + \tilde{\beta}_{ij}| < 1$, $\forall i, j = 1, \cdots, n$. Therefore, it is sufficient to impose that $\hat{\alpha}_{ii} + \hat{\beta}_{ii} < 1$, $\forall i = 1, \cdots, n$ to ensure positive semi-definiteness of the conditional covariance matrix.

A drawback of this approach is that there is no straightforward way to compute the standard errors of the parameter estimates. The reason is that the new matrices $\tilde{A}$, $\tilde{B}$, and $\tilde{D}$ are very nonlinear transformations of the initial matrices $\hat{A}$, $\hat{B}$, and $\hat{D}$ for which standard errors are available. Ledoit, Santa-Clara, and Wolf (2003) suggest the use of the bootstrap procedure to obtain standard errors. It should be noticed that for large-scale multivariate GARCH models, the need for such standard errors is not clear.

6.1.3 Modeling conditional correlation

The models described in the first section can be viewed as natural extensions of the baseline univariate GARCH model. In particular, they all propose a specification for the conditional covariances that is similar to the one adopted for modeling variances in the univariate GARCH model. Unfortunately, they have some drawbacks. First, the number of unknown parameters is a power function of the number of variables, so that the estimation of these models becomes extremely difficult as n grows. Second, the derivation of the restrictions ensuring that the covariance matrix is positive definite is often difficult (except for the BEKK model).

Second-generation models focus on the dynamics of correlations rather than on the dynamics of covariances. At first sight, this task is more demanding, because it cannot be constructed as a natural generalization of the univariate GARCH model. However, due to the critical role of correlations in finance, this shift from covariances to correlations was needed.

The Constant Conditional Correlation (CCC) model

Bollerslev (1990) has suggested that the time-varying conditional covariances be parameterized in order to be proportional to the product of the corresponding conditional standard deviations. The intuition for this model is the following. Assume that $\sigma_{ij,t}$ is the covariance between two assets i and j to be modeled. Also, let $\sigma_{i,t}^2$ be the conditional variance modeled by some univariate GARCH model. Under the assumption of keeping correlation constant, denoting ρ_{ij} the constant correlation between the assets i and j, it follows that

$$\rho_{ij} = \frac{\sigma_{ij,t}}{\sigma_{i,t}\sigma_{j,t}} \qquad \Rightarrow \qquad \sigma_{ij,t} = \rho_{ij}\sigma_{i,t}\sigma_{j,t}.$$

Thus, knowledge of ρ_{ij} that can be computed using standardized innovations, and knowledge of the marginal GARCH models, yields a description of the time-varying covariance.

The extension to a general model with n assets is straightforward. Bollerslev (1990) introduces a time-invariant (n,n) correlation matrix with unit diagonal elements

$$R = \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{12} & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho_{n-1,n} \\ \rho_{1n} & \cdots & \rho_{n-1,n} & 1 \end{pmatrix}. \tag{6.5}$$

Therefore, the temporal variation in Σ_t is determined solely by the conditional variances

$$\Sigma_t = D_t^{1/2} R D_t^{1/2}, \tag{6.6}$$

where D_t is, as before, the diagonal matrix of conditional variances. We only need to model the dynamics of the n conditional variances and to estimate the constant correlation matrix, so that the number of parameters to estimate reduces itself to $n(1+p+q)+n(n+1)/2$. An advantage of this approach is that if the conditional variances in the D_t matrices are all positive and the conditional correlation matrix R is positive definite, the sequence of conditional covariance matrices Σ_t is guaranteed to be positive definite for all t. Note, in addition, that as argued before, the correlation matrix R may be estimated in a preliminary step using the sample correlation matrix of normalized residuals.

The CCC model has been widely used in the empirical literature because of its computational simplicity. However, at least for financial returns, the assumption of constant correlation is not supported by the data. There is a huge empirical evidence that correlations vary over time (see Chapter 2). Several extensions of the CCC model have been recently proposed to allow time-varying conditional correlations.

The first models with time-varying conditional correlations have been proposed by Engle (2002), Engle and Sheppard (2001), and Tse and Tsui (2002). The basic idea is that the conditional correlation matrix R_t is in fact time varying, so that the conditional covariance matrix previously defined by (6.6) is now defined as

$$\Sigma_t = D_t^{1/2} R_t D_t^{1/2}. \tag{6.7}$$

Since the two models of Engle (2002) and Tse and Tsui (2002) are conceptually different, we present both of them in turn.

The dynamic conditional correlation (DCC) model

Engle (2002) and Engle and Sheppard (2001) have developed a model in which the conditional correlation matrix in (6.7) is defined by

$$\begin{aligned} R_t &= \operatorname{diag}(Q_t)^{-1/2} \times Q_t \times \operatorname{diag}(Q_t)^{-1/2}, \\ Q_t &= (1-\delta_1-\delta_2)\bar{Q} + \delta_1 \left(u_{t-1}u_{t-1}'\right) + \delta_2 Q_{t-1}, \end{aligned}$$

where $\bar{Q}$ is the unconditional covariance matrix of $u_t = \{\varepsilon_{i,t}/\sigma_{i,t}\}_{i=1,\cdots,n}$ and diag(Q_t) is the (n,n) matrix with the diagonal of Q_t on the diagonal and zeros off-diagonal. The matrix $\bar{Q}$ may be estimated by the sample analogue $\frac{1}{T}\sum_{t=1}^{T}\hat{u}_t\hat{u}_t'$. Parameters δ_1 and δ_2 are assumed to satisfy $0 \le \delta_1, \delta_2 \le 1$ and $\delta_1+\delta_2 \le 1$. Once these restrictions are imposed, the conditional correlation matrix R_t is guaranteed to be positive definite during the estimation.

A drawback of this approach is that only two additional parameters δ_1 and δ_2 drive the dynamics of all the correlations. Cappiello, Engle, and Sheppard (2003) have recently suggested an extension of this model in which each element of the correlation matrix has an autonomous and asymmetric dynamics

$$\begin{aligned} R_t &= \operatorname{diag}(Q_t)^{-1/2} \times Q_t \times \operatorname{diag}(Q_t)^{-1/2}, \\ Q_t &= \left(\bar{Q} - A'\bar{Q}A - B'\bar{Q}B - G'\bar{N}G\right) + A'\left(u_{t-1}u_{t-1}'\right)A \\ &\quad + B'Q_{t-1}B + G'\left(n_{t-1}n_{t-1}'\right)G, \end{aligned}$$

where A, B and G are (n,n) diagonal parameter matrices and $n_t = 1_{\{u_t<0\}} \odot u_t$ is the $(n,1)$ vector that contains the normalized residual if it is negative and 0 otherwise. $\bar{Q} = E[u_t u_t']$ and $\bar{N} = E[n_t n_t']$ are estimated by their sample analogues $\frac{1}{T}\sum_{t=1}^{T} \hat{u}_t \hat{u}_t'$ and $\frac{1}{T}\sum_{t=1}^{T} \hat{n}_t \hat{n}_t'$.

The conditional covariance matrix $\Sigma_t = D_t^{1/2} R_t D_t^{1/2}$ is guaranteed to be positive definite if $\left(\bar{Q} - A'\bar{Q}A - B'\bar{Q}B - G'\bar{N}G\right)$ is positive definite. The DCC model of Engle (2002) is obtained as a special case if $G = 0$ and if A and B have $\sqrt{\delta_1}$ and $\sqrt{\delta_2}$ on the diagonal, respectively.

The time-varying correlation (TVC) model

The idea of Tse and Tsui (2002) is to specify the dynamic of the conditional correlation matrix R_t as an ARMA process

$$R_t = (1 - \theta_1 - \theta_2)\, R + \theta_1 R_{t-1} + \theta_2 \Psi_{t-1}, \tag{6.8}$$

where $R = \left\{\rho_{ij}\right\}$ is a time-invariant (n,n) matrix of correlations as in (6.5). The key idea of Tse and Tsui (2002) is to model the (n,n) matrix $\Psi_t = \left\{\psi_{ij,t}\right\}$ using a set of past normalized observations

$$\psi_{ij,t} = \frac{\sum_{h=0}^{m-1} u_{i,t-h} u_{j,t-h}}{\sqrt{\left(\sum_{h=0}^{m-1} u_{i,t-h}^2\right)\left(\sum_{h=0}^{m-1} u_{j,t-h}^2\right)}} \qquad \text{for } 1 \le i < j \le n,$$

where $u_t = D_t^{-1}\varepsilon_t = \left(\varepsilon_{i,t}/\sigma_{i,t}\right)_{i=1,\cdots,n}$. Stated differently, Ψ_t is the sample correlation matrix of normalized residuals $E_t = \left(u_t, \cdots, u_{t-m+1}\right)'$. Therefore, if we define B_t the (n,n) diagonal matrix with $\left(\sum_{h=0}^{m-1} u_{i,t-h}^2\right)^{1/2}$ as ith diagonal element, we can rewrite Ψ_t as

$$\Psi_t = B_t^{-1} E_t E_t' B_t^{-1} = \left\{\psi_{ij,t}\right\}.$$

As long as $m \ge n$, the matrix Ψ_t will be in general positive definite, of course if the $u_{i,t}$ are not linearly dependent. To ensure positive definiteness of the matrix R_t, the parameters θ_1 and θ_2 have to satisfy $0 \le \theta_1, \theta_2 \le 1$ and $\theta_1 + \theta_2 \le 1$. Note that the constant conditional correlation model is nested in this model, because it corresponds to the case where $\theta_1 = \theta_2 = 0$. Time-variability in the conditional correlation matrix is therefore obtained at the cost of only two additional parameters.

While the model proposed by Tse and Tsui (2002) builds on an ARMA process for the dynamics of the correlation matrix, the model of Engle (2002) is based on a GARCH-type specification for the dynamics of the covariance matrix. An advantage of the latter approach is that the dynamic of Q_t is based on a single lag of the $u_{i,t}$ terms (as in the standard GARCH$(1,1)$ model), whereas (6.8) requires an arbitrary number m of observations to compute sample correlations.

General dynamic covariance model

A generalization of most of the previous models has been developed by Kroner and Ng (1998). The so-called general dynamic covariance (GDC) model nests many of the existing models while including asymmetric effects. The model is defined as follows

$$
\begin{aligned}
\Sigma_t &= D_t R D_t + \Phi \odot \Theta_t, \\
D_t &= \{d_{ij,t}\} \text{ with } d_{ii,t} = \sqrt{\theta_{ii,t}}, \forall i \text{ and } d_{ij,t} = 0, \forall i \neq j, \\
\Theta_t &= \{\theta_{ij,t}\}, \\
\theta_{ij,t} &= \omega_{ij} + a_i' \varepsilon_{t-1} \varepsilon_{t-1}' a_i + b_i \Sigma_{t-1} b_i' \qquad \forall i, j, \\
R &= \{\rho_{ij}\} \text{ with } \rho_{ii} = 1 \qquad \forall i, \\
\Phi &= \{\varphi_{ij}\} \text{ with } \varphi_{ii} = 0, \forall i \text{ and } \varphi_{ij} = \varphi_{ji}, \forall i, j,
\end{aligned}
$$

where a_i, b_i, $i = 1, \cdots, n$, are $(n, 1)$ parameter vectors, ρ_{ij}, φ_{ij}, and ω_{ij} are scalars with $\Omega = \{\omega_{ij}\}$ and R positive definite and symmetric matrices.

The GDC model has two components: the first term $D_t R D_t$ is like the constant correlation model but with the variance functions given by that of the BEKK model. The second term $\Phi \odot \Theta_t$ has zero diagonal elements but has off-diagonal elements given by the BEKK-type covariance functions. Note that the elements of Σ_t can be written as

$$
\begin{aligned}
\sigma_{ii,t} &= \theta_{ii,t} \qquad \forall i, \\
\sigma_{ij,t} &= \rho_{ij} \sqrt{\theta_{ii,t}} \sqrt{\theta_{jj,t}} + \varphi_{ij} \theta_{ij,t} \qquad \forall i \neq j.
\end{aligned}
$$

Thus, the GDC model is a hybrid of the constant conditional correlation model and the BEKK model.

Proposition 6.1. *(Kroner and Ng, 1998) Consider the following set of conditions:*

1. $\rho_{ij} = 0 \qquad \forall i \neq j$.
2. $a_i = \alpha_i e_i$ *and* $b_i = \beta_i e_i$ $\forall i$, *where* e_i *is the ith column of an* (n, n) *identity matrix, and* α_i *and* β_i, $i = 1, \cdots, n$, *are scalars.*
3. $\varphi_{ij} = 0 \qquad \forall i \neq j$.
4. $\varphi_{ij} = 1 \qquad \forall i \neq j$.
5. $A = \alpha(\omega\lambda')$ *and* $B = \beta(\omega\lambda')$ *where* $A = \{a_i\}_{i=1}^n$, $B = \{b_i\}_{i=1}^n$, ω *and* λ *are* $(n, 1)$ *vectors and* α *and* β *are scalars.*

The GDC model reduces to several multivariate GARCH models under different combinations of these conditions. Specifically, the GDC model becomes a restricted vech model (with the restrictions $\beta_{ij} = \beta_{ii}\beta_{jj}$ *and* $\alpha_{ij} = \alpha_{ii}\alpha_{jj}$*) under conditions (i) and (ii), the constant conditional correlation model under conditions (ii) ad (iii), the BEKK model under conditions (i), and (iv) and the F-GARCH model under conditions (i), (iv), and (v).*

Kroner and Ng (1998) also develop an extension to the GDC that allows for asymmetry. The asymmetric dynamic covariance (ADC) model has the same structure as the GDC model, except that the equation for $\theta_{ij,t}$ incorporates the leverage effect in the BEKK model, in a way close to the GJR model

$$\theta_{ij,t} = \omega_{ij} + a_i' \varepsilon_{t-1} \varepsilon_{t-1}' a_i + b_i \Sigma_{t-1} b_i' + g_i' \eta_{t-1} \eta_{t-1}' g_i \qquad \forall i, j,$$

where $\eta_t = \left(\eta_{1,t}, \cdots, \eta_{n,t}\right)'$ with $\eta_{i,t} = \max\left(-\varepsilon_{i,t}, 0\right)$ and g_i, $i = 1, \cdots, n$, are scalars.

6.1.4 Estimation issues

Maximum likelihood

We suppose now that we have a sample of size T of the $(n, 1)$ vector of observations written as $\underline{x}_T = \{x_t\}_{t=1}^T$ with conditional mean and conditional variance given by (6.1) and (6.2). Unknown parameters are regrouped in θ. Under the assumption of conditional multivariate normal distribution, the log-likelihood function for $\underline{x}_T$ is

$$L_T\left(\theta | \underline{x}_T\right) = \sum_{t=1}^{T} \ell_t\left(\theta\right),$$

with

$$\ell_t\left(\theta\right) = -\frac{n}{2} \log\left(2\pi\right) - \frac{1}{2}\left(\log\left|\Sigma_t\left(\theta\right)\right| + \left(x_t - \mu_t\left(\theta\right)\right)' \Sigma_t^{-1}\left(\theta\right)\left(x_t - \mu_t\left(\theta\right)\right)\right), \tag{6.9}$$

where the term $-\frac{1}{2} \log\left|\Sigma_t\left(\theta\right)\right|$ comes from the Jacobian of the transformation from the innovation process to the observed variables.

Then, the ML estimator $\hat{\theta}_{ML}$, that maximizes (6.9), is asymptotically normal with asymptotic distribution

$$\sqrt{T}\left(\hat{\theta}_{ML} - \theta_0\right) \Longrightarrow \mathcal{N}\left(0, A_0^{-1}\right),$$

where A_0 is the information matrix evaluated at the true parameter vector θ_0. See Section 4.3.3 for further details on the construction and estimation of A_0.

Estimation of the DCC model

For the DCC model, Engle (2002) and Engle and Sheppard (2001) have also proposed a two-step estimation, based on the idea that parameters of the conditional variances (denoted θ_V) and of the conditional correlations (denoted θ_C) can be estimated separately with $\theta = \left(\theta_V', \theta_C'\right)'$.[6] A justification is

[6] Bollerslev (1990) proposed a similar two-step approach for the estimation of the CCC model. We present here the case of the DCC model, because it nests the CCC model.

that the log-likelihood can be written as the sum of a volatility part and a correlation part. Since $\Sigma_t = D_t^{1/2} R_t D_t^{1/2}$, we have

$$\log |\Sigma_t| = \log |D_t| + \log |R_t| ,$$

because D_t is diagonal, and

$$\begin{aligned}(x_t - \mu_t)' \Sigma_t^{-1} (x_t - \mu_t) &= (x_t - \mu_t)' \left(D_t^{1/2} R_t D_t^{1/2}\right)^{-1} (x_t - \mu_t) \\ &= u_t' R_t^{-1} u_t + (x_t - \mu_t)' D_t^{-1} (x_t - \mu_t) - u_t' u_t,\end{aligned}$$

where $u_t = D_t^{-1/2} \varepsilon_t$ is a vector of normalized innovations. The last two terms of the second equality are clearly equal, but this expression allows one to break down the log-likelihood in the two following terms

$$\ell (\theta_V, \theta_C | \underline{x}_T) = \ell_V (\theta_V | \underline{x}_T) + \ell_C (\theta_V, \theta_C | \underline{x}_T) ,$$

with

$$\begin{aligned}\ell_V (\theta_V | \underline{x}_T) &= -\frac{1}{2} \sum_{t=1}^{T} \left[\frac{n}{2} \log (2\pi) + \log |D_t| + (x_t - \mu_t)' D_t^{-1} (x_t - \mu_t)\right] \\ &= -\sum_{i=1}^{n} \left[\frac{T}{2} \log (2\pi) + \frac{1}{2} \sum_{t=1}^{T} \left(\log \left(\sigma_{i,t}^2\right) + \left(\frac{x_{it} - \mu_{i,t}}{\sigma_{i,t}}\right)^2\right)\right], \\ \ell_C (\theta_V, \theta_C | \underline{x}_T) &= -\frac{1}{2} \sum_{t=1}^{T} \left(\log |R_t| + u_t' R_t^{-1} u_t - u_t' u_t\right) .\end{aligned}$$

Notice that $\ell_V (\theta_V | \underline{x}_T)$ is simply the sum of log-likelihoods of the individual GARCH equations for each series. The second step consists in estimating the parameters pertaining to the correlation matrix, conditionally on the parameters estimated in the first stage.

Therefore, since squared residuals are not dependent on correlation parameters, these parameters can be ignored for the estimation of the conditional volatility dynamics. The two-step estimation then relies on maximizing the log-likelihood as follows. First, we estimate the volatility parameters through

$$\hat{\theta}_V \in \arg\max_{\{\theta_V\}} \qquad \ell_V (\theta_V | x_t, t = 1, \cdots, T) ,$$

and then

$$\hat{\theta}_C \in \arg\max_{\{\theta_C\}} \qquad \ell_C \left(\hat{\theta}_V, \theta_C | \underline{x}_T\right) = -\frac{1}{2} \sum_{t=1}^{T} \left(\log |R_t| + \hat{u}_t' R_t^{-1} \hat{u}_t - \hat{u}_t' \hat{u}_t\right) ,$$

where $\hat{u}_{i,t} = \left(r_{i,t} - \hat{\mu}_{i,t}\right) / \hat{\sigma}_{i,t}$.

Engle and Sheppard (2001) show that the two-step estimator $\hat{\theta}_{TS} = (\hat{\theta}_V', \hat{\theta}_C')'$ is consistent and asymptotically normal, with distribution

$$\sqrt{T}\left(\hat{\theta}_{TS} - \theta_0\right) \Longrightarrow \mathcal{N}\left(0, A_0^{-1} B_0 A_0^{\prime -1}\right),$$

where

$$A_0 = \begin{bmatrix} \frac{\partial^2 \ell_V(\theta_{V0})}{\partial\theta_V \partial\theta_V'} & 0 \\ \frac{\partial^2 \ell_C(\theta_0)}{\partial\theta_V \partial\theta_C'} & \frac{\partial^2 \ell_C(\theta_0)}{\partial\theta_C \partial\theta_C'} \end{bmatrix},$$

$$B_0 = V\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial \ell_V'(\theta_{V0})}{\partial\theta_V}, \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial \ell_C'(\theta_0)}{\partial\theta_C}\right].$$

Due to the structure of A_0, the asymptotic variances of the GARCH parameters $\hat{\theta}_V$ for each series are the standard robust covariance matrix estimators. For the second-stage parameters, however, the asymptotic variance involves all parameters.

Since the two sets of parameters are estimated using limited information, the estimator $\hat{\theta}_{TS}$ is not fully efficient. However, as argued by Pagan (1986), if we perform an additional iteration of the Newton-Raphson algorithm to the log-likelihood, starting at $\hat{\theta}_{TS}$, then we obtain an asymptotically efficient estimator.

6.1.5 Specification tests

Due to its computational burden, it is important to test whether a multivariate GARCH model is able to fit the data correctly. A first preliminary test would be to evaluate the ability of univariate GARCH models to correctly describe the data for each series. However, such an approach cannot be recommended, because it does not take into account the possible correlation between the series, which is actually the main characteristic of multivariate models. Therefore, multivariate tests are needed.

We do not focus in this section on the detection of non-normality in a multivariate GARCH context. The issue of dealing with non-normal distributions is addressed in the next section devoted to the modeling of multivariate models with non-normal distributions. We thus consider tests of GARCH effects, or alternatively tests of serial correlation in square standardized residuals. There are basically two types of tests. The first series of tests is based on the properties of standardized residuals, and the second test is a Portmanteau test.

Residual-based statistics

A natural diagnostic test for the multivariate GARCH model, proposed by Tse (2002), is based on the regression of cross-products of the standardized residuals $\hat{z}_{i,t}\hat{z}_{j,t}$ on own lags as follows

$$\hat{z}_{i,t}\hat{z}_{j,t} - \hat{\rho}_{ij,t} = \sum_{k=1}^{p} \alpha_{ij}^{k} \left(\hat{z}_{i,t-k}\hat{z}_{j,t-k} - \hat{\rho}_{ij,t-k}\right) + \nu_{ij,t},$$

for $i, j = 1, \cdots, n$, $j > i$. The choice of explanatory variables in these regressions may theoretically differ from one exercise to the other, depending on the type of dependency we want to test. If we denote $d_{ij,t} = (\hat{z}_{i,t-1}\hat{z}_{j,t-1}, \cdots, \hat{z}_{i,t-p}\hat{z}_{j,t-p})'$, the test statistic is

$$RB\,(p) = T\hat{\alpha}_{ij}'\hat{L}_{ij}\hat{\Omega}_{ij}^{-1}\hat{L}_{ij}\hat{\alpha}_{ij},$$

with

$$\begin{aligned} L_{ij} &= \operatorname*{plim}_{T\to\infty} \left(\frac{1}{T}\sum_{t=1}^{T} d_{ij,t}d_{ij,t}'\right), \\ \Omega_{ij} &= E\left[\left(z_{i,t}z_{j,t} - \rho_{ij,t}\right)^2\right] L_{ij} - Q_{ij}GQ_{ij}', \\ Q_{ij} &= \operatorname*{plim}_{T\to\infty} \left(\frac{1}{T}\sum_{t=1}^{T} d_{ij,t}\frac{\partial\left(z_{i,t}z_{j,t} - \rho_{ij,t}\right)}{\partial\theta'}\right), \end{aligned}$$

where θ is the vector of parameters in the GARCH model and G is the asymptotic covariance matrix of θ, such that $\sqrt{T}\left(\hat{\theta} - \theta\right) \Longrightarrow \mathcal{N}\,(0, G)$. Tse (2002) shows that, under the null hypothesis that the specification is correct, the statistic $RB\,(p)$ is asymptotically distributed as a χ^2 with p degrees of freedom.

Evidently, a similar test can be performed for squared standardized residuals $\hat{z}_{i,t}^2$, using the following regression

$$\hat{z}_{i,t}^2 - 1 = \sum_{k=1}^{p} \alpha_i^k \left(\hat{z}_{i,t-k}^2 - 1\right) + \nu_{i,t} \qquad \text{for } i = 1, \cdots, n.$$

Portmanteau statistics

The Ljung-Box portmanteau test for serial correlation has been extended to a multivariate context by Baillie and Bollerslev (1990) (see also Hosking, 1980). It is written as

$$H\,(p) = T^2 \sum_{i=1}^{p} \left(\frac{1}{T-i}\right) \operatorname{tr}\left(\hat{C}_i\hat{C}_0^{-1}\hat{C}_i\hat{C}_0^{-1}\right),$$

where $\hat{C}_i = \frac{1}{T}\sum_{t=i+1}^{T}\left(x_t - \bar{x}\right)\left(x_{t-i} - \bar{x}\right)'$ is the sample autocovariance matrix of order i of x_t. Under the null hypothesis of constant correlation, the statistics $H\,(p)$ is distributed as a $\chi^2\left(n^2p\right)$.

This test statistic may be used to test the presence of ARCH effects in the squared returns. In this case, we define the $n\,(n+1)\,/2$ vector of cross-products as $y_t = vech\left(\left(x_t - \bar{x}\right)\left(x_t - \bar{x}\right)'\right)$. Then, the test statistic $H\,(p)$ is estimated with $\hat{C}_i = \frac{1}{T}\sum_{t=i+1}^{T}(y_t - \bar{y})(y_{t-i} - \bar{y})'$ as a sample autocovariance matrix for the covariances. Under the null hypothesis of no ARCH effect, the test statistic $H(p)$ is asymptotically distributed as a $\chi^2(n^2 p)$.

6.1.6 Test of constant conditional correlation matrix

Due to its computational simplicity, the CCC–GARCH model is very popular among empirical researchers. However, there are several problems that seem to be overlooked in empirical applications. First, the assumption of constant correlation is often taken for granted and seldom analyzed or tested. Second, the issue of how the assumption of a constant conditional correlation affects the dynamics of the conditional variance is rarely considered.

Alternative tests can be based on a specific parametric specification of the conditional correlation. In this case, however, implementing the test procedure is more demanding, because we have to estimate, or at least to specify, the dynamics of correlation. Therefore, we loose the main advantage of the constant correlation GARCH model, for which the sample correlation matrix is a consistent estimator of the conditional correlation matrix. Several tests have been proposed that rely on different dynamics of the correlation matrix under the alternative. The dynamic correlation models presented above evidently provide a suitable setup for testing the constant correlation hypothesis.[7] In the following, we focus on tests that do not require the estimation of a complete GARCH model with timescale correlations.

Test based on the information matrix

Bera and Kim (2002) suggested an Information Matrix (IM) test for the constant-correlation hypothesis in a bivariate GARCH model. The basic idea is to derive a score test of the hypothesis that the variances of the parameters of interest are 0, so that it does not require the explicit specification of an alternative hypothesis. A definite advantage of this approach is thus that the test statistic does not depend on a particular specification of correlation variations.

Under the null hypothesis of constant correlation, the conditional covariance matrix Σ_t can be written as $\Sigma_t = D_t^{1/2} R D_t^{1/2}$ with

$$D_t = \begin{pmatrix} \sigma_{1,t}^2 & 0 \\ 0 & \sigma_{2,t}^2 \end{pmatrix}, \qquad \text{and} \qquad R = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

The matrix Σ_t is positive definite for all t, if each of the conditional variances $\sigma_{i,t}$ are positive, and if the conditional correlation matrix is positive definite, i.e., $|\rho| < 1$.

[7] For instance, in the model by Tse and Tsui (2002), the null of constant correlation simply corresponds to $\theta_1 = \theta_2 = 0$.

The conditional log-likelihood is then defined as

$$\ell(\theta) = \frac{1}{T}\sum_{t=1}^{T} \ell_t(\theta),$$

with θ the vector of unknown parameters and

$$\begin{aligned}\ell_t(\theta) &= -\log(2\pi) - \frac{1}{2}\log|\Sigma_t| - \frac{1}{2}\varepsilon_t' \Sigma_t^{-1}\varepsilon_t \\ &= -\log(2\pi) - \frac{1}{2}\log\left(1-\rho^2\right) - \frac{1}{2}u_t' R^{-1} u_t,\end{aligned}$$

with $u_t = D_t^{1/2}\varepsilon_t$ denoting normalized innovations.

Then, first and second derivatives with respect to ρ are

$$\begin{aligned}\frac{\partial \ell_t(\theta)}{\partial \rho} &= \frac{1}{1-\rho^2}\left(v_{1,t}v_{2,t}+\rho\right), \\ \frac{\partial^2 \ell_t(\theta)}{\partial \rho^2} &= \frac{1}{\left(1-\rho^2\right)^2}\left(-v_{1,t}^2 - v_{2,t}^2 + 2\rho v_{1,t}v_{2,t} + 1 + \rho^2\right),\end{aligned}$$

where $v_t = \left(u_{1,t} - \rho u_{2,t}, u_{2,t} - \rho u_{1,t}\right)/\sqrt{1-\rho^2}$. This implies that the score is

$$\begin{aligned}s_t(\theta) &= \left(\frac{\partial \ell_t(\theta)}{\partial \rho}\right)^2 + \frac{\partial^2 \ell_t(\theta)}{\partial \rho^2} \\ &= \frac{1}{\left(1-\rho^2\right)^2}\left(v_{1,t}^2 v_{2,t}^2 + 4\rho v_{1,t}v_{2,t} - v_{1,t}^2 - v_{2,t}^2 + 1 + 2\rho^2\right).\end{aligned}$$

Suppose now that we have the ML estimator $\hat{\theta}$ of the parameter vector, so that $\hat{v}_{i,t}$ is the estimate of $\nu_{i,t}$ with θ replaced by $\hat{\theta}$. In particular, we have $\hat{\rho} = \frac{1}{T}\sum_{t=1}^{T} \hat{u}_{1,t}\hat{u}_{2,t}$. Note that

$$\frac{1}{T}\sum_{t=1}^{T} \hat{v}_{i,t}\hat{v}_{j,t} = \begin{cases} 1 & \text{if } i = j, \\ -\hat{\rho} & \text{if } i \neq j. \end{cases}$$

Therefore, we obtain the indicator $d\left(\hat{\theta}\right)$ in the IM test for the constancy of ρ_t, when $\hat{\theta}$ is the MLE of θ, as

$$d\left(\hat{\theta}\right) = \left(1-\rho^2\right)^2 \sum_{t=1}^{T} s_t(\hat{\theta}) = \frac{1}{T}\sum_{t=1}^{T}\left(\hat{v}_{1,t}^2\hat{v}_{2,t}^2 - 1 - 2\hat{\rho}^2\right).$$

Bera and Kim (2002) also show that the asymptotic variance of $d(\hat{\theta})$ is

$$V\left[\sqrt{T}d(\hat{\theta})\right] = 4\left(1 + 4\rho^2 + \rho^4\right),$$

which can be consistently estimated by substituting the MLE $\hat{\rho}$ in place of ρ.

Therefore the test statistic is

$$IMC = \frac{Td(\hat{\theta})^2}{\hat{V}\left[\sqrt{T}d(\hat{\theta})\right]} = \frac{\left(\sum_{t=1}^{T}\left(\hat{v}_{1,t}^2\hat{v}_{2,t}^2 - 1 - 2\hat{\rho}^2\right)\right)^2}{4T\left(1 + 4\hat{\rho}^2 + \hat{\rho}^4\right)}.$$

It is asymptotically distributed as a $\chi^2(1)$ under the null of constant correlation.

Test based on the LM statistic

Tse (2000) proposed a test for the constant-correlation hypothesis based on the LM approach. The idea is to extend the constant-correlation model to one that includes time-varying correlations. When certain parameters in the extended model are imposed to be zero, the constant-correlation model is obtained. The extension proposed by Tse is simply

$$\begin{aligned}\rho_{ij,t} &= \rho_{ij} + \delta_{ij}\varepsilon_{i,t-1}\varepsilon_{j,t-1},\\ \sigma_{ij,t} &= \rho_{ij,t}\sigma_{i,t}\sigma_{j,t},\end{aligned}$$

where δ_{ij} are additional parameters.

The constant-correlation hypothesis can be tested by examining the hypothesis $H_0 : \delta_{ij} = 0$ for $1 \leq i < j \leq 1$. Under the null hypothesis, there are $n(n-1)/2$ independent restrictions.

To ensure that the alternative model provides well-defined positive definite conditional covariance matrices, further restrictions have to be imposed on the parameters δ_{ij}. It is assumed that within a neighborhood of $\delta_{ij} = 0$, the optimal properties of the LM test hold under some regularity conditions.[8]

We denote θ the vector of unknown parameters, including δ_{ij}, $1 \leq i < j \leq n$. We define the $(n, 1)$ score vector

$$s = \frac{\partial \ell(\theta)}{\partial \theta} = \sum_{t=1}^{T} \frac{\partial \ell_t(\theta)}{\partial \theta},$$

and the (n, n) information matrix

$$V = E\left[-\frac{\partial^2 \ell(\theta)}{\partial \theta \partial \theta'}\right].$$

[8] An alternative specification is to allow correlation to depend on the products of the lagged standardized residuals

$$\rho_{ij,t} = \rho_{ij} + \delta_{ij} z_{i,t-1} z_{j,t-1}.$$

In this case, however, because $z_{i,t}$ depends on other parameters of the model through $\sigma_{i,t}$, an analytic derivation of the LM statistic is intractable.

Then, the LM statistic for the null of constant correlation is $\hat{s}'\hat{V}^{-1}\hat{s}$, where the hats denote evaluation at $\hat{\theta}$ under the null hypothesis. Tse suggests to replace V by the sum of cross-products of the first derivatives of ℓ_t, so that

$$LMC = \hat{s}' \left(\hat{S}'\hat{S}\right)^{-1} \hat{s},$$

where $\hat{S}$ is the (T, n) matrix the rows of which are the partial derivatives of $\partial \ell_t (\theta) / \partial \theta'$. Under the regularity conditions, LMC is asymptotically distributed as a $\chi^2 (n(n-1)/2)$.

Monte-Carlo simulations have been performed by Tse (2000) to compare the two test statistics IMC and LMC. Although both test statistics appear to be correctly sized under the null, the LM test is found to have better power against some alternatives and to be more robust to non-normality. It should be noticed, however, that the alternatives considered by Tse (2000) are close to the specification he proposes to model time-variability in correlations. It is unclear which statistic would perform best for alternatives that are different from those envisaged by Tse (2000).

6.1.7 Illustration

Several papers have investigated the variability of the dependency parameter over time. Hamao, Masulis, and Ng (1990), Susmel and Engle (1994), and Bekaert and Harvey (1995) measured the interdependence of returns and volatilities across stock markets. More specifically, Longin and Solnik (1995) tested the hypothesis of a constant conditional correlation between a large number of stock markets. They found that correlation generally increases in periods of high volatility of the U.S. market. Recent contributions by Kroner and Ng (1998), as well as Engle and Sheppard (2001) develop GARCH models capable of estimating and testing hypotheses of time-varying covariance matrices. Ang and Chen (2002) document that dependency between U.S. stocks and the aggregate U.S. market increases more during downside movements than during upside movements.

To illustrate some properties of the multivariate GARCH models described above, we estimate several specifications for two pairs of time series: the SP500 and DAX daily returns, and the SP500 and FT-SE returns, over the period from 1980 to 2004. We begin with first-generation models. We estimate several versions of the bivariate BEKK model, the full, diagonal, and scalar versions. Parameter estimates and log-likelihoods of these models are reported in Table 6.1. Parameter estimates indicate that, in the full BEKK model, cross-effects (α_{12}, α_{21}, β_{12} and β_{21}) are in general rather small and insignificant. This provides a rationale for the diagonal BEKK model. Now all parameters are significant. As already highlighted for the univariate GARCH models, the sum of parameters (here, $\alpha_{11}^2 + \beta_{11}^2$ and $\alpha_{22}^2 + \beta_{22}^2$, due to the structure of the model) is close to (yet smaller than) one. In addition, we notice that α_{11} and α_{22} on one hand and β_{11} and β_{22} on the other hand are rather close, suggesting a further reduction of the number of parameters, assuming $\alpha_{11} = \alpha_{22} = \alpha$ and $\beta_{11} = \beta_{22} = \beta$. The so-called scalar BEKK model assumes that the variances and covariances have the same structure and the same parameters α and β

$$\sigma_{ij,t} = \omega_{ij} + \alpha^2 \varepsilon_{i,t-1} \varepsilon_{j,t-1} + \beta^2 \sigma_{ij,t-1}.$$

Now, we compare the dynamics of correlations given by first-generation models with those obtained with models specifically designed to capture time-varying correlation, namely the Engle's (2002) DCC model and the Tse and Tsui's (2002) TVC model. Table 6.2 reports parameter estimates corresponding to the CCC, DCC, and TVC models. First, they are estimated using the two-step approach proposed in Section 6.1.3. Therefore, the parameter estimates for the SP500 are obviously the same whatever the second return of the pair. Second, the parameters α and β in this table can be interpreted as the square of the parameters in the previous table. The dynamics of variances are very close to those obtained with BEKK models. The difference relies in the estimation of the covariances / correlations. In the BEKK models, covariances are rather smoothed, whereas correlations are very erratic (see Figures 6.1 and 6.3). With the CCC model, the correlation parameter is constant by definition. In contrast, the DCC and TVC models provide very similar and smoothed dynamics for the correlation parameter.

Figures 6.1 and 6.3 display the conditional correlation between the SP500 and the DAX, and between the SP500 and the FT-SE, respectively, as implied by the full BEKK model. As it appears clearly, these conditional correlations are very erratic. Figures 6.2 and 6.4 display the conditional correlation obtained with the CCC, DCC, and TVC models. First, we notice that the conditional correlations are now much smoother than those obtained with BEKK models. Second, we observe that the correlations implied by the DCC and TVC models are barely distinguishable from each other. Therefore, in many empirical applications, the direct modeling of the conditional correlation (through a DCC or a TVC model) is likely to provide a rather sensible estimate of the evolution in the correlation parameter.

The multivariate GARCH approach described so far assumes that the joint distribution of innovations is normal. Obviously, it is not likely to be the case in practice, because the univariate distribution of returns has generally been found to be non-normal. We therefore turn now to the modeling of the joint distribution. More precisely, we are interested in identifying some multivariate distributions that may help capturing the asymmetry and fat-tailedness of the return distribution.

Table 6.1. *Parameter estimates of various specifications of the BEKK model*

	SP500–DAX		SP500–FT-SE	
	Estimate	Std error	Estimate	Std error
Full BEKK model				
ω_{11}	0.0745	(0.0185)	0.0602	(0.0428)
ω_{12}	−0.0081	(0.0251)	−0.1099	(0.1213)
ω_{22}	0.1603	(0.0244)	0.1215	(0.0715)
a_{11}	0.1818	(0.0204)	0.1408	(0.0367)
a_{12}	−0.0319	(0.0628)	−0.2017	(0.0655)
a_{21}	0.0258	(0.0185)	0.0842	(0.0497)
a_{22}	0.2803	(0.0264)	0.3078	(0.0347)
b_{11}	0.9805	(0.0049)	0.9882	(0.0089)
b_{12}	0.0105	(0.0097)	0.0531	(0.0230)
b_{21}	−0.0059	(0.0045)	−0.0223	(0.0187)
b_{22}	0.9524	(0.0079)	0.9109	(0.0298)
log-lik.	−18113.3	–	−15983.0	–
Diagonal BEKK model				
ω_{11}	0.0863	(0.0139)	0.0796	(0.0144)
ω_{12}	0.028	(0.0111)	0.0395	(0.0200)
ω_{22}	0.1505	(0.0229)	0.1170	(0.0248)
a_{11}	0.1996	(0.0168)	0.2063	(0.0234)
a_{22}	0.2620	(0.0223)	0.2381	(0.0490)
b_{11}	0.9761	(0.0040)	0.9758	(0.0051)
b_{22}	0.9582	(0.0062)	0.9617	(0.0171)
log-lik.	−18120.3	–	−16100.2	–
Scalar BEKK model				
ω_{11}	0.1083	(0.0124)	0.0982	(0.0187)
ω_{12}	0.0215	(0.0089)	0.026	(0.0064)
ω_{22}	0.134	(0.0193)	0.0911	(0.0116)
a	0.2326	(0.0159)	0.2186	(0.0177)
b	0.9669	(0.0043)	0.9711	(0.0056)
log-lik.	−18131.1	–	−16106.3	–

Table 6.2. *Parameter estimates of the CCC, DCC, and TVC models*

	SP500–DAX		SP500–FT-SE	
	Estimate	Std error	Estimate	Std error
CCC model				
ω_1	0.0066	(0.0025)	0.0066	(0.0025)
a_1	0.0459	(0.0082)	0.0459	(0.0082)
b_1	0.9480	(0.0094)	0.9480	(0.0094)
ω_2	0.0296	(0.0113)	0.0201	(0.0059)
a_2	0.1019	(0.0221)	0.0894	(0.0135)
b_2	0.8834	(0.0213)	0.8875	(0.0168)
ρ	0.2576	(0.0174)	0.3058	(0.0258)
log-lik.	−18144.5	–	−16047.1	–
DCC model				
ω_1	0.0066	(0.0025)	0.0066	(0.0025)
a_1	0.0459	(0.0082)	0.0459	(0.0082)
b_1	0.9480	(0.0094)	0.9480	(0.0094)
ω_2	0.0296	(0.0113)	0.0201	(0.0058)
a_2	0.1019	(0.0221)	0.0894	(0.0135)
b_2	0.8834	(0.0213)	0.8875	(0.0167)
δ_1	0.0079	(0.0026)	0.0073	(0.0030)
δ_2	0.9909	(0.0032)	0.9867	(0.0055)
log-lik.	−18027.1	–	−16030.1	–
TVC model				
ω_1	0.0066	(0.0025)	0.0066	(0.0025)
a_1	0.0459	(0.0082)	0.0459	(0.0082)
b_1	0.9480	(0.0094)	0.9480	(0.0094)
ω_2	0.0296	(0.0113)	0.0201	(0.0058)
a_2	0.1019	(0.0221)	0.0894	(0.0135)
b_2	0.8834	(0.0213)	0.8875	(0.0168)
θ_1	0.0084	(0.0026)	0.0068	(0.0031)
θ_2	0.9909	(0.0029)	0.9896	(0.0055)
log-lik.	−18028.2	–	−16022.1	–

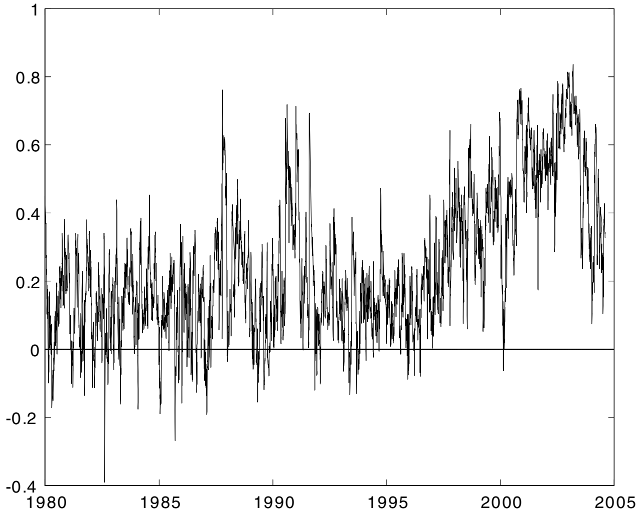

Fig. 6.1. *SP500–DAX. Conditional correlation implied by the full BEKK model.*

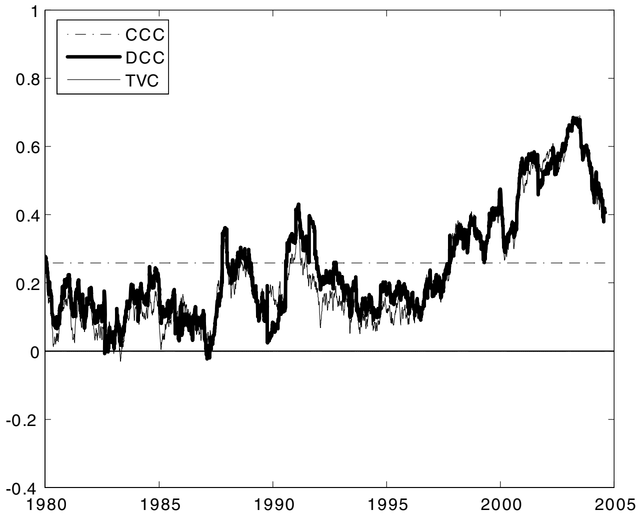

Fig. 6.2. *SP500–DAX. Conditional correlation implied by the CCC, DCC and TVC models.*

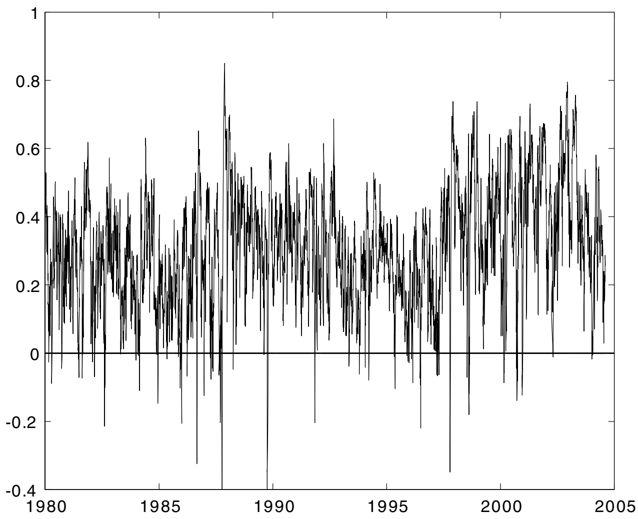

Fig. 6.3. *SP500–FT-SE. Conditional correlation implied by the full BEKK model.*

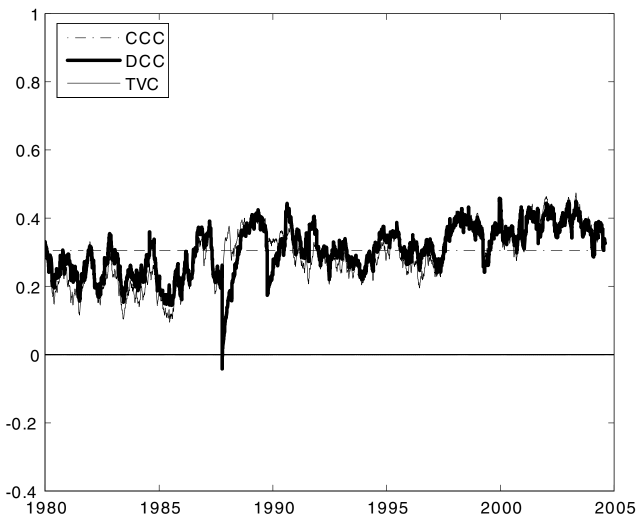

Fig. 6.4. *SP500–FT-SE. Conditional correlation implied by the CCC, DCC, and TVC models.*

6.2 Modeling the multivariate distribution

We now assume that the dynamic of the time-series $x_t = (x_{1,t}, \cdots, x_{n,t})'$ is given by the following multivariate dynamic regression model, with time-varying means, variances, and covariances

$$x_t = \mu_t + \varepsilon_t, \tag{6.10}$$

$$\varepsilon_t = \Sigma_t^{1/2} z_t, \tag{6.11}$$

$$\mu_t = E[x_t | \mathcal{F}_{t-1}] = \mu(\theta, \mathcal{F}_{t-1}), \tag{6.12}$$

$$\Sigma_t = V[x_t | \mathcal{F}_{t-1}] = \Sigma(\theta, \mathcal{F}_{t-1}), \tag{6.13}$$

$$z_t \sim g(z_t|\eta). \tag{6.14}$$

The dynamics of the $(n, 1)$ conditional mean vector μ_t is given by (6.12), whereas the dynamics of the (n, n) conditional variance matrix Σ_t is given by (6.13). Last, standardized residuals z_t, defined as $\Sigma_t^{-1/2}(x_t - \mu_t)$, is the *iid* random vector of dimension $(n, 1)$ with a zero mean and identity variance matrix. There are several possibilities to obtain $\Sigma_t^{1/2}$. The first and probably most common one is the Cholesky decomposition, where $\Sigma_t^{1/2}$ is a lower triangular matrix. Another possibility is based on the eigenvector decomposition, $\Sigma_t = \Omega_t D_t \Omega_t$, where Ω_t is the matrix of eigenvectors, standardized to unit length, and D_t is the diagonal matrix of eigenvalues. By construction, we have $\Omega_t \Omega_t = I_t$. Then, $\Sigma_t^{1/2} = \Omega_t D_t^{1/2}$, where $D_t^{1/2}$ is the matrix whose diagonal elements are the square roots of the eigenvalues. Vector θ includes all parameters of the conditional mean and variance equations. As specified in (6.14), the conditional distribution is g, with shape parameters η.

As in the univariate case, the choice of the conditional distribution $g(.)$ is crucial. Engle and González-Rivera (1991) and Newey and Steigerwald (1997) have shown that the following results hold: (i) Under the assumption of a correct specification of the conditional mean and variance matrix, the ML estimation, assuming z_t to be *iid* with a Gaussian distribution, provides consistent estimators, even when the Gaussian assumption does not hold. (ii) The ML estimator relying on a Gaussian distribution is inefficient, however, with the degree of inefficiency increasing with the degree of departure from normality. (iii) The ML estimation, assuming z_t to be *iid* with a non-Gaussian distribution, provides more efficient estimators than the Gaussian ML, when the assumption made on the innovation process holds. (iv) When the assumption made on the innovation process does not hold, the ML estimation relying on a non-Gaussian distribution provides inconsistent estimators.

Another difficulty in the multivariate case comes from the way dependency between variables is introduced. In the Gaussian framework, dependency is introduced through the covariance matrix. The multivariate Gaussian distribution with zero mean and identity covariance matrix is defined as

$$g(z_t) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2} z_t' z_t\right).$$

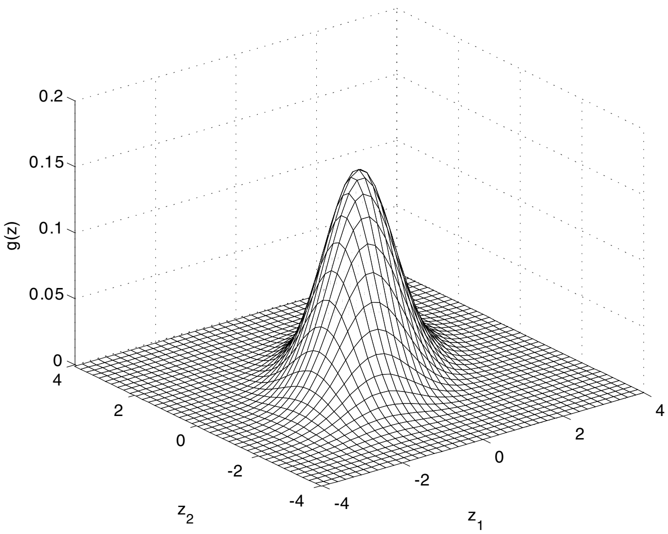

Fig. 6.5. *Pdf of the bivariate Gaussian distribution, with* $\rho = 0.5$.

Therefore, the Gaussian distribution of the observations x_t with conditional mean μ_t and covariance matrix Σ_t is given by

$$f(x_t) = \frac{1}{(2\pi)^{n/2}} \frac{1}{|\Sigma_t|^{1/2}} \exp\left(-\frac{1}{2}(x_t - \mu_t)' \Sigma_t^{-1} (x_t - \mu_t)\right).$$

Given that the Gaussian distribution only depends on μ_t and Σ_t, the natural measure of dependency is Σ_t. For the general case, dependency may be introduced in various ways.

Figure 6.5 shows the *pdf* of the bivariate Gaussian distribution. The two variables are supposed to have a zero mean and unit variance. The correlation is 0.5. The contour of the distribution is displayed in Figure 6.6.

One of the strong limitations of the multiviarate Gaussian distribution is that it does not allow any dependence between the two variables in the tails (for more details, see Section 6.3), a feature that has been found to play a central role in the joint modeling of asset returns, especially in the context of VaR applications. In addition, as seen in the univariate context, the Gaussian distribution is symmetric, so that it is unable to capture the observed asymmetry in the return distribution.

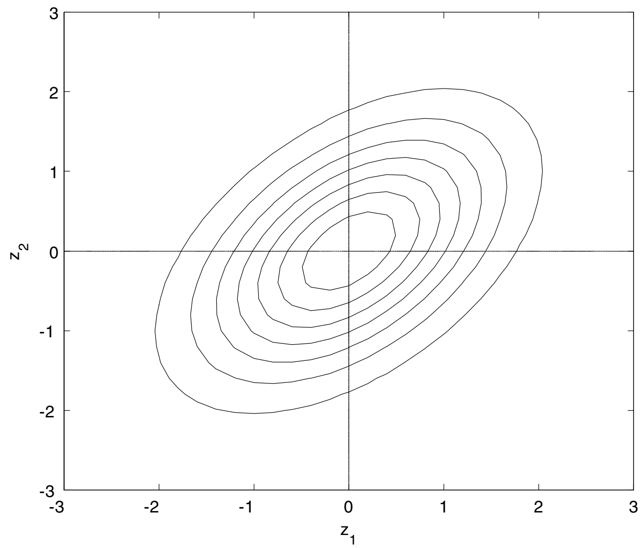

Fig. 6.6. *Contour of the pdf of the bivariate Gaussian distribution, with* $\rho = 0.5$.

6.2.1 Standard multivariate distributions

Multivariate Student *t* distribution

The multivariate Student t distribution is a distribution that naturally generates dependency in the tails. Yet, this extension to the Student t distribution depends on the way in which dependency between innovations is introduced. Assume that the n innovations are expressed as $Z_i = U_i/\sqrt{S_i^2 \nu_i}$, $i = 1, \cdots, n$, where the U_is and S_is are all mutually independent, with U_i having a univariate Gaussian distribution, and S_i^2 a χ^2 distribution with ν_i degrees of freedom. In this case, because the Z_is are independent, the joint density function is simply the product of the individual Student t density functions.

Now, there are basically two ways to introduce dependence into the joint distribution of returns X:

First, dependence can be introduced in the innovation process Z, by assuming that the χ^2-variables S_i^2 that appear in the definition of the Z_is are the same for each component, with ν degrees of freedom, so that $Z_i = U_i/\sqrt{S^2\nu}$, $i = 1, \cdots, n$. We may also assume that $U_1, \cdots, U_n$ have a joint multivariate distribution with covariance matrix R. Since the U_is are normalized, $R = \{\rho_{ij}\}$ is also the correlation matrix of $(U_1, \cdots, U_n)$. In this case, the multivariate Student t distribution, with dependent components and degree-of-freedom parameter ν, is

$$g(z|\nu) = \frac{\Gamma\left(\frac{\nu+n}{2}\right)}{(\pi(\nu-2))^{n/2}\,\Gamma\left(\frac{\nu}{2}\right)|R|^{1/2}}\left(1+\frac{1}{\nu-2}z'R^{-1}z\right)^{-\frac{\nu+n}{2}}.$$

When the U_is are mutually independent ($R = I_n$), the distribution reduces to

$$g(z|\nu) = \frac{\Gamma\left(\frac{\nu+n}{2}\right)}{(\pi(\nu-2))^{n/2}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{z'z}{\nu-2}\right)^{-\frac{\nu+n}{2}}. \tag{6.15}$$

Each marginal distribution is symmetric, with zero mean, unit variance (this is obtained by using the terms $(\nu-2)$ rather than ν in 6.15), zero skewness, and a kurtosis equal to $3\,(\nu-2)\,/\,(\nu-4)$. When the Student t distribution is introduced to capture fat-tailedness of the series, however, there is neither theoretical nor empirical reason for the degree-of-freedom parameters to be equal for all series. Assuming a unique degree of freedom ν for all components would therefore introduce an artificial dependence in the system.

Alternatively, the Gaussian variables U_i and the χ^2-distributed variables S_i^2 can be assumed to be independent components, with dependence introduced for the Xs via some linear structure involving a Choleski decomposition of the covariance matrix. If the degrees of freedom ν_i are different for each variable, the Student t distribution with independent components is written as

$$g(z|\nu_1,\cdots,\nu_n) = \prod_{i=1}^{n}\frac{\Gamma\left(\frac{\nu_i+1}{2}\right)}{\sqrt{\pi(\nu_i-2)}\Gamma\left(\frac{\nu_i}{2}\right)}\left(1+\frac{z_i^2}{\nu_i-2}\right)^{-\frac{\nu_i+1}{2}}, \tag{6.16}$$

while the multivariate return process X is assumed to be correlated through a non-diagonal covariance matrix Σ, such that the distribution of X, denoted $f(x|\nu_1,\cdots,\nu_n)$, is deduced from the relation $X = \Sigma^{1/2}Z$.

Now, each marginal distribution is symmetric, with zero mean, unit variance, zero skewness, and a kurtosis equal to $3\,(\nu_i-2)\,/\,(\nu_i-4)$. Clearly, in empirical applications, the equality of the ν_is can be explicitly tested in a second step. An undesirable property of this distribution is that it is not a member of the elliptical familly described below (Section 6.2.1). As a consequence, as described in Section 6.2.4, in a multivariate GARCH model with skewed Student t innovations, the estimation of the GARCH parameters and of the shape parameters cannot be performed separately.

Figures 6.7 and 6.9 shows the *pdf* of the two types of bivariate Student t distributions. The first one corresponds to the distribution with dependent components (equation (6.15)) with degree of freedom $\nu = 6$ and correlation between the z_is equal to $\rho_{12} = 0.5$. The second is the distribution with independent components (equation (6.16)) with degrees of freedom $\nu_1 = \nu_2 = 6$ and correlation between the x_is equal to $\rho_{12} = 0.5$. In the two cases, the two variables are supposed to have a zero mean and unit variance. The contours of the two distributions are displayed in Figures 6.8 and 6.10 respectively.

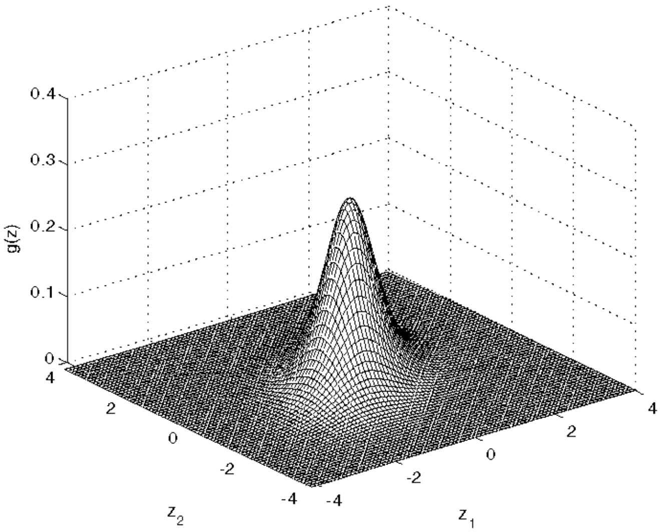

Fig. 6.7. *Pdf of the Student t distribution with dependent components and* $\rho_{12} = 0.5$.

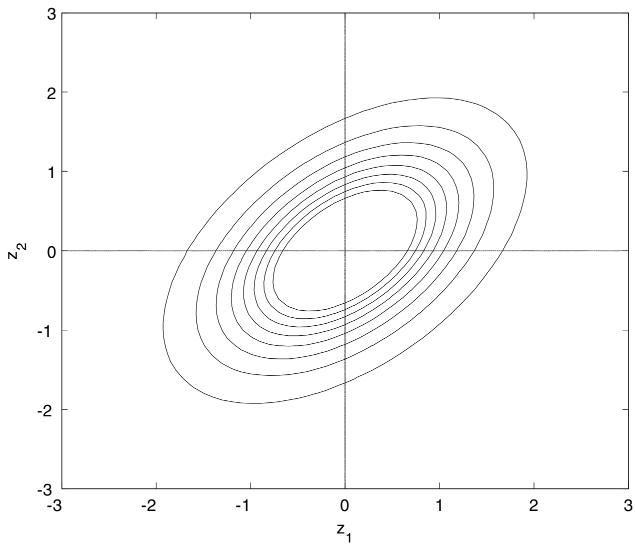

Fig. 6.8. *Contour of the pdf of the Student t distribution with dependent components and* $\rho_{12} = 0.5$.

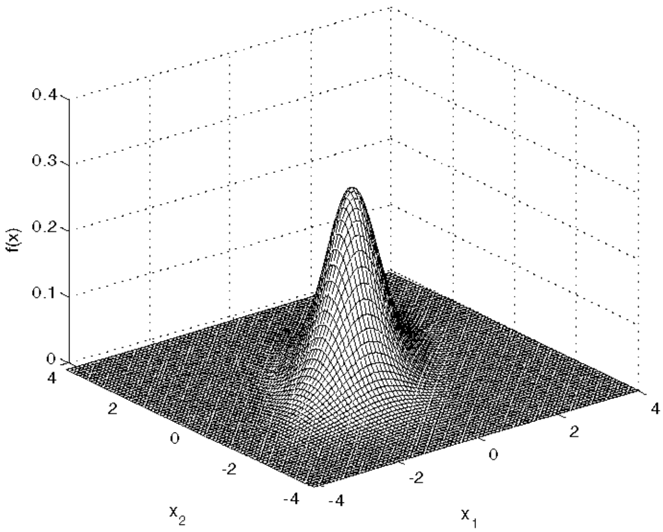

Fig. 6.9. *Pdf of the Student t distribution with independent components and* $\rho_{12} = 0.5$.

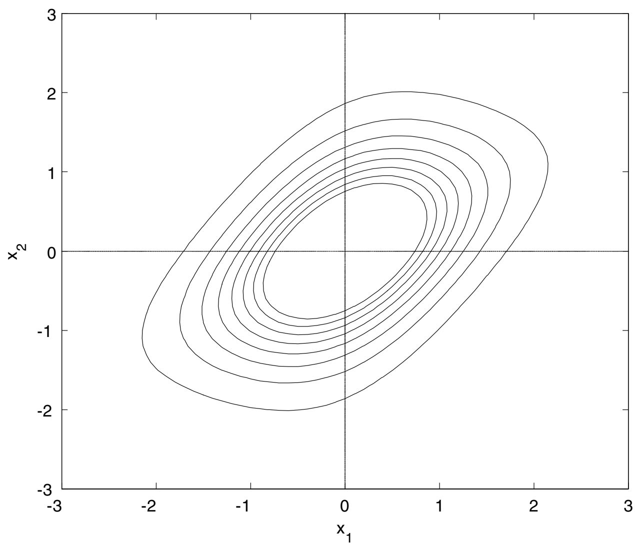

Fig. 6.10. *Contour of the pdf of the Student t distribution with independent components and* $\rho_{12} = 0.5$.

Multivariate Student t distributions are able to capture the fat-tailedness of the empirical distribution of asset returns. However, they are not designed to take asymmetry into account. Before turning to asymmetric distributions, we briefly describe the class of multivariate elliptical distributions, which includes the Gaussian and Student t distributions. Indeed, a series of papers have proposed a general class of multivariate distributions with asymmetry, based on elliptical distributions (see Section 6.2.2).

Elliptical distributions

A well-studied class of multivariate distributions is the class of elliptical distributions. For more detail, see Fang, Kotz, and Ng (1990). An n-dimensional vector Z is said to be elliptically distributed with location vector μ and (n,n) dispersion matrix Σ, if the density is

$$g\left(z\,|\mu,\Sigma\right) = |\Sigma|^{-1/2} f^{(n)}\left(\left(z-\mu\right)' \Sigma^{-1}\left(z-\mu\right)\right), \tag{6.17}$$

for some density generating function $f^{(n)}(u)$, $u \geq 0$, such that

$$\int_0^\infty u^{n/2-1} f^{(n)}(u)\,du = \frac{\Gamma(n/2)}{\pi^{n/2}},$$

so that $f^{(n)}$ is a spherical n-dimensional density. We denote this function $Z \sim El_n\left(\mu;\Sigma;f^{(n)}\right)$ with *pdf* $g_{f^{(n)}}(.)$ and *cdf* $G_{f^{(n)}}(.)$.

Elliptical distributions have several interesting properties. In particular, if $Z \sim El_n\left(\mu;\Sigma;f^{(n)}\right)$, then for any (k,n) matrix A with rank $k \leq n$ and any $(k,1)$ vector b, we have $AZ+b \sim El_k\left(A\mu+b;A\Sigma A';f^{(k)}\right)$. Another property is that if Z is partitioned as

$$Z = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim El_n\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}; \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}; f^{(n)}\right),$$

with Z_i an $(n_i,1)$ vector (with $n = n_1+n_2$), then Z_i is also elliptically distributed with $Z_i \sim El_n\left(\mu_i;\Sigma_i;f^{(n_i)}\right)$.

Instances of elliptical distribution are the multivariate Gaussian distribution, for which

$$f^{(n)}(u) = \frac{e^{-u/2}}{(2\pi)^{n/2}},$$

so that

$$g\left(z\,|\mu,\Sigma\right) = \frac{1}{(2\pi)^{n/2}}\,|\Sigma|^{-1/2} \exp\left(-\frac{1}{2}\left(z-\mu\right)' \Sigma^{-1}\left(z-\mu\right)\right).$$

Similarly, the multivariate Student t distribution is defined by

$$f^{(n)}(u) = \left(1+\frac{u}{\nu}\right)^{-(\nu+n)/2} \qquad \nu > 0,$$

so that

$$g(z\,|\mu,\Sigma) = \frac{\Gamma\left(\frac{\nu+n}{2}\right)}{(\pi\nu)^{n/2}\,\Gamma\left(\frac{\nu}{2}\right)}\,|\Sigma|^{-1/2}\left(1+\frac{(z-\mu)'\,\Sigma^{-1}(z-\mu)}{\nu}\right)^{-\frac{\nu+n}{2}}.$$

Notice that this specification slightly differs from (6.15). As explained above, (6.15) ensures that the variance of Z is by construction equal to one, while here we have $V[Z]=\nu/(\nu-2)$.

6.2.2 Skewed elliptical distribution

The multivariate skewed Gaussian distribution has been first studied by Azzalini and Dalla Valle (1996), Azzalini and Capitanio (1999), and extended to the Student t distribution by Azzalini and Capitanio (2003). Branco and Dey (2001) have introduced a general class of multivariate skew-elliptical distributions.

This approach builds on the univariate skewed Gaussian distribution, defined by Azzalini (1985). A random variable Z has a so-called *skewed Gaussian distribution* if its *pdf* is defined by

$$g(z) = 2\varphi(z)\,\Phi(\lambda z),$$

where $\varphi(z)$ is the standard Gaussian *pdf*, and $\Phi(z)$ is the corresponding *cdf*. Some intuition for this specification is provided below. When $\lambda=0$, $g(z)$ is the standard normal distribution. When $\lambda\to-\infty$ $(+\infty)$, we obtain a Gaussian distribution truncated from above (below) at zero. The parameter λ plays the role of shape parameter.

Skewed Gaussian distribution

This distribution has been extended in several ways. In particular, Azzalini and Dalla Valle (1996) have considered the case of the multivariate skewed Gaussian distribution. Branco and Dey (2001) have extended this approach to the elliptical distribution in a multivariate framework.

The multivariate extension of the skewed Gaussian distribution has been proposed by Azzalini and Dalla Valle (1996), and Azzalini and Capitanio (1999). If Z is a $(n,1)$ random vector with mean zero and, unit variances and correlation matrix R, its *pdf* is defined by

$$g(z) = 2\varphi_n(z|R)\,\Phi\left(\lambda' z\right),$$

where $\varphi_n(z|R)$ is the multivariate Gaussian *pdf* with covariance matrix R, and λ is a $(n,1)$ vector.

When the random vector Z is not assumed to have zero means and unit variances, we can define the skewed Gaussian distribution as follows. Let

$\Sigma = \{\sigma_{ij}\}$ be the (n, n) covariance matrix of Z, $D =$ diag$(\sigma_{11}, \cdots, \sigma_{nn})$ the matrix containing variances, and $R = D^{-1/2} \Sigma D^{-1/2}$ the associated correlation matrix. Let ξ be the $(n, 1)$ location parameter vector. Then, the random variable Z has a skew normal distribution if its *pdf* may be written as

$$g(z) = 2\varphi_n(z - \xi|\Sigma)\,\Phi\left(\lambda' D^{-1/2}(z - \xi)\right).$$

We then write $Z \sim \mathcal{SN}_n(\xi, \Sigma, \lambda)$, referring to ξ, Σ, and λ as the location, dispersion, and shape (or skewness) parameters, respectively.

Such a distribution can be derived in several ways, described by Azzalini and Capitanio (1999). A first way of generating a skewed Gaussian distribution is *conditioning*. Suppose that U_0 is a scalar random variable and U is an n-dimensional variable, such that the joint distribution is a multivariate Gaussian distribution

$$\begin{pmatrix} U_0 \\ U \end{pmatrix} \sim \mathcal{N}_{n+1}(0, \Sigma^*) \quad \text{with} \quad \Sigma^* = \begin{pmatrix} 1 & \delta' \\ \delta & R \end{pmatrix},$$

where Σ^* is a full-rank covariance matrix. Then, the distribution of $U\,|U_0 > 0$ is $\mathcal{SN}_n(0, R, \lambda)$, where λ is defined as

$$\lambda' = \frac{\delta' R^{-1}}{\left(1 - \delta' R^{-1}\delta\right)^{1/2}}.$$

Alternatively, the random variable defined by

$$Z = \begin{cases} U & \text{if } U_0 > 0, \\ -U & \text{if } U_0 < 0, \end{cases}$$

is also distributed as a skewed Gaussian variate.

The skewed Gaussian distribution can also be obtained by *transformation*. Suppose that

$$\begin{pmatrix} U_0 \\ U \end{pmatrix} \sim \mathcal{N}_{n+1}(0, \Sigma^*) \qquad \text{with} \qquad \Sigma^* = \begin{pmatrix} 1 & 0 \\ 0 & \Psi \end{pmatrix},$$

where Ψ is a full-rank covariance matrix. Also define the $(n, 1)$ vector

$$Z_j = \delta_j\,|U_0| + \sqrt{1 - \delta_j^2}\,U_j',$$

with weights given by $-1 < \delta_j < 1$ for $j = 1, \cdots, n$. Then, $(Z_1, \cdots, Z_n)$ has an n-dimensional skewed Gaussian distribution with parameters that are functions of the δ's and Ψ.

Skewed elliptical distribution

Branco and Dey (2001), Sahu, Dey, and Branco (2003), and Azzalini and Capitanio (2003) have proposed a general class of multivariate skewed elliptical distributions, based on the approach developed by Azzalini and Dalla Valle (1996) and Azzalini and Capitanio (1999). The standard multivariate elliptical distribution is defined in Section 6.2.1.

Suppose that U_0 is a scalar random variable and U is a n-dimensional variable, such that the joint distribution is a multivariate elliptical distribution

$$U^* = \begin{pmatrix} U_0 \\ U \end{pmatrix} \rightarrow El_{n+1}\left(\mu^*, \Sigma^*; f^{(n+1)}\right),$$

with

$$\mu^* = \begin{pmatrix} 0 \\ \mu \end{pmatrix} \quad \text{and} \quad \Sigma^* = \begin{pmatrix} 1 & \delta' \\ \delta & \Sigma \end{pmatrix},$$

where Σ^* is a full-rank covariance matrix. Then, $Y = U \,|U_0 > 0$ has a skewed elliptical distribution, such that $Y \sim SE_n\left(\mu, \Sigma, \delta; f^{(n+1)}\right)$, with location μ, scale Σ, characteristic function φ, and skewness parameter δ. The *pdf* of Y is denoted

$$g_Y(y) = 2g_{f^{(n)}}(y)\, G_{f_{q(y)}}\left(\lambda'(y-\mu)\right),$$

where $g_{f^{(n)}}(.)$ is the *pdf* of $El_n\left(\mu, \Sigma; f^{(n)}\right)$ with generator function $f^{(n)}(.)$ and $G_{f_{q(z)}}(.)$ is the *cdf* of a univariate elliptical distribution $El_1\left(0, 1; f_{q(z)}\right)$ with $f_{q(z)}$ as the generator function. In addition, λ is defined as

$$\lambda' = \frac{\delta' \Sigma^{-1}}{\left(1 - \delta' \Sigma^{-1} \delta\right)^{1/2}},$$

and

$$f_{q(y)}(u) = \frac{f^{(n+1)}(u + q(y))}{f^{(n)}(q(y))},$$

with $q(y) = (y-\mu)' \Sigma^{-1} (y-\mu)$.

A special case is the skewed Student t distribution. The *pdf* of Y can be written as

$$g_Y(y) = 2t_n(y, \nu)\, T_1\left(\delta' D^{-1/2}(y-\mu)\left(\frac{\nu+n}{q(y)+\nu}\right)^{1/2}; \nu+n\right),$$

where

$$t_n(y, \nu) = \frac{1}{|\Sigma|^{1/2}} g_n(q(y); \nu)$$

is the density function of a n-dimensional Student t variate with ν degrees of freedom and $T_1(x; \nu+n)$ denotes the univariate Student t distribution function with $\nu+n$ degrees of freedom.

6.2.3 Skewed Student t distribution

There are other ways to generate multivariate skewed distributions. For instance, Bauwens and Laurent (2005) have generalized the approach proposed by Fernández and Steel (1998) in the univariate context, presented in Section 5.2.4. The method consists in changing the shape of the distribution at each side of the mode. If $h(y)$ is a symmetric multivariate distribution with zero mean and identity covariance matrix, the new $(n, 1)$ random vector Z has an asymmetric distribution with the same mode as $h(y)$

$$g(z \,|\xi) = \left(\prod_{i=1}^{n} \frac{2}{\xi_i + \frac{1}{\xi_i}}\right) h(y),$$

where $y = (y_1, \cdots, y_n)'$, with

$$y_i = \begin{cases} z_i^* \xi_i & \text{if } z_i < 0, \\ z_i^* / \xi_i & \text{if } z_i \geq 0. \end{cases} \tag{6.18}$$

Such a specification has several interesting properties. First, the marginal densities have the same patterns as the ones defined in the univariate case by Fernández and Steel (1998) (see Section 5.2.4). Second, ξ_i is a measure of the asymmetry of the marginal density of Z_i^*. Third, the rth moment of Z_i^* is given by

$$E\left[(Z_i^*)^r \,|\xi_i\right] = \frac{\xi_i^{r+1} + \frac{(-1)^r}{\xi_i^{r+1}}}{\xi_i + \frac{1}{\xi_i}} 2E\left[(Z_i^*)^r \,|Z_i > 0\right].$$

Finally, the components of Z^* are uncorrelated, because those of Y are uncorrelated by assumption. Consequently, if a is the vector of means and b the vector of standard deviations of Z^*, then Z^* can be standardized by the transformation $Z = (Z^* - a) \div b$, where $\div$ denotes the element-wise division.

This general formulation has been specialized by Bauwens and Laurent (2005) and Jondeau and Rockinger (2005) to the case of the multivariate skewed Student t distribution. This distribution is therefore a multivariate extension of the distribution proposed by Hansen (1994), Fernández and Steel (1998), and Jondeau and Rockinger (2003a), and is presented in Section 5.2.4.[9]

The case with dependent components

Following the two approaches described in Section 6.2.1, we now define the multivariate distribution assuming dependent or independent components. We begin with the specification with dependent components, which is given as

[9] As discussed in Section 5.2, the skewed t distribution proposed by Fernández and Steel (1998) is directly related to the distribution proposed by Hansen (1994) through a change of notation of the asymmetry parameter. In this section, we use the notation of Fernández and Steel (1998) for the asymmetry parameter.

$$g(z|\nu,\xi_1,\cdots,\xi_n) = c\left(\prod_{i=1}^{n}\frac{2b_i}{\xi_i+\frac{1}{\xi_i}}\right)\left(1+\frac{y'y}{\nu-2}\right)^{-\frac{\nu+n}{2}}, \tag{6.19}$$

where $y=(y_1,\cdots,y_n)'$ is defined in (6.18) with $z_i^*=(b_i z_i+a_i)$, and

$$c = \frac{\Gamma\left(\frac{\nu+n}{2}\right)}{\left(\pi(\nu-2)\right)^{n/2}\Gamma\left(\frac{\nu}{2}\right)},$$
$$a_i = \frac{\Gamma\left(\frac{\nu-1}{2}\right)\sqrt{\nu-2}}{\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)}\left(\xi_i-\frac{1}{\xi_i}\right),$$
$$b_i^2 = \left(\xi_i^2+\frac{1}{\xi_i^2}-1\right)-a_i^2.$$

Notice that the standardized variables $Z_i=\left(Z_i^*-a_i\right)/b_i$ have zero mean and unit variance. This formulation is an extension of the multivariate Student t distribution with dependent components, defined in (6.15).

The case with independent components

We now describe the asymmetric extension of the multivariate Student t distribution with independent components, defined in (6.16). It is expressed as

$$g(z|\nu_1,\cdots,\nu_n,\xi_1,\cdots,\xi_n) = \prod_{i=1}^{n}\frac{2b_i c_i}{\xi_i+\frac{1}{\xi_i}}\left(1+\frac{y_i^2}{\nu_i-2}\right)^{-\frac{\nu_i+1}{2}}, \tag{6.20}$$

where y_i is given by (6.18) with $z_i^*=(b_i z_i+a_i)$, and

$$c_i = \frac{\Gamma\left(\frac{\nu_i+1}{2}\right)}{\sqrt{\pi(\nu_i-2)}\Gamma\left(\frac{\nu_i}{2}\right)},$$
$$a_i = \frac{\Gamma\left(\frac{\nu_i-1}{2}\right)\sqrt{\nu_i-2}}{\sqrt{\pi}\Gamma\left(\frac{\nu_i}{2}\right)}\left(\xi_i-\frac{1}{\xi_i}\right),$$
$$b_i^2 = \left(\xi_i^2+\frac{1}{\xi_i^2}-1\right)-a_i^2.$$

It is worth emphasizing that, although the standardized variables $Z_i = \left(Z_i^*-a_i\right)/b_i$ are independent by construction, returns can be modeled as dependent through their covariance matrix. Assume that returns are defined as $X=\Sigma^{1/2}Z^*$. The moments of returns can, therefore, be computed in the following way. We denote $\Sigma^{1/2}=\left(\omega_{ij}\right)_{i,j=1,\cdots,n}$ the Choleski decomposition of the covariance matrix of returns, so that $X_i=\sum_{r=1}^{n}\omega_{ir}Z_r^*$.[10] In addition,

[10] For ease of exposition, we assume that $E\left[X\right]=\mu=0$ and we intentionally omit the dependence of the covariance matrix Σ and hence of the ω_{ij}s with respect to time.

we recall that because innovations Z^* are assumed to be independent, their first moments are given by $\mu_i^{(1)} = 0$, $\mu_i^{(2)} = 1$ and (see Section 5.2.4)

$$\mu_i^{(3)} = \left(M_{i,3} - 3a_i M_{i,2} + 2a_i^3\right)/b_i^3,$$
$$\mu_i^{(4)} = \left(M_{i,4} - 4a_i M_{i,3} + 6a_i^2 M_{i,2} - 3a_i^4\right)/b_i^4.$$

The (i,j) component of the covariance matrix of returns is given by

$$E\left[X_i X_j\right] = \sum_{r=1}^{n} \omega_{ir}\omega_{jr}.$$

The (i,j,k) component of the third central moments of returns is

$$E\left[X_i X_j X_k\right] = \sum_{r=1}^{n} \omega_{ir}\omega_{jr}\omega_{kr}\mu_r^{(3)}.$$

Finally, the (i,j,k,l) component of the fourth central moments of returns is

$$E\left[X_i X_j X_k X_l\right] = \sum_{r=1}^{n} \omega_{ir}\omega_{jr}\omega_{kr}\omega_{lr}\mu_r^{(4)} + \sum_{r=1}^{n}\sum_{\substack{s=1 \\ s\neq r}}^{n} \psi_{rs}\mu_r^{(2)}\mu_s^{(2)},$$

where $\psi_{rs} = \omega_{ir}\omega_{jr}\omega_{ks}\omega_{ls} + \omega_{ir}\omega_{js}\omega_{kr}\omega_{ls} + \omega_{ir}\omega_{js}\omega_{kr}\omega_{ls}$.

Analytical expressions of moments are quite cumbersome to derive, yet their numerical computation is very fast, because only matrix manipulations are required. For an n-variable system, the dimension of the covariance matrix is (n,n), but only $n(n+1)/2$ elements have to be computed. Similarly, the co-skewness matrix has dimension (n,n,n), but only $n(n+1)(n+2)/6$ elements have to be computed. Finally, the co-kurtosis matrix has dimension (n,n,n,n), but only $n(n+1)(n+2)(n+3)/24$ elements have to be computed.[11]

As it clearly appears from these expressions, co-skewness between excess returns depends on individual skewness of innovations and correlations between returns (through the ω_{ij}s). Co-kurtosis between excess returns depends on individual kurtosis and volatilities of innovations and correlations between returns. Such a time-variability is likely to have two sources. On one hand, the covariance matrix between excess returns is time-varying, so that the ω_{ij} elements themselves are time-varying. On the other hand, skewness and kurtosis of innovations may be time-varying, for instance, if the degree-of-freedom parameter (ν) or the asymmetry parameter (ξ) vary over time. In our illustration (Section 6.2.6), we do not consider the latter case, but persistence in conditional higher moments has been found, in a univariate context, by Jondeau and Rockinger (2003a).

[11] For $n = 5$, one has 15 different elements for the covariance matrix, 35 elements for the co-skewness matrix, and 70 elements for the co-kurtosis matrix (whereas these matrices have 25, 125, and 625 elements, respectively).

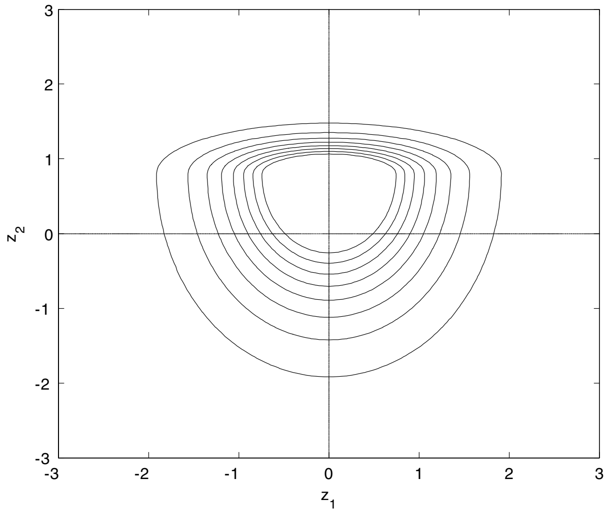

Fig. 6.11. *Contour plot of the bivariate skewed Student t distribution with dependent components and* $\rho = 0$.

Comparison

Figures 6.11 to 6.13 show the *pdf* and the contour plot of the two types of bivariate skewed Student t distributions (6.19) and (6.20). The two variables have a zero mean and unit variance. The degree-of-freedom parameter ν is equal to 6 in all cases. The asymmetry parameters are equal to $\xi = (1; 0.5)$ so that the first component is symmetric but the second one is markedly asymmetric. In Figure 6.11, the correlation coefficient ρ_{12} is assumed to be 0, whereas in Figures 6.12 and 6.13, it is equal to 0.5.

6.2.4 Estimation

The estimation of the multivariate model with non-Gaussian multivariate distribution, given by (6.10)–(6.14), raises some additional difficulties as compared with the Gaussian case. Unknown parameters are now θ (for the dynamics of the conditional mean and conditional covariance matrix) and η (for the conditional distribution). Define $\xi = \left(\theta', \eta'\right)'$ the vector of unknown parameters. The ML estimator of the parameter vector ξ is obtained by maximizing

$$L_T\left(\xi | \underline{x}_T\right) = \sum_{t=1}^{T} \ell_t\left(\xi\right),$$

where

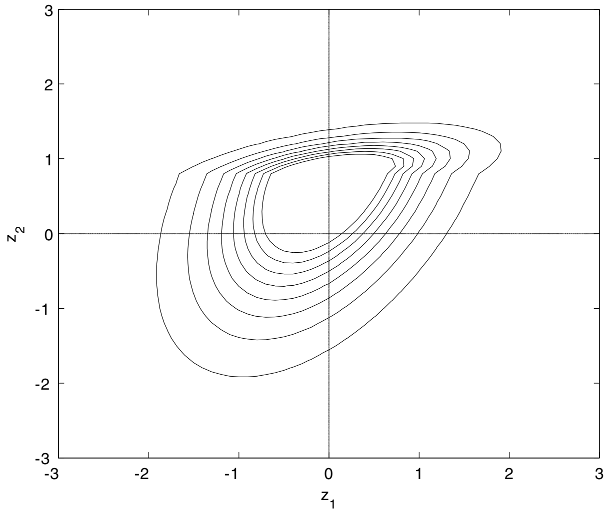

Fig. 6.12. *Contour plot of the bivariate skewed Student t distribution with dependent components and* $\rho = 0.5$.

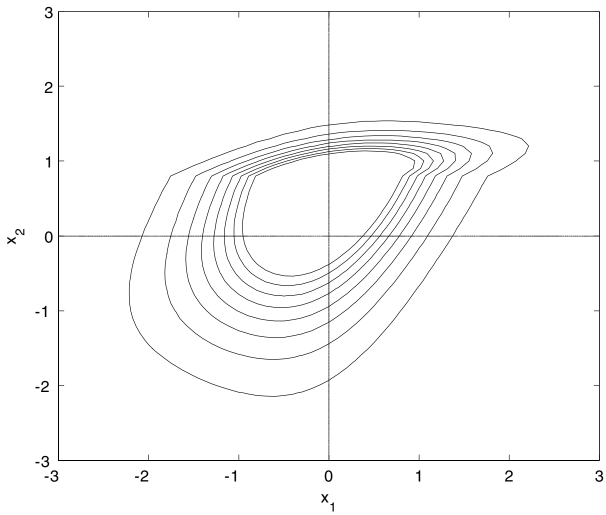

Fig. 6.13. *Contour plot of the bivariate skewed Student t distribution with independent components and* $\rho = 0.5$.

$$\ell_t(\xi) = -\frac{1}{2}\log|\Sigma_t(\theta)| + g\left(\Sigma_t(\theta)^{-1/2}(x_t - \mu_t(\theta))|\xi\right).$$

In general, the maximization of the log-likelihood has to be performed in one step, because the two sets of parameters θ and η interact.

A noticeable exception is the class of elliptical distributions. In this case, the distribution only involves $(x_t - \mu_t(\theta))' \Sigma_t^{-1}(\theta)(x_t - \mu_t(\theta))$, which can be rewritten as $z_t(\theta)'z_t(\theta)$ with $z_t(\theta) = \Sigma_t(\theta)^{-1/2}(x_t - \mu_t(\theta))$. Therefore, a two-step estimation can be performed, in which the two sets of parameters are estimated separately. The parameters θ are estimated using QML, assuming normality of the innovation process. Therefore, it is obtained by maximizing

$$L_T(\theta|\underline{x}_T) = \sum_{t=1}^{T}\ell_t(\theta),$$

where

$$\ell_t(\theta) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\left(\log|\Sigma_t(\theta)| + (x_t - \mu_t(\theta))' \Sigma_t^{-1}(\theta)(x_t - \mu_t(\theta))\right).$$

This procedure yields a consistent estimator of θ even if the true distribution is not normal. This consistency result has been proven by Bollerslev and Wooldridge (1992) and Jeantheau (1998). Once the parameter vector $\hat{\theta}$ has been estimated, the standardized residuals are obtained as $\hat{z}_t(\hat{\theta}) = \hat{\Sigma}_t(\hat{\theta})^{-1/2}\left(x_t - \hat{\mu}_t(\hat{\theta})\right)$. Then, parameters η are estimated by maximizing the log-likelihood function, assuming now that innovations are drawn from the multivariate distribution $g(z_t|\eta)$

$$\tilde{L}_T\left(\hat{\theta}, \eta|\underline{x}_T\right) = \sum_{t=1}^{T}\hat{\ell}_t\left(\hat{\theta}, \eta\right),$$

where

$$\hat{\ell}_t\left(\hat{\theta}, \eta\right) = -\frac{1}{2}\log\left|\Sigma_t\left(\hat{\theta}\right)\right| - \frac{1}{2}\sum_{t=1}^{T}\log\left(g(\hat{z}_t\left(\hat{\theta}\right))|\eta\right).$$

Such an approach can be adopted for the estimation of GARCH models with Gaussian or Student t innovations. However, it would not apply for the skewed Student t distribution, because it does not belong to the elliptical family. In such case, a joint estimation is required.[12] See Section 6.1.4 for further details on the estimation of the covariance matrix of $\hat{\theta}$.

Hafner and Rombouts (2003) have recently proposed an extension of the semi-parametric estimation technique of Engle and González-Rivera (1991) to

[12] Notice that, in the case of the Student t distribution, parameters θ and η can be estimated separately. However, if we estimate a DCC model, it is not possible to estimate the parameters pertaining to the variances and to the correlations separately anymore. The reason is that the log-likelihood cannot be broken down into variances and correlations components.

multivariate GARCH models. As in the univariate framework (Section 8.3), this approach is based on the following steps. The first step consists in estimating the model by QML, providing an estimate of the parameters θ, say $\tilde{\theta}$. Then, the fitted residuals $\hat{\varepsilon}_t$ and the fitted variances $\hat{\sigma}_t^2(\tilde{\theta})$ are used to compute the standardized residuals $\hat{z}_t(\tilde{\theta}) = \hat{\varepsilon}_t / \hat{\sigma}_t(\tilde{\theta})$, which should have zero mean and unit variance. In a third step, the density $g(\hat{z}_t(\tilde{\theta}))$ is estimated using a non-parametric method. The estimated density is denoted $\hat{g}$. Finally, the parameters of the GARCH model are estimated by maximizing the log-likelihood function, with $\hat{g}$ held as fixed.

6.2.5 Adequacy tests

In Section 5.3, we described some tests aimed at assessing the ability of a given univariate distribution to capture the stylized facts of the empirical distribution. Most of these tests have been extended to the multivariate context. For instance, Diebold, Hahn, and Tay (1999) proposed an extension of the test introduced by Diebold, Gunter, and Tay (1998). The basic idea consists in writing the joint distribution $g_t\left(z_{1,t}, \cdots, z_{n,t}\right)$ as the product of conditional distributions, as in

$$g_t\left(z_{1,t}, \cdots, z_{n,t}\right) = g_t\left(z_{n,t} | z_{n-1,t}, \cdots, z_{1,t}\right) \times \cdots \times g_t\left(z_{2,t} | z_{1,t}\right) \times g_t\left(z_{1,t}\right).$$

Then, at each period, it is possible to transform each component of the vector $\left(z_{1,t}, \cdots, z_{n,t}\right)'$ by its corresponding conditional distribution. We then obtain n series of $u_{i,t} = \int_{-\infty}^{z_{i,t}} g_t\left(y_{i,t} | y_{i-1,t}, \cdots, y_{1,t}\right) \mathrm{d}y_{i,t}$ that should be found to be *iid* $U(0,1)$, both individually and jointly, if the model is correct.

One difficulty occurs in the multivariate case, because there are $n!$ ways to factor the joint distribution in terms of conditional distributions. For instance, in the bivariate case, we can write the joint distribution as

$$g_t\left(z_{1,t}, z_{2,t}\right) = g_t\left(z_{2,t} | z_{1,t}\right) \times g_t\left(z_{1,t}\right),$$

or

$$g_t\left(z_{1,t}, z_{2,t}\right) = g_t\left(z_{1,t} | z_{2,t}\right) \times g_t\left(z_{2,t}\right).$$

The first decomposition would provide us with probability integral transforms $u_{1,t}$ and $u_{2|1,t}$ and the second would provide $u_{2,t}$ and $u_{1|2,t}$, where we denote for instance

$$u_{2|1,t} = \int_{-\infty}^{z_{2,t}} g_t\left(y_{2,t} | y_{1,t}\right) dy_{2,t}.$$

Finally, we have to test if u_1, $u_{2|1}$, u_2 and $u_{1|2}$ are each *iid* $U(0,1)$ and if $\left(u_1, u_{2|1}\right)$ and $\left(u_2, u_{1|2}\right)$ are also *iid* $U(0,1)$. Consequently, for large numbers of variables, implementing the adequacy test may be quite cumbersome.

6.2.6 Illustration

To illustrate the working of the Student t and skewed Student t distributions in the context of multivariate GARCH models, we use once again two pairs of time series: the SP500 and DAX daily returns and the SP500 and FT-SE returns, over the period from January 1980 to August 2004. We consider the DCC model of Engle (2002), whose estimation under normality is reported in Table 6.2, and we estimate the same specification assuming Student t and skewed Student t innovations (with dependent components) in turn. Table 6.3 reports parameter estimates corresponding to these various models.[13] The log-likelihood indicates that the model with Student t innovations very markedly dominates the model with Gaussian innovations (compare with Table 6.2). The degree-of-freedom parameter ν is in the same range of values as found previously, in the univariate context, in Section 5.4.

Introducing an asymmetric t distribution also helps improving the fit of the empirical joint distribution, although less significantly. This finding can be explained by the weak asymmetry in the distribution of the SP500 return. Such evidence has been already observed in the univariate estimation. Finally, we do not display the dynamics of the conditional correlation, because there is no noticeable difference with the one obtained under normality (Section 6.1).

6.3 Copula functions

In many situations where marginal distributions are not Gaussian, it is simply impossible to define a joint distribution. This is the case when we want to link two variables that have different marginal distributions (for instance, a Student t variate and a Pareto variate). This is also the case for a large number of marginal distributions, for which a multivariate extension does not exist. In such contexts, a solution is to use copula functions. These functions have the property to relate two marginal distributions instead of the two series directly. Therefore, once margins have been computed, no reference is made to their true functional form. This is the reason why copula functions are able to relate any kind of margin.

The textbooks for the analysis of copula functions are Joe (1997) and Nelsen (1999). Some surveys also provide valuable information on copulas, see in particular Riboulet, Roncalli, and Bouyé (2000), and Embrechts, Lindskog, and McNeil (2003). It should be noticed at this point that copula functions have been abundantly used to investigate the behavior of the tails of a multivariate distribution. This issue is addressed in detail in Section 7.2.

[13] Notice that the parameter estimates of the SP500 obtained for the models with Student t and skewed Student t innovations differ for the two pairs of returns. The reason is that in such cases, two-step estimation is precluded, because the log-likelihood cannot be broken down into the variance part and the correlation part. However, the difference remains barely noticeable.

Table 6.3. *Parameter estimates of the DCC model under various distributions*

	SP500–DAX		SP500–FT-SE	
	Estimate	Std. err.	Estimate	Std. err.
DCC with Student t innovations				
ω_1	0.0058	(0.0017)	0.0061	(0.0018)
α_1	0.0407	(0.0053)	0.0414	(0.0060)
β_1	0.9527	(0.0062)	0.9512	(0.0070)
ω_2	0.0152	(0.0033)	0.0148	(0.0031)
α_2	0.0730	(0.0082)	0.0727	(0.0084)
β_2	0.9185	(0.0088)	0.9087	(0.0107)
ν	7.5728	(0.5586)	8.1771	(0.6486)
δ_1	0.0090	(0.0026)	0.0078	(0.0078)
δ_2	0.9898	(0.0031)	0.9848	(0.0188)
log-lik.	−17599.1	–	−15639.1	–
DCC with skewed Student t innovations				
ω_1	0.0058	(0.0019)	0.0062	(0.0020)
α_1	0.0409	(0.0056)	0.0417	(0.0064)
β_1	0.9524	(0.0068)	0.9509	(0.0077)
ω_2	0.0152	(0.0040)	0.0145	(0.0034)
α_2	0.0722	(0.0095)	0.0719	(0.0090)
β_2	0.9193	(0.0108)	0.9099	(0.0118)
ν	7.6718	(0.5665)	8.2845	(0.6580)
ξ_1	0.9790	(0.0141)	0.9882	(0.0139)
ξ_2	0.9412	(0.0156)	0.9180	(0.0159)
δ_1	0.0090	(0.0026)	0.0073	(0.0087)
δ_2	0.9898	(0.0032)	0.9858	(0.0213)
log-lik.	−17593.4	–	−15626.5	–

6.3.1 Definitions and properties

The study of copulas is quite a recent phenomenon in statistics. Hence, it is not astonishing that copulas have only recently found their way into empirical finance. In order to understand their usefulness, consider two random variables X and Y with marginal distributions, or *margins*, $F(x) = \Pr[X \leq x]$ and $G(y) = \Pr[Y \leq y]$. In this paper, we assume that the cumulative distribution functions (*cdf*) are continuous. The random variables may also have joint distribution function, $H(x,y) = \Pr[X \leq x, Y \leq y]$. All the distribution functions, $F(\cdot)$, $G(\cdot)$, and $H(\cdot,\cdot)$ have as range the interval $[0,1]$. In some cases, a multivariate distribution exists, so that the function $H(\cdot,\cdot)$ has an explicit expression. One such case is the multivariate normal distribution. In many cases, however, a description of the margins $F(\cdot)$ and $G(\cdot)$ is relatively easy to obtain, whereas an explicit expression of the joint distribution $H(\cdot,\cdot)$ may be difficult to obtain. This is where copulas are useful because they link margins into a multivariate distribution function.

Definition 6.2. *A bivariate copula is a function* $C : [0,1] \times [0,1] \to [0,1]$ *with the three following properties:*

1. *$C(u,v)$ is increasing in u and v.*
2. *$C(0,v) = C(u,0) = 0$, $C(1,v) = v$, $C(u,1) = u$.*
3. *$C(u_2,v_2) - C(u_2,v_1) - C(u_1,v_2) + C(u_1,v_1) \geq 0$, $\forall u_1,u_2,v_1,v_2$ in $[0,1]$ such that $u_1 \leq u_2$ and $v_1 \leq v_2$.*

Property 1 states that, when one marginal distribution is constant, the joint probability will increase provided that the other marginal distribution increases. Property 2 states that if one margin has zero probability to occur, then it must be the same for the joint occurrence. Also, if on the contrary, one margin is certain to occur, then the probability of a joint occurrence is determined by the remaining margin probability. Property 3 indicates that, if u and v both increase, then the joint probability also increases. This property is, therefore, a multivariate extension of the condition that a *cdf* is increasing.

Furthermore, if we set $u = F(x)$ and $v = G(y)$, then $C(F(x), G(y))$ yields a description of the joint distribution of X and Y. Having obtained this intuitive definition, we now propose the following important theorem, proven in Sklar (1959) and Schweizer and Sklar (1974).

Theorem 6.3 (Sklar's theorem). *Let F and G be the marginal distributions of X and Y, respectively, and let H be the joint distribution function of (X,Y). Then, there exists a copula C such that, for all real numbers (x,y),*

$$H(x,y) = C(F(x), G(y)). \tag{6.21}$$

Furthermore, if F and G are continuous, then C is unique. Conversely, if F and G are the distributions of X and Y, respectively, then the function H defined by (6.21) is a joint distribution function with marginal distributions F and G.

The density of a copula is related to its *cdf* through the following relation

$$c\,(u,v) \equiv \frac{\partial^2 C(u,v)}{\partial u \partial v}.$$

Similarly, the density h of the distribution H is defined by the relationship

$$h\,(x,y) = c\,(F\,(x)\,, G\,(y)) \times f\,(x) \times g\,(y)\,.$$

Notice that many results developed in this section extend to a higher dimensional framework. Some of the results, however, hold in the bivariate framework only. In many cases, the ease of interpretation of the dependency parameter does not hold when there are more than two margins.

6.3.2 Measures of concordance

For a number of standard distributions (namely, the elliptical family, which includes the Gaussian and the Student t distributions), dependency is naturally measured by Pearson's correlation coefficient. However, when other distributions are considered, alternative measures are needed to characterize the link between time series. Most of the results in this section are drawn from Nelsen (1999) and Embrechts, Lindskog, and McNeil (2003).

Following Nelsen (1999, p. 136), two pairs of random variables (X, Y) and $(\tilde{X}, \tilde{Y})$ are concordant if $X < \tilde{X}$ implies $Y < \tilde{Y}$ or if $X > \tilde{X}$ implies $Y > \tilde{Y}$ and discordant if $X < \tilde{X}$ implies $Y > \tilde{Y}$ or if $X > \tilde{X}$ implies $Y < \tilde{Y}$.

Kendall's tau for the random variables X and Y is defined as the probability of concordance minus the probability of discordance (or non-covariation) of two independent pairs of random variables (X, Y) and $(\tilde{X}, \tilde{Y})$

$$\tau\,[X, Y] = \Pr\left[(X - \tilde{X})(Y - \tilde{Y}) > 0\right] - \Pr\left[(X - \tilde{X})(Y - \tilde{Y}) < 0\right].$$

Spearman's rho for the random variables X and Y is defined as

$$\varrho_S\,[X, Y] = 3\left(\Pr\left[(X - \tilde{X})(Y - Y') > 0\right] - \Pr\left[(X - \tilde{X})(Y - Y') < 0\right]\right),$$

where (X, Y), $(\tilde{X}, \tilde{Y})$, and (X', Y') are three independent copies. Since $\tilde{X}$ and Y' are independent, Spearman's rho can be viewed as the distance between the distribution of X and Y and independence. Spearman's rho was first proposed in 1904 by the psychologist C. Spearman. Similar to Kendall's tau, it is related to the probabilities of concordance and discordance. The distinction is that one pair (X, Y) has distribution $H(x, y)$, and the second pair, say $(\tilde{X}, \tilde{Y})$, is a pair of independent random variables with same margins as X and Y, meaning that $(\tilde{X}, \tilde{Y})$ has distribution function $F(x)\,G(y)$. To render this concept operational, it is convenient to consider three independent pairs of random variables (X, Y), $(\tilde{X}, \tilde{Y})$ and (X', Y') each drawn from $H(x, y)$. The assumption that all pairs are independent means that $\tilde{X}$ can be viewed as drawn from $F(x)$ and Y' from $G(y)$. The issue then is to see if (X, Y) and $(\tilde{X}, Y')$ are concordant.

Theorem 6.4. *(Schweizer and Wolff, 1981) Let X and Y be continuous random variables whose copula is C. Then Kendall's tau and Spearman's rho for X and Y are defined as*

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

$$\varrho_S(X, Y) = 12\int\int_{[0,1]^2} u\,v\; dC(u, v) - 3 = 12E\left[U, V\right] - 3.$$

Note that Spearman's rho can be written as $\varrho_S\,[X, Y] = \rho\left[F(X), G(Y)\right]$, and may thus be viewed as Pearson's correlation of the marginal distributions $F(X)$ and $G(Y)$.

Nelsen (1999) shows that Kendall's tau and Spearman's rho satisfy conditions required to be concordance measures. If $\kappa_{X,Y}$ denote Kendall's tau or

Spearman's rho between X and Y, we have in particular that $-1 \leq \kappa_{X,Y} \leq 1$, $\kappa_{X,X} = 1$, $\kappa_{X,-X} = -1$, $\kappa_{X,Y} = \kappa_{Y,X}$ and $\kappa_{-X,Y} = \kappa_{X,-Y} = -\kappa_{X,Y}$. For continuous variables, all values of Kendall's tau and Spearman's rho in the interval $[-1, 1]$ can be obtained by a suitable choice of the underlying copula.

While Pearson's correlation is a natural scalar measure of dependence in elliptical distributions (for instance, the Gaussian or the Student t distributions), using it as a measure of dependence in more general situations might prove misleading. Indeed, Pearson's correlation has some undesirable properties. First, it is possible that the correlation $\rho[X, Y]$ between two random variables X and Y is equal to zero while the two variables are not independent. Second, the range of permissible values for the correlation is not necessarily $[-1, 1]$. Indeed, depending on the marginal distributions of the two variables, the actual range may be much smaller than $[-1, 1]$. Third, Pearson's correlation is not invariant for an increasing transform of X and Y. Finally, we may have that $\rho[X, Y] > 0$, while X and Y do not have necessarily a positive dependence.

6.3.3 Non-parametric copulas

A first approach to the modeling of non-linear dependence consists in non-parametrically estimating the unrestricted joint density (Silverman, 1986, Härdle, 1990, Scott, 1992). This method has been used by Deheuvels (1979, 1981) and Fermanian and Scaillet (2003) to deduce a non-parametric estimate of the associated unrestricted copula. The advantage of this approach is that it does not require any additional assumption on the non-linear dependence. However, it suffers from the drawbacks of any non-parametric approach. In particular, the interpretation of the patterns of non-linear dependence is often complicated, and it is likely to provide inaccurate and erratic results, even in the bivariate case. This approach has been recently improved by the work of Gagliardini and Gouriéroux (2006), who propose an intermediate approach in which the joint density is constrained and depends on a small number of one-dimensional functional parameters, yielding efficient non-parametric estimators for the one-dimensional functional parameters, which characterize non-linear dependence.

The non-parametric copula has been proposed by Deheuvels (1979, 1981). It is defined as

$$C_T\left(\frac{t_1}{T}, \frac{t_2}{T}\right) = \frac{1}{T}\sum_{t=1}^{T} 1_{\left\{x_t \leq x_{t_1,T}, y_t \leq y_{t_2,T}\right\}} \qquad \text{with } 1 \leq t_1, t_2 \leq T,$$

where $x_{t,T}$ denotes the order statistics, i.e., $x_{1,T} \leq \cdots \leq x_{T,T}$ are ordered realizations. The empirical copula frequency is then defined as

$$c_T\left(\frac{t_1}{T}, \frac{t_2}{T}\right) = \begin{cases} \frac{1}{T} \text{ if } (x_{t_1,T}, y_{t_2,T}) \text{ belongs to the sample} \\ 0 \text{ otherwise,} \end{cases}$$

so that C_T and c_T are related through the relation

$$C_T\left(\frac{t_1}{T},\frac{t_2}{T}\right)=\sum_{p=1}^{t_1}\sum_{q=1}^{t_2}c_T\left(\frac{p}{T},\frac{q}{T}\right).$$

Sample versions of Kendall's tau and Spearman's rho are given by

$$\tau=\frac{2T}{T-1}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\sum_{p=1}^{t_1}\sum_{q=1}^{t_2}\left[c_T\left(\frac{t_1}{T},\frac{t_2}{T}\right)\cdot c_T\left(\frac{p}{T},\frac{q}{T}\right)\right.$$
$$\left.-c_T\left(\frac{t_1}{T},\frac{q}{T}\right)\cdot c_T\left(\frac{p}{T},\frac{t_2}{T}\right)\right],$$

and

$$\varrho_S=\frac{12}{T^2-1}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\left[C_T\left(\frac{t_1}{T},\frac{t_2}{T}\right)-\frac{t_1}{T}\cdot\frac{t_2}{T}\right].$$

A non-parametric estimation of copulas has been developed by Fermanian and Scaillet (2003), using a kernel-based approach. Such an approach has the advantage to provide a smooth reconstitution of the copula function without assuming any particular parametric structure on the dependence structure between margins.

6.3.4 Review of some copula families

We provide in this section a few examples of copula functions that have been studied and estimated in the empirical finance literature. Such copulas can be found, for instance, in Riboulet, Roncalli, and Bouyé (2000), Jondeau and Rockinger (2006b), and Patton (2006). The copula functions described in this section cover some of the most used distribution in the literature. For a more general description of copula functions, see Joe (1997) and Nelsen (1999). We do not consider in this review how margins are defined. They may be parametric as well as non-parametric.

Table 6.4 reports, for the copula functions considered below, the range of possible values that can be reached by the concordance measures. κ denotes the Kendall's tau or the Spearman's rho. This table indicates that some copula functions may be able to reproduce only positive or only negative dependence. For instance, the Clayton , Gumbel or Marshall-Olkin copula only have positive dependence.

Elliptical copulas

The class of elliptical distributions includes several well-known multivariate distributions, such as the Gaussian and the Student t distributions. Let X be a n-dimensional random vector, $\mu \in \mathbb{R}^n$ and Σ a (n, n) non-negative definite,

Table 6.4. *Description of some usual copula functions*

Copula	Parameter space	Range for κ
Gaussian copula	$-1 < \rho < 1$	$-1 < \kappa < 1$
Student t copula	$-1 < \rho < 1$ and $n > 2$	$-1 < \kappa < 1$
Frank copula	$-\infty < \theta < \infty$	$-1 < \kappa < 1$
Clayton copula	$0 < \theta < \infty$	$0 < \kappa \leq 1$
Gumbel copula	$1 \leq \theta < \infty$	$0 \leq \kappa < 1$
Plackett copula	$0 < \theta < \infty$ and $\theta \neq 1$	$-1 < \kappa < 1$
Marshall-Olkin copula	$0 \leq \alpha \leq 1$ and $0 \leq \beta \leq 1$	$0 < \kappa \leq 1$

symmetric matrix. Then, if the characteristic function $\varphi_{X-\mu}(t)$ of $(X-\mu)$ is a function of the quadratic form $t'\Sigma t$, i.e., $\varphi_{X-\mu}(t) = \phi(t'\Sigma t)$, then X has an elliptical distribution with parameters μ, Σ, and ϕ. In addition, if X has a density, it is of the form $|\Sigma|^{-1/2} g\left((X-\mu)'\Sigma^{-1}(X-\mu)\right)$, for some non-negative function g of one scalar variable. Therefore, the contours of equal density form ellipsoids in $\mathbb{R}^n$.

Gaussian copula

The Gaussian copula is defined by the following *cdf*

$$\begin{aligned} C_\rho(u,v) &= \Phi_\rho\left(\Phi^{-1}(u), \Phi^{-1}(v)\right) \\ &= \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{s^2 - 2\rho st + t^2}{2(1-\rho^2)}\right) dsdt, \end{aligned}$$

where Φ_ρ is the bivariate standardized Gaussian *cdf* with Pearson's correlation $\rho \in [-1;1]$ and Φ^{-1} denotes the inverse of the distribution function of the univariate standard normal distribution.

The density of the Gaussian copula is given by

$$c_\rho(u,v) = \frac{1}{|R|^{1/2}} \exp\left(-\frac{1}{2}\psi'\left(R^{-1} - I_2\right)\psi\right),$$

where $\psi = \left(\Phi^{-1}(u), \Phi^{-1}(v)\right)'$, and R is the $(2,2)$ correlation matrix between u and v with ρ as correlation parameter.

Kendall's tau and Spearman's rho are given respectively by

$$\begin{aligned} \tau(C_\rho) &= \frac{2}{\pi}\arcsin(\rho), \\ \varrho_S(C_\rho) &= \frac{6}{\pi}\arcsin(\rho/2). \end{aligned}$$

Figure 6.14 displays the Gaussian copula $c_\rho(u,v)$ for $\rho = 0.5$. Figure 6.15 presents the density h of the two-dimensional distribution defined as: $h(x,y) = c_\rho(F(x), G(y)) \times f(x) \times g(y)$. In the first case, the two variables are supposed to be distributed as $\mathcal{N}(0,1)$, so that the Gaussian copula is equivalent to the multivariate Gaussian distribution. In the second case, X is a t_3 while Y is a $\mathcal{N}(0,1)$.

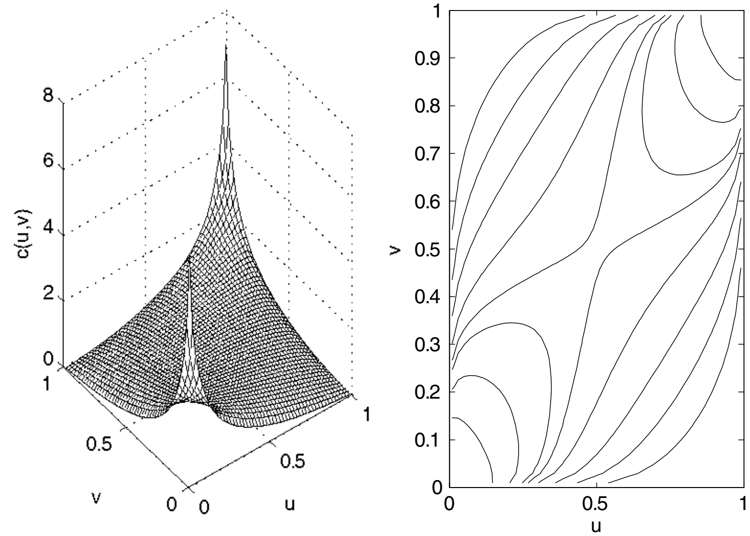

Fig. 6.14. *Gaussian copula and its contour plot.*

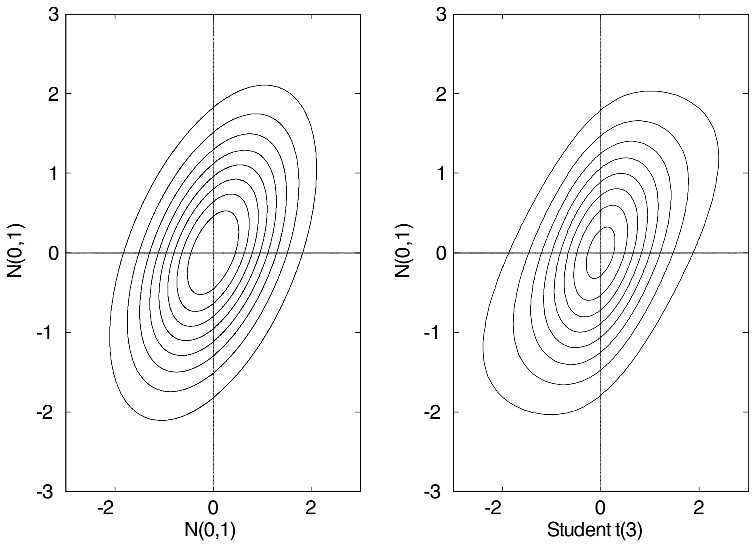

Fig. 6.15. *Contour plot of the density of the Gaussian copula.*

Student t copula

The Student t copula is defined by

$$C_{\rho,n}(u,v) = t_{\rho,n}\left(t_n^{-1}(u), t_n^{-1}(v)\right)$$
$$= \int_{-\infty}^{t_n^{-1}(u)} \int_{-\infty}^{t_n^{-1}(v)} \frac{\Gamma\left(\frac{n+2}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\pi n\sqrt{1-\rho^2}} \left(1 + \frac{\psi' R^{-1}\psi}{n}\right)^{-\frac{n+2}{2}} d\psi,$$

where $\psi = \left(t_n^{-1}(u), t_n^{-1}(v)\right)'$, $t_{\rho,n}$ is the bivariate Student t *cdf* with n degrees of freedom and correlation ρ, and t_n^{-1} denotes the inverse of the distribution function of the univariate Student t distribution with n degrees of freedom. The density of the Student t copula is given by

$$c_{\rho,n}(u,v) = \frac{1}{\sqrt{|R|}} \frac{\Gamma\left(\frac{n+2}{2}\right)\Gamma\left(\frac{n}{2}\right)}{\left(\Gamma\left(\frac{n+1}{2}\right)\right)^2} \frac{\left(1 + \frac{1}{n}\psi' R^{-1}\psi\right)^{-\frac{n+2}{2}}}{\prod_{i=1}^{2}\left(1 + \frac{1}{n}\psi_i^2\right)^{-\frac{n+1}{2}}}.$$

As for the Gaussian copula, Kendall's tau is given by

$$\tau(C_\rho) = \frac{2}{\pi}\arcsin(\rho).$$

The analytic expression for Spearman's rho is unknown. Consequently, it must be computed numerically.

Figure 6.16 represents the density of the Student t copula for $\nu = 3$ and $\rho = 0.5$, corresponding to a Kendall's tau equal to 0.333. The left figure corresponds to Gaussian margins, whereas the right figure corresponds to a Gaussian and a Student t with $\nu = 3$ margins.

Archimedean copulas

Unlike the copulas described so far, the Archimedean ones are not derived from multivariate distribution functions. An advantage is that most Archimedean copulas have closed form expressions. A disadvantage is that multivariate extensions of Archimedean copulas are somewhat difficult to establish.

Theorem 6.5. *(Nelsen, 1999, p. 91) Let φ be a continuous, strictly decreasing function from $[0,1]$ to $[0,\infty)$ such that $\varphi(1) = 0$ and let φ^{-1} be the inverse of φ. Then, the function from $[0,1]^2$ to $[0,1]$ given by*

$$C(u,v) = \varphi^{-1}(\varphi(u) + \varphi(v))$$

is a copula if and only if φ is convex.

The function φ is called a generator of the copula. Assuming that φ^{-1} is twice continuously differentiable, the copula density is

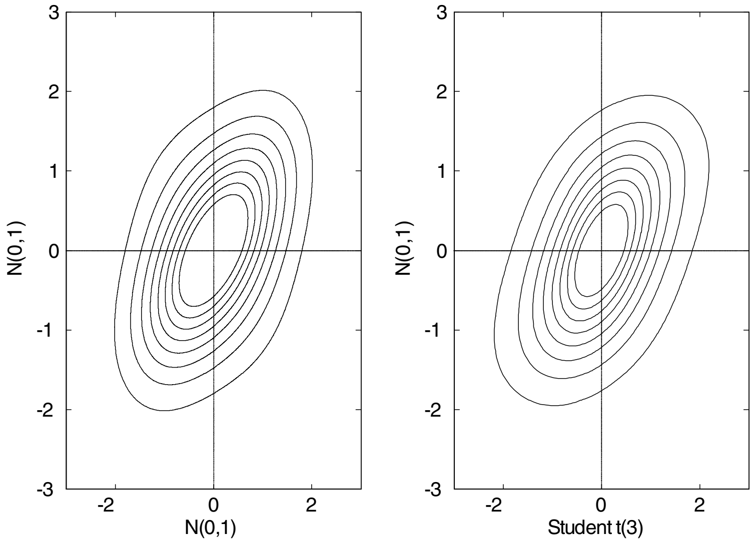

Fig. 6.16. *Contour plot of the density of the Student t copula.*

$$c(u,v) = \frac{\varphi^{-1\prime\prime}(\varphi(u)+\varphi(v))}{\varphi^{-1\prime}(\varphi'(u))\,\varphi^{-1\prime}(\varphi'(v))}.$$

In this case, Kendall's tau is given by

$$\tau(C) = 1 + 4\int_0^1 \frac{\varphi(t)}{\varphi'(t)}dt.$$

Many additional details on Archimedean copulas are provided by Nelsen (1999) and Embrechts, Lindskog, and McNeil (2003).

Frank copula

When $\varphi(t) = \log\left(e^{-\theta}-1\right) - \log\left(e^{-\theta t}-1\right)$, for $\theta \neq 0$, we obtain the Frank family of copulas. The Frank copula is defined by

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

and

$$c_\theta(u,v) = \frac{\theta\left(1-e^{-\theta}\right)e^{-\theta(u+v)}}{\left[\left(1-e^{-\theta}\right)-\left(1-e^{-\theta u}\right)\left(1-e^{-\theta v}\right)\right]^2}.$$

The copula is defined for $\theta \neq 0$. Kendall's tau is given by

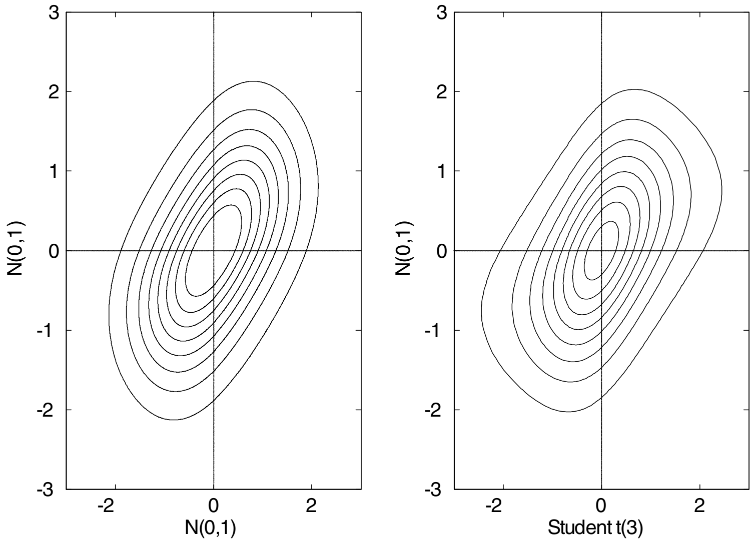

Fig. 6.17. *Contour plot of the density of the Frank copula.*

$$\tau\left(C_\theta\right)=1-4\frac{1-D_1\left(\theta\right)}{\theta},$$

and Spearman's rho is

$$\varrho_S\left(C_\theta\right)=1-12\frac{D_1\left(\theta\right)-D_2\left(\theta\right)}{\theta},$$

where

$$D_k\left(x\right)=\begin{cases}\frac{k}{x^k}\int_0^x\frac{t^k}{e^t-1}dt & \text{if } x\geq 0,\\ \frac{k|x|}{1+k}+\frac{k}{|x|^k}\int_x^0\frac{t^k}{e^t-1}dt & \text{if } x<0\end{cases}$$

is the Debye function (see Abramowitz and Stegun, 1970).

Figure 6.17 represents the density of the Clayton copula for $\theta = 3$, corresponding to a Kendall's tau equal to 0.307.

Clayton copula

When $\varphi\left(t\right)=\left(t^{-\theta}-1\right)/\theta$, for $\theta\in\left[-1;\infty\right)\setminus\{0\}$, we obtain the Clayton family of copulas. The Clayton copula is defined by

$$C_\theta\left(u,v\right)=\max\left(\left(u^{-\theta}+v^{-\theta}-1\right)^{-1/\theta},0\right).$$

For $\theta > 0$, the copula simplifies to

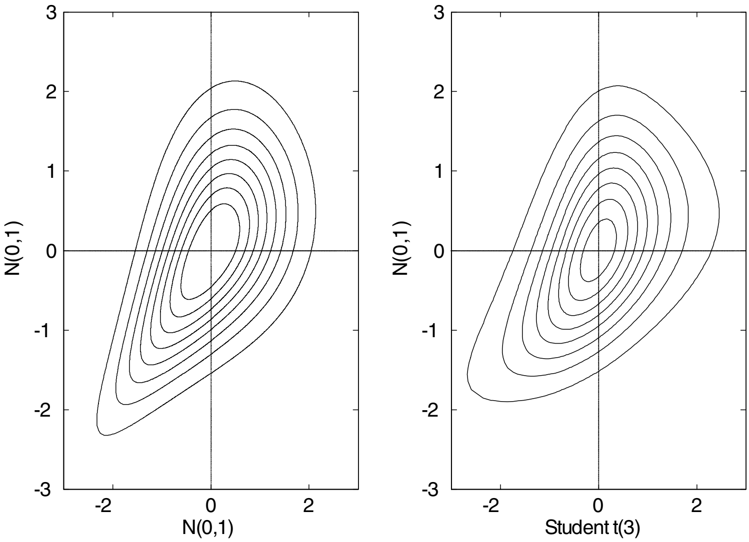

Fig. 6.18. *Contour plot of the density of the Clayton copula.*

$$C_\theta(u,v) = \left(u^{-\theta} + v^{-\theta} - 1\right)^{-1/\theta}.$$

The density of the copula is given by

$$c_\theta(u,v) = (1+\theta)(uv)^{-\theta-1}\left(u^{-\theta} + v^{-\theta} - 1\right)^{-2-1/\theta}.$$

Kendall's tau is given by

$$\tau(C_\theta) = \frac{\theta}{\theta+2}.$$

Figure 6.18 represents the density of the Clayton copula for $\theta = 1$, corresponding to a Kendall's tau equal to 0.333. The figure confirms that the Clayton copula generates dependence in the lower-tail but not in the upper tail. In order to also generate some dependence in the upper tail, we can use the so-called rotated copula, defined as follows: if (U,V) has copula $C_\theta(u,v)$, then $(1-U, 1-V)$ is distributed according to the rotated copula $C_\theta^R(u,v)$. For the Clayton copula, we have

$$C_\theta^R(u,v) = u + v - C_\theta(1-u, 1-v),$$

with density

$$c_\theta^R(u,v) = c_\theta(1-u, 1-v).$$

The rotated copula has dependence in the upper tail.

Gumbel copula

When $\varphi(t) = (-\log t)^{\theta}$, for $\theta \in [1;\infty)$, we obtain the Gumbel family of copulas. The Gumbel copula is defined by

$$C_\theta(u,v) = \exp\left\{-\left[(-\log u)^{\theta} + (-\log v)^{\theta}\right]^{1/\theta}\right\},$$

and

$$c_\theta(u,v) = \frac{C_\theta(u,v)\,[\log u \times \log v]^{\theta-1}}{uv\left[(-\log u)^{\theta} + (-\log v)^{\theta}\right]^{2-1/\theta}} \times \left\{\left[(-\log u)^{\theta} + (-\log v)^{\theta}\right]^{1/\theta} + \theta - 1\right\}.$$

Kendall's tau is given by

$$\tau(C_\theta) = 1 - \frac{1}{\theta}.$$

Figure 6.19 represents the density of the Gumbel copula for $\theta = 1.5$, corresponding to a Kendall's tau equal to 0.333. As for the Clayton copula, it is possible to define a rotated Gumbel copula with

$$C_\theta^R(u,v) = u + v - 1 + C_\theta(1-u, 1-v),$$

with density

$$c_\theta^R(u,v) = c_\theta(1-u, 1-v).$$

Plackett copula

The Plackett copula, proposed by Plackett (1965), is defined by

$$C_\theta(u,v) = \frac{1}{2(\theta-1)} \times \left[1 + (\theta-1)(u+v) - \sqrt{[1+(\theta-1)(u+v)]^2 - 4uv\theta(\theta-1)},\right]$$

and

$$c_\theta(u,v) = \frac{\theta[1 + (u - 2uv + v)(\theta-1)]}{\left([1+(\theta-1)(u+v)]^2 - 4uv\theta(\theta-1)\right)^{\frac{3}{2}}}.$$

The copula is defined for $\theta \in [0;\infty)$. Note that the Spearman's rho of the Plackett copula is simply derived from the dependence parameter θ as

$$\rho_\theta = \frac{\theta+1}{\theta-1} - \frac{2\theta\log(\theta)}{(\theta-1)^2}. \tag{6.22}$$

Figure 6.20 represents the density of the Plackett copula for $\theta = 3$, corresponding to a Spearman's rho equal to 0.352.

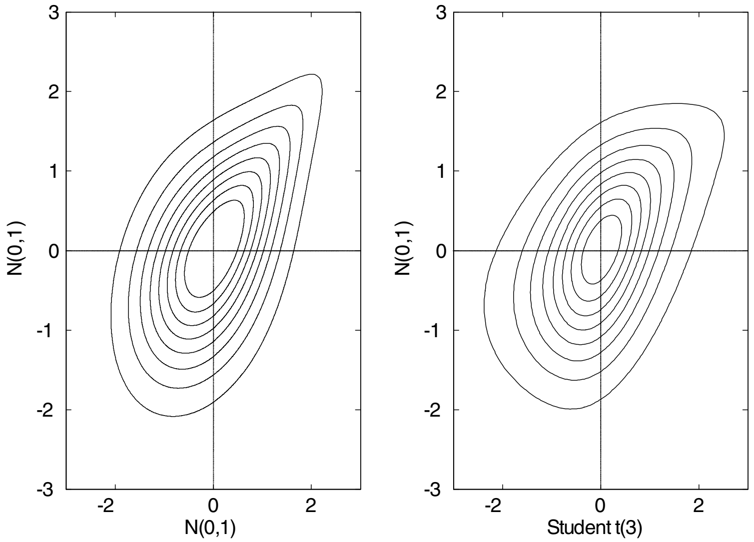

Fig. 6.19. *Contour plot of the density of the Gumbel copula.*

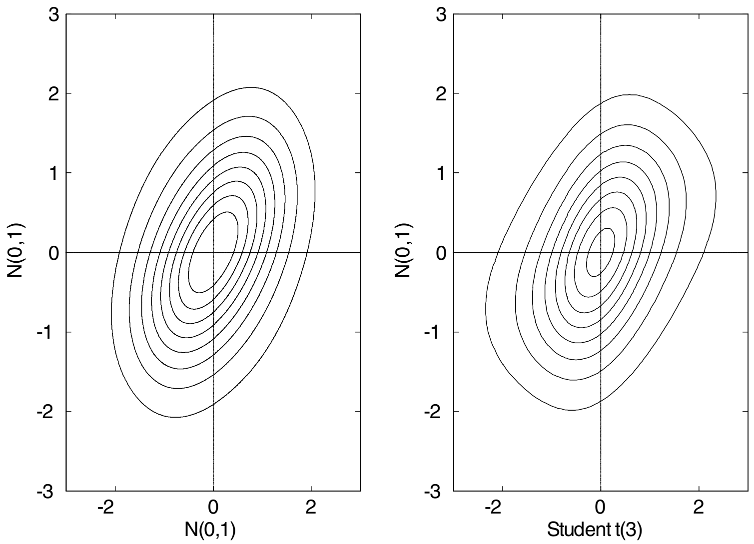

Fig. 6.20. *Contour plot of the density of the Plackett copula.*

Marshall-Olkin copula

The Marshall-Olkin copula is defined by the function

$$C_{\alpha,\beta}(u,v) = \min\left(u^{1-\alpha}v, uv^{1-\beta}\right) = \begin{cases} u^{1-\alpha}v & \text{if } u^{\alpha} \geq v^{\beta}, \\ uv^{1-\beta} & \text{if } u^{\alpha} \leq v^{\beta}, \end{cases}$$

with $u = \bar{F}(x)$ and $v = \bar{G}(y)$, where $\bar{F}(x) = \Pr[X > x]$ and $\bar{G}(y) = \Pr[Y > y]$ denote the marginal survival functions. The domain of definition of the parameters α and β is: $0 \leq \alpha, \beta \leq 1$.

The density of the copula is defined by

$$c_{\alpha,\beta}(u,v) \equiv \frac{\partial^2 C_{\alpha,\beta}(u,v)}{\partial u \partial v} = \begin{cases} u^{-\alpha} \text{ if } u^{\alpha} > v^{\beta}, \\ v^{-\beta} \text{ if } u^{\alpha} < v^{\beta}, \end{cases}$$

so that the Marshall-Olkin copulas have a singular component, with a mass concentrated on the curve $u^{-\alpha} = v^{-\beta}$.

Kendall's tau and Spearman's rho are obtained, applying the results of the previous section (Embrechts, Lindskog, and McNeil, 2003)

$$\tau(C_{\alpha,\beta}) = 4\int\int_{[0,1]^2} C(u,v)\, \mathrm{d}C(u,v) - 1 = \frac{\alpha\beta}{\alpha+\beta-\alpha\beta},$$
$$\varrho_S(C_{\alpha,\beta}) = 12\int\int_{[0,1]^2} u\,v\,\mathrm{d}C(u,v) - 3 = \frac{3\alpha\beta}{2\alpha+2\beta-\alpha\beta}.$$

All values in the interval $[0,1]$ can be obtained for both the Kendall's tau and the Spearman's rho.

Figure 6.21 represents the density of the Marshall-Olkin copula for $\lambda_1 = 0.5$, $\lambda_2 = 0.1$ and $\lambda_{12} = 1$, so that $\alpha = 0.667$ and $\beta = 0.909$. These values correspond to a Kendall's tau equal to 0.375.

6.3.5 Estimation

Several approaches have been proposed to estimate the parameters of a copula function. In addition to the standard ML estimation, a two-step estimation procedure is readily available, because the log-likelihood of the model can be written as the sum of two components, the margins and the dependence structure. Therefore, it is natural to estimate the parameters of the margins and the parameters of the copula function separately (Shih and Louis, 1995, Joe and Xu, 1996). The copula parameters can also be semi-parametrically estimated, using the marginal empirical distribution to compute the copula (Genest, Ghoudi, and Rivest, 1995).

Other alternative estimation procedures may be considered as well. For instance, as it has been shown in Section 6.3.4, the parameters of the copula function are in general related to one of the concordance measures (Kendall's

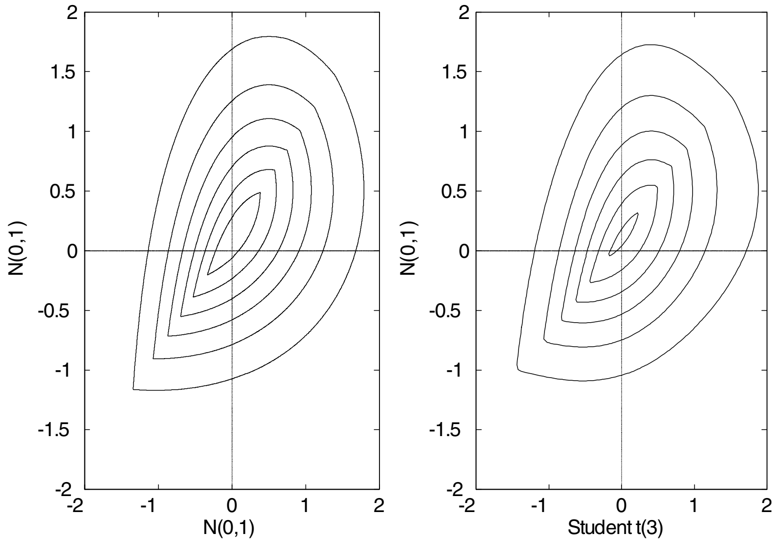

Fig. 6.21. *Contour plot of the density of the Marshall-Olkin copula.*

tau or Spearman's rho). It is therefore possible to estimate the unknown parameters using the method of moments, in such a way that the Kendall's tau (say) of the theoretical copula fits the empirical Kendall's tau. Genest and Rivest (1993) have developed an estimation procedure adapted for Archimedean copula functions.

ML estimator

We assume now that the unknown parameters associated with the marginal densities f and g are denoted θ_x and θ_y, respectively, and that the unknown parameters associated with the copula function c are denoted θ_γ. We denote the $(K, 1)$ vector $\theta = (\theta_x, \theta_y, \theta_\gamma)$. The maximum likelihood estimate (MLE) of a model is obtained by maximizing the conditional log-likelihood function, which is defined as

$$\begin{aligned} L_T\left(\theta | \underline{x}_t, \underline{y}_t\right) &= \sum_{t=1}^{T} \log\left(c_{\theta_\gamma}\left(F\left(x_t, \theta_x\right), G\left(y_t, \theta_y\right)\right) \times f\left(x_t, \theta_x\right) \times g\left(y_t, \theta_y\right)\right) \\ &= \sum_{t=1}^{T} \log c_{\theta_\gamma}\left(F\left(x_t, \theta_x\right), G\left(y_t, \theta_y\right)\right) \\ &\quad + \sum_{t=1}^{T} \left(\log\left[f\left(x_t, \theta_x\right)\right] + \log\left[g\left(y_t, \theta_y\right)\right]\right). \end{aligned} \tag{6.23}$$

As previously, the ML estimator $\hat{\theta}_{ML}$ is asymptotically normal with asymptotic distribution

$$\sqrt{T}\left(\hat{\theta}_{ML}-\theta_0\right) \Longrightarrow \mathcal{N}\left(0, A_0^{-1}\right),$$

where A_0 is the information matrix of Fisher. See Section 4.3.3 for further details on the construction and estimation of A_0. See also Section 5.1.2 for the computation of the covariance matrix of the QML estimator.

Two-step estimator (or inference functions for margins method)

In practical applications, the ML estimation may be difficult. First, the dimension of the problem can be large, because it requires one to estimate jointly the parameters of the margins and of the copula function. Second, the dependency parameter of the copula function may be a convoluted expression of the margin parameters. Therefore, an analytical expression of the gradient of the likelihood might not exist. Only numerical gradients may be computable, implying a slowing down of the numerical procedure.

In some cases, it is possible to split the vector of parameters into two parts: those associated with the marginal distributions and those associated with the copula function. This is the case, in particular, when there is no cross-restriction between the marginal distributions and the copula function.

A first two-step estimator, initially proposed by Shih and Louis (1995), Joe and Xu (1996), and Joe (1997) and used in a conditional setup by Jondeau and Rockinger (2006b) and Patton (2006), corresponds to the case when the parameters of the two marginal distributions can be estimated separately. The first step is the estimation of the marginal models. In the bivariate setting, we have

$$\tilde{\theta}_x \in \arg\max_{\{\theta_x \in \Theta_x\}} \sum_{t=1}^{T} \log[f(x_t, \theta_x)], \tag{6.24}$$

$$\tilde{\theta}_y \in \arg\max_{\{\theta_y \in \Theta_y\}} \sum_{t=1}^{T} \log[g(y_t, \theta_y)]. \tag{6.25}$$

Then, in a second step, the parameters θ_γ of the copula function can be estimated conditionally on the margin parameters

$$\tilde{\theta}_\gamma \in \arg\max_{\{\theta_\gamma \in \Theta_\gamma\}} \sum_{t=1}^{T} \log\left[c_{\theta_\gamma}\left(F\left(x_t, \tilde{\theta}_x\right), G\left(y_t, \tilde{\theta}_y\right)\right)\right].$$

If the model is correctly specified, then under rather mild assumptions, $\tilde{\theta}_x$, $\tilde{\theta}_y$, and $\tilde{\theta}_\gamma$ are consistent and asymptotically normal estimators (Patton, 2006).

Another estimator corresponds to the case when we cannot separate the parameters of the two marginal distributions. This may be the case, for instance, when the two variables are related by a multivariate GARCH model.

Although the two variables are still assumed to be independent, parameters of the two marginal distributions cannot be estimated separately. In this context, the first-step estimators are obtained by solving

$$\left(\breve{\theta}_x, \breve{\theta}_y\right) \in \underset{\{\theta_x \in \Theta_x, \theta_y \in \Theta_y\}}{\arg\max} \sum_{t=1}^{T} \log[f(x_t, \theta_x, \theta_y)] + \sum_{t=1}^{T} \log[g(y_t, \theta_x, \theta_y)],$$

while the estimators of the dependence parameters are obtained by solving

$$\breve{\theta}_\gamma \in \arg \max_{\{\theta_\gamma \in \Theta_\gamma\}} \sum_{t=1}^{T} \log \left[c_{\theta_\gamma}\left(F\left(x_t, \breve{\theta}_x\right), G\left(y_t, \breve{\theta}_y\right)\right)\right].$$

As above, estimators $\breve{\theta}_x$, $\breve{\theta}_y$, and $\breve{\theta}_\gamma$ are shown to be consistent and asymptotically normal.

Semi-parametric ML

Genest, Ghoudi, and Rivest (1995) have proposed an estimation procedure that avoids specifying the marginal distribution. Instead of using a parametric marginal distribution, they suggest the use of the marginal empirical distribution function. The empirical distribution function of X is

$$\hat{u}_T(\tau) = \hat{F}_T(x_\tau) = \frac{1}{T} \sum_{t=1}^{T} 1_{\{x_t \leq x_{\tau,T}\}},$$

where $x_{1,T} \leq \cdots \leq x_{T,T}$ is the ordered sample of observations. In other words, $\hat{u}_T(\tau)$ represents the frequency of observations below or equal to x_τ in the sample $\{x_t\}_{t=1}^{T}$. Genest, Ghoudi, and Rivest (1995) suggest that we redefine $\hat{u}_T(\tau)$ as $\frac{T}{T+1}\hat{u}_T(\tau)$ to avoid difficulties arising from the possible unboundedness of the log-likelihood when some $\hat{u}_T$s tend to one.

Considering the empirical margins directly avoids assuming any theoretical distribution for margins and therefore avoids the estimation of the parameters of the marginal distributions. Then, the parameters θ_γ of the copula function are estimated by maximizing the pseudo log-likelihood

$$\breve{\theta}_\gamma \in \arg \max_{\{\theta_\gamma \in \Theta_\gamma\}} \sum_{t=1}^{T} \log c_{\theta_\gamma}\left(\hat{F}_T(x_t), \hat{G}_T(y_t)\right).$$

Genest, Ghoudi, and Rivest (1995) show that the estimator $\breve{\theta}_\gamma$ is asymptotically normal, with a larger asymptotic variance than the ML estimator (obtained assuming that the margins are known).[14]

[14] In the case where $\breve{\theta}_\gamma$ is multidimensional, this implies that the difference between the asymptotic covariance matrices of $\breve{\theta}_\gamma$ and $\hat{\theta}_{\gamma,ML}$ is semi-definite positive.

6.3.6 Adequacy tests

One reason for the success of the copula functions is that they are able to model almost any kind of relationship between time series. But one drawback is that the class of copula functions is now very large, so that it is often difficult to select the "right" one, i.e., the copula function that best fits the data at hand. There is therefore a need for adequacy tests.

Several papers have recently proposed goodness-of-fit tests for copula models. Genest and Rivest (1993) propose a test for the Archimedean family. More recent papers are by Breyman, Dias, and Embrechts (2003), Fermanian (2003), and Malavergne and Sornette (2003). As argued by Fermanian and Scaillet (2004), there are some difficulties in designing a statistical test for the adequacy of copula functions to the data. The reason is that the initial variables have to be transformed in their probability integral transforms through the marginal distribution functions. Since the margins are unknown, they have to be treated as nuisance parameters.

A natural way is to adapt the test proposed in the multivariate distribution context by Diebold, Hahn, and Tay (1999) for copula functions. Such an adaptation has been proposed by Breymann, Dias, and Embrechts (2003). As in Section 6.2.4, we refer to the Rosenblatt (1952) transform and define the probability integral transform in terms of conditional distributions. Consider n random variables X_i whose joint distribution is given by $F_X(x_1, \cdots, x_n)$ and marginal distributions by $F_{X_i}(x_i) = \Pr[X_i \leq x_i]$. We define the probability integral transform $U_i = T(X_i)$ where

$$
\begin{aligned}
T(X_1) &= \Pr[X_1 \leq x_1] = F_{X_1}(x_1), \\
T(X_2) &= \Pr[X_2 \leq x_2 | X_1 = x_1] = F_{X_2|X_1}(x_2|x_1), \\
&\vdots \\
T(X_n) &= \Pr[X_n \leq x_n | X_1 = x_1, \cdots, X_{n-1} = x_{n-1}] \\
&= F_{X_n|X_1,\cdots,X_{n-1}}(x_n|x_1, \cdots, x_{n-1}),
\end{aligned}
$$

Therefore, the variables U_i for $i = 1, \cdots, n$ are *iid* $U(0,1)$ individually and jointly.

Suppose now that the copula C is such that $C(F_{X_1}(x_1), \cdots, F_{X_n}(x_n)) = F_X(x_1, \cdots, x_n)$. If we denote $C_i(u_1, \cdots, u_i) = C(u_1, \cdots, u_i, 1, \cdots, 1)$ the joint marginal distribution for the first i variables $(U_1, \cdots, U_i)$, then the conditional distribution of U_i given $(U_1, \cdots, U_{i-1})$ is

$$
C_i(u_i|u_1, \cdots, u_{i-1}) = \frac{\partial^{i-1} C_i(u_1, \cdots, u_i)}{\partial u_1 \cdots \partial u_{i-1}} \Big/ \frac{\partial^{i-1} C_{i-1}(u_1, \cdots, u_{i-1})}{\partial u_1 \cdots \partial u_{i-1}}.
$$

Since we have

$$
C_i\left(F_{X_i}(x_i)|F_{X_1}(x_1), \cdots, F_{X_{i-1}}(x_{i-1})\right) = F_{X_i|X_1,\cdots,X_{i-1}}(x_i|x_1, \cdots, x_{i-1}),
$$

we deduce that

$$U_i = C_i \left(F_{X_i}(x_i) | F_{X_1}(x_1), \cdots, F_{X_{i-1}}(x_{i-1}) \right) \qquad i = 2, \cdots, n,$$

(with $U_1 = C_1 \left(F_{X_1}(x_1) \right)$). Since the copula is a multivariate distribution function, it follows that the U_is are *iid* $U(0,1)$ (individually and jointly) if C is the true distribution.

These results are used by Breymann, Dias, and Embrechts (2003) to construct the following test. They propose to transform the U_is using the inverse of the univariate normal distribution $\Phi^{-1}(U_i)$, $i = 1, \cdots, n$, then the transformed variables $\Phi^{-1}(U_i)$ should be *iid* $\mathcal{N}(0,1)$. The proposed test statistic is

$$S = \sum_{i=1}^{n} \left(\Phi^{-1}(U_i) \right)^2,$$

that is asymptotically distributed, under the null, as a $\chi^2(n)$. This test is in fact an Anderson-Darling test.

A strong limitation of this test is that it is based on the empirical marginal distributions in order to transform the marginal data. Therefore, they should be treated as infinite dimensional nuisance parameters. This obviously affects the critical values of the test. Malevergne and Sornette (2003) suggest the use of bootstrap to compute the empirical critical values of the Anderson-Darling test.

Another difficulty, already mentioned for the test of Diebold, Hahn, and Tay (1999), is that the Rosenblatt transform allows $n!$ ways to factor the joint distribution in terms of conditional distributions.

To deal with the fact that the empirical distribution is unknown and has to be treated as nuisance parameters, we may adopt a nonparametric approach. For instance, Fermanian (2003) considers a kernel estimation of the empirical copula density to circumvent the direct use of the empirical copula process. He proposes a goodness-of-fit test based on the difference between the kernel estimator of the copula and the assumed copula.

6.3.7 Modeling the conditional dependency parameter

For notational convenience, we set $u_t \equiv F(x_t, w_x)$ and $v_t \equiv G(y_t, w_y)$. We denote by ρ the dependency parameter. It may be for instance the correlation parameter in the Gaussian or Student t copulas. The conditioning can be achieved by expressing ρ as a function of explanatory variables, for instance lagged values of u_t and v_t, or some other predetermined variable z_t. A rather general specification for ρ_t is

$$\rho_t = \Gamma(\underline{u}_{t-1}, \underline{v}_{t-1}, \underline{z}_{t-1}, \underline{\gamma}_{t-1}; w_\rho),$$

where Γ is a function depending on the parameter vector w_ρ and $\underline{u}_{t-1}$ denotes $\{u_{t-1}, u_{t-2}, \cdots\}$.

Many different specifications of the dependency parameter are possible in this context. As a first approach, we may follow Gouriéroux and Monfort (1992) and adopt a specification in which ρ_t depends on the position of past joint realizations in the unit square. This means that we decompose the unit square of joint past realizations into a grid. The parameter ρ_t will be constant for each element of the grid. More precisely, our basic model is

$$\log(\rho_t) = \sum_{j=1}^{16} d_j \, 1_{[(u_{t-1}, v_{t-1}) \in \mathcal{A}_j]},$$

where $\mathcal{A}_j$ is the jth element of the unit-square grid. To each parameter d_j, an area $\mathcal{A}_j$ is associated. Figure 6.22 illustrates the position of the areas d_js. In the figure, we have set equally spaced threshold levels, i.e., p_1, p_2, and p_3 take the values 0.25, 0.5, and 0.75. The same for q_1, q_2, and q_3. For instance, $\mathcal{A}_1 = [0, p_1[\times [0, q_1[$ and $\mathcal{A}_2 = [p_1, p_2[\times [0, q_1[$. The choice of 16 subintervals is somewhat arbitrary. This choice of parameterization has the advantage to provide an easy testing of several conjectures concerning the impact of past joint returns on subsequent dependency while still allowing for a large number of observations per area.

This specification does not allow the measurement of persistence in ρ_t, however. The difficulty is to derive an adequate model to capture the dynamic of the dependency parameter. As alternative approaches, we may adopt a specification close to the one proposed by Tse and Tsui (2002) or Engle (2002) in their modeling of the Pearson's correlation in a GARCH context. For instance, in the case of the time-varying conditional correlation model of Tse and Tsui, the dynamic of the Spearman's rho would be given by

$$\rho_t = (1 - \theta_1 - \theta_2)\, \rho + \theta_1 \, \rho_{t-1} + \theta_2 \, \psi_{t-1},$$

where $\psi_t = \left(\sum_{h=0}^{m-1} u_{t-h} v_{t-h} \right) / \left(\sum_{h=0}^{m-1} u_{t-h}^2 \sum_{h=0}^{m-1} v_{t-h}^2 \right)^{1/2}$ represents the correlation between the margins over the recent period. We impose that $0 \leq \theta_1, \theta_2 \leq 1$ and $\theta_1 + \theta_2 \leq 1$. The null hypothesis $\theta_1 = \theta_2 = 0$ can be tested using a standard Wald statistic.

Another alternative approach may be that parameters of the Student t copula, with degree-of-freedom parameter n and correlation ρ, are driven by a Markov-switching model of the type

$$\rho_t = \underline{\rho} S_t + \overline{\rho} \, (1 - S_t) \,, \tag{6.26}$$

$$n_t = \underline{n} S_t + \overline{n} \, (1 - S_t) \,, \tag{6.27}$$

where S_t denotes the unobserved regime of the system at time t. S_t is assumed to follow a two-state Markov process, with transition probability matrix given by

$$\begin{pmatrix} p & 1-p \\ 1-q & q \end{pmatrix},$$

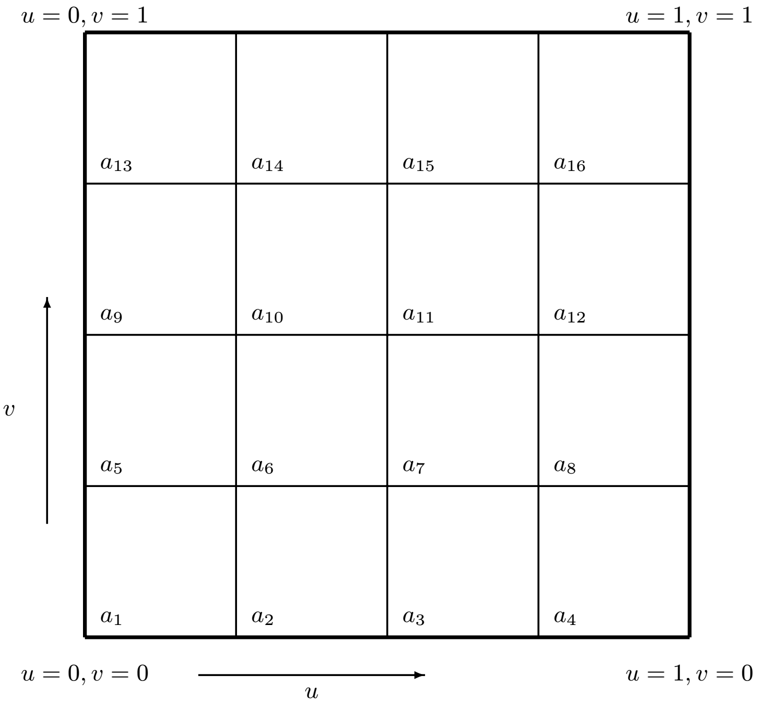

Fig. 6.22. *Unit-square grid for the dependency parameter.*

with

$$p = \Pr\left[S_t = 0|S_{t-1} = 0\right],$$
$$q = \Pr\left[S_t = 1|S_{t-1} = 1\right].$$

Note that, in this model, we do not necessarily assume that univariate characteristics of returns also shift. Quasi Maximum-Likelihood estimation of this model can be easily obtained using the approach developed by Hamilton (1989) and Gray (1995). For the degree-of-freedom parameter, we may investigate several hypotheses. For instance, we may test whether it is regime independent ($\underline{n} = \overline{n}$) or whether it is infinite, so that the Gaussian copula would prevail for a given regime.

6.3.8 Illustration

We now consider the estimation of copula functions for our two pairs of daily returns (SP500 and DAX on one hand and SP500 and FT-SE on the other hand) over the period from January 1980 to August 2004. To compare the

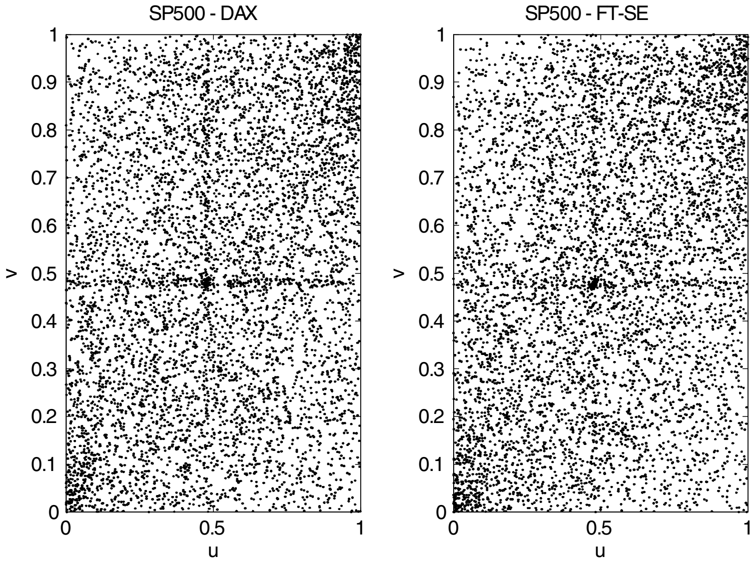

Fig. 6.23. *Scatterplot of margins.*

various copula functions at hand, we adopt a relatively simple specification for the margins. We assume that the dynamic of each daily return is given by a constant conditional mean and a GARCH$(1,1)$ conditional volatility. The marginal distribution is assumed to be either the normal or the Student t distribution. The estimation is performed using the two-step procedure described in Section 6.3.5: parameters of the marginal distributions are estimated in the first step (giving $\tilde{\theta}_x$ and $\tilde{\theta}_y$), then the margins $\hat{u}_t = F(x_t, \tilde{\theta}_x)$ and $\hat{v}_t = G(y_t, \tilde{\theta}_y)$ are computed and the parameters θ_γ of the copula $c_{\theta_\gamma}(\hat{u}_t, \hat{v}_t)$ are estimated (giving $\tilde{\theta}_\gamma$). Parameter estimates of the marginal distributions are not reported. They are very close to those presented in Chapter 4. We present in Figure 6.23 scatterplots of the marginal *cdf*s u_t and v_t for the SP500–DAX and for the SP500–FT-SE, respectively, when the margins are supposed to be Student t. We notice that, except for the region where one margin is large and the other small, the unit square is rather uniformly filled with realizations. In both figures, there is a higher concentration in the corners along the diagonal. This clustering corresponds to the observation that correlation may be higher in the tails.

Tables 6.5 and 6.6 report parameter estimates corresponding to these various models. For each copula function, we report the copula component of the log-likelihood ($\log c_{\theta_\gamma}(F(x_t, \theta_x), G(y_t, \theta_y))$), as well as concordance measures (Kendall's tau, Spearman's rho, and the lower-tail and upper-tail dependence measures). We first notice that the estimates of the dependence parameters are rather similar whatever the marginal distribution and the copula function. The parameters τ and ϱ_S are typically around 0.17 and 0.25 for the pair SP500–DAX and around 0.2 and 0.3 for the pair SP500–FT-SE. We notice that dependence is found to be more pronounced for the symmetric copulas than for the asymmetric ones. We notice that the Student t copula which has both lower-tail and upper-tail dependence is characterized by quite large degree-of-freedom parameters, so that no tail dependence is obtained. For this date, this finding can be confirmed using a test based on the extreme value theory, as illustrated in the next chapter. Finally, we may compare the log-likelihoods obtained with the various copula functions.[15] We observe that the copula that provides the best fit of the data at hand is the Student t copula.

Table 6.5. *Estimation of various copula functions (SP500–DAX)*

	log-lik.	Parameter estimate	Kendall's tau	Spearman's rho
SP500–DAX (Gaussian margins)				
Gaussian	220.9177	0.2576	0.1659	0.2467
Student t	259.2342	0.2718	0.1752	0.2588
		16.7154		
Frank	213.7218	1.7728	0.1898	0.2876
Plackett	226.8283	2.4512	0.1971	0.2911
Clayton	109.0424	0.1704	0.0785	0.1174
Rotated Clayton	177.7902	0.2973	0.1294	0.1926
Gumbel	224.3033	1.1914	0.1607	0.2369
Rotated Gumbel	101.5494	1.1617	0.1392	0.2058
SP500–DAX (Student t margins)				
Gaussian	217.5797	0.2594	0.1670	0.2484
Student t	249.0502	0.2592	0.1669	0.2464
		9.9642		
Frank	196.1103	1.5477	0.1663	0.2551
Plackett	206.1987	2.2003	0.1738	0.2575
Clayton	171.3745	0.2768	0.1216	0.1811
Rotated Clayton	182.6102	0.3185	0.1374	0.2043
Gumbel	221.9475	1.1884	0.1585	0.2338
Rotated Gumbel	203.9667	1.1725	0.1471	0.2174

[15] Since the estimation is performed using the two-step approach, the reported copula component of the log-likelihood can be compared for the various copula functions. However, it is not surprising that there is no systematic pattern when we compare the Gaussian margins with the Student t margins. Inspection of the total log-likelihood (including the margins) clearly indicates that the model with Student t margins performs better than the model with Gaussian margins.

Table 6.6. *Estimation of various copula functions (SP500–FT-SE)*

	log-lik.	Parameter estimate	Kendall's tau	Spearman's rho
SP500–FT-SE (Gaussian margins)				
Gaussian	323.1311	0.3263	0.2116	0.3129
Student t	367.9556	0.3294	0.2137	0.3141
		18.3911		
Frank	323.3922	2.1610	0.2289	0.3421
Plackett	337.0456	2.8816	0.2317	0.3401
Clayton	223.2368	0.3027	0.1314	0.1956
Rotated Clayton	210.9462	0.3479	0.1482	0.2201
Gumbel	284.0590	1.2376	0.1920	0.2818
Rotated Gumbel	255.8926	1.2148	0.1768	0.2602
SP500–FT-SE (Student t margins)				
Gaussian	336.0140	0.3185	0.2063	0.3054
Student t	364.5217	0.3202	0.2075	0.3044
		11.4373		
Frank	305.6729	1.9384	0.2067	0.3112
Plackett	317.3299	2.6161	0.2111	0.3110
Clayton	290.9546	0.3745	0.1577	0.2339
Rotated Clayton	242.3389	0.3848	0.1613	0.2392
Gumbel	303.7355	1.2374	0.1918	0.2816
Rotated Gumbel	330.7893	1.2266	0.1847	0.2714