Chapter 9
Multivariate Models

In this chapter, we consider multivariate models for the joint distribution of several risk factors such as returns or log returns for different assets, zero rate changes for different maturity times, changes in implied volatility, and losses due to defaults on risky loans. Our aim is to specify a good model for the future value $g(\mathbf{X})$ of a portfolio, where the function g is known and its argument $\mathbf{X}$ is a random vector of, for instance, log returns and zero rate changes over a given future time period. Since the function g is known, what remains is to make a good choice of probability distribution for random vector $\mathbf{X}$.

The first sections, Sects. 9.1–9.3, present spherical and elliptical distributions and their applicability in a wide range of problems in risk management. Elliptical distributions provide convenient and flexible multivariate models. This set of models includes the multivariate normal model but allows for a much wider range of tail behavior and dependence properties.

Elliptical distributions have the following important property: if $\mathbf{X}$ has an elliptical distribution, then the distribution of any linear combination $\mathbf{w}^T\mathbf{X}$ of its components is known. This property is useful because if $\mathbf{X}$ represents the returns of the financial assets in a portfolio, then we know the distribution of every linear portfolio. The property is useful even if we do not model the returns directly with an elliptical distribution. Suppose that $\mathbf{X}$ represents a vector of log returns, zero rate changes, etc. and is modeled by an elliptical distribution. If the portfolio value at some future time is given by $g(\mathbf{X})$, then a first order Taylor approximation of g around the mean vector $\boldsymbol{\mu} = \mathrm{E}[\mathbf{X}]$ leads to the first-order approximation

$$g(\mathbf{X}) \approx g(\boldsymbol{\mu}) + \sum_{k=1}^{d} \frac{\partial g}{\partial x_k}(\boldsymbol{\mu})(X_k - \mu_k).$$

The right-hand side is a linear combination of the components of $\mathbf{X}$ whose distribution therefore is known. Thus, whenever linearization of the nonlinear function g is justified, we can approximate the probability distribution of $g(\mathbf{X})$ analytically.

An important property of spherically distributed random vectors is that they can be decomposed into a product of a radial part and an independent angular part that is uniformly distributed on a sphere. This property makes it easy to simulate from a spherical (elliptical) distribution in any dimension. In particular, we can approximate the probability distribution of $g(\mathbf{X})$ arbitrarily well by simulating a large enough sample from $\mathbf{X}$ and consider the resulting empirical distribution of the simulated outcomes of $g(\mathbf{X})$.

A series of applications of elliptical distribution in risk management, including risk aggregation, solvency computations for an insurance company, and hedging of options, is presented in Sect. 9.3.

Then we turn our attention to multivariate models for random vectors that do not show signs of elliptical symmetry, and the notion of copula is introduced in Sect. 9.4. On the one hand, the copula is just a multivariate distribution function appearing in the representation of a multivariate distribution function in terms of its (continuous) marginal distribution functions. On the other hand, the copula may be identified as the dependence structure of a multivariate distribution, and by varying the copula for a random vector $\mathbf{X}$ for which the distributions of the components X_k are held fixed, we may understand better the effect of the dependence between the X_k on the distribution for the future portfolio value $g(\mathbf{X})$. We rarely have sufficient information to accurately specify the copula of a random vector $\mathbf{X}$, and by varying the copula within a set of copula functions, we may study the robustness of the distribution of the portfolio value $g(\mathbf{X})$ to misspecifications of the dependence between the components of $\mathbf{X}$. Moreover, the representation of a multivariate model for $\mathbf{X}$ in terms of a copula and distribution functions for the X_k is useful for simulation from the distribution of $\mathbf{X}$: an outcome from $\mathbf{X}$ is constructed as an outcome from the copula together with an application of the quantile transform.

Finally, in Sect. 9.5, we consider the effect of dependence modeling for large homogeneous portfolios. We consider a high-dimensional random vector $\mathbf{X}$ with equally distributed components and study the effect of the dependence between the components on the distribution of the sum of the components of $\mathbf{X}$.

9.1   Spherical Distributions

A random vector $\mathbf{Y}$ has a spherical distribution in $\mathbb{R}^d$ if its distribution is spherically symmetric. In other words, its distribution is invariant under rotations and reflections. Linear transformations that represent rotations and reflections correspond to multiplication by orthogonal matrices. Recall that a matrix $\mathbf{O}$ is orthogonal if it has real entries and $\mathbf{O}\mathbf{O}^{\mathrm{T}} = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix. Formally, $\mathbf{Y}$ has a spherical distribution if

$$\mathbf{O}\mathbf{Y} \stackrel{\mathrm{d}}{=} \mathbf{Y} \quad \text{for every orthogonal matrix } \mathbf{O}. \tag{9.1}$$

Figure 9.1 shows scatter plots of samples from two spherical distributions.
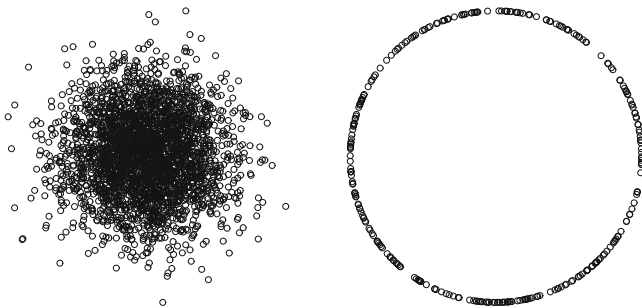
Fig. 9.1 *Left plot*: sample of size 3,000 from bivariate standard normal distribution. *Right plot*: sample of size 300 from uniform distribution on the unit circle

Three examples of spherical distributions are presented below. Before presenting the examples, let us recall the definition of the multivariate normal distribution.

(1) A random vector $\mathbf{Z}$ has standard normal distribution $N_d(\mathbf{0}, \mathbf{I})$ if $\mathbf{Z} = (Z_1, \ldots, Z_d)^{\mathrm{T}}$, where $Z_1, \ldots, Z_d$ are independent and $N(0, 1)$-distributed.
(2) A random vector $\mathbf{X}$ is $N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$-distributed if $\mathbf{X} \stackrel{\mathrm{d}}{=} \boldsymbol{\mu} + \mathbf{A}\mathbf{Z}$, where $\mathbf{A}\mathbf{A}^{\mathrm{T}} = \boldsymbol{\Sigma}$ and $\mathbf{Z}$ is $N_d(\mathbf{0}, \mathbf{I})$-distributed.

Example 9.1 (Standard normal distribution). The first example of a spherical distribution is the standard normal distribution $N_d(\mathbf{0}, \mathbf{I})$. Let $\mathbf{Z}$ have a $N_d(\mathbf{0}, \mathbf{I})$ distribution, and let $\mathbf{O}$ be an arbitrary orthogonal matrix. By property (2) above, $\mathbf{O}\mathbf{Z}$ has the distribution $N_d(\mathbf{0}, \mathbf{O}\mathbf{O}^{\mathrm{T}})$. Since $\mathbf{O}\mathbf{O}^{\mathrm{T}} = \mathbf{I}$, we conclude that $\mathbf{Z}$ satisfies (9.1). The left plot in Fig. 9.1 shows a scatter plot of a sample from $N_d(\mathbf{0}, \mathbf{I})$.

Example 9.2 (Standard normal variance mixture). Another example of a spherical distribution is obtained by multiplying a $N_d(\mathbf{0}, \mathbf{I})$-distributed random vector $\mathbf{Z}$ by an independent nonnegative random variable W. Notice that, for any orthogonal matrix $\mathbf{O}$,
$$\mathbf{O}W\mathbf{Z} = W\mathbf{O}\mathbf{Z} \stackrel{\mathrm{d}}{=} W\mathbf{Z},$$
where the last equality follows since $\mathbf{Z}$ is spherically distributed.

The uniform distribution on the unit sphere $\mathbb{S}^{d-1} = \{\mathbf{x} \in \mathbb{R}^d : |\mathbf{x}| = 1\}$, where $|\mathbf{x}|^2 = \mathbf{x}^{\mathrm{T}}\mathbf{x}$, assigns equal probability to any two subsets of $\mathbb{S}^{d-1}$ with the same surface area.

Example 9.3 (Uniform distribution on the unit sphere). A third example of a spherical distribution is the uniform distribution on the unit sphere, i.e., the probability mass is distributed uniformly on the unit sphere $\mathbb{S}^{d-1}$. Let $\mathbf{U}$ be uniformly distributed on the unit sphere and consider a subset A of the unit sphere. For any orthogonal matrix $\mathbf{O}$ it holds that

$$P(\mathbf{O}\mathbf{U} \in A) = P(\mathbf{U} \in \mathbf{O}^{-1}A) = P(\mathbf{U} \in \mathbf{O}^{\mathrm{T}}A) = P(\mathbf{U} \in A),$$

where the last equality holds because $\mathbf{O}$ is an orthogonal matrix and therefore A and $\mathbf{O}^T A$ have the same surface area. Therefore, $\mathbf{U}$ is spherically distributed. The right plot in Fig. 9.1 shows a sample from the uniform distribution on the unit circle.

The following property is a key property of spherical distributions.

Proposition 9.1. *If $\mathbf{a}$ is an arbitrary vector in $\mathbb{R}^d$ and $\mathbf{Y}$ is spherically distributed and of the same dimension, then $\mathbf{a}^T\mathbf{Y} \overset{\mathrm{d}}{=} |\mathbf{a}|Y_1$.*

Proof. Take $\mathbf{a} \neq \mathbf{0}$, let $\mathbf{u} = \mathbf{a}/|\mathbf{a}|$, and pick an orthogonal matrix $\mathbf{O}$ whose first row is equal to $\mathbf{u}^T$. Since $\mathbf{OY} \overset{\mathrm{d}}{=} \mathbf{Y}$, it follows that $\mathbf{a}^T\mathbf{Y} = |\mathbf{a}|\mathbf{u}^T\mathbf{Y} = |\mathbf{a}|(\mathbf{OY})_1 \overset{\mathrm{d}}{=} |\mathbf{a}|Y_1$. $\qquad\square$

The following property is another key property of spherical distributions.

Proposition 9.2. *If $\mathbf{Y}$ is spherically distributed, then $\mathbf{Y} \overset{\mathrm{d}}{=} R\mathbf{U}$, where $R \overset{\mathrm{d}}{=} |\mathbf{Y}|$, $\mathbf{U}$ is uniformly distributed on the unit sphere and R and $\mathbf{U}$ are independent. Moreover, $\mathrm{P}(\mathbf{Y}/|\mathbf{Y}| \in \cdot \mid |\mathbf{Y}| > 0) = \mathrm{P}(\mathbf{U} \in \cdot)$.*

The proposition provides a way to simulate from a spherical distribution. First draw a vector from the uniform distribution on the unit sphere by sampling from a standard normal distribution and dividing by its norm. Then draw the radial part by sampling from the distribution of $|\mathbf{Y}|$.

To prove Proposition 9.2, we first state and prove the following lemma.

Lemma 9.1. *The uniform distribution on the unit sphere is the unique spherical distribution on the unit sphere.*

Proof. Let $\mathbf{Z}$ have a spherical distribution on the unit sphere. For any orthogonal matrix $\mathbf{O}$ and subset A of $\mathbb{S}^{d-1}$ it holds that $\mathrm{P}(\mathbf{Z} \in \mathbf{O}A) = \mathrm{P}(\mathbf{O}^T\mathbf{Z} \in \mathbf{O}^T\mathbf{O}A) = \mathrm{P}(\mathbf{Z} \in A)$ since $\mathbf{Z}$ is spherically distributed and $\mathbf{O}^T$ is an orthogonal matrix. If $\mathbf{Z}$ were not uniformly distributed on $\mathbb{S}^{d-1}$, then there would exist a subset A_0 of $\mathbb{S}^{d-1}$ and an orthogonal matrix $\mathbf{O}_0$ such that $\mathrm{P}(\mathbf{Z} \in A) \neq \mathrm{P}(\mathbf{Z} \in \mathbf{O}_0 A_0)$, which contradicts that $\mathbf{Z}$ is spherically distributed. $\qquad\square$

Proof of Proposition 9.2. It is sufficient to show that $\mathrm{P}(|\mathbf{Y}| > r, \mathbf{Y}/|\mathbf{Y}| \in A) = \mathrm{P}(|\mathbf{Y}| > r)\,\mathrm{P}(\mathbf{U} \in A)$ for any $r \geq 0$ and any subset A of $\mathbb{S}^{d-1}$, where $\mathbf{U}$ is uniformly distributed on the unit sphere.

We claim that, for any $r \geq 0$, $I\{|\mathbf{Y}| > r\}\mathbf{Y}/|\mathbf{Y}|$ is spherically distributed. To prove the claim, note that for any orthogonal matrix $\mathbf{O}$ it holds that $|\mathbf{OY}| = |\mathbf{Y}|$ and $\mathbf{OY} \overset{\mathrm{d}}{=} \mathbf{Y}$ and therefore

$$\mathbf{O}I\{|\mathbf{Y}| > r\}\mathbf{Y}/|\mathbf{Y}| = I\{|\mathbf{OY}| > r\}\mathbf{OY}/|\mathbf{OY}| \overset{\mathrm{d}}{=} I\{|\mathbf{Y}| > r\}\mathbf{Y}/|\mathbf{Y}|.$$

To complete the proof of Proposition 9.2, we may without loss of generality take r such that $\mathrm{P}(|\mathbf{Y}| > r) > 0$ and note that

$$P(\mathbf{Y}/|\mathbf{Y}| \in A \mid |\mathbf{Y}| > r) = P(I\{|\mathbf{Y}| > r\}\mathbf{Y}/|\mathbf{Y}| \in A)/P(|\mathbf{Y}| > r)$$

$$= P(I\{|\mathbf{Y}| > r\}\mathbf{Y}/|\mathbf{Y}| \in \mathbf{O}A)/P(|\mathbf{Y}| > r)$$

$$= P(\mathbf{Y}/|\mathbf{Y}| \in \mathbf{O}A \mid |\mathbf{Y}| > r).$$

It now follows from Lemma 9.1 that $P(\mathbf{Y}/|\mathbf{Y}| \in A \mid |\mathbf{Y}| > r) = P(\mathbf{U} \in A)$, and therefore $P(|\mathbf{Y}| > r, \mathbf{Y}/|\mathbf{Y}| \in A) = P(|\mathbf{Y}| > r)\,P(\mathbf{U} \in A)$. $\qquad\square$

9.2   Elliptical Distributions

The multivariate normal distribution is very useful in the construction of multivariate models. Its popularity derives primarily from the fact that it is tractable, allowing for explicit calculations, and it that can be motivated asymptotically by the central limit theorem. For univariate data that show clear signs of symmetry the univariate normal distribution does not necessarily give a good fit to the data. Typically normal tails do not match empirical tails particularly well. Similarly, the multivariate normal distribution is often at best a reasonable first approximation for samples of multivariate observations with clear signs of elliptical symmetry.

A random vector $\mathbf{X}$ has a $N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution if

$$\mathbf{X} \stackrel{\mathrm{d}}{=} \boldsymbol{\mu} + \mathbf{A}\mathbf{Z}, \tag{9.2}$$

where $\mathbf{A}\mathbf{A}^{\mathrm{T}} = \boldsymbol{\Sigma}$ and $\mathbf{Z}$ has a $N_d(\mathbf{0}, \mathbf{I})$ distribution. An easy way to obtain a richer class of multivariate distributions, which share many of the tractable properties of the multivariate normal distribution, is to replace the standard normal vector $\mathbf{Z}$ in (9.2) by an arbitrary spherically distributed random vector $\mathbf{Y}$. Formally, a random vector $\mathbf{X}$ has an elliptical distribution if there exist a vector $\boldsymbol{\mu}$, a matrix $\mathbf{A}$, and a spherically distributed vector $\mathbf{Y}$ such that

$$\mathbf{X} \stackrel{\mathrm{d}}{=} \boldsymbol{\mu} + \mathbf{A}\mathbf{Y}. \tag{9.3}$$

The matrix $\mathbf{A}$ and the spherical distribution of $\mathbf{Y}$ in (9.3) are not determined by the distribution of $\mathbf{X}$: we may replace the pair $(\mathbf{A}, \mathbf{Y})$ in (9.3) by $(c\mathbf{A}, c^{-1}\mathbf{Y})$ for any constant $c \in (0, \infty)$. A matrix $\boldsymbol{\Sigma}$ satisfying $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^{\mathrm{T}}$ is called a dispersion matrix of the elliptically distributed vector $\mathbf{X}$. If the covariance matrix $\mathrm{Cov}(\mathbf{X})$ exists finitely, then $\mathrm{Cov}(\mathbf{X}) = c\,\boldsymbol{\Sigma}$ for some constant $c \in (0, \infty)$. To verify this claim, we note that, by (9.3) and Proposition 9.2,

$$\mathrm{Cov}(\mathbf{X}) = \mathrm{E}[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^{\mathrm{T}}] = \mathrm{E}[R^2]\mathbf{A}\,\mathrm{E}[\mathbf{U}\mathbf{U}^{\mathrm{T}}]\mathbf{A}^{\mathrm{T}} = \frac{\mathrm{E}[R^2]}{d}\boldsymbol{\Sigma}.$$

The last equality above can be proven as follows. Consider a standard normally distributed vector $\mathbf{Z}$ and recall that $\mathbf{Z}/|\mathbf{Z}|$ is uniformly distributed on the unit sphere and $\mathrm{E}[|\mathbf{Z}|^2] = d$. Therefore,

$$\mathbf{I} = \mathrm{Cov}(\mathbf{Z}) = \mathrm{E}[|\mathbf{Z}|^2]\,\mathrm{E}[\mathbf{U}\mathbf{U}^{\mathrm{T}}] = d\,\mathrm{E}[\mathbf{U}\mathbf{U}^{\mathrm{T}}].$$

For a dispersion matrix $\boldsymbol{\Sigma}$ with nonzero diagonal entries we define the linear correlation parameter $\rho_{ij} = \Sigma_{ij}/(\Sigma_{ii}\Sigma_{jj})^{1/2}$. If $\mathrm{Cov}(\mathbf{X})$ exists finitely, then $\rho_{ij} = \mathrm{Cor}(X_i, X_j)$, i.e., the linear correlation parameter coincides with the ordinary linear correlation coefficient.

The normal variance mixture distributions are the distributions of random vectors with stochastic representation

$$\mathbf{X} \stackrel{\mathrm{d}}{=} \boldsymbol{\mu} + W\mathbf{AZ}, \tag{9.4}$$

where $\mathbf{A}$ and $\mathbf{Z}$ are the same as in (9.2) and W is a nonnegative random variable independent of $\mathbf{Z}$. From Example 9.2 it follows that a normal variance mixture distribution is an elliptical distribution. By conditioning on $W = w$, we see that $\mathbf{X}|W = w$ is $\mathrm{N}_d(\boldsymbol{\mu}, w^2\boldsymbol{\Sigma})$-distributed, which explains the name normal variance mixture. If $\mathrm{E}[W^2] < \infty$, then $\mathbf{X}$ has a well-defined mean vector $\boldsymbol{\mu} = \mathrm{E}[\mathbf{X}]$ and covariance matrix

$$\mathrm{Cov}(\mathbf{X}) = \mathrm{E}[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^{\mathrm{T}}] = \mathrm{E}[W^2]\mathbf{A}\,\mathrm{E}[\mathbf{ZZ}^{\mathrm{T}}]\mathbf{A}^{\mathrm{T}} = \mathrm{E}[W^2]\boldsymbol{\Sigma}.$$

Example 9.4 (Multivariate Student's t). If we take $W^2 \stackrel{\mathrm{d}}{=} \nu/S_\nu$, where S_ν has a Chi-square distribution with ν degrees of freedom, then the resulting distribution of $\mathbf{X} = \boldsymbol{\mu} + W\mathbf{AZ}$ is called a multivariate Student's t distribution with ν degrees of freedom, written $t_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$. Note that $\boldsymbol{\Sigma}$ is not the covariance matrix of $\mathbf{X}$. Since $\mathrm{E}[W^2] = \nu/(\nu-2)$ if $\nu > 2$, it follows that $\mathrm{Cov}(\mathbf{X}) = (\nu/(\nu-2))\boldsymbol{\Sigma}$.

For a normally distributed random vector $\mathbf{X} \stackrel{\mathrm{d}}{=} \boldsymbol{\mu} + \mathbf{AZ}$, where $\mathbf{AA}^{\mathrm{T}} = \boldsymbol{\Sigma}$, any linear combination of the components of $\mathbf{X}$ is again normally distributed. That is, for any nonrandom vector $\mathbf{w}$ of the same dimension,

$$\begin{aligned}
\mathbf{w}^{\mathrm{T}}\mathbf{X} &\stackrel{\mathrm{d}}{=} \mathbf{w}^{\mathrm{T}}\boldsymbol{\mu} + \mathbf{w}^{\mathrm{T}}\mathbf{AZ} \\
&= \mathbf{w}^{\mathrm{T}}\boldsymbol{\mu} + (\mathbf{A}^{\mathrm{T}}\mathbf{w})^{\mathrm{T}}\mathbf{Z} \\
&\stackrel{\mathrm{d}}{=} \mathbf{w}^{\mathrm{T}}\boldsymbol{\mu} + (\mathbf{w}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{w})^{1/2}Z_1.
\end{aligned}$$

A similar property holds for arbitrary elliptical distributions.

Proposition 9.3. *If $\mathbf{X}$ has an elliptical distribution with stochastic representation $\mathbf{X} \stackrel{\mathrm{d}}{=} \boldsymbol{\mu} + \mathbf{AY}$, where $\mathbf{Y}$ is spherically distributed, then for any vector $\mathbf{a}$ of the same dimension $\mathbf{a}^{\mathrm{T}}\mathbf{X} \stackrel{\mathrm{d}}{=} \mathbf{a}^{\mathrm{T}}\boldsymbol{\mu} + (\mathbf{a}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{a})^{1/2}Y_1$, where $\boldsymbol{\Sigma} = \mathbf{AA}^{\mathrm{T}}$.*

The proof is omitted since the result follows immediately from Proposition 9.1 and the defining property (9.3) of elliptical distributions.

As was previously mentioned, normal variance mixture distributions and, more generally, elliptical distributions share many of the attractive properties of normal distributions. However, there are important exceptions. Recall that the components of the $\mathrm{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$-distributed vector $\boldsymbol{\mu} + \mathbf{AZ}$ are independent if and only if $\mathbf{AA}^{\mathrm{T}} = \boldsymbol{\Sigma}$ is a diagonal matrix, that is, if the components are uncorrelated. This property does not hold for arbitrary normal variance mixture distributions. If $\mathbf{X} \stackrel{\mathrm{d}}{=} \boldsymbol{\mu} + W\mathbf{AZ}$

with $\mathbf{A}\mathbf{A}^{\mathrm{T}} = \boldsymbol{\Sigma}$ a diagonal matrix, then the components of $\mathbf{X}$ are uncorrelated. If $\boldsymbol{\Sigma}$ is a diagonal matrix with strictly positive diagonal entries, then $(X_k, X_l) \stackrel{\mathrm{d}}{=} (\mu_k + WA_{k,k}Z_k, \mu_l + WA_{l,l}Z_l)$, where $A_{k,k}, A_{l,l} > 0$. Clearly, X_k and X_l are not independent unless W is a constant.

The sum of independent elliptically distributed random vectors with the same (up to a constant factor) dispersion matrix is elliptically distributed.

Proposition 9.4. *If the random vectors $\mathbf{X}_1$ and $\mathbf{X}_2$ in $\mathbb{R}^d$ are independent and elliptically distributed with common dispersion matrix $\boldsymbol{\Sigma}$, then $\mathbf{X}_1 + \mathbf{X}_2$ is elliptically distributed.*

Proof. For a matrix $\mathbf{A}$ such that $\mathbf{A}\mathbf{A}^{\mathrm{T}} = \boldsymbol{\Sigma}$ we may write $\mathbf{X}_1 + \mathbf{X}_2 \stackrel{\mathrm{d}}{=} \boldsymbol{\mu}_1 + \boldsymbol{\mu}_2 + \mathbf{A}(\mathbf{Y}_1 + \mathbf{Y}_2)$ for some independent spherically distributed vectors $\mathbf{Y}_1$ and $\mathbf{Y}_2$. It remains to show that $\mathbf{Y}_1 + \mathbf{Y}_2$ is spherically distributed. For every orthogonal matrix $\mathbf{O}$ and $\mathbf{y}$ in $\mathbb{R}^d$,

$$P(\mathbf{O}(\mathbf{Y}_1 + \mathbf{Y}_2) \le \mathbf{y}) = \int P(\mathbf{O}\mathbf{Y}_1 + \mathbf{z} \le \mathbf{y} \mid \mathbf{O}\mathbf{Y}_2 = \mathbf{z}) dF_{\mathbf{O}\mathbf{Y}_2}(\mathbf{z}) d\mathbf{z}$$

$$= \int P(\mathbf{Y}_1 + \mathbf{z} \le \mathbf{y}) dF_{\mathbf{Y}_2}(\mathbf{z}) d\mathbf{z}$$

$$= P(\mathbf{Y}_1 + \mathbf{Y}_2 \le \mathbf{y}),$$

i.e., $\mathbf{O}(\mathbf{Y}_1 + \mathbf{Y}_2) \stackrel{\mathrm{d}}{=} \mathbf{Y}_1 + \mathbf{Y}_2$, from which the conclusion follows. □

Example 9.5 (Summation of log returns). Consider a set of identically distributed and uncorrelated random variables $X_1, \ldots, X_n$ that represent future daily log returns for some asset. Suppose that each log return has a finite mean μ and standard deviation σ. If the log returns are independent, then by the central limit theorem, $X_1 + \cdots + X_n$ is approximately $\mathrm{N}(n\mu, n\sigma^2)$-distributed for n large. If the vector $\mathbf{X} = (X_1, \ldots, X_n)^{\mathrm{T}}$ of log returns has an elliptical distribution, then Proposition 9.3 implies that

$$X_1 + \cdots + X_n = \mathbf{1}^{\mathrm{T}}\mathbf{X} \stackrel{\mathrm{d}}{=} n\mu + n^{1/2}(X_1 - \mu).$$

We see that the n-day log return and the 1-day log return belong to the same location-scale family of distributions. For instance, if the 1-day log return has a heavy-tailed Student's t distribution with a low-degree-of-freedom parameter, then so does the n-day log return.

9.2.1 Goodness of Fit of an Elliptical Model

Consider a random vector $\mathbf{X}$ with an elliptical distribution with representation $\mathbf{X} = \boldsymbol{\mu} + \mathbf{A}\mathbf{Y}$, where $\mathbf{Y}$ has a spherical distribution and $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^{\mathrm{T}}$ is invertible. By Proposition 9.3,

$$\mathbf{w}^{\mathrm{T}}\mathbf{X} \stackrel{\mathrm{d}}{=} \mathbf{w}^{\mathrm{T}}\boldsymbol{\mu} + (\mathbf{w}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{w})^{1/2}Y_1 \quad \text{for all nonrandom vectors } \mathbf{w} \ne \mathbf{0}$$

or, equivalently,

$$\frac{\mathbf{w}^{\mathrm{T}}\mathbf{X} - \mathbf{w}^{\mathrm{T}}\boldsymbol{\mu}}{(\mathbf{w}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{w})^{1/2}} \stackrel{\mathrm{d}}{=} Y_1 \quad \text{for all nonrandom vectors } \mathbf{w} \neq \mathbf{0}. \tag{9.5}$$

The property (9.5) can be used to investigate whether or not a multivariate sample is likely to come from an elliptical distribution. Let us illustrate the procedure by an example.

Example 9.6 (Estimation and fit of an elliptical model). Consider a sample of size 500 of pairs of daily log returns for the Dow Jones Industrial Average (DJIA) and Nasdaq Composite indices (index values from November 11, 2008 until November 4, 2010). The scatter plot of the pairs of log returns is shown in the upper left plot in Fig. 9.2. The log-return sample is denoted $\{\mathbf{x}_1, \ldots, \mathbf{x}_{500}\}$. We assume initially that the sample can be seen as outcomes from an elliptically distributed vector $\mathbf{X}$ and investigate whether this assumption can be rejected or not. If it is not rejected, then we also want to determine the elliptical distribution of $\mathbf{X}$. We assume that the location parameter $\boldsymbol{\mu}$ and a scalar multiple of the shape parameter $\mathbf{C} = c\boldsymbol{\Sigma}$, which is assumed invertible, can be estimated. Note that (9.5) can be expressed as

$$\frac{\mathbf{w}^{\mathrm{T}}\mathbf{X} - \mathbf{w}^{\mathrm{T}}\boldsymbol{\mu}}{(\mathbf{w}^{\mathrm{T}}\mathbf{C}\mathbf{w})^{1/2}} \stackrel{\mathrm{d}}{=} c^{-1/2}Y_1 \quad \text{for all nonrandom vectors } \mathbf{w} \neq \mathbf{0}.$$

If the covariance matrix $\mathrm{Cov}(\mathbf{X})$ exists finitely, then $\boldsymbol{\mu} = \mathrm{E}[\mathbf{X}]$, and we may take $\mathbf{C} = \mathrm{Cov}(\mathbf{X})$. Here we estimate $\boldsymbol{\mu}$ and $\mathbf{C}$ by the sample mean and sample covariance, respectively. The estimates are denoted $\widehat{\boldsymbol{\mu}}$ and $\widehat{\mathbf{C}}$. Consider a large set of vectors $\{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$ of unit length. For each $\mathbf{w}_k$ we construct the sample $\{y_{k,1}, \ldots, y_{k,500}\}$ by

$$y_{k,l} = \frac{\mathbf{w}_k^{\mathrm{T}}\mathbf{x}_l - \mathbf{w}_k^{\mathrm{T}}\widehat{\boldsymbol{\mu}}}{(\mathbf{w}_k^{\mathrm{T}}\widehat{\mathbf{C}}\mathbf{w}_k)^{1/2}} \quad \text{for } k = 1, \ldots, n, \quad l = 1, \ldots, 500.$$

Each such sample can be viewed as a sample from $c^{-1/2}Y_1$. If the data were generated by the elliptical distribution of $\mathbf{X}$, then all the n constructed samples must come from the same distribution, the distribution of $c^{-1/2}Y_1$. By overlaying the n q–q plots of the empirical quantiles for the n samples against the quantiles of a chosen reference distribution, we can check graphically whether the data appear to be consistent with an elliptical distribution or not. Moreover, the distribution of $c^{-1/2}Y_1$ can be estimated from the q–q plots.

Here we take $n = 100$ and sample the $\mathbf{w}_k$ randomly from the uniform distribution of the unit sphere by setting $\mathbf{w}_k = \mathbf{z}_k/|\mathbf{z}_k|$, where the $\mathbf{z}_k$ are outcomes of independent $\mathrm{N}_2(\mathbf{0}, \mathbf{I})$-distributed random vectors. The upper left plot in Fig. 9.2 is a scatter plot of the sample $\{\mathbf{x}_1, \ldots, \mathbf{x}_{500}\}$. The upper right plot in Fig. 9.2 shows the $n = 100$ q–q plots of the empirical quantiles of the samples $\{y_{k,1}, \ldots, y_{k,500}\}$ (y-axis) against the quantiles of the standard normal distribution (x-axis). The q–q plots indicate a reasonable fit to a common distribution with heavier tails than the normal distribution.
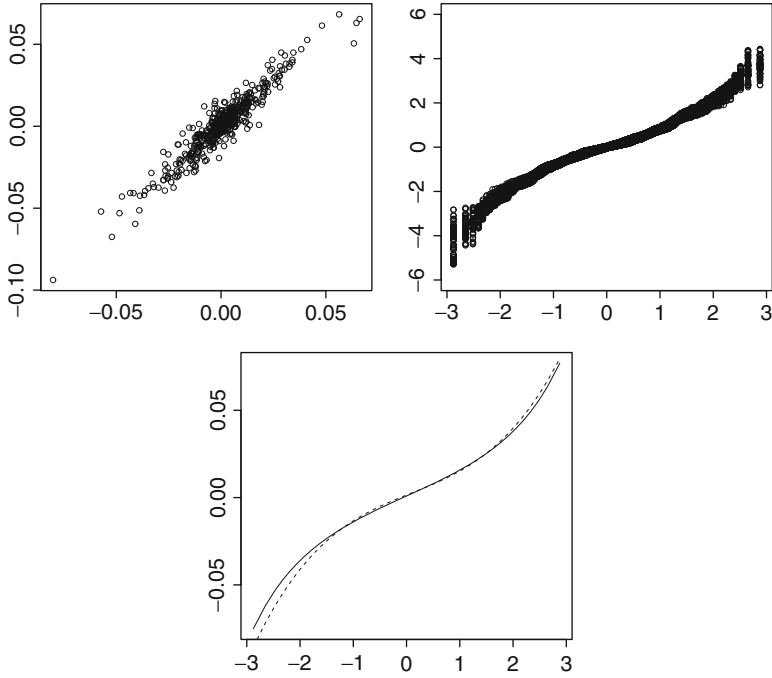
Fig. 9.2 *Scatter plot* showing pairs (x_D, x_N) of DJIA and Nasdaq log returns. *Upper right plot*: 100 overlaid q–q plots for empirical quantiles for each of 100 samples $\{y_{k,1}, \ldots, y_{k,500}\}$ (y-axis) against standard normal quantiles (x-axis). The *solid curve* in the *lower plot* shows the quantiles of the model for the Nasdaq log returns based on the fitted bivariate Student's t model (y-axis) against standard normal quantiles (x-axis). The *dashed curve* in the *lower plot* shows the polynomial normal quantiles (Example 8.11) fitted to the Nasdaq log returns (y-axis) against the standard normal quantiles (x-axis)

Under the assumption that $\{\mathbf{x}_1, \ldots, \mathbf{x}_{500}\}$ is a sample from the bivariate Student's t_ν distribution with $\nu > 2$ (otherwise it does not make sense to use the sample covariance matrix) and under the assumption that $\widehat{\boldsymbol{\mu}} = \boldsymbol{\mu}$ and $\widehat{\mathbf{C}} = \mathrm{Cov}(\mathbf{X})$, it holds that all the samples $\{y_{k,1}, \ldots, y_{k,500}\}$ are samples from the distribution of $((\nu-2)/\nu)^{1/2}Z$, where Z is standard t_ν-distributed. Least-squares estimation based on all 100 univariate samples gives the estimate $\widehat{\nu} \approx 4.09$. The selected model for the sample $\{\mathbf{x}_1, \ldots, \mathbf{x}_{500}\}$ is the distribution $t_2(\widehat{\boldsymbol{\mu}}, ((\widehat{\nu}-2)/\widehat{\nu})\widehat{\mathbf{C}}, \widehat{\nu})$.

The second marginal distribution (a univariate Student's t distribution) of the bivariate Student's t distribution for the pair of DJIA and Nasdaq log returns provides a model for the Nasdaq log returns. In the lower plot in Fig. 9.2, we compare this model to the polynomial normal model in Example 8.11. The solid curve in the lower plot is a q–q plot of the quantiles for the model for the Nasdaq log returns (y-axis) against standard normal quantiles (x-axis). The quantiles of the

fitted polynomial normal model in Example 8.11 are plotted against the standard
normal quantiles as the dashed curve in the lower plot in Fig. 9.2. It is hard to
distinguish between the two models.

9.2.2   Asymptotic Dependence and Rank Correlation

We now introduce general notions of dependence and study them in the context of
elliptical distributions.

The first notion of dependence measures the dependence of extreme values and
is called tail dependence or asymptotic dependence. Consider a pair (X_1, X_2) of
random variables with equally distributed components. We say that X_1 and X_2
are asymptotically dependent in the lower left tail if the limit $\lim_{x \to -\infty} P(X_2 \leq x \mid X_1 \leq x)$, the coefficient of lower tail dependence, is strictly positive and
asymptotically independent if the limit is zero.

Proposition 9.5. *If (X_1, X_2) has a bivariate standard normal distribution with
linear correlation coefficient $\rho < 1$, then $\lim_{x \to -\infty} P(X_2 \leq x \mid X_1 \leq x) = 0$.*

Proof. First note that $P(X_2 \leq x \mid X_1 \leq x) = P(X_1 \leq x, X_2 \leq x)/\Phi(x)$ and that
$(X_1, X_2) \stackrel{\mathrm{d}}{=} (Z_1, \rho Z_1 + (1 - \rho^2)^{1/2} Z_2)$, where Z_1, Z_2 are independent and standard
normally distributed. If $\rho = -1$, then the statement of the proposition holds, so we
may without loss of generality assume that $|\rho| < 1$. We may write

$$P(X_1 \leq x, X_2 \leq x) = \int_{-\infty}^{\infty} P\left(Z_1 \leq x, \rho Z_1 + (1 - \rho^2)^{1/2} t \leq x\right) \phi(t) dt$$

$$= \int_{-\infty}^{a(x)} \Phi(x) \phi(t) dt + \int_{a(x)}^{\infty} \Phi((x - (1 - \rho^2)^{1/2} t)/\rho) \phi(t) dt,$$

where $a(x) = ((1 - \rho)/(1 + \rho))^{1/2} x$. Therefore,

$$\lim_{x \to -\infty} P(X_2 \leq x \mid X_1 \leq x) = \lim_{x \to -\infty} \left(\Phi(a(x)) + \frac{\int_{a(x)}^{\infty} \Phi((x - (1 - \rho^2)^{1/2} t)/\rho) \phi(t) dt}{\Phi(x)} \right)$$

$$= \lim_{x \to -\infty} \frac{\int_{a(x)}^{\infty} \Phi((x - (1 - \rho^2)^{1/2} t)/\rho) \phi(t) dt}{\Phi(x)}.$$

Applying l'Hôpital's rule gives

$$\lim_{x \to -\infty} P(X_2 \leq x \mid X_1 \leq x)$$

$$= - \lim_{x \to -\infty} \frac{\Phi(x)}{\phi(x)} \left(\frac{1 - \rho}{1 + \rho} \right)^{1/2} + \lim_{x \to -\infty} \frac{1}{\rho \phi(x)} \int_{a(x)}^{\infty} \phi((x - (1 - \rho^2)^{1/2} t)/\rho) \phi(t) dt.$$

We saw in Example 8.1 that $\Phi(x) \sim -\phi(x)/x$ as $x \to -\infty$, so we only need to compute the last limit given above. By writing up explicitly the standard normal densities and making a substitution of integration variable, we arrive at

$$\frac{1}{\rho\phi(x)} \int_{a(x)}^{\infty} \phi((x - (1 - \rho^2)^{1/2}t)/\rho)\phi(t)dt = \int_{-\infty}^{a(x)} \phi(u)du = \Phi(a(x)),$$

which tends to 0 as $x \to -\infty$. □

Unlike the components of a normally distributed random vector, the components of a vector with a bivariate Student's t distribution are asymptotically dependent. We omit the proof of the following proposition and refer the reader to Sect. 9.6 for further details.

Proposition 9.6. *Let (X_1, X_2) have an elliptical distribution with linear correlation parameter ρ. If X_1 and X_2 are equally distributed, and if $P(X_1 \leq x)$ is regularly varying at $-\infty$ with index $-\alpha$, then*

$$\lim_{x \to -\infty} P(X_2 \leq x \mid X_1 \leq x) = \frac{\int_{(\pi/2-\arcsin\rho)/2}^{\pi/2}(\cos t)^{\alpha}dt}{\int_{0}^{\pi/2}(\cos t)^{\alpha}dt}.$$

Zero correlation does not imply asymptotic independence, and covariances and correlations do not provide sufficient information to assess dependence between extreme values. For example, a quadratic hedge—based on a covariance structure—may perform poorly when it matters the most if the liability and the hedging instruments are asymptotically dependent. There are many examples from financial markets of simultaneous extreme price movements for assets whose log returns are weakly correlated between the assets.

Consider an elliptically distributed random vector (X_1, X_2) with a dispersion matrix Σ. Recall that any matrix $\Sigma_c = c\Sigma$ is a dispersion matrix for (X_1, X_2). However, the linear correlation parameter $\rho = \Sigma_{1,2}/(\Sigma_{1,1}\Sigma_{2,2})^{1/2}$ is uniquely determined by the elliptical distribution. Since $\rho = \text{Cor}(X_1, X_2)$, whenever $\text{Cor}(X_1, X_2)$ exists [the variances $\text{Var}(X_1)$ and $\text{Var}(X_2)$ are nonzero and finite], we may estimate ρ as the sample correlation coefficient. However, for heavy-tailed data (corresponding to distributions with finite variances) the sample correlation coefficient is an estimator of ρ with a large—or infinite—variance. An alternative approach to estimating the linear correlation parameter ρ is based on estimating another (rank) correlation coefficient called Kendall's tau, whose value for an elliptical distribution can be expressed in terms of the linear correlation parameter ρ. This approach allows for estimation of ρ also for elliptical distributions whose marginal distributions have infinite variances.

Kendall's tau for the random vector (X_1, X_2) is defined as

$$\tau(X_1, X_2) = P((X_1 - X_1')(X_2 - X_2') > 0) - P((X_1 - X_1')(X_2 - X_2') < 0), \quad (9.6)$$

where (X_1', X_2') is an independent copy of (X_1, X_2).

Proposition 9.7. *Let (X_1, X_2) have an elliptical distribution with location parameter (μ_1, μ_2) and linear correlation parameter ρ. If $P(X_1 = \mu_1) = P(X_2 = \mu_2) = 0$, then*

$$\tau(X_1, X_2) = \frac{2}{\pi} \arcsin \rho. \tag{9.7}$$

Proof. Without loss of generality we may consider the case $|\rho| < 1$. Since $P((X_1 - X_1')(X_2 - X_2') = 0) = 0$, we find that

$$\tau(X_1, X_2) = 2 P((X_1 - X_1')(X_2 - X_2') > 0) - 1.$$

The independence of $\mathbf{X} = (X_1, X_2)^T$ and $\mathbf{X}' = (X_1', X_2')^T$ and representation (9.3) imply that

$$(\mathbf{X}, \mathbf{X}') \stackrel{d}{=} (\boldsymbol{\mu}, \boldsymbol{\mu}) + \mathbf{A}(R\mathbf{U}, R'\mathbf{U}'),$$

where $R, R', \mathbf{U}, \mathbf{U}'$ are independent. From Proposition 9.4 we know that $\mathbf{X} - \mathbf{X}' \stackrel{d}{=} \mathbf{A}R^*\mathbf{U}^*$, where R^* and $\mathbf{U}^*$ are independent, and the assumption $P(X_1 = \mu_1) = P(X_2 = \mu_2) = 0$ implies that $P(R^* = 0) = 0$. With $\mathbf{W} = \mathbf{A}\mathbf{U}^*$ we have found that

$$\tau(X_1, X_2) = 2 P(R^* W_1 W_2 > 0) - 1 = 2 P(W_1 W_2 > 0) - 1.$$

Write

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_1 \sigma_2 \rho \\ \sigma_1 \sigma_2 \rho & \sigma_2^2 \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} \sigma_1 (1 - \rho^2)^{1/2} & \sigma_1 \rho \\ 0 & \sigma_2 \end{pmatrix}, \quad \mathbf{U}^* \stackrel{d}{=} \begin{pmatrix} \cos U \\ \sin U \end{pmatrix},$$

where U is uniformly distributed on $[-\pi, \pi)$. Then

$$
\begin{aligned}
P(W_1 W_2 > 0) &= 2 P(W_1 > 0, W_2 > 0) \\
&= 2 P(\sigma_1 (1 - \rho^2)^{1/2} \cos U + \sigma_1 \rho \sin U > 0, \sigma_2 \sin U > 0) \\
&= 2 P((1 - \rho^2)^{1/2} \cos U + \rho \sin U > 0, \sin U > 0) \\
&= 2 P(\cos y \cos U + \sin y \sin U > 0, \sin U > 0),
\end{aligned}
$$

where $y = \arcsin \rho \in [-\pi/2, \pi/2]$. Clearly, $\sin U > 0$ is here equivalent to $U \in (0, \pi)$. Since $\cos y \cos U + \sin y \sin U = \cos(U - y)$ and $\cos(U - y) > 0$ is here equivalent to $U \in (y - \pi/2, y + \pi/2)$, we find that

$$
\begin{aligned}
P(\cos y \cos U + \sin y \sin U > 0, \sin U > 0) &= P(U \in (y - \pi/2, y + \pi/2) \cap (0, \pi)) \\
&= P(U \in (0, y + \pi/2)).
\end{aligned}
$$

Putting the pieces together gives

$$\tau(X_1, X_2) = 4 \frac{\arcsin \rho + \pi/2}{2\pi} - 1 = \frac{2}{\pi} \arcsin \rho. \qquad \square$$

Consider the function sign(x) with value 0 for $x = 0$ and the value $x/|x|$ otherwise. Kendall's tau in (9.6) can be written as

$$\tau(X_1, X_2) = \mathrm{E}\left[\mathrm{sign}\left((X_1 - X_1')(X_2 - X_2')\right)\right]. \tag{9.8}$$

Given a sample $\{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$ of identically distributed vectors $\mathbf{X}_k = (X_{k,1}, X_{k,2})^{\mathrm{T}}$, we estimate (9.8) by the number of index pairs (j, k), where $j < k$ such that $(X_{j,1} - X_{k,1})(X_{j,2} - X_{k,2}) > 0$ minus the number of index pairs such that $(X_{j,1} - X_{k,1})(X_{j,2} - X_{k,2}) < 0$ divided by the total number of index pairs:

$$\widehat{\tau} = \binom{n}{2}^{-1} \sum_{j<k} \mathrm{sign}\left((X_{j,1} - X_{k,1})(X_{j,2} - X_{k,2})\right).$$

Finally, if the $\mathbf{X}_k$ are elliptically distributed such that the condition in Proposition 9.7 holds, then the estimator of the linear correlation parameter ρ is chosen as

$$\widehat{\rho} = \sin\left(\frac{\pi}{2}\widehat{\tau}\right). \tag{9.9}$$

To assess the accuracy of the estimator in (9.9) and compare it to the sample correlation coefficient, we consider a simulation study that is summarized in Fig. 9.3. For samples from a bivariate normal distribution the two estimators perform similarly. For samples from a bivariate Student's t distribution with three degrees of freedom we find that the estimator in (9.9), a nonlinear transformation of Kendall's tau estimator, performs much better than the sample correlation coefficient and similarly to its performance on data from a bivariate normal distribution.

9.2.3   Linearization and Elliptical Distributions

Suppose that the future value of a financial portfolio can be expressed as $g(\mathbf{X})$, where the function g is a known function and $\mathbf{X}$ is a random vector whose components represent, e.g., log returns for a given set of assets over a given future time period. If the time period is rather short and if $\mathbf{X}$ is likely to take a value that is not too far from its expected value $\boldsymbol{\mu} = \mathrm{E}[\mathbf{X}]$, then the first-order approximation

$$g(\mathbf{X}) \approx g(\boldsymbol{\mu}) + \nabla g^{\mathrm{T}}(\boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu}) = g(\boldsymbol{\mu}) + \sum_{k=1}^{d} \frac{\partial g}{\partial x_k}(\boldsymbol{\mu})(X_k - \mu_k)$$

can be assumed to be accurate. The approximation replaces the nonlinear expression in the components of $\mathbf{X}$ by a weighted sum of the components translated by a constant. However, it is typically hard to determine the probability distribution of a
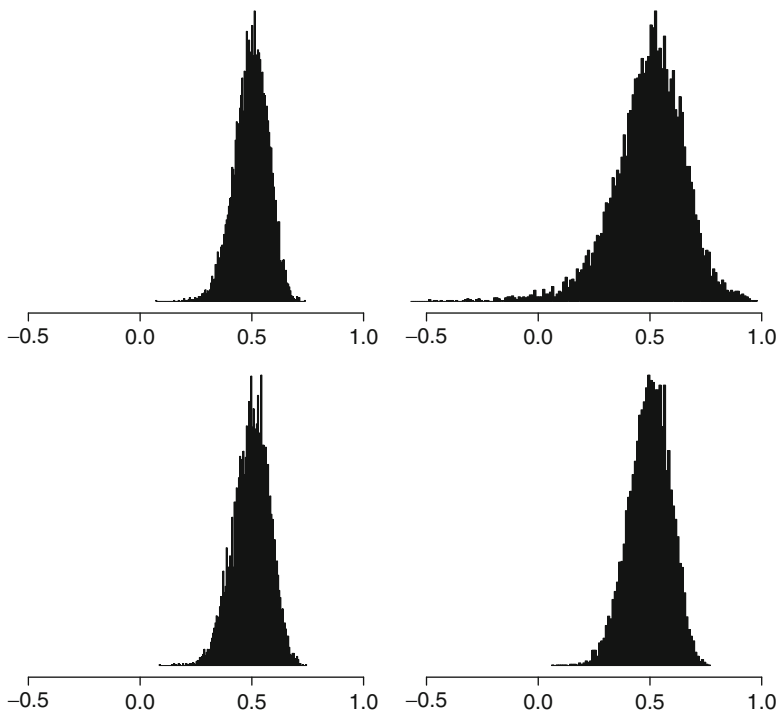
Fig. 9.3 Histograms based on 10,000 estimates of linear correlation parameter, where each estimate is based on a sample of size 100 from a bivariate elliptical distribution with linear correlation parameter 0.5. *Plots* to *left* show estimates based on samples from a bivariate normal distribution. *Plots* to *right* show estimates based on sample from a bivariate Student's *t* distribution with three degrees of freedom. The estimates in the *upper plots* are ordinary sample correlations. The estimates in the *lower plots* are transformations of Kendall's tau estimates as in (9.9)

sum of dependent random variables. An important exception is when $\mathbf{X}$ is elliptically distributed. In this case, $\mathbf{X}$ has the stochastic representation $\mathbf{X} \stackrel{\mathrm{d}}{=} \boldsymbol{\mu} + \mathbf{A}\mathbf{Y}$, where $\mathbf{Y}$ has a spherical distribution, so Proposition 9.3 gives

$$g(\mathbf{X}) \approx g(\boldsymbol{\mu}) + \nabla g^{\mathrm{T}}(\boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu}) \stackrel{\mathrm{d}}{=} g(\boldsymbol{\mu}) + \left(\nabla g^{\mathrm{T}}(\boldsymbol{\mu})\,\boldsymbol{\Sigma}\,\nabla g(\boldsymbol{\mu})\right)^{1/2} Y_1, \quad (9.10)$$

where $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^{\mathrm{T}}$ or, more explicitly,

$$g(\mathbf{X}) \stackrel{\mathrm{d}}{\approx} g(\boldsymbol{\mu}) + \left(\sum_{j,k=1}^{d} \frac{\partial g}{\partial x_j}(\boldsymbol{\mu})\frac{\partial g}{\partial x_k}(\boldsymbol{\mu})\Sigma_{j,k}\right)^{1/2} Y_1.$$

The accuracy of this approximation clearly depends strongly on how concentrated the probability mass of $\mathbf{X}$ is around its expected value $\boldsymbol{\mu}$. We illustrate the accuracy of the linearization with an example for one-dimensional elliptical distributions and a specific function g.
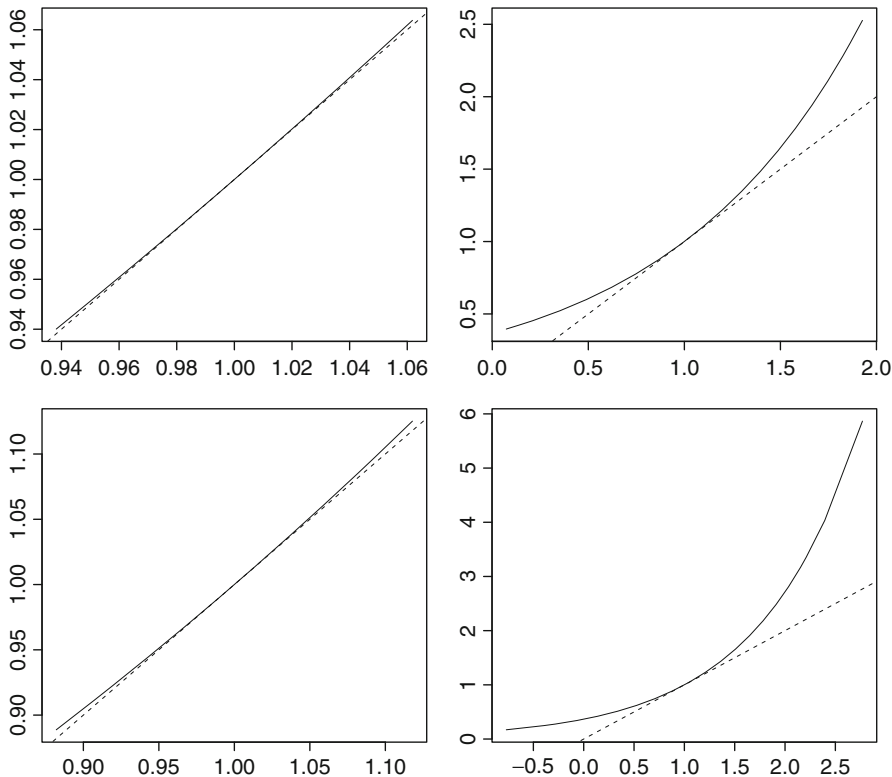
Fig. 9.4 These four q–q plots illustrate the approximation error from linearization. The *plots* show the quantiles of e^X (*y*-axis) against the quantiles of $1 + X$ (*x*-axis). The *upper plots* correspond to X's being $N(0, 0.02^2)$-distributed (*left*) and $N(0, 0.3^2)$-distributed (*right*). The *lower plots* correspond to X's having a Student's t distribution with three degrees of freedom and standard deviation 0.02 (*left*) and 0.3 (*right*)

Example 9.7 (Linearization). Let $g(x) = e^x$ and consider a random variable X with a spherical distribution with distribution function F. The quantile function of $g(X)$ is $g(F^{-1}(p))$, whereas the that of $g(0) + g'(0)X = 1 + X$ is $1 + F^{-1}(p)$. Figure 9.4 plots the quantiles of e^X (*y*-axis) against the quantiles of $1 + X$ (*x*-axis) together with the dashed straight line corresponding to a perfect fit. The upper plots correspond to X's being normally distributed with standard deviation 0.02 (left) and 0.3 (right). The lower plots correspond to X's having a Student's t distribution with three degrees of freedom and standard deviation 0.02 (left) and 0.3 (right). We see that the smaller the standard deviation is and the lighter the tails are, the more accurate is the linear approximation.

Example 9.8 (Linearization and risk measures). Suppose that $g(\mathbf{X})$ represents the value at time T of a portfolio of financial assets, where $\mathbf{X}$ has an elliptical distribution with stochastic representation $\mathbf{X} \stackrel{\mathrm{d}}{=} \boldsymbol{\mu} + \mathbf{A}\mathbf{Y}$. Consider a risk measure

ρ and the approximation of $g(\mathbf{X})$ in (9.10). If B_0 is the discount factor giving the current value of money at time T, and if ρ is translation invariant and positively homogeneous, then

$$\rho(g(\mathbf{X})) \approx \rho\left(g(\boldsymbol{\mu}) + \left(\nabla g^{\mathrm{T}}(\boldsymbol{\mu})\,\boldsymbol{\Sigma}\,\nabla g(\boldsymbol{\mu})\right)^{1/2} Y_1\right)$$

$$= -B_0 g(\boldsymbol{\mu}) + \left(\nabla g^{\mathrm{T}}(\boldsymbol{\mu})\,\boldsymbol{\Sigma}\,\nabla g(\boldsymbol{\mu})\right)^{1/2} \rho(Y_1).$$

For ρ chosen as value-at-risk (VaR) or expected shortfall (ES) and for Y_1 normally distributed or Student's t-distributed, the quantity $\rho(Y_1)$ can be computed as in Example 6.13. If Y_1 is standard normally distributed, then

$$\mathrm{VaR}_p(Y_1) = B_0 \Phi^{-1}(1-p) \quad \text{and} \quad \mathrm{ES}_p(Y_1) = B_0 \frac{\phi(\Phi^{-1}(1-p))}{p}.$$

If Y_1 has a standard Student's t distribution with ν degrees of freedom, then

$$\mathrm{VaR}_p(Y_1) = B_0 t_\nu^{-1}(1-p) \quad \text{and} \quad \mathrm{ES}_p(Y_1) = B_0 \frac{g_\nu(t_\nu^{-1}(1-p))}{p} \left(\frac{\nu + (t_\nu^{-1}(p))^2}{\nu-1}\right),$$

where g_ν and t_ν are the density and distribution functions, respectively, of Y_1.

If ρ is a monotone risk measure and g is a convex function, then it follows from Proposition 2.2 that

$$\rho(g(\mathbf{X})) \leq \rho\left(g(\boldsymbol{\mu}) + \nabla g^{\mathrm{T}}(\boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})\right),$$

i.e., linearization overestimates the risk. If ρ is also translation invariant and positively homogeneous, then

$$\rho(g(\mathbf{X})) \leq \rho\left(g(\boldsymbol{\mu}) + \nabla g^{\mathrm{T}}(\boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})\right)$$

$$= -B_0 g(\boldsymbol{\mu}) + \left(\nabla g^{\mathrm{T}}(\boldsymbol{\mu})\,\boldsymbol{\Sigma}\,\nabla g(\boldsymbol{\mu})\right)^{1/2} \rho(Y_1).$$

As an illustration, let $\mathbf{X}$ be a vector of log returns of d assets and consider a linear portfolio consisting of a long position of current value $w_k \geq 0$ in the kth asset, for every k. Then the future portfolio value is $g(\mathbf{X}) = w_1 e^{X_1} + \cdots + w_d e^{X_d}$ and g is convex.

Example 9.8 illustrates how linearization and an elliptical approximation can be used to construct explicit approximation formulas for risk measures. This approach must be used with caution. The accuracy of the first-order approximation of g around $\boldsymbol{\mu}$ evaluated at $\mathbf{X}$ is best around $\boldsymbol{\mu}$. However, risk measures of $g(\mathbf{X})$, such as VaR and ES, typically depend on the behavior of $\mathbf{X}$ far from $\boldsymbol{\mu}$.

Example 9.9 (Linearization over a short time horizon). Considering a portfolio of shares of two stocks. The portfolio contains h_1 and h_2 shares of the two stocks.

The spot prices at time t are given by S_t^1 and S_t^2, respectively. Suppose that we want to compute $\text{VaR}_p(V_T - V_0/B_0)$, where $V_T - V_0/B_0$ is the change in portfolio value from now until time T, measured in money at time T. We have

$$V_T - V_0/B_0 = h_1(S_T^1 - S_0^1/B_0) + h_2(S_T^2 - S_0^2/B_0)$$
$$= h_1 S_0^1(e^{X_1} - 1/B_0) + h_2 S_0^2(e^{X_2} - 1/B_0)$$
$$= g(X_1, X_2),$$

where $(X_1, X_2) = (\log(S_T^1/S_0^1), \log(S_T^2/S_0^2))$ is the log-return pair from now until time T. If T is small (a couple of days, say), then it may be reasonable to set $\mu_1 = \mu_2 = 0$ and $B_0 = 1$, which yields

$$g(X_1, X_2) \approx g(\mu_1, \mu_2) + \sum_{k=1}^{2} \frac{\partial g}{\partial x_k}(\mu_1, \mu_2)(X_k - \mu_k)$$

$$= \sum_{k=1}^{2} h_k S_0^k(e^{\mu_k} - 1/B_0) + \sum_{k=1}^{2} h_k S_0^k e^{\mu_k}(X_k - \mu_k)$$

$$= h_1 S_0^1 X_1 + h_2 S_0^2 X_2.$$

If $\mathbf{X} = (X_1, X_2)^{\mathrm{T}}$ has an elliptical distribution with representation $\mathbf{X} = \mathbf{AY}$, where $\mathbf{AA}^{\mathrm{T}} = \boldsymbol{\Sigma}$, then

$$\text{VaR}_p(V_1 - V_0/B_0) \approx V_0 + \text{VaR}_p((\mathbf{w}^{\mathrm{T}} \boldsymbol{\Sigma} \mathbf{w})^{1/2} Y_1)$$
$$= V_0 + (\mathbf{w}^{\mathrm{T}} \boldsymbol{\Sigma} \mathbf{w})^{1/2} F_{Y_1}^{-1}(1 - p),$$

where $\mathbf{w}^{\mathrm{T}} = (h_1 S_0^1, h_2 S_0^2)$.

Example 9.10 (Linearization over a long time horizon). Suppose that we want to compute $\text{VaR}_p(V_T - V_0/B_0)$ for a portfolio over a T-day period. Suppose further that V_T can be expressed as a function g of the T-day log returns and that the vectors $\mathbf{X}_1, \ldots, \mathbf{X}_T$ of 1-day log returns are independent and identically elliptically distributed with mean $\boldsymbol{\mu} = \mathrm{E}[\mathbf{X}_1]$ and covariance matrix $\boldsymbol{\Sigma} = \text{Cov}(\mathbf{X}_1)$. Set $\mathbf{W} = \mathbf{X}_1 + \cdots + \mathbf{X}_T$ and note that $\mathbf{W}$, with $\mathrm{E}[\mathbf{W}] = T\boldsymbol{\mu}$ and $\text{Cov}(\mathbf{W}) = T\boldsymbol{\Sigma}$, is the vector of log returns for the entire T-day period. From Proposition 9.4 we know that $\mathbf{W}$ is elliptically distributed, however, in general (unless $\mathbf{W}$ is normally distributed) of a different type than $\mathbf{X}_1$. The elliptical distribution of $\mathbf{W}$ is not easily inferred from the distribution of $\mathbf{X}_1$. However, if T is sufficiently large, then it may be reasonable (based on the central limit theorem) to assume that $\mathbf{W}$ is approximately normally distributed. However, one should be aware that the convergence in distribution to the normal distribution in the central limit theorem is slow in the tail regions.

Linearization, together with the normal approximation, gives

$$g(\mathbf{W}) = \sum_{k=1}^{d} h_k S_0^k (e^{W_k} - 1/B_0)$$

$$\approx \sum_{k=1}^{d} h_k S_0^k (e^{T\mu_k} - 1/B_0) + \sum_{k=1}^{d} h_k S_0^k e^{T\mu_k} (W_k - T\mu_k)$$

$$\overset{\mathrm{d}}{\approx} \sum_{k=1}^{d} h_k S_0^k (e^{T\mu_k} - 1/B_0) + T^{1/2} \left(\sum_{j,k=1}^{d} h_j h_k S_0^j S_0^k e^{T(\mu_j + \mu_k)} \Sigma_{j,k} \right)^{1/2} Z,$$

where Z is standard normally distributed. In particular,

$$\mathrm{VaR}_p(V_T - V_0/B_0) \approx \sum_{k=1}^{d} h_k S_0^k \left(1 - B_0 e^{T\mu_k}\right)$$

$$+ T^{1/2} B_0 \left(\sum_{j,k=1}^{d} h_j h_k S_0^j S_0^k e^{T(\mu_j + \mu_k)} \Sigma_{j,k} \right)^{1/2} \Phi^{-1}(1-p).$$

If $B_0 e^{T\mu_k} \approx 1$ for all k, then the estimate of $\mathrm{VaR}_p(V_T - V_0/B_0)$ is approximately proportional to the square root of the length T of the time period.

As an illustration, we consider the situation where $\mathbf{X}_1$ has a ten-dimensional Student's t distribution with three degrees of freedom, with zero mean and standard deviations 0.01 and pairwise linear correlation coefficients of 0.4. Moreover, we assume that we hold one share of each stock ($h_k = 1$), that the current share price is 10 for each stock ($S_0^k = 10$), and that interest rates can be ignored ($B_0 = 1$). This gives

$$\mathrm{VaR}_p(V_T - V_0/B_0) \approx T^{1/2}(d(1 + 0.4(d - 1)))^{1/2} \Phi^{-1}(1 - p)$$

$$= (27T)^{1/2} \Phi^{-1}(1 - p).$$

We now compare this estimate to the empirical estimate based on a large simulated sample of independent copies of $V_T - V_0/B_0$. The results are shown in Fig. 9.5. It is interesting to note that for T small, the underestimation of $\mathrm{VaR}_p(V_T - V_0/B_0)$ for p small due to the lighter tails of the normal distribution is offset by the overestimation of $\mathrm{VaR}_p(V_T - V_0/B_0)$ due to linearization.

Fig. 9.5 Illustration of
accuracy of estimates of
$\text{VaR}_{0.05}(V_T - V_0)$ and
$\text{VaR}_{0.01}(V_T - V_0)$ based on
linearization and a normal
approximation, as functions
of $T \in \{1, \ldots, 100\}$ (*dashed
curves*). The *solid curves*
show the empirical VaR
estimates based on simulated
samples of size 10^5
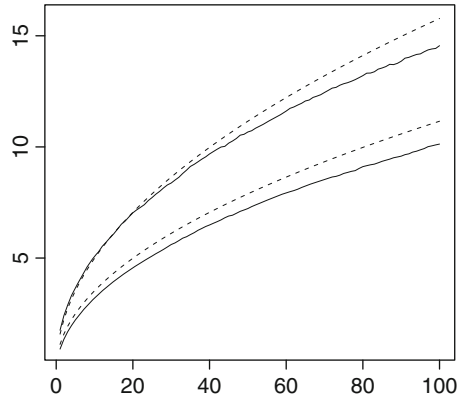
9.3 Applications of Elliptical Distributions in Risk Management

In this section, we consider five applications of elliptical distributions in risk management. In the first application, we derive a risk-aggregation formula that relates the risk, in terms of a translation-invariant and positively homogeneous risk measure, for a sum of jointly elliptically distributed random values to the risk of the terms in the sum. The second application shows how linearization and a normal approximation can be used to approximate the risk measure $\text{VaR}_{0.005}(A-L)$ used to determine the solvency of an insurance company. This application presents the idea behind the so-called standard formula that is used in the measurement of risk in the insurance industry. The third application suggests a hedging approach to European call options that is more appropriate than delta hedging if the joint distribution for the log return of the underlying asset value and the change in the implied volatility can be assumed to be elliptical. The fourth application presents how a trader might design a bet on changes in implied volatility for two maturity times and considers ways to investigate the risk of such a bet. The fifth application illustrates that if the vector of returns on a set of risky assets can be assumed to be elliptically distributed, then portfolio investment problems can often be reduced to the trade-off investment problem (4.7).

9.3.1 Risk Aggregation with Elliptical Distributions

Consider a company divided into n business units with future net values of assets and liabilities given by $X_1, \ldots, X_n$. Suppose that each business unit is able to accurately estimate $\text{E}[X_k]$ and $\rho(X_k)$, where ρ is some translation-invariant and positively homogeneous risk measure. The company wants to compute $\rho(X_1 + \cdots + X_n)$ to get a measurement on the aggregate risk for the whole company. There is no straightforward way to combine the individual risk estimates $\rho(X_k)$

and the expected values $E[X_k]$ into an aggregate risk estimate. However, there is a convenient risk-aggregation formula that is valid under the assumption that $(X_1, \ldots, X_n)^T$ has an elliptical distribution.

Suppose that $\mathbf{X} = (X_1, \ldots, X_n)^T$ has an elliptical distribution so that $\mathbf{X} \stackrel{d}{=} \boldsymbol{\mu} + \mathbf{AY}$, with $\mathbf{AA}^T = \boldsymbol{\Sigma}$, and $\mathbf{Y}$ has a spherical distribution. Matrix $\boldsymbol{\Sigma}$ can always be expressed as the product $\mathbf{DCD}$, where $\mathbf{D}$ is a diagonal matrix with diagonal entries $D_{k,k} = \Sigma_{k,k}^{1/2}$ and $\mathbf{C}$ is a correlation matrix (the linear correlation matrix of $\mathbf{X}$ if it exists). Note that

$$\rho(X_1 + \cdots + X_n) = -B_0 \sum_{k=1}^n \mu_k + \rho\left(\sum_{k=1}^n (X_k - \mu_k)\right),$$

where B_0 is the discount factor between now and the considered future time, and

$$\sum_{k=1}^n (X_k - \mu_k) = \mathbf{1}^T \mathbf{AY} \stackrel{d}{=} (\mathbf{1}^T \boldsymbol{\Sigma} \mathbf{1})^{1/2} Y_1.$$

Since $\mathbf{1}^T \boldsymbol{\Sigma} \mathbf{1} = \Sigma_{1,1} + \Sigma_{1,2} + \cdots + \Sigma_{n,n}$ and $\Sigma_{j,k} = D_{j,j} C_{j,k} D_{k,k}$, it holds that

$$\rho\left(\sum_{k=1}^n (X_k - \mu_k)\right) = \rho\left(\left(\sum_{j,k} \Sigma_{j,k}\right)^{1/2} Y_1\right)$$

$$= \left(\sum_{j,k} C_{j,k} D_{j,j} D_{k,k}\right)^{1/2} \rho(Y_1)$$

$$= \left(\sum_{j,k} C_{k,l} D_{j,j} D_{k,k} \rho(Y_1)^2\right)^{1/2}$$

$$= \left(\sum_{j,k} C_{j,k} \rho(D_{j,j} Y_j) \rho(D_{k,k} Y_k)\right)^{1/2}$$

$$= \left(\sum_{j,k} C_{j,k} \rho(X_j - \mu_j) \rho(X_k - \mu_k)\right)^{1/2}.$$

We have found that if $\mathbf{X} = (X_1, \ldots, X_n)^T$ has an elliptical distribution and if ρ is a translation-invariant and positively homogeneous risk measure, then

$$\rho(X_1 + \cdots + X_n) = \left(\sum_{j,k} C_{j,k} \{B_0 \mu_j + \rho(X_j)\}\{B_0 \mu_k + \rho(X_k)\}\right)^{1/2} - B_0 \sum_k \mu_k.$$

The only additional input needed, besides the individual risk estimates $\rho(X_k)$ and the means μ_k, are the linear correlation coefficients $C_{j,k}$.

9.3.2   Solvency of an Insurance Company

In this section, we present another example of linearization and normal approximation in the context of the solvency of an insurance company.

Consider an insurance company with assets and liabilities. Let A and L denote the time 1 (1 year from now) values of the assets and liabilities, respectively. We consider the insurance company to be solvent if

$$\mathrm{VaR}_{0.005}(A - L) \leq 0.$$

If r_1 is the current risk-free, 1-year zero rate, then we may write

$$\mathrm{VaR}_{0.005}(A - L) = F^{-1}_{e^{-r_1}(L-A)}(0.995).$$

We consider a stylized model for the assets and liabilities and assume that the liabilities correspond to the stochastic cash flow $(C_1, \ldots, C_n)$, where C_k is the amount the insurer has to pay at the end of year k due to the occurrence of claims before the end of year 1. Each written contract offers a protection for the insured over a 1-year period. Operating expenses for the insurer could be included in the C_k or dealt with in other ways. The expectation $\mathrm{E}[C_k]$ is the expected claim amount to be paid at time k, and $e^{-r_k}\mathrm{E}[C_k]$ is the present value of this amount. The expected claim amount $\mathrm{E}[C_k]$ could be determined by some stochastic claim-reserving method, such as the chain ladder method presented in Sect. 7.6.1. The best estimate, at time 0, of the present value of the liabilities is

$$L_0 = \sum_{k=1}^{n} \mathrm{E}[C_k] e^{-r_k k}.$$

At time 1 we observe C_1 and receive new information about the future payments C_k. If $\mathbf{I}_1$ denotes the information available at time 1, then $\mathrm{E}[C_k \mid \mathbf{I}_1]$ is the updated prediction of the payment due at time k. The time 1 value of the liabilities is therefore given by

$$L = \sum_{k=1}^{n} \mathrm{E}[C_k \mid \mathbf{I}_1] e^{-(r_{k-1} + \Delta r_{k-1})(k-1)},$$

where $\Delta \mathbf{r}$ is the vector of zero rate changes from time 0 to 1. Suppose for simplicity that the assets of the insurer consist of a bond portfolio designed to match future claim payments and K units of cash on a bank account. The time 0 value A_0 of the assets and the time 1 value A of the assets are given by

$$A_0 = \sum_{k=1}^{n} \mathrm{E}[C_k] e^{-r_k k} + K,$$

$$A = \sum_{k=1}^{n} \mathrm{E}[C_k] e^{-(r_{k-1} + \Delta r_{k-1})(k-1)} + K e^{r_1}.$$

The time 0 value of the bond portfolio precisely matches the time 0 value of the liability, $A_0 - K = L_0$. Moreover,

$$e^{-r_1}(L - A) = e^{-r_1} \sum_{k=1}^{n} (E[C_k \mid I_1] - E[C_k]) e^{-(r_{k-1} + \Delta r_{k-1})(k-1)} - K$$

$$= \sum_{k=1}^{n} e^{-r_k k} E[C_k] Y_k e^{X_k} - K$$

$$= g(X_1, \ldots, X_n, Y_1, \ldots, Y_n),$$

where $X_k = -r_1 - (r_{k-1} + \Delta r_{k-1})(k-1) + r_k k$, $Y_k = (E[C_k \mid I_1] - E[C_k])/ E[C_k]$ for $k = 1, \ldots, n$, and

$$g(\mathbf{x}, \mathbf{y}) = \sum_{k=1}^{n} e^{-r_k k} E[C_k] y_k e^{x_k} - K.$$

The quantity $Y_1 = (C_1 - E[C_1])/ E[C_1]$ measures the relative deviation of the actual amount paid at the end of the year from the current prediction. For $k \geq 2$, $Y_k = (E[C_k \mid I_1] - E[C_k])/ E[C_k]$ measures the relative deviation of the updated prediction at the end of the year of the claim payments at time k, for claims incurred before the end of the year, from the current prediction.

Since g is a nonlinear function of the risk factors $(X_1, \ldots, X_n, Y_1, \ldots, Y_n)$, the computation of VaR is simplified substantially by linearization. Let $\mu_k = E[X_k]$, and note that $E[Y_k] = 0$. Therefore, it makes sense to consider the first-order approximation of g around $(\mu_1, \ldots, \mu_n, 0, \ldots, 0)$, which gives

$$g(X_1, \ldots X_n, Y_1, \ldots, Y_n) \approx g(\mu_1, \ldots, \mu_k, 0, \ldots, 0) + \sum_{k=1}^{n} e^{-r_k k} E[C_k] Y_k e^{\mu_k}$$

$$= -K + \sum_{k=1}^{n} e^{-r_k k} E[C_k] Y_k e^{\mu_k}$$

$$= -K + \mathbf{w}^{\mathsf{T}} \mathbf{Y},$$

where $w_k = e^{-r_k k} E[C_k] e^{\mu_k}$. Because of the linearization, the effect of the X_k vanishes. The contributions to the risk coming from changes in the zero rates are second-order effects and do not show up in the linearized version of g. Although ignoring second-order effects is convenient for explicit computations, it leads to a crude approximation.

If $\mathbf{Y}$ is $N(\mathbf{0}, \boldsymbol{\Sigma})$-distributed, then we find that

$$\mathrm{VaR}_{0.005}(A - L) \approx F_{-K + \mathbf{w}^{\mathsf{T}} \mathbf{Y}}^{-1}(0.995) = -K + (\mathbf{w}^{\mathsf{T}} \boldsymbol{\Sigma} \mathbf{w})^{1/2} \Phi^{-1}(0.995).$$

Taking this approximation as an equality we find that the solvency condition $\mathrm{VaR}_{0.005}(A - L) \le 0$ is equivalent to

$$K \ge (\mathbf{w}^{\mathrm{T}} \boldsymbol{\Sigma} \mathbf{w})^{1/2} \Phi^{-1}(0.995).$$

The outlined procedure is the basic idea behind the standard formula in the Solvency II framework for the computation of sufficient buffer capital for an insurance company. Of course, in practice, many more risk factors need to be included and the insurer's asset portfolio is more complex. Nevertheless, the linearization approach and the normal approximation is at the heart of the standard formula. To compensate for the inaccuracies of linearization and the normal approximation, the covariance matrix $\boldsymbol{\Sigma}$ is not estimated from data but given exogenously by the regulators.

9.3.3   Hedging of a Call Option When the Volatility Is Stochastic

Suppose that now at time 0 we have issued a European call option with strike price K on the value S_T of a stock market index at time T. Suppose also that we want to hedge against changes in the option price from now until time $t < T$ by taking a position in the underlying index and deposit cash to minimize

$$\mathrm{E}[(h_0 + h_1 S_t - C_t)^2],$$

where C_t is the call option price at time t. If t is small, then the delta-hedging approach in Sect. 3.5 gives an approximative solution to the quadratic hedging problem. Suppose that the option price is expressed in terms of the Black–Scholes formula (1.7) as a function $C_t = C(S_t, \sigma_t, r_t, t, T - t)$, where the arguments correspond to the value of the underlying index at time t, the option's implied volatility at time t, interest rate prevailing between time t and the maturity time T of the option, and the remaining time to maturity. The delta-hedging approach relies on the first-order approximation

$$C_t \approx C_0 + \frac{\partial C_0}{\partial S_0}(S_t - S_0),$$

which gives the delta-hedge position $(h_0^{\delta}, h_1^{\delta}) \approx (h_0, h_1)$, where

$$h_1^{\delta} = \frac{\partial C_0}{\partial S_0} \quad \text{and} \quad h_0^{\delta} = C_0 - \frac{\partial C_0}{\partial S_0} S_0.$$

The Black–Scholes formula reads

$$C_t = S_t \Phi(d_1) - K e^{-r_t(T-t)} \Phi(d_2),$$

$$d_1 = \frac{\log(S_t/K) + (r_t + \sigma_t^2/2)(T - t)}{\sigma_t \sqrt{T - t}} \quad \text{and} \quad d_2 = d_1 - \sigma_t \sqrt{T - t}$$

and gives

$$\frac{\partial C_0}{\partial S_0} = \Phi(d_1), \quad d_1 = \frac{\log(S_0/K) + (r_0 + \sigma_0^2/2)T}{\sigma_0 \sqrt{T}}.$$

The hedging error at time t is

$$h_0^\delta + h_1^\delta S_t - C_t = C_0 - C_t + \Phi(d_1)(S_t - S_0).$$

The change in the interest rate from r_0 to r_t typically does not contribute much to the hedging error, and therefore we may approximate $r_t \approx r_0$. We may therefore view the hedging error as a function of the changes in the index value and in the implied volatility or, equivalently, as a function $g(\mathbf{z})$ evaluated at $\mathbf{Z} = (Z_1, Z_2)$, where $Z_1 = \log(S_t/S_0)$ and $Z_2 = \sigma_t - \sigma_0$. Therefore, a model for (Z_1, Z_2) implies a model for the hedging error and the latter model can be analyzed by, e.g., simulation from (Z_1, Z_2). The sample from (Z_1, Z_2) can then be converted to a sample from the distribution of the hedging error whose empirical distribution can be studied. Alternatively, we could linearize the nonlinear function $g(\mathbf{z})$ and evaluate the linear approximation at $\mathbf{Z} = (Z_1, Z_2)$. The linearization approach may give an approximation of the distribution for the hedging error that can be analyzed analytically, without simulation. Consider the first-order approximation

$$C_t \approx g(\mathbf{0}) + \frac{\partial g}{\partial z_1}(\mathbf{0}) Z_1 + \frac{\partial g}{\partial z_2}(\mathbf{0}) Z_2,$$

where $g(\mathbf{z}) = g_1(g_2(z_1), g_3(z_2))$ with $g_2(z_1) = S_0 e^{z_1}$, $g_3(z_2) = z_2 + \sigma_0$, and

$$g_1(s, \sigma) = s\Phi(d_1) - Ke^{-r_0(T-t)}\Phi(d_2),$$
$$d_1 = \frac{\log(s/K) + (r_0 + \sigma^2/2)(T - t)}{\sigma\sqrt{T - t}} \quad \text{and} \quad d_2 = d_1 - \sigma\sqrt{T - t}.$$

The chain rule, together with the expressions for the partial derivatives of the Black–Scholes formula (Sect. 1.2.2), gives

$$\frac{\partial g}{\partial z_1}(\mathbf{0}) = \frac{\partial g_1}{\partial s}(S_0, \sigma_0)\frac{dg_2}{dz_1}(0) = \Phi(d_1)S_0,$$

$$\frac{\partial g}{\partial z_2}(\mathbf{0}) = \frac{\partial g_1}{\partial \sigma}(S_0, \sigma_0)\frac{dg_3}{dz_2}(0) = \phi(d_1)S_0\sqrt{T - t}.$$

Summing up, we arrive at the following approximation of the hedging error:

$$\begin{aligned}
h_0^\delta + h_1^\delta S_t - C_t &= C_0 - C_t + \Phi(d_1)(S_t - S_0)\\
&\approx C_0 - C_0 - \Phi(d_1)S_0 Z_1 - \phi(d_1)S_0\sqrt{T - t}Z_2\\
&\quad + \Phi(d_1)(S_0(1 + Z_1) - S_0)\\
&= -\phi(d_1)S_0\sqrt{T - t}(\sigma_t - \sigma_0).
\end{aligned}$$

We see that the position, the delta hedge and the issued call option, is immune against changes in the index value (approximately, over a short time period) and that the hedging error is due to changes in the implied volatility. We also find that the variance of the hedging error is

$$\text{Var}(h_0^\delta + h_1^\delta S_t - C_t) \approx \phi(d_1)^2 S_0^2 (T - t) \, \text{Var}(\sigma_t).$$

We now want to reduce the hedging error by replacing the delta hedge by a similar hedge that also takes changes in the implied volatility into account. The position in the underlying index and in cash for the optimal quadratic hedge is

$$h_1 = \frac{\text{Cov}(S_t, C_t)}{\text{Var}(S_t)} \quad \text{and} \quad h_0 = \text{E}[C_t] - h_1 \, \text{E}[S_t].$$

Here we approximate

$$\begin{aligned}
\text{Cov}(S_t, C_t) &\approx \text{Cov}(S_0 Z_1, \Phi(d_1) S_0 Z_1 + \phi(d_1) S_0 \sqrt{T - t} Z_2) \\
&= S_0^2 \Phi(d_1) \, \text{Var}(Z_1) + S_0^2 \phi(d_1) \sqrt{T - t} \, \text{Cov}(Z_1, Z_2), \\
\text{Var}(S_t) &\approx S_0^2 \, \text{Var}(Z_1), \\
\text{E}[C_t] &\approx C_0, \\
\text{E}[S_t] &\approx S_0.
\end{aligned}$$

This gives the hedge $(h_0^*, h_1^*) \approx (h_0, h_1)$, where

$$\begin{aligned}
h_1^* &= \Phi(d_1) + \phi(d_1) \sqrt{T - t} \frac{\text{Cov}(Z_1, Z_2)}{\text{Var}(Z_1)} \\
&= \Phi(d_1) + \phi(d_1) \sqrt{T - t} \frac{\sigma_{Z_2}}{\sigma_{Z_1}} \rho, \\
h_0^* &= C_0 - h_1^* S_0,
\end{aligned}$$

where $\sigma_{Z_k} = \text{Var}(Z_k)^{1/2}$ and $\rho = \text{Cor}(Z_1, Z_2)$. We observe that the position h_1^* in the underlying index corresponds to the delta-hedge position h_1^δ plus a correction term. We get the following approximation of the hedging error:

$$\begin{aligned}
h_0^* + h_1^* S_t - C_t &\approx C_0 + \left(\Phi(d_1) + \phi(d_1) \sqrt{T - t} \frac{\sigma_{Z_2}}{\sigma_{Z_1}} \rho \right) S_0 Z_1 \\
&\quad - C_0 - \Phi(d_1) S_0 Z_1 - \phi(d_1) S_0 \sqrt{T - t} Z_2 \\
&= \phi(d_1) S_0 \sqrt{T - t} \left(\frac{\sigma_{Z_2}}{\sigma_{Z_1}} \rho Z_1 - Z_2 \right).
\end{aligned}$$

In particular, the variance of the hedging error is approximately

$$\text{Var}(h_0^* + h_1^* S_t - C_t) \approx \text{Var}\left(\phi(d_1)S_0\sqrt{T-t}\left(\frac{\sigma_{Z_2}}{\sigma_{Z_1}}\rho Z_1 - Z_2\right)\right)$$

$$= \phi(d_1)^2 S_0^2 (T-t)\,\text{Var}(\sigma_t)(1-\rho^2),$$

where the last equality can be verified by straightforward computations of the variance of the sum of two correlated terms. Notice that taking changes in implied volatility into account when computing the approximation of the quadratic hedge makes the variance of the hedging error smaller by a factor of $(1-\rho^2)$.

9.3.4 Betting on Changes in Volatility

Suppose that a trader is betting on changes in implied volatility from time 0 today until time $t > 0$ in the future for two future maturity times and that we want to analyze the riskiness of this volatility bet. Consider two call options on the values of an index at two future times $0 < T_1 < T_2$. The trader believes that over a short period of time the change in implied volatility $\sigma_t^1 - \sigma_0^1$ for the nearer maturity time T_1 will be greater than that for the more distant maturity time T_2, $\sigma_t^2 - \sigma_0^2$. The trader wants to capitalize on this belief but at the same time not bet on other potential movements of the underlying index value. We first determine the particular portfolio corresponding to the volatility bet.

Consider a long position of size h_2 in a call option with strike K_1 maturing at time T_1 and a short position of size h_3 in a call option with strike K_2 maturing at time T_2. The future value of this position is, to a first-order approximation and with the expressions for the partial derivatives of the Black–Scholes formula,

$$h_2 C_t^1 - h_3 C_t^2 \approx h_2 C_0^1 - h_3 C_0^2$$

$$+h_2\left(\Phi(d_1^1)(S_t - S_0) + \phi(d_1^1)S_0\sqrt{T_1}(\sigma_t^1 - \sigma_0^1)\right)$$

$$-h_3\left(\Phi(d_1^2)(S_t - S_0) + \phi(d_1^2)S_0\sqrt{T_2}(\sigma_t^2 - \sigma_0^2)\right),$$

where

$$d_1^j = \frac{\log(S_0/K_j) + (r_j + (\sigma_0^j)^2/2)T_j}{\sigma_0^j\sqrt{T_j}} \quad \text{for } j = 1, 2.$$

With $Z_1 = \log(S_t/S_0)$, $Z_2 = \sigma_t^1 - \sigma_0^1$, and $Z_3 = \sigma_t^2 - \sigma_0^2$, and the approximation $S_t - S_0 \approx S_0 Z_1$, we get

$$h_2 C_t^1 - h_3 C_t^2 \approx h_2 C_0^1 - h_3 C_0^2 + (h_2\Phi(d_1^1) - h_3\Phi(d_1^2))S_0 Z_1$$

$$+h_2 S_0\phi(d_1^1)\sqrt{T_1}Z_2 - h_3 S_0\phi(d_1^2)\sqrt{T_2}Z_3.$$

The volatility bet is a bet on the occurrence of the event $Z_2 > Z_3$, and on nothing else. Therefore, the trader chooses h_2 and h_3 so that

$$h_2 \phi(d_1^1) \sqrt{T_1} - h_3 \phi(d_1^2) \sqrt{T_2} = 0,$$

meaning that the impact of a parallel shift in the implied volatility should be approximately zero. Moreover, the trader wants the bet to be immune to changes in the value of the underlying index. Therefore, the trader takes the position

$$h_1 = -(h_2 \Phi(d_1^1) - h_3 \Phi(d_1^2))$$

in the index and a position

$$h_0 = -h_1 S_0 - h_2 C_0^1 + h_3 C_0^2$$

in cash. Summing up, we find that the volatility bet corresponds to the portfolio weights h_0, h_1, h_2, h_3 and the future portfolio value

$$h_0 + h_1 S_t + h_2 C_t^1 - h_3 C_t^2 \approx h_2 S_0 \phi(d_1^1) \sqrt{T_1} Z_2 - h_2 \frac{\phi(d_1^1) \sqrt{T_1}}{\phi(d_1^2) \sqrt{T_2}} S_0 \phi(d_1^2) \sqrt{T_2} Z_3$$

$$= h_2 S_0 \phi(d_1^1) \sqrt{T_1} (Z_2 - Z_3).$$

To estimate the risk of holding this portfolio until time t, we could now assign a bivariate elliptical distribution to (Z_2, Z_3), determine the corresponding univariate elliptical distribution of $Z_2 - Z_3$, and finally compute $\rho(h_2 S_0 \phi(d_1^1) \sqrt{T_1}(Z_2 - Z_3))$ for a suitable choice of risk measure ρ. However, this apparent straightforward approach to measuring the riskiness of the volatility bet is not unproblematic. Assigning a bivariate model to (Z_2, Z_3) can at best be guided by historical data on implied volatility changes but will to a large extent be based on subjective beliefs. Moreover, if the sizes of the option positions are large, then it may be unrealistic to assume that the positions can be closed at time t if t is small. In that case, we need a longer time period for the risk modeling, and this makes the whole linearization approach questionable.

9.3.5 Portfolio Optimization with Elliptical Distributions

Suppose vector $\mathbf{R}$ of returns on a collection of risky assets can be modeled by a normal variance mixture distribution so that $\mathbf{R} \stackrel{d}{=} \boldsymbol{\mu} + W \mathbf{A} \mathbf{Z}$, where $\mathbf{Z}$ is $N_d(\mathbf{0}, \mathbf{I})$-distributed and independent of $W \geq 0$, and $\mathbf{A}\mathbf{A}^T = \boldsymbol{\Sigma}$. If R_0 is the return on a risk-free asset, then the future value of a portfolio with monetary portfolio weights $\mathbf{w}$ in the risky assets and w_0 in the risk-free asset can be expressed as

$$V_1 = w_0 R_0 + \mathbf{w}^T \mathbf{R}$$
$$\stackrel{d}{=} w_0 R_0 + \mathbf{w}^T \boldsymbol{\mu} + (\mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w})^{1/2} W Z_1. \tag{9.11}$$

Suppose the variance $\text{Var}(V_1) = \sigma^2 \mathbf{w}^{\mathsf{T}} \boldsymbol{\Sigma} \mathbf{w}$, where $\sigma^2 = \text{Var}(WZ_1)$, exists. Then the solution to the investment problem

$$\text{maximize} \ \ w_0 R_0 + \mathbf{w}^{\mathsf{T}} \boldsymbol{\mu} - \tfrac{c}{2V_0} \sigma^2 \mathbf{w}^{\mathsf{T}} \boldsymbol{\Sigma} \mathbf{w}$$
$$\text{subject to} \ \ w_0 + \mathbf{w}^{\mathsf{T}} \mathbf{1} \leq V_0$$

follows from the solution to the trade-off investment problem (4.7) by replacing $\boldsymbol{\Sigma}$ in (4.7) by $\sigma^2 \boldsymbol{\Sigma}$ and is given by

$$\mathbf{w} = \frac{V_0}{c} (\sigma^2 \boldsymbol{\Sigma})^{-1} (\boldsymbol{\mu} - R_0 \mathbf{1}) \quad \text{and} \quad w_0 = V_0 - \mathbf{w}^{\mathsf{T}} \mathbf{1}.$$

A convenient feature of having an elliptical distribution for vector $\mathbf{R}$ of returns is that portfolio optimization problems often reduce to the trade-off investment problem (4.7). Consider the problem of portfolio optimization in the context of a spectral risk measure.

Example 9.11 (Spectral risk measures). Portfolio optimization with respect to a spectral risk measure (Sect. 6.5) amounts to minimizing a spectral risk measure $\rho_\phi(X)$, where X denotes a future portfolio value, under a budget constraint (and possibly additional constraints). By the stochastic representation (9.11), we can express the quantile function of V_1 as

$$F_{V_1}^{-1}(p) = w_0 R_0 + \mathbf{w}^{\mathsf{T}} \boldsymbol{\mu} + (\mathbf{w}^{\mathsf{T}} \boldsymbol{\Sigma} \mathbf{w})^{1/2} F_{WZ_1}^{-1}(p).$$

Therefore, the spectral risk measure

$$\rho_\phi(X) = -\int_0^1 \phi(p) F_{X/R_0}^{-1}(p) dp,$$

applied to $X = V_1 - V_0 R_0$, can be expressed as

$$\rho_\phi(V_1 - V_0 R_0) = -\int_0^1 \phi(p) F_{V_1/R_0}^{-1}(p) dp + V_0$$

$$= \frac{1}{R_0} \left(-w_0 R_0 - \mathbf{w}^{\mathsf{T}} \boldsymbol{\mu} - (\mathbf{w}^{\mathsf{T}} \boldsymbol{\Sigma} \mathbf{w})^{1/2} \int_0^1 \phi(p) F_{WZ_1}^{-1}(p) dp \right) + V_0.$$

In particular, we can formulate the portfolio optimization problem

$$\text{minimize} \ \ \rho_\phi(w_0 R_0 + \mathbf{w}^{\mathsf{T}} \mathbf{R} - V_0 R_0)$$
$$\text{subject to} \ \ w_0 + \mathbf{w}^{\mathsf{T}} \mathbf{1} \leq V_0$$

as the trade-off problem

$$\text{maximize} \ \ w_0 R_0 + \mathbf{w}^{\mathsf{T}} \boldsymbol{\mu} - \tfrac{c}{2V_0} (\mathbf{w}^{\mathsf{T}} \boldsymbol{\Sigma} \mathbf{w})^{1/2}$$
$$\text{subject to} \ \ w_0 + \mathbf{w}^{\mathsf{T}} \mathbf{1} \leq V_0,$$

where

$$c = -2V_0 \int_0^1 \phi(p) F_{WZ_1}^{-1}(p) dp.$$

We conclude that, for an elliptical model for vector $\mathbf{R}$ of returns, minimizing the spectral risk measure of the future portfolio value subject to a budget constraint is equivalent to solving a trade-off problem with the trade-off parameter given above.

9.4 Copulas

A rather common situation arises when we search for a multivariate model for a set of random variables $Y_1, \ldots, Y_d$ whose univariate distributions are rather well understood but whose joint distribution is only partially understood. A useful approach to the construction of a multivariate distribution for $\mathbf{Y} = (Y_1, \ldots, Y_d)$ with specified univariate marginal distribution functions $G_1, \ldots, G_d$, the distribution functions of the vector's components, is obtained by combining the so-called probability and quantile transforms. The probability transform says that if X is a random variable with a continuous distribution function F, then $F(X)$ is uniformly distributed on the interval $(0, 1)$. The quantile transform says that if U is uniformly distributed and if G is any distribution function, then $G^{-1}(U)$ has distribution function G. This implies that for any random vector $\mathbf{X} = (X_1, \ldots, X_d)$ whose components have continuous distribution functions $F_1, \ldots, F_d$, the random vector $\mathbf{Y} = (G_1^{-1}(F_1(X_1)), \ldots, G_d^{-1}(F_d(X_d)))$ corresponds to a multivariate model with prespecified univariate marginal distributions. If all F_k and G_k are both continuous and strictly increasing, then the preceding statement is actually straightforward to verify:

$$P(G_k^{-1}(F_k(X_k)) \le y) = P(F_k(X_k) \le G_k(y)) = G_k(y),$$

which shows that Y_k has distribution function G_k. The difficulty when it comes to constructing a good multivariate model for $\mathbf{Y}$ using this approach clearly lies in the choice of the distribution for vector $\mathbf{X}$ since the dependence between the X_k will be inherited by the Y_k.

Example 9.12. Consider the two scatter plots in Fig. 9.6. The left scatter plot shows a sample of size 2,000 from a bivariate standard normal distribution with linear correlation 0.5. The right scatter plot shows a sample of size 2,000 from a bivariate distribution with standard normal marginal distributions and a dependence structure inherited from a bivariate Student's t distribution with one degree of freedom. The points of the right scatter plot were obtained from the points of the left scatter plot as follows. Write $\mathbf{Z}_1, \ldots, \mathbf{Z}_{2000}$ for the independent bivariate normal random vectors whose outcomes are shown in the left plot. Let $S_1, \ldots, S_{2000}$ be independent χ_1^2-distributed random variables independent of the sample from the bivariate normal distribution. A sample of independent bivariate Student's t_1-distributed vectors was obtained by setting $\mathbf{X}_k = S_k^{-1/2} \mathbf{Z}_k$ for $k = 1, \ldots, 2000$. Finally, the random vectors whose outcomes are shown in the plot to the right were constructed as $\mathbf{Y}_k = (\Phi^{-1}(t_1(X_{k,1})), \Phi^{-1}(t_1(X_{k,2})))^\mathsf{T}$ for $k = 1, \ldots, 2000$.
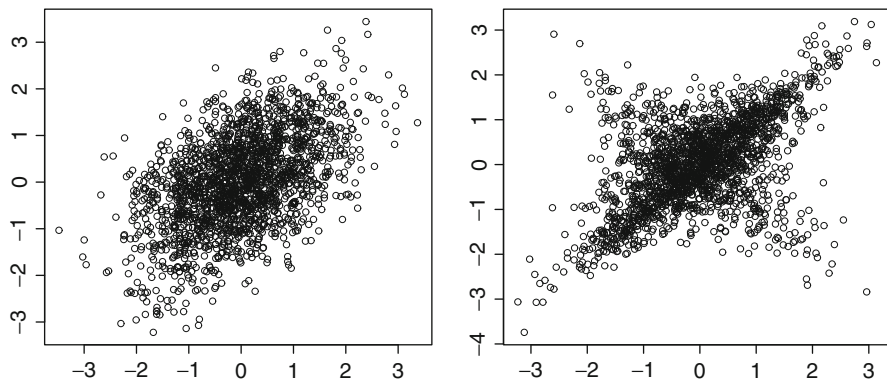
Fig. 9.6 Samples of size 2,000 from two bivariate distributions with standard normal marginal distributions. *Left plot*: sample from a bivariate standard normal with linear correlation 0.5. *Right plot*: sample from a bivariate standard Student's t distribution with one degree of freedom, with marginal distributions transformed to a standard normal

Suppose that we want to build a multivariate model corresponding to a random vector $\mathbf{X} = (X_1, \ldots, X_d)$ with a nontrivial dependence between its components and certain marginal distribution functions $F_1, \ldots, F_d$. Then the quantile transform says that we may start with a suitable vector $\mathbf{U} = (U_1, \ldots, U_d)$ whose components are uniformly distributed on $(0, 1)$ and specify $\mathbf{X}$ as

$$\mathbf{X} = (F_1^{-1}(U_1), \ldots, F_d^{-1}(U_d)).$$

The random vector $\mathbf{X}$ inherits the dependence among its components from vector $\mathbf{U}$. The distribution function C of a random vector $\mathbf{U}$ whose components U_k are uniformly distributed on $(0, 1)$ is called a copula, i.e.,

$$C(u_1, \ldots, u_d) = P(U_1 \leq u_1, \ldots, U_d \leq u_d), \quad (u_1, \ldots, u_d) \in (0, 1)^d.$$

Let $(X_1, \ldots, X_d)$ be a random vector with distribution function $F(x_1, \ldots, x_d) = P(X_1 \leq x_1, \ldots, X_d \leq x_d)$ and suppose that $F_k(x) = P(X_k \leq x)$ is a continuous function for every k. The probability transform, statement (iv) of Proposition 6.1, implies that the components of the vector $\mathbf{U} = (U_1, \ldots, U_d) = (F_1(X_1), \ldots, F_d(X_d))$ are uniformly distributed on $(0, 1)$. In particular, the distribution function C of $\mathbf{U}$ is a copula and we call it the copula of $\mathbf{X}$. Using statement (i) of Proposition 6.1 we find that

$$\begin{aligned}
C(F_1(x_1), \ldots, F_d(x_d)) &= P(U_1 \leq F_1(x_1), \ldots, U_d \leq F_d(x_d)) \\
&= P(F_1^{-1}(U_1) \leq x_1, \ldots, F_d^{-1}(U_d) \leq x_d) \\
&= F(x_1, \ldots, x_d).
\end{aligned}$$

This representation of the joint distribution function F in terms of the copula C and the marginal distribution functions $F_1, \ldots, F_d$ explains the name "copula": a function that "couples" the joint distribution function to its univariate marginal distribution functions.

Example 9.13 (Gaussian and Student's t copulas). The copula C_R^{Ga} of a d-dimensional standard normal distribution, with linear correlation matrix R, is the distribution function of the random vector $(\Phi(X_1), \ldots, \Phi(X_d))$, where Φ is the univariate standard normal distribution function and $\mathbf{X}$ is $\mathrm{N}_d(\mathbf{0}, R)$-distributed. Hence,

$$C_R^{\mathrm{Ga}}(\mathbf{u}) = \mathrm{P}(\Phi(X_1) \leq u_1, \ldots, \Phi(X_d) \leq u_d) = \Phi_R^d(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)),$$

where Φ_R^d is the distribution function of $\mathbf{X}$. Copulas of the preceding form are called Gaussian copulas.

The copula $C_{\nu,R}^t$ of a d-dimensional standard Student's t distribution with $\nu > 0$ degrees of freedom and linear correlation matrix R is the distribution of the random vector $(t_\nu(X_1), \ldots, t_\nu(X_d))$, where $\mathbf{X}$ has a $t_d(\mathbf{0}, R, \nu)$ distribution and t_ν is the univariate standard Student's t_ν distribution function. Hence,

$$C_{\nu,R}^t(\mathbf{u}) = \mathrm{P}(t_\nu(X_1) \leq u_1, \ldots, t_\nu(X_d) \leq u_d) = t_{\nu,R}^d(t_\nu^{-1}(u_1), \ldots, t_\nu^{-1}(u_d)),$$

where $t_{\nu,R}^d$ the distribution function of $\mathbf{X}$. Copulas of the preceding form are called Student's t copulas.

Consider a random vector (Y_1, Y_2) with continuous strictly increasing marginal distribution functions G_1 and G_2 and the copula of a Student's t distribution with linear correlation parameter ρ. We consider here the question of how ρ can be estimated from a sample from the distribution of (Y_1, Y_2). We may write $(Y_1, Y_2) = (G_1^{-1}(F_1(X_1)), G_2^{-1}(F_2(X_2)))$, where (X_1, X_2) has a Student's t distribution with linear correlation parameter ρ. In particular, the functions T_1 and T_2 given by $T_k(x) = G_k^{-1}(F_k(x))$ are continuous and strictly increasing, so for an independent copy (X_1', X_2') of (X_1, X_2) it holds that

$$\begin{aligned}
\tau(Y_1, Y_2) &= \tau(T_1(X_1), T_2(X_2)) \\
&= 2\,\mathrm{P}((T_1(X_1) - T_1(X_1'))(T_2(X_2) - T_2(X_2')) > 0) - 1 \\
&= 2\,\mathrm{P}((X_1 - X_1')(X_2 - X_2') > 0) - 1 \\
&= \tau(X_1, X_2).
\end{aligned}$$

It follows immediately from (9.7) that $\rho = \sin(\pi \tau(Y_1, Y_2)/2)$. Therefore, the estimate $\widehat{\tau}$ of $\tau(Y_1, Y_2)$ from the sample from the distribution of (Y_1, Y_2) gives an estimate $\widehat{\rho} = \sin(\pi \widehat{\tau}/2)$ of ρ.

Example 9.14 (Investments in foreign stocks). Consider a Swedish investor about to invest Swedish kronor (SEK) in foreign telecom stocks. The current share prices of British Telecom (BT) and Deutsche Telekom (DT) are 185.5 British pounds (GBP)
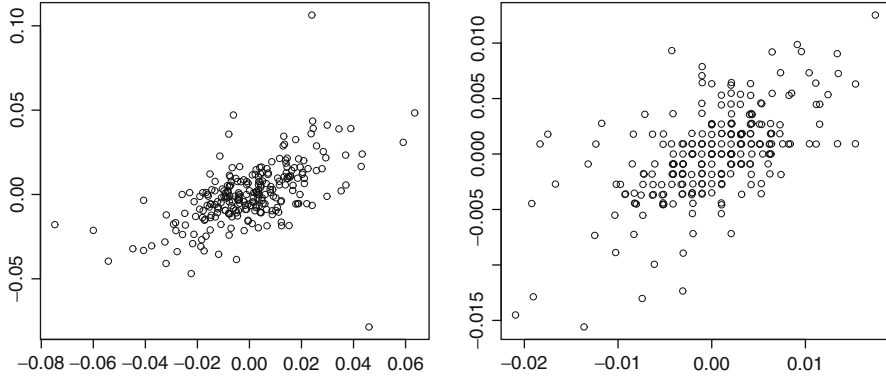
Fig. 9.7 *Scatter plot* to *left* shows log-return pairs, British Telecom in pounds on the *x*-axis and Deutsche Telekom in euros on the *y*-axis. The *scatter plot* to the *right shows* log-return pairs, SEK/GBP on the *x*-axis and SEK/EUR on the *y*-axis

and 9.26 euros (EUR), respectively. The current SEK/GBP exchange rate is 0.0942 (*x* kronor can be exchanged for 0.0942*x* pounds). The current SEK/EUR exchange rate is 0.1098.

The investor has obtained a sample of four-dimensional vectors of share prices, in the local currencies, and exchange rates from the 249 most recent (trading) days. We assume that the investor believes that the information in the data is relevant for assessing future portfolio values, and that no additional information on which to base model selection is available. The scatter plots for the stock log-return pairs and for the exchange-rate log-return pairs are shown in Fig. 9.7.

The investor is about to invest the amounts w_1 and w_2 kronor in the two foreign telecom stocks and wants to model the portfolio value V_1 in kronor tomorrow. Let A_t, B_t, C_t, D_t denote the time t share prices (BT and DT) and exchange rates (SEK/GBP and SEK/EUR). Let $X_A = \log(A_1/A_0)$ be the log return from today until tomorrow for BT in GBP and similarly for X_B, X_C, X_D. If h_1 and h_2 are the number of shares of BT and DT bought, then

$$\frac{A_0}{C_0} h_1 = w_1 \quad \text{and} \quad \frac{B_0}{D_0} h_2 = w_2.$$

The portfolio value in kronor tomorrow is therefore

$$\begin{aligned}
V_1 &= h_1 \frac{A_1}{C_1} + h_2 \frac{B_1}{D_1} \\
&= w_1 \frac{A_1}{A_0} \left(\frac{C_1}{C_0}\right)^{-1} + w_2 \frac{B_1}{B_0} \left(\frac{D_1}{D_0}\right)^{-1} \\
&= w_1 \exp\{X_A - X_C\} + w_2 \exp\{X_B - X_D\}.
\end{aligned}$$

If the investor has already decided on a particular portfolio, i.e., has chosen the portfolio weights w_1 and w_2, then the log-return data may be used to generate a sample from the distribution of V_1 by viewing V_1 as a function of (X_A, X_B, X_C, X_D). This sample can be transformed into a sample from the distribution of the portfolio log return $\log(V_1/V_0)$, where $V_0 = w_1 + w_2$, and a parametric model can be chosen for the portfolio log return.

Here we want to allow the investor to vary the portfolio weights in order to choose an optimal (according to some criterion left unspecified) portfolio. Therefore, instead of setting up a model for V_1 directly, we set a model for the joint log-return distribution of (X_A, X_B, X_C, X_D) from which the model for V_1 is easily inferred.

The Student's t location-scale family of distributions is a natural choice of parametric family for log returns. Maximum-likelihood estimation of the parameter triple (μ, σ, ν) of the Student's t location-scale family on the samples of daily log returns gives the following estimates:

$$
\begin{aligned}
(-6 \cdot 10^{-4},\ 0.013,\ 3.7) &\quad \text{(British Telecom in pounds)}, \\
(2 \cdot 10^{-4},\ 0.015,\ 7.7) &\quad \text{(Deutsche Telekom in euros)}, \\
(2 \cdot 10^{-4},\ 0.006,\ 9.6) &\quad \text{(SEK/GBP)}, \\
(8 \cdot 10^{-5},\ 0.004,\ 8.6) &\quad \text{(SEK/EUR)}.
\end{aligned}
$$

There is no a priori reason for the log-return distributions to be symmetric; the polynomial normal model in Example 8.10 is also a natural model for the log returns. The estimated parameters $(\theta_0, \theta_1, \theta_2, \theta_3)$ based on the samples of daily log returns are

$$
\begin{aligned}
(3.1,\ 142.6,\ -1.4,\ 15.5) \cdot 10^{-4} &\quad \text{(British Telecom in pounds)}, \\
(-8.8,\ 120.1,\ 9.2,\ 22.1) \cdot 10^{-4} &\quad \text{(Deutsche Telekom in euros)}, \\
(4.0,\ 53.5,\ -3.7,\ 3.6) \cdot 10^{-4} &\quad \text{(SEK/GBP)}, \\
(1.9,\ 31.1,\ -2.0,\ 3.5) \cdot 10^{-4} &\quad \text{(SEK/EUR)}.
\end{aligned}
$$

The conditions $\theta_3 > 0$ and $3\theta_1\theta_3 - \theta_2^2 > 0$ ensuring that the third-degree polynomial is strictly increasing is satisfied for estimated parameter vectors. Figure 9.8 shows the empirical quantiles of the log returns of BT and DT against those of the fitted parametric distributions. By comparing the two upper q–q plots we find that the polynomial normal model captures the asymmetry between the left and right tails in BT log-return data, whereas the Student's t model does not.

We now proceed to the modeling of the dependence between the log returns. The sample correlations between log returns of the stocks and log returns of the exchange rates is approximately zero, and there are no obvious economic reasons not to assume independence between the log-return pairs (X_A, X_B) and (X_C, X_D) of stocks and exchange rates, respectively. We therefore assume that the log-return pairs (X_A, X_B) and (X_C, X_D) are independent and that the distribution functions of the two log-return pairs are of the form, with subscripts s for stocks and e for exchange rates,

$$
P(X_A \le x_A, X_B \le x_B) = C^t_{\nu_s, \rho_s}(F_A(x_A), F_B(x_B)),
$$

$$
P(X_C \le x_C, X_D \le x_D) = C^t_{\nu_e, \rho_e}(F_C(x_C), F_D(x_D)),
$$
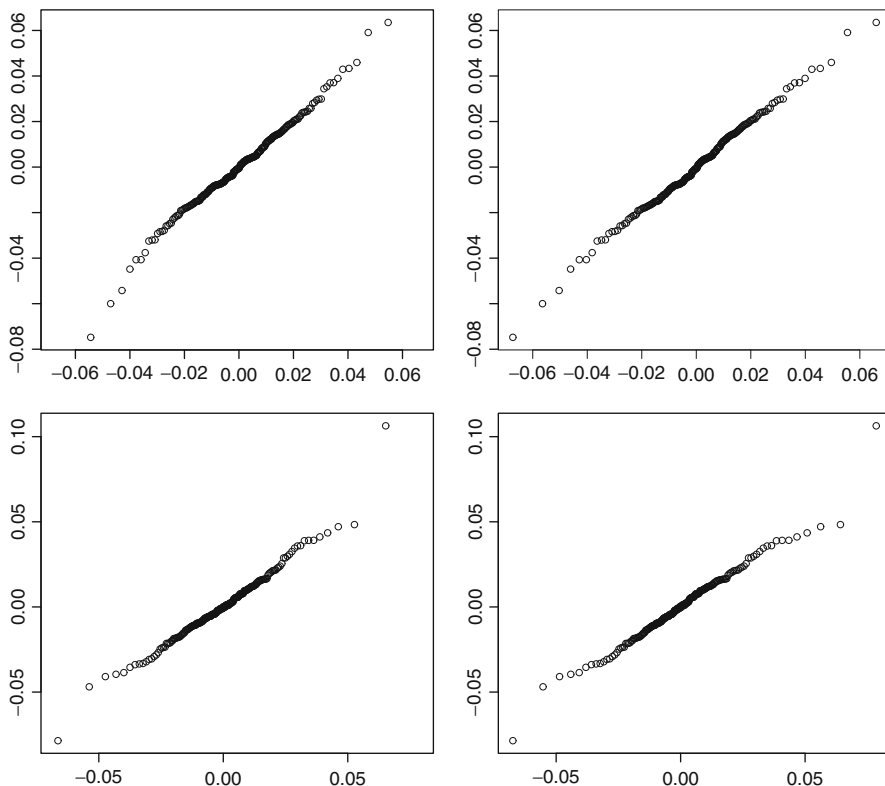
Fig. 9.8 *Upper plots*: empirical quantiles of British Telecom log-return data (y-axes) against quantiles of fitted distributions (x-axes): Student's t model to the *left* and polynomial normal model to the *right*. *Lower plots*: empirical quantiles of Deutsche Telekom log-return data (y-axes) against quantiles of fitted distributions (x-axes): Student's t model to the *left* and polynomial normal model to the *right*

where F_A, F_B, F_C, F_D denote the distribution functions of X_A, X_B, X_C, X_D. Student's t copula is a flexible parametric family for the dependence structure of the log-return pairs. Set $U_A = F_A(X_A)$ and similarly for U_B, U_C, U_D. The assumption of Student's t copulas as models for the dependence structure for the log-return pairs requires that $(U_A, U_B) \overset{\mathrm{d}}{=} (1 - U_A, 1 - U_B)$ and $(U_C, U_D) \overset{\mathrm{d}}{=} (1 - U_C, 1 - U_D)$. Whatever choice of models for the individual log returns X_A, X_B, X_C, X_D among the sets of models given above, the log-return data give no reasons to reject the hypothesis that $(U_A, U_B) \overset{\mathrm{d}}{=} (1 - U_A, 1 - U_B)$ and $(U_C, U_D) \overset{\mathrm{d}}{=} (1 - U_C, 1 - U_D)$ (Fig. 9.9).

We may now estimate ρ_s and ρ_e by $\widehat{\rho}_s = \sin(\pi \widehat{\tau}_s / 2)$ and $\widehat{\rho}_e = \sin(\pi \widehat{\tau}_e / 2)$, and the estimate of (ρ_s, ρ_e) is approximately $(0.62, 0.61)$. Under the assumption that the marginal distribution functions F_A, F_B, F_C, F_D of the joint log-return
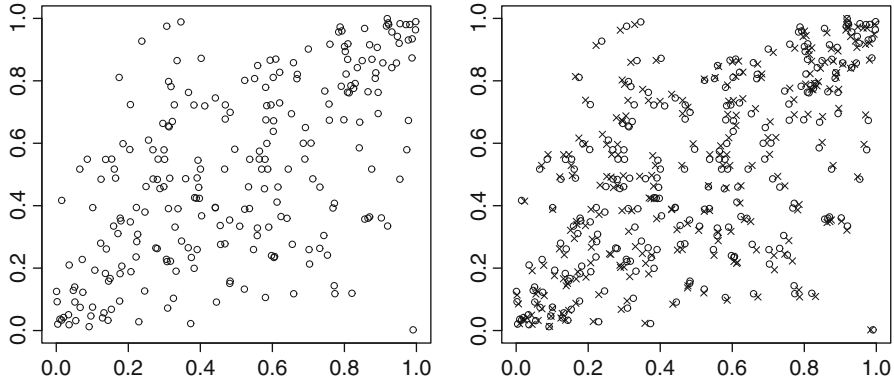
Fig. 9.9 *Left scatter plot*: sample points in (9.12) obtained by componentwise transformation of original log-return pairs for stocks by fitted Student's *t* location-scale distribution functions. *Right scatter plot*: with corresponding sample points for componentwise transformation by distribution functions of fitted polynomial normal models added, marked by *times symbol*, to illustrate the effect of the componentwise transformations

distribution equal the estimated marginal distribution functions $\widehat{F}_A, \widehat{F}_B, \widehat{F}_C, \widehat{F}_D$, we may transform the samples

$$\{(X_A^1, X_B^1), \ldots, (X_A^{248}, X_B^{248})\} \quad \text{and} \quad \{(X_C^1, X_D^1), \ldots, (X_C^{248}, X_D^{248})\}$$

into the samples

$$\{(U_A^1, U_B^1), \ldots, (U_A^{248}, U_B^{248})\} \quad \text{and} \quad \{(U_C^1, U_D^1), \ldots, (U_C^{248}, U_D^{248})\} \qquad (9.12)$$

from Student's *t* copulas, where $U_A^k = \widehat{F}_A(X_A^k)$, and similarly for U_B^k, U_C^k, U_D^k. In the case of a polynomial normal model choice, dropping subscripts for notational convenience, $\widehat{F}(x) = \Phi(\widehat{g}^{-1}(x))$, and $\widehat{g}^{-1}(x)$ is obtained as the (here unique real) solution y to the polynomial equation $\widehat{\theta}_0 + \widehat{\theta}_1 y + \widehat{\theta}_2 y^2 + \widehat{\theta}_3 y^3 = x$. Under the further assumption that the linear correlation parameters ρ_s, ρ_e equal the estimates $\widehat{\rho}_s, \widehat{\rho}_e$, the two samples in (9.12) are samples from two Student's *t* copulas whose parameters are known except for the degree-of-freedom parameters ν_s and ν_e. The unknown parameters can be estimated by maximum likelihood, and the bivariate density function of Student's *t* copula corresponding to the pair of log returns for stocks is given by

$$c_{\nu_s, \widehat{\rho}_s}^t(u_1, u_2) = \frac{\partial^2}{\partial u_1 \partial u_2} t_{\nu_s, \widehat{\rho}_s}^2(t_{\nu_s}^{-1}(u_1), t_{\nu_s}^{-1}(u_2)) = \frac{g_{\nu_s, \widehat{\rho}_s}^2(t_{\nu_s}^{-1}(u_1), t_{\nu_s}^{-1}(u_2))}{g_{\nu_s}(t_{\nu_s}^{-1}(u_1)) g_{\nu_s}(t_{\nu_s}^{-1}(u_2))},$$

where $t^2_{\nu_s,\widehat{\rho}_s}$ and $g^2_{\nu_s,\widehat{\rho}_s}$ denote the distribution and density function, respectively, of the bivariate Student's t distribution with degree-of-freedom parameter ν_s and linear correlation parameter $\widehat{\rho}_s$, and t_{ν_s} and g_{ν_s} denote the distribution and density function, respectively, of the univariate Student's t distribution with degree-of-freedom parameter ν_s. The procedure is similar for the pair of log returns for the exchange rates.

The samples in (9.12) depend on the choice of parametric models for the log returns and the corresponding parameter estimates. Therefore, we will here obtain two pairs of estimates $(\widehat{\nu}_s, \widehat{\nu}_e)$ of the copula parameters ν_s and ν_e. If the log-return distributions are assumed to be Student's t distributions and the parameters are estimated by maximum likelihood, then we obtain the copula parameter estimates $(\widehat{\nu}_s, \widehat{\nu}_e) \approx (5.1, 6.8)$. If the log-return distributions are assumed to be given by the polynomial normal model, then we obtain the copula parameter estimates $(\widehat{\nu}_s, \widehat{\nu}_e) \approx (3.6, 5.5)$.

Now that the two models for the joint log-return distribution of the vector (X_A, X_B, X_C, X_D) are set up and their parameters estimated, we evaluate the models in terms of how close the resulting distribution of the portfolio log return

$$\log(V_1/V_0), \quad V_1 = \frac{V_0}{2} \exp\{X_A - X_C\} + \frac{V_0}{2} \exp\{X_B - X_D\}$$

is to the empirical distribution of the portfolio log return. The joint log-return models do not give closed-form expressions for the distributions of the portfolio log return. However, the portfolio log-return distributions are straightforward to simulate from. We simulate 10^5 outcomes of $\log(V_1/V_0)$, according to the chosen model, by simulating outcomes (Z_A, Z_B) and (Z_C, Z_D) of independent Student's t-distributed random vectors and using the formula

$$\log\left(\frac{1}{2} \exp\{\widehat{F}_A^{-1}(t_{\widehat{\nu}_s}(Z_A)) - \widehat{F}_C^{-1}(t_{\widehat{\nu}_e}(Z_C))\} + \frac{1}{2} \exp\{\widehat{F}_B^{-1}(t_{\widehat{\nu}_s}(Z_B)) - \widehat{F}_D^{-1}(t_{\widehat{\nu}_e}(Z_D))\}\right),$$

where (Z_A, Z_B) has a bivariate standard Student's t distribution with degree-of-freedom parameter $\widehat{\nu}_s$ and linear correlation parameter $\widehat{\rho}_s$, and (Z_C, Z_D) has a bivariate standard Student's t distribution with degree-of-freedom parameter $\widehat{\nu}_e$ and linear correlation parameter $\widehat{\rho}_e$. Finally, we compare the empirical distributions of the simulated samples of size 10^5 to the empirical distribution based on the original log-return sample. The result is shown in Fig. 9.10. Both models give a good fit to the log-return data.

If $\mathbf{X} = (X_1, \ldots, X_d)$ is a random vector with continuous marginal distribution functions $F_1, \ldots, F_d$, and if $G_1, \ldots, G_d$ are any given distribution functions, then the random vector $\mathbf{Y} = (G_1^{-1}(F_1(X_1)), \ldots, G_d^{-1}(F_d(X_d)))$ has marginal distribution functions $G_1, \ldots, G_d$ and has inherited the dependence structure or copula from vector $\mathbf{X}$. However, it may happen that the distribution functions $F_1, \ldots, F_d$ cannot be determined explicitly. Another option is to consider a family of models for vectors $(U_1, \ldots, U_d)$ with components that are uniformly distributed on $(0, 1)$ and consider models of the form $\mathbf{Y} = (G_1^{-1}(U_1), \ldots, G_d^{-1}(U_d))$.
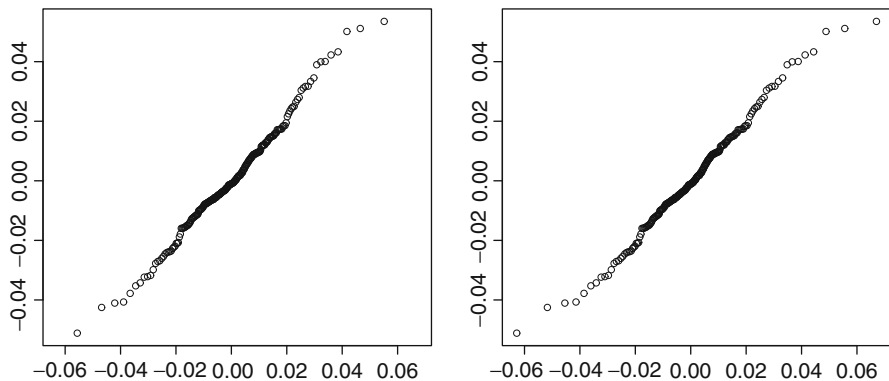
Fig. 9.10 These two q–q plots show the empirical quantiles of the portfolio log returns (y-axes) for $w_1 = w_2 = V_0/2$ against the quantiles of two models for $\log(V_1/V_0)$ (x-axes). The plot to the *left* corresponds to the model for (X_A, X_B, X_C, X_D) with Student's t marginal distributions, and the plot to the *right* corresponds to the model for (X_A, X_B, X_C, X_D) with polynomial normal marginal distributions

Example 9.15 (Archimedean copulas). Consider a strictly positive random variable X with a density f and Laplace transform $\Psi(t) = \mathrm{E}[e^{-tX}]$. A useful family of copulas called Archimedean copulas is based on the fact that $\Psi(-\log(V)/X)$ is uniformly distributed on $(0,1)$ if V is uniformly distributed on $(0,1)$ and independent of X. To verify this claim we first note that

$$\Psi(t) = \int_0^\infty e^{-tx} f(x)dx \quad \text{and} \quad \Psi'(t) = -\int_0^\infty xe^{-tx} f(x)dx < 0,$$

so Ψ is nonnegative, continuous, and strictly decreasing on $[0,\infty)$. For any $u \in (0,1)$ we can now verify that

$$
\begin{aligned}
\mathrm{P}\left(\Psi\left(\frac{-\log V}{X}\right) \le u\right) &= \mathrm{E}\left[\mathrm{P}\left(\Psi\left(\frac{-\log V}{X}\right) \le u \mid X\right)\right] \\
&= \mathrm{E}\left[\mathrm{P}\left(V \le e^{-\Psi^{-1}(u)X} \mid X\right)\right] \\
&= \mathrm{E}\left[e^{-\Psi^{-1}(u)X}\right] \\
&= \Psi(\Psi^{-1}(u)) = u.
\end{aligned}
$$

It follows that if $V_1, \ldots, V_d$ are uniformly distributed on $(0,1)$ and independent of X, then the distribution function C of

$$\mathbf{U} = \left(\Psi\left(\frac{-\log V_1}{X}\right), \ldots, \Psi\left(\frac{-\log V_d}{X}\right)\right) \tag{9.13}$$

is a copula. We should always aim to understand a multivariate model through its stochastic representation. Here Ψ is decreasing with $\Psi(0) = 1$ and $\lim_{t\to\infty} \Psi(t) = 0$. Therefore, we observe that if X takes a small value, then the random variables $-\log(V_k)/X$, for $k = 1, \ldots, d$, are all likely to take large values, which implies small values for the random variables $U_k = \Psi(-\log(V_k)/X)$. In particular, choosing a random variable X that has a relatively high probability of taking very small values is likely to lead to asymptotic dependence, in the sense that small values for one component are likely to imply small values for other components, for a model with the stochastic representation $(G_1^{-1}(U_1), \ldots, G_d^{-1}(U_d))$. Simulation from an Archimedean copula C as above is straightforward: just independently simulate standard uniform variates $V_1, \ldots, V_d$ and X, and set $\mathbf{U}$ according to (9.13). Note that the copula can be expressed explicitly as

$$
\begin{aligned}
C(u_1, \ldots, u_d) &= \mathrm{P}(U_1 \leq u_1, \ldots, U_d \leq u_d) \\
&= \mathrm{E}\left[\mathrm{P}\left(V_1 \leq e^{-\Psi^{-1}(u_1)X}, \ldots, V_d \leq e^{-\Psi^{-1}(u_d)X} \mid X\right)\right] \\
&= \mathrm{E}\left[e^{-(\Psi^{-1}(u_1)+\cdots+\Psi^{-1}(u_d))X}\right] \\
&= \Psi(\Psi^{-1}(u_1) + \cdots + \Psi^{-1}(u_d)).
\end{aligned}
\tag{9.14}
$$

Example 9.16 (Clayton copula). If X has a Gamma$(1/\theta, 1)$ distribution, then X has density function $f(x) = x^{1/\theta-1}e^{-x}/\Gamma(1/\theta)$ and Laplace transform

$$
\Psi(t) = \mathrm{E}[e^{-tX}] = \int_0^\infty e^{-tx} \frac{1}{\Gamma(1/\theta)} x^{1/\theta-1}e^{-x}\mathrm{d}x = (t + 1)^{-1/\theta}.
$$

This choice of Ψ gives the Clayton copula. Solving $\Psi(\Psi^{-1}(u)) = u$ for $\Psi^{-1}(u)$ gives $\Psi^{-1}(u) = u^{-\theta} - 1$. Therefore, the copula expression (9.14) takes the form

$$
C_\theta^{\mathrm{Cl}}(\mathbf{u}) = (u_1^{-\theta} + \cdots + u_d^{-\theta} - d + 1)^{-1/\theta}.
$$

Applying l'Hôpital's rule shows that the Clayton copula has lower tail dependence in the sense that

$$
\begin{aligned}
\lim_{u\to0} \mathrm{P}(U_k \leq u \mid U_j \leq u) &= \lim_{u\to0} \frac{(2u^{-\theta} - 1)^{-1/\theta}}{u} \\
&= \lim_{u\to0} \frac{\frac{d}{du}(2u^{-\theta} - 1)^{-1/\theta}}{\frac{d}{du}u} \\
&= \lim_{u\to0} 2u^{-\theta-1}(2u^{-\theta} - 1)^{-1/\theta-1} \\
&= 2^{-1/\theta}.
\end{aligned}
$$

If $\theta = 1$, then both X, and the random variables $-\log V_k$ are standard exponentially distributed. In particular, we may write

$$(U_1, \ldots, U_d) \stackrel{\mathrm{d}}{=} \left(\frac{E_0}{E_0 + E_1}, \ldots, \frac{E_0}{E_0 + E_d} \right),$$

where $E_0, E_1, \ldots, E_d$ are independent and standard exponentially distributed. We see that for all the U_k to take small values, we need E_0 to take a small value. However, for all the U_k to take large values, we need E_0 to take a large value and for all $E_1, \ldots, E_d$ to take small values. The latter is less likely, and therefore a reasonable guess is that the Clayton copula does not have upper tail dependence: $\lim_{u \to 0} \mathrm{P}(U_k > u \mid U_j > u) = 0$. An application of l'Hôpital's rule verifies this claim. Samples from the Clayton copula are illustrated graphically in Fig. 9.11.

9.4.1 Misconceptions of Correlation and Dependence

Now we turn to common misconceptions of linear correlation. We have seen that given any two univariate distribution functions F_1 and F_2 and copula function C, $F(x_1, x_2) = C(F_1(x_1), F_2(x_2))$ is a bivariate distribution function with marginal distribution functions F_1 and F_2. It is typically hard to know which copula C to choose, and it is therefore tempting to ask for a bivariate distribution with given marginal distribution functions F_1 and F_2 and a given linear correlation coefficient ρ. However, we will see that this question is ill-posed in the sense that the set of bivariate distributions fulfilling the requirement may be empty.

To this end we first consider an integral representation of the covariance between two random variables in terms of their joint distribution function and their marginal distribution functions.

Proposition 9.8. *If (X_1, X_2) has distribution function F and marginal distribution functions F_1 and F_2 and the covariance $\mathrm{Cov}(X_1, X_2)$ exists finitely, then*

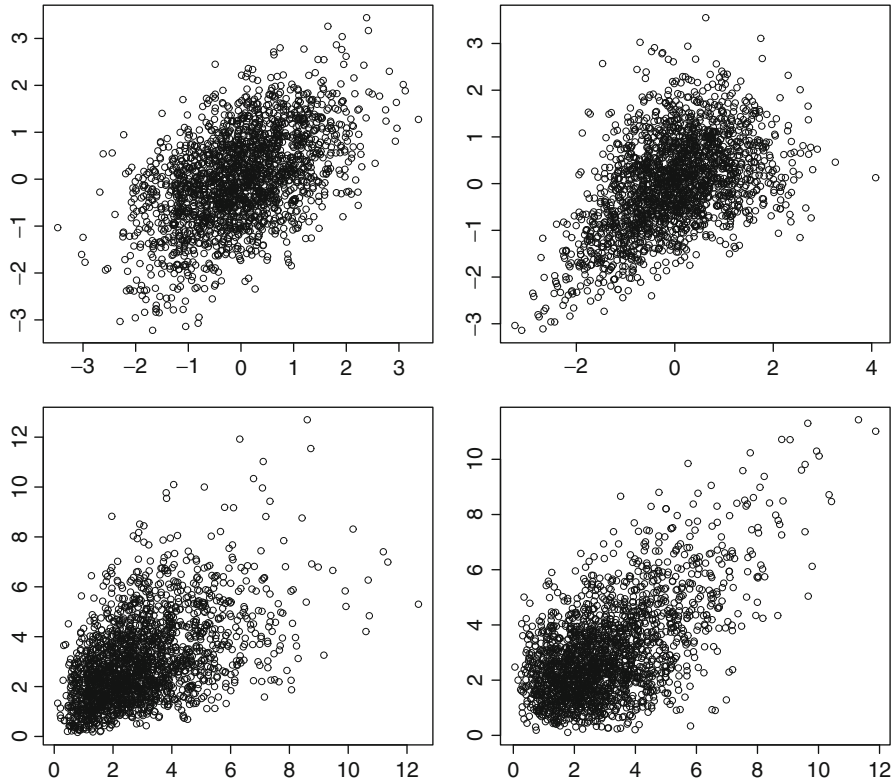$$\mathrm{Cov}(X_1, X_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (F(x_1, x_2) - F_1(x_1) F_2(x_2)) dx_1 dx_2.$$

Proof. Let (Y_1, Y_2) be an independent copy of (X_1, X_2), and note that

$$\mathrm{E}[(X_1 - Y_1)(X_2 - Y_2)] = \mathrm{E}[X_1 X_2] - \mathrm{E}[X_1 Y_2] + \mathrm{E}[Y_1 Y_2] - \mathrm{E}[Y_1 X_2] = 2\,\mathrm{Cov}(X_1, X_2).$$

Writing

$$(X_1 - Y_1) = \int_{-\infty}^{\infty} (I\{Y_1 \leq x_1\} - I\{X_1 \leq x_1\}) dx_1,$$

Fig. 9.11 The *upper two scatter plots* show samples of size 2,000 from two bivariate distributions with standard normal marginal distributions. The *left plot* shows a sample from the bivariate standard normal distribution with linear correlation coefficient 0.5 and the *right plot* shows a sample from a bivariate Clayton copula with parameter $\theta = 1$, componentwise transformed to standard normal marginal distributions. The *two lower scatter plots* show samples of size 2,000 from two bivariate distributions with Gamma(3, 1) marginal distributions. The *left plot* corresponds to the copula of a bivariate standard normal distribution with linear correlation 0.5, and the *right plot* corresponds to the copula of the vector (U_1, U_2) such that $(1 - U_1, 1 - U_2)$ has a bivariate Clayton copula with parameter $\theta = 1$

and similarly for $(X_2 - Y_2)$, we find that

$$
\begin{aligned}
&\mathrm{E}[(X_1 - Y_1)(X_2 - Y_2)] \\
&= \mathrm{E}\left[\int_{-\infty}^{\infty} (I\{Y_1 \leq x_1\} - I\{X_1 \leq x_1\}) dx_1 \int_{-\infty}^{\infty} (I\{Y_2 \leq x_2\} - I\{X_2 \leq x_2\}) dx_2 \right] \\
&= \mathrm{E}\left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (I\{Y_1 \leq x_1\} - I\{X_1 \leq x_1\})(I\{Y_2 \leq x_2\} - I\{X_2 \leq x_2\}) dx_1 dx_2 \right]
\end{aligned}
$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E[I\{Y_1 \le x_1\} - I\{X_1 \le x_1\}]\,E[I\{Y_2 \le x_2\} - I\{X_2 \le x_2\}]dx_1 dx_2$$

$$= 2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(F(x_1, x_2) - F_1(x_1)F_2(x_2))dx_1 dx_2,$$

from which the conclusion follows. □

To determine which joint distribution function gives the minimal and maximal covariance (and therefore also linear correlation), we need to determine sharp upper and lower bounds on F in terms of F_1 and F_2. Note that

$$\min(P(X_1 \le x_1), P(X_2 \le x_2)) \ge P(X_1 \le x_1, X_2 \le x_2)$$
$$= 1 - P(X_1 > x_1 \text{ or } X_2 > x_2)$$
$$\ge 1 - (P(X_1 > x_1) + P(X_2 > x_2))$$
$$= P(X_1 \le x_1) + P(X_2 \le x_2) - 1,$$

so

$$\max(F_1(x_1) + F_2(x_2) - 1, 0) \le F(x_1, x_2) \le \min(F_1(x_1), F_2(x_2)). \qquad (9.15)$$

If $(X_1, X_2) = (F_1^{-1}(U), F_2^{-1}(U))$, then statement (i) of Proposition 6.1 implies that

$$F(x_1, x_2) = P(F_1^{-1}(U) \le x_1, F_2^{-1}(U) \le x_2)$$
$$= P(U \le F_1(x_1), U \le F_2(x_2))$$
$$= \min(F_1(x_1), F_2(x_2)),$$

so the upper bound is attained. In this case, X_1 and X_2 are said to be comonotonic. If $(X_1, X_2) = (F_1^{-1}(U), F_2^{-1}(1 - U))$, then statement (i) of Proposition 6.1 implies that

$$F(x_1, x_2) = P(F_1^{-1}(U) \le x_1, F_2^{-1}(1 - U) \le x_2)$$
$$= P(U \le F_1(x_1), 1 - U \le F_2(x_2))$$
$$= \max(F_1(x_1) + F_2(x_2) - 1, 0),$$

so the lower bound is also attained. In this case, X_1 and X_2 are said to be countermonotonic.

Proposition 9.9. *Let F_1 and F_2 be distribution functions for random variables with nonzero finite variances. The set of linear correlation coefficients $\rho(F)$ for the set of bivariate distribution functions F with marginal distribution functions F_1 and F_2 form a closed interval $[\rho_{\min}, \rho_{\max}]$ with $0 \in (\rho_{\min}, \rho_{\max})$ such that $\rho(F) = \rho_{\min}$ if and only if $F(x_1, x_2) = \max(F_1(x_1) + F_2(x_2) - 1, 0)$ and $\rho = \rho_{\max}$ if and only if $F(x_1, x_2) = \min(F_1(x_1), F_2(x_2))$.*

Proof. The existence of attainable minimum and maximum linear correlation values ρ_{min}, ρ_{max} follows immediately from Proposition 9.8 and the bounds in (9.15). Taking $F(x_1, x_2) = F_1(x_1)F_2(x_2)$ shows that $0 \in [\rho_{min}, \rho_{max}]$. By Proposition 9.8, $\rho_{max} = 0$ would imply that $\min(F_1(x_1), F_2(x_2)) = F_1(x_1)F_2(x_2)$ for all x_1, x_2, which in turn implies that either F_1 or F_2 takes only the values 0 and 1. Such distribution functions correspond to constant random variables for which the variance is zero. We conclude that $\rho_{max} > 0$. A similar argument shows that $\rho_{min} < 0$. It remains to show that any value in $[\rho_{min}, \rho_{max}]$ is attainable. For $\lambda \in [0, 1]$ the function

$$F_\lambda(x_1, x_2) = \lambda \max(F_1(x_1) + F_2(x_2) - 1, 0) + (1 - \lambda) \min(F_1(x_1), F_2(x_2))$$

is a distribution function since it is the distribution function of the random vector

$$I(F_1^{-1}(U), F_2^{-1}(1 - U)) + (1 - I)(F_1^{-1}(U), F_2^{-1}(U)),$$

where I and U are independent, I takes the value 1 with probability λ and the value 0 otherwise, and U is uniformly distributed on $(0, 1)$. Moreover, $F_\lambda(x_1, 1) = F_1(x_1)$ and $F_\lambda(1, x_2) = F_2(x_2)$. Varying $\lambda \in [0, 1]$ shows that all values in the interval $[\rho_{min}, \rho_{max}]$ are attainable correlation values. $\square$
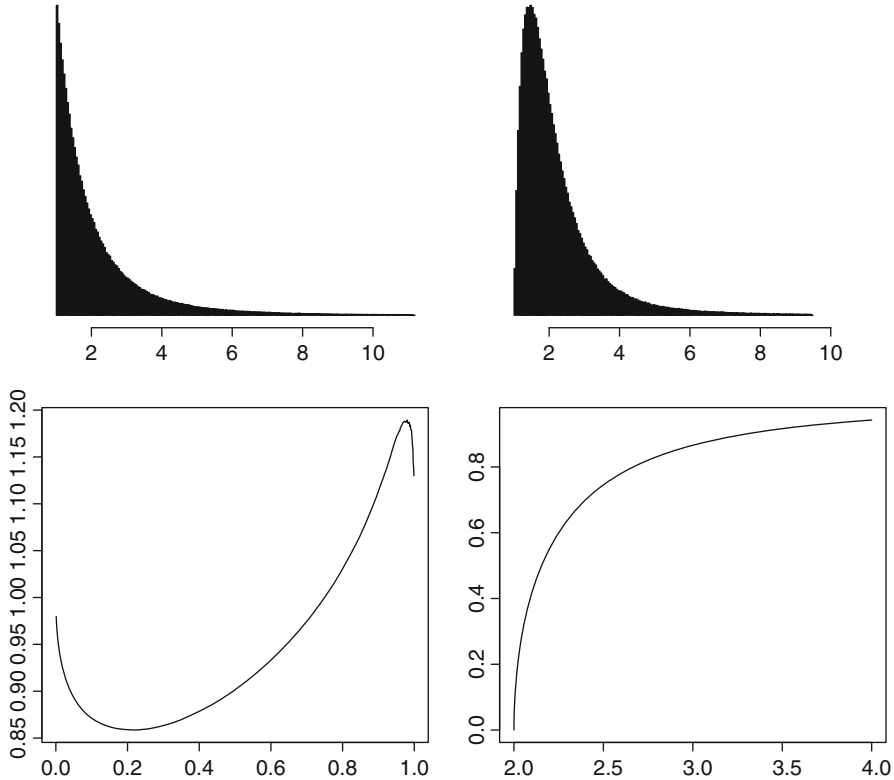
Example 9.17 (A bad stress test). Consider potential aggregate losses X and Y in two lines of business for an insurance company. Suppose that X is $\text{Exp}(\alpha)$-distributed and that Y is $\text{Pa}(\alpha)$-distributed with an unspecified dependence structure. To perform a stress test, the chief risk officer asks an actuary to assign a high linear correlation to the pair (X, Y) and study the effect on the quantile values for the sum $X + Y$. This problem is ill-posed. The correlation coefficient does not exist for $\alpha \leq 2$ since

$$\text{E}[Y^2] = \lim_{x \to \infty} \int_1^x y^2 \alpha y^{-\alpha-1} dy = \lim_{x \to \infty} \frac{\alpha}{2 - \alpha}(x^{2-\alpha} - 1)$$

does not exist finitely for $\alpha \leq 2$. Moreover, for $\alpha > 2$ not all correlation values are possible for the pair (X, Y), and for each attainable correlation value there are infinitely many possible joint distributions for (X, Y) that may produce very different distributions for $X + Y$.

To compute the upper bound ρ_{max} of the attainable correlation values, we note that $Y \overset{d}{=} e^X$ and that X and Y are comonotonic if $Y = e^X$. In particular, $\rho_{max} = \text{Cor}(X, e^X)$. The means and variances of the $\text{Exp}(\alpha)$ and $\text{Pa}(\alpha)$ distributions are given by

$$\text{E}[X] = \frac{1}{\alpha}, \quad \text{Var}(X) = \frac{1}{\alpha^2}, \quad \text{E}[e^X] = \frac{\alpha}{\alpha - 1}, \quad \text{Var}(e^X) = \frac{\alpha}{(\alpha - 1)^2(\alpha - 2)},$$

Fig. 9.12 The *upper two plots* show histograms of the distribution of $X + Y$, based on samples of size 10^6 and with the values corresponding to very high quantiles omitted, where X is $\text{Exp}(\alpha)$-distributed and Y is $\text{Pa}(\alpha)$-distributed with $\alpha = 2.1$. The *left histogram* corresponds to $Y = e^X$, and the *right histogram* corresponds to X and Y independent. *Lower right plot*: empirical quantiles of $X + e^X$ divided by empirical quantiles of $X + Y$ with X and Y independent. *Lower right plot*: $\rho_{\max}$ as a function of α

and integration by parts can be used to compute the covariance

$$\text{Cov}(X, e^X) = \text{E}[Xe^X] - \text{E}[X]\,\text{E}[e^X] = \int_0^\infty x\alpha e^{(1-\alpha)x}\,dx - \frac{1}{\alpha - 1} = \frac{1}{(\alpha - 1)^2}.$$

We find that

$$\rho_{\max} = \frac{\text{Cov}(X, e^X)}{\text{Var}(X)^{1/2}\,\text{Var}(e^X)^{1/2}} = \frac{(\alpha^2 - 2\alpha)^{1/2}}{\alpha - 1}.$$

The lower right plot in Fig. 9.12 shows $\rho_{\max}$ as a function of α. For instance, $\alpha = 2.1$ gives $\rho_{\max} \approx 0.4$, which may indicate weak dependence, although it corresponds to comonotonicity. The histograms in Fig. 9.12 show the distribution of $X + Y$ in the case of comonotonicity (left plot) and independence (right plot) for X and Y for $\alpha = 2.1$.

Example 9.18 (Correlation and causality). If we analyze quarterly data of changes in the 3-month, zero-coupon bond rate for government bonds and quarterly data of log returns of the country's stock market index, then it is likely that a bivariate autoregressive model of order 1, AR(1), gives a rather good fit. With X_t^1 and X_t^2 denoting the change in the 3-month rate and the index log return, respectively, from quarter $t-1$ to t, consider the model

$$\begin{pmatrix} X_t^1 \\ X_t^2 \end{pmatrix} = \begin{pmatrix} 0.45 & 0.02 \\ -9.2 & 0.35 \end{pmatrix} \begin{pmatrix} X_{t-1}^1 \\ X_{t-1}^2 \end{pmatrix} + \begin{pmatrix} Z_t^1 \\ Z_t^2 \end{pmatrix},$$

or in matrix form $\mathbf{X}_t = \mathbf{A}\mathbf{X}_{t-1} + \mathbf{Z}_t$, where the $\mathbf{Z}_k$ are independent and identically distributed and

$$\mathrm{Cov}(\mathbf{Z}_t) = \begin{pmatrix} 2 \cdot 10^{-5} & 0 \\ 0 & 10^{-2} \end{pmatrix}.$$

We find that

$$\mathbf{X}_t = \mathbf{A}\mathbf{X}_{t-1} + \mathbf{Z}_t = \mathbf{A}(\mathbf{A}\mathbf{X}_{t-2} + \mathbf{Z}_{t-1}) + \mathbf{Z}_t = \cdots = \sum_{k=0}^{\infty} \mathbf{A}^k \mathbf{Z}_{t-k}.$$

In particular,

$$\mathrm{Cov}(\mathbf{X}_t) = \sum_{k=0}^{\infty} \mathbf{A}^k \, \mathrm{Cov}(\mathbf{Z}_t)(\mathbf{A}^k)^{\mathsf{T}} \approx \begin{pmatrix} 3.18 \cdot 10^{-5} & -2.82 \cdot 10^{-5} \\ -2.82 \cdot 10^{-5} & 1.47 \cdot 10^{-2} \end{pmatrix},$$

which corresponds to a linear correlation coefficient $\mathrm{Cor}(X_t^1, X_t^2) \approx -0.04$. However,

$$\mathrm{Cov}(X_t^2, X_{t-1}^1) = \mathrm{Cov}(-9.2 X_{t-1}^1 + 0.35 X_{t-1}^2 + Z_t^2, X_{t-1}^1)$$
$$= -9.2 \, \mathrm{Var}(X_t^1) + 0.35 \, \mathrm{Cov}(X_t^1, X_t^2),$$

which gives

$$\mathrm{Cor}(X_t^2, X_{t-1}^1) = -9.2 \left(\frac{\mathrm{Var}(X_t^1)}{\mathrm{Var}(X_t^2)} \right)^{1/2} + 0.35 \, \mathrm{Cor}(X_t^1, X_t^2) \approx -0.44$$

reflecting the fact that the stock market typically reacts negatively to increasing interest rates (the present value of future dividends decreases) and positively to decreasing interest rates. Similarly, $\mathrm{Cor}(X_t^1, X_{t-1}^2) \approx 0.41$, which may reflect the fact that central banks raise interest rates to cool down an overheated economy and lower interest rates to boost a struggling economy. The main point is that the linear correlation coefficient $\mathrm{Cor}(X_t^1, X_t^2) \approx -0.04$ that could be estimated on the pairs of interest rate changes and index log returns says very little about the dependencies between interest rate changes and index log returns. Here we have two rather strong causal dependencies that essentially net out when only considering the dependence among the components of the random vector (X_t^1, X_t^2).
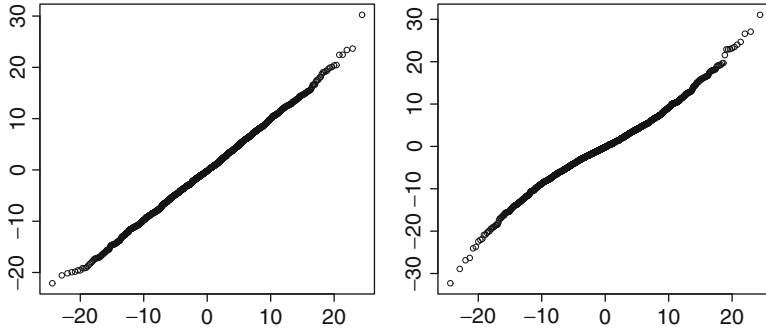
Fig. 9.13 q–q plots of simulated samples of size 2,000 against a normal distribution with zero mean and variance 55. The first distribution (*left*) is the sum of the components of a ten-dimensional standard normal distributed vector with pairwise linear correlation 0.5. The second distribution (*right*) is the sum of the components on a ten-dimensional random vector with standard normal univariate marginal distributions and the dependence structure of a ten-dimensional Student's *t* distribution with one degree of freedom and pairwise linear correlation 0.5

Example 9.19 (Asymptotic dependence). We know from Proposition 9.5 that the components of a bivariate standard normally distributed vector (X_1, X_2) with linear correlation $\rho < 1$ are asymptotically independent in the sense that $\lim_{x \to -\infty} P(X_2 \leq x \mid X_1 \leq x) = 0$. In this case, an extreme value for one component is not likely to make the other component take an extreme value. Combining Proposition 9.6 and Example 8.2 implies that the components of a bivariate standard Student's t_ν-distributed vector (Y_1, Y_2) with linear correlation $\rho \in (0, 1)$ are asymptotically dependent in the sense that $\lim_{x \to -\infty} P(Y_2 \leq x \mid Y_1 \leq x) = \lambda > 0$. In this case, an extreme value for one component makes it likely that the other component will take an extreme value.

Consider the random vector $(U_1, U_2) = (\Phi(X_1), \Phi(X_2))$, whose distribution function is called a Gaussian copula, and the random vector $(V_1, V_2) = (t_\nu(Y_1), t_\nu(Y_2))$, whose distribution function is called a t_ν copula. If G is a distribution function and $p \in (0, 1)$ is small, then the probability that both components of the vector $(Z_1, Z_2) = (G^{-1}(V_1), G^{-1}(V_2))$ take values smaller than $G^{-1}(p)$ is approximately

$$P(Z_1 \leq G^{-1}(p), Z_2 \leq G^{-1}(p)) = P(V_1 \leq p) P(V_2 \leq p \mid V_1 \leq p) \approx \lambda p,$$

whereas the corresponding probability of joint extremes for the vector $(W_1, W_2) = (G^{-1}(U_1), G^{-1}(U_2))$ is of the order p^2. As a consequence, the left tail of $Z_1 + Z_2$ will be heavier than that of $W_1 + W_2$. The influence of the (lack of) asymptotic dependence of the (Gaussian) t_ν copula of a random vector on the tail behavior of the sum of its components is valid for vectors of arbitrary dimension. Figure 9.13 illustrates this effect graphically in terms of q–q plots for 10-dimensional random vectors **Z** and **W**, where $G = \Phi$ is the standard normal distribution function and the

underlying multivariate standard normally distributed $\mathbf{X}$ and Student's t_1-distributed $\mathbf{Y}$ both have pairwise linear correlation parameter $\rho = 0.5$. With R denoting the linear correlation matrix with off-diagonal entries 0.5,

$$Z_1 + \cdots + Z_{10} \overset{\text{d}}{=} (\mathbf{1}^{\mathrm{T}} R \mathbf{1})^{1/2} Z_1$$

is $N(0, 55)$-distributed.

Example 9.20 (Default risk). Consider a portfolio of corporate loans of a retail bank. Suppose there are n loans and let, for $k = 1, \ldots, n$, X_k be an indicator that takes the value 1 if the kth obligor has defaulted on its loan at the end of the year, and 0 otherwise. Suppose also that the default probabilities $p_k = P(X_k = 1)$ can be accurately estimated and may be considered as known. A common estimation approach is to divide the obligors into m homogeneous groups so that all obligors belonging to the same group have the same default probability. The estimates of default probabilities can then be based on the relative frequencies of defaults over the years for the different groups.

The random variable $N = X_1 + \cdots + X_n$, representing the total number of defaults within the current year, is likely to be of interest to the bank. However, the default probabilities only determine the marginal distributions and not the full multivariate distribution of the random vector $(X_1, \ldots, X_n)$. To specify a multivariate model for the default indicators, it is common to consider a vector $(Y_1, \ldots, Y_n)$ of so-called latent variables. The latent variable Y_k may represent the difference between the values of the assets and liabilities of the kth obligor at the end of the year, and a threshold d_k is determined so that $Y_k \leq d_k$ corresponds to default for obligor k. We may now express the probability that the first k among the n loans default as, assuming that the Y_k have continuous distribution functions,

$$\begin{aligned}
p_{1 \ldots k} &= P(Y_1 \leq d_1, \ldots, Y_k \leq d_k) \\
&= C(P(Y_1 \leq d_1), \ldots, P(Y_k \leq d_k), 1, \ldots, 1) \\
&= C(p_1, \ldots, p_k, 1, \ldots, 1),
\end{aligned}$$

where C denotes the copula of $(Y_1, \ldots, Y_n)$. Joint default probabilities of this type will depend heavily on the choice of copula C. To illustrate this point, we consider a numerical example.

Consider a loan portfolio with $n = 1{,}000$ obligors, and suppose that the default probability of each obligor is equal to $p = 0.05$, i.e., $p_k = 0.05$ for each k. We consider four different copula models for the latent variable vector: (a) C is a Gaussian copula with pairwise correlation parameter $\rho = 0$, (b) C is a Gaussian copula with pairwise correlation parameter $\rho = 0.1$, (c) C is a Student's t_3 copula with pairwise correlation parameter $\rho = 0$, and (d) C is a Student's t_3 copula with pairwise correlation parameter $\rho = 0.1$.

For each model we generate a sample of size 10^5 from the resulting model for N, the total number of defaults, and illustrate the distribution of N in terms of the histograms shown in Fig. 9.14. The histograms show clearly that zero correlation for
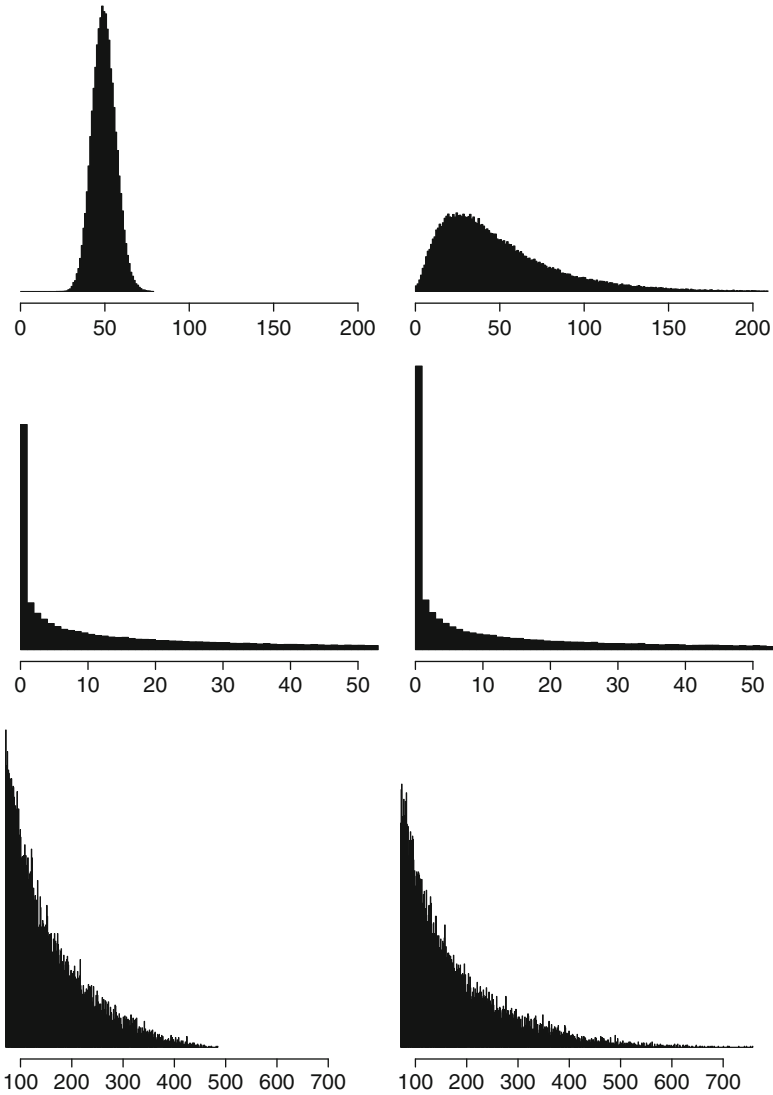
Fig. 9.14 The distribution of the number of defaults is illustrated in histograms based on samples of size 10^5 for the sum of 10^3 default indicators. The *histograms* correspond to the following latent variable models: Gaussian with $\rho = 0$ (*upper left*), Gaussian with $\rho = 0.1$ (*upper right*), Student's t_3 with $\rho = 0$ (*middle and lower left*), Student's t_3 with $\rho = 0.1$ (*middle and lower right*)

the underlying Student's t distribution is far from independence. For the Gaussian copula, zero correlation is equivalent to independence. The histograms also show the impact on the distribution of N of the small change in the correlation parameter ρ from 0 to 0.1.

9.5 Models for Large Portfolios

In this section we investigate models for the aggregated loss $S_n = X_1 + \cdots + X_n$ for a large homogeneous portfolio over a specified time period. Here X_k represent the loss from an investment in the kth asset. As an example we consider the aggregate loss of a bank's portfolio of loans to small and medium size firms due to failure of borrowers to honor their contracted obligations to the lender (the bank). The number of assets, n, is thought of as very large, and we do not have enough information to accurately specify an n-dimensional distribution for $(X_1, \ldots, X_n)$. We will present a cruder approach based on conditional independence.

In many cases, it is not reasonable to assume that the X_k are independent because the losses may depend on the state of the economy. However, it may be reasonable to assume that the X_k are conditionally independent, given the values of a set of economic indicators (e.g., current and future values of interest rates for different maturities, capacity utilization in the industry, GDP growth). Let the components of random vector $\mathbf{Z}$ represent the future values of the economic indicators, and let $f_n(\mathbf{Z}) = \mathrm{E}[S_n/n \mid \mathbf{Z}]$ be the expected average loss conditional on the economic indicators. When n is large, it seems plausible that the diversification effect causes the idiosyncratic risks to be small and the main risk drivers are captured by vector $\mathbf{Z}$. This motivates the approximation $S_n \approx n f_n(\mathbf{Z})$. A mathematical motivation for the approximation $S_n \approx n f_n(\mathbf{Z})$ is given in the following result.

Proposition 9.10. *Let $X_1, \ldots, X_n$ be random variables that are conditionally independent given random vector $\mathbf{Z}$. Write $S_n = X_1 + \cdots + X_n$ and $f_n(\mathbf{Z}) = \mathrm{E}[S_n/n \mid \mathbf{Z}]$. Then*

$$\mathrm{P}(|S_n/n - f_n(\mathbf{Z})| > \varepsilon) \leq \frac{\sum_{k=1}^{n} \mathrm{E}[\mathrm{Var}(X_k \mid \mathbf{Z})]}{(n\varepsilon)^2}, \quad \varepsilon > 0.$$

If, in addition, the X_k are identically distributed, then $f = f_n$ does not depend on n and

$$\mathrm{P}(|S_n/n - f(\mathbf{Z})| > \varepsilon) \leq \frac{\mathrm{E}[X_1^2] - \mathrm{E}[f(\mathbf{Z})^2]}{n\varepsilon^2}, \quad \varepsilon > 0.$$

If, further, the X_k take values in $\{0, 1\}$, then

$$\mathrm{P}(|S_n/n - f_n(\mathbf{Z})| > \varepsilon) \leq \frac{\mathrm{E}[f(\mathbf{Z})] - \mathrm{E}[f(\mathbf{Z})^2]}{n\varepsilon^2}.$$

Proof. An application of Chebyshev's inequality gives

$$\mathrm{P}(|S_n/n - f_n(\mathbf{Z})| > \varepsilon) = \mathrm{E}[\mathrm{P}(|S_n - n f_n(\mathbf{Z})| > n\varepsilon \mid \mathbf{Z})]$$
$$\leq \frac{\mathrm{E}[\mathrm{Var}(S_n \mid \mathbf{Z})]}{n^2 \varepsilon^2}.$$

Because the X_k are conditionally independent given $\mathbf{Z}$, it follows that

$$E[\text{Var}(S_n \mid \mathbf{Z})] = \sum_{k=1}^{n} E[\text{Var}(X_k \mid \mathbf{Z})],$$

which proves the first claim. The second claim follows from the first claim because

$$E[\text{Var}(X_k \mid \mathbf{Z})] = E[\text{Var}(X_1 \mid \mathbf{Z})] = E[E[X_1^2 \mid \mathbf{Z}] - (E[X_1 \mid \mathbf{Z}])^2] = E[X_1^2] - E[f(\mathbf{Z})^2].$$

Moreover, if X_1 takes a value in $\{0, 1\}$, then $E[X_1^2 \mid \mathbf{Z}] = E[X_1 \mid \mathbf{Z}] = f(\mathbf{Z})$. This completes the proof. □

Proposition 9.10 not only motivates the approximation $S_n \approx n f_n(\mathbf{Z})$; it also provides an upper bound for tail probabilities for the aggregated loss S_n. For instance, combining Proposition 9.10 and the inequality

$$P(S_n > s) = P(S_n > s, |S_n - n f_n(\mathbf{Z})| \le \varepsilon n) + P(S_n > s, |S_n - n f_n(\mathbf{Z})| > \varepsilon n)$$
$$\le P(n f_n(\mathbf{Z}) > s - \varepsilon n) + P(|S_n - n f_n(\mathbf{Z})| > \varepsilon n), \quad \varepsilon > 0,$$

gives an upper bound for $P(S_n > s)$. The upper bound for the tail probability gives an upper bound for the quantile. If the X_k are identically distributed and conditionally independent given $\mathbf{Z}$, then, with $C = E[X_1^2] - E[f(\mathbf{Z})^2]$,

$$F_{S_n}^{-1}(q) = \min\{s : F_{S_n}(s) \ge q\}$$
$$= \min\{s : P(S_n > s) \le 1 - q\}$$
$$\le \min\{s : P(n f(\mathbf{Z}) > s - \varepsilon n) + C/(\varepsilon^2 n) \le 1 - q\}$$
$$= n(\varepsilon + F_{f(\mathbf{Z})}^{-1}(q + C/(\varepsilon^2 n))), \quad \varepsilon > 0.$$

In particular,

$$F_{S_n}^{-1}(q) \le n \min_{\varepsilon > 0} \left(\varepsilon + F_{f(\mathbf{Z})}^{-1}(q + C/(\varepsilon^2 n)) \right), \quad C = E[X_1^2] - E[f(\mathbf{Z})^2]. \quad (9.16)$$

The upper bound for quantile (9.16) can be used to derive upper bounds for risk measures such as VaR and ES.

Example 9.21 (A large homogeneous loan portfolio). Consider a large portfolio of loans to small and medium size firms and suppose that we want to analyze the distribution of aggregate losses from now until 1 year from now due to defaults. Write X_k for the loss on the kth loan and $S_n = X_1 + \cdots + X_n$ for the aggregated loss. In this case, X_k can be written as $X_k = L_k I_k$, where I_k is the default indicator that takes the value 1 if the kth obligor defaults and 0 otherwise, and L_k is the amount of money lost if the kth obligor defaults. If the default probabilities $P(I_k = 1)$ are of similar size and the loss given default variables L_k are statistically similar, then the loan portfolio can be considered homogeneous.

A particularly nice situation is where the L_k are identical and deterministic, $L_k = l$, for each $k = 1, \ldots, n$. In this case

$$f_n(\mathbf{Z}) = \mathrm{E}\left[\frac{S_n}{n} \mid \mathbf{Z}\right] = \mathrm{E}\left[\frac{1}{n}\sum_{k=1}^{n} L_k I_k \mid \mathbf{Z}\right] = l\,\mathrm{E}\left[\frac{N_n}{n} \mid \mathbf{Z}\right],$$

where $N_n = I_1 + \cdots + I_n$ is the number of defaults. The fraction of defaults, given the economic indicators, is written as $p_n(\mathbf{Z}) = \mathrm{E}[N_n/n \mid \mathbf{Z}]$. That is, $f_n(\mathbf{Z}) = l p_n(\mathbf{Z})$, and the aggregated loss can be approximated by $S_n \approx n l p_n(\mathbf{Z})$. If, in addition, the default indicators are identically distributed, then $p_n(\mathbf{Z}) = p(\mathbf{Z})$ does not depend on n, and the last statement of Proposition 9.10 leads to

$$\mathrm{P}(|N_n/n - p(\mathbf{Z})| > \varepsilon) \le \frac{\mathrm{E}[p(\mathbf{Z})(1 - p(\mathbf{Z}))]}{n\varepsilon^2}.$$

9.5.1 Beta Mixture Model

In this section, we will illustrate the modeling approach presented in the previous example for a specific choice of model for $N = N_n$ defaults and $p(\mathbf{Z})$ fraction of defaults. Write $N = I_1 + \cdots + I_n$, where the I_k are identically distributed, independent, and Bernoulli distributed with parameter Z conditional on the random variable $Z = f(\mathbf{Z})$, which we take to be Beta(a, b)-distributed. We do not give any economic interpretation of the Beta(a, b)-distributed Z and choose this model only because it is a particularly simple model to work with in terms of both analytical and numerical computations.

The assumption that Z is Beta(a, b)-distributed implies that Z has the density function
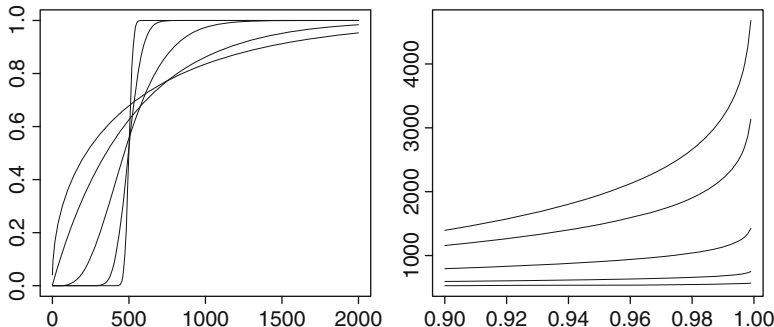
$$g(z) = \frac{1}{\beta(a, b)} z^{a-1}(1 - z)^{b-1}, \quad a, b > 0, z \in (0, 1),$$

where $\beta(a, b)$ can be expressed in terms of the Gamma function as

$$\beta(a, b) = \int_0^1 z^{a-1}(1 - z)^{b-1}\mathrm{d}z = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a + b)}.$$

Using the property $\Gamma(z + 1) = z\Gamma(z)$ of the Gamma function we find that

$$\mathrm{E}[Z] = \frac{1}{\beta(a, b)}\int_0^1 z^a(1 - z)^{b-1}\mathrm{d}z = \frac{\beta(a + 1, b)}{\beta(a, b)} = \frac{a}{a + b},$$

$$\mathrm{E}[Z^2] = \frac{\beta(a + 2, b)}{\beta(a, b)} = \frac{a(a + 1)}{(a + b)(a + b + 1)}.$$

Fig. 9.15 Distribution functions (*left*) and quantile functions (*right*) for beta-binomial distributions with $n = 10^4$, $p = 0.05$, and $(a, b) = ((1-c)/c)(p, 1-p)$ for $c = 0, 0.001, 0.01, 0.05, 0.1$ ($c = 0$ gives the Bin(n, p) distribution)

Conditional on Z, the number of defaults N has a Bin(n, Z) distribution, and therefore the distribution of N is given by

$$P(N = k) = \binom{n}{k} \int_0^1 z^k (1 - z)^{n-k} g(z) dz$$

$$= \binom{n}{k} \frac{1}{\beta(a, b)} \int_0^1 z^{a+k-1} (1 - z)^{n-k+b-1} dz$$

$$= \binom{n}{k} \frac{\beta(a + k, b + n - k)}{\beta(a, b)},$$

which is called the beta-binomial distribution. The distribution function of the beta-binomial distribution is illustrated in Fig. 9.15. The expected number of defaults is easily computed:

$$E[N] = E[E[N \mid Z]] = E[nZ] = n\frac{a}{a + b}.$$

In addition, the individual default probability is $P(I_1 = 1) = E[E[I_1 \mid Z]] = E[Z]$, the pairwise default probability is $P(I_1 = I_2 = 1) = E[Z^2]$, and the default correlation is

$$\text{Cor}(I_1, I_2) = \frac{E[I_1 I_2] - E[I_1]^2}{E[I_1^2] - E[I_1]^2} = \frac{E[Z^2] - E[Z]^2}{E[Z] - E[Z]^2} = \frac{1}{a + b + 1}.$$

To analyze the model, we fix the common individual default probability at $p = P(I_1 = 1)$. This implies that we allow only parameter pairs (a, b) for which $p = a/(a + b)$, i.e., pairs (a, b) satisfying

$$(a, b) = \frac{1 - c}{c}(p, 1 - p), \quad c \in (0, 1),$$

where c are possible values for the default correlation $\text{Cor}(I_1, I_2)$. We can now study the beta-binomial distribution and compare the quantile $F_N^{-1}(q)$ with its estimate $nF_Z^{-1}(q)$. We find that for $q \in [0.9, 0.99]$, the values of $F_N^{-1}(q)/F_{nZ}^{-1}(q)$ are in the intervals

$$
F_N^{-1}(q)/F_{nZ}^{-1}(q) \in \begin{cases} (1.006675, 1.015625) & \text{for } c = 0.001, \\ (1.001286, 1.003138) & \text{for } c = 0.01, \\ (1.000199, 1.000966) & \text{for } c = 0.05, \\ (1.000023, 1.000610) & \text{for } c = 0.1. \end{cases}
$$

In particular, the approximation $N \approx nZ$ is very accurate. We also find that a small change in the common default correlation coefficient between the I_k has a huge effect on the distribution of $N = I_1 + \cdots + I_n$. This is seen in Fig. 9.15, which shows distribution functions and quantile functions for beta-binomial models with $p = 0.05$, $n = 10^4$, and different correlation coefficients. Figure 9.15 illustrates clearly that only specifying the individual default probability p says very little about the distribution of N. Every choice of $(a, b) = ((1 - c)/c)(p, 1 - p)$, $c > 0$, gives default probability p. Let $Z_{c,p}$ be Beta-distributed with the parameters a, b above. Then, for every $\varepsilon > 0$,

$$
P(|Z_{c,p} - p| > \varepsilon) \le \frac{\text{Var}(Z_{c,p})}{\varepsilon^2} = \frac{p(1 - p)c}{\varepsilon^2}.
$$

In particular, if $N_{c,p}$ is beta-binomially distributed with mixture variable $Z_{c,p}$, then

$$
\begin{aligned}
P(N_{c,p} = k) &= \text{E}\left[\binom{n}{k} Z_{c,p}^k (1 - Z_{c,p})^{n-k} \right] \\
&= \text{E}\left[\binom{n}{k} Z_{c,p}^k (1 - Z_{c,p})^{n-k}; |Z_{c,p} - p| \le c^{1/3} \right] \\
&\quad + \text{E}\left[\binom{n}{k} Z_{c,p}^k (1 - Z_{c,p})^{n-k}; |Z_{c,p} - p| > c^{1/3} \right] \\
&\le \max_{|t| \le c^{1/3}} \binom{n}{k} (p + t)^k (1 - (p + t))^{n-k} + p(1 - p)c^{1/3} \\
&\to \binom{n}{k} p^k (1 - p)^{n-k} \quad \text{as } c \to 0.
\end{aligned}
$$

The lower bound is constructed similarly. We conclude that $N_{c,p}$ converges in distribution to $\text{Bin}(n, p)$ as $c \to 0$. This is also seen in Fig. 9.15.

9.6   Notes and Comments

Much more material on elliptical distributions can be found in the book [16] by Kai-Tai Fang, Samuel Kotz, and Kai Wang Ng.

For further material on multivariate elliptical and copula-based models, dependence concepts, and applications in financial risk management we refer the reader to the book [31] by Alexander McNeil, Rüdiger Frey, and Paul Embrechts. Much material on models and methods for portfolio credit risk, which we have only touched upon here, can be found in [31]. Moreover, techniques for parameter estimation for copula models, a topic we have not considered at all, are presented and illustrated in [31].

A statement equivalent to Proposition 9.6 appears in the book Chap. [12] by Paul Embrechts, Alexander McNeil and Daniel Straumann It can be proved by considering the conditional density of one component of a bivariate Student's t-distributed vector given a value of its other component. However, the asymptotic dependence (or tail dependence) property of the Student's t distribution is a consequence of a more general fact that says that pairs of components of an elliptically distributed random vector are asymptotically dependent if the distribution functions of its components are regularly varying. A proof of this more general fact, which also applies to Proposition 9.6, can be found in the article [22] by Henrik Hult and Filip Lindskog. The statement in Proposition 9.7 appears in the book Chap. [25] by Filip Lindskog, Alexander McNeil, and Uwe Schmock. and in the article [15] by Hong-Bin Fang, Kai-Tai Fang, and Samuel Kotz.

The reader seeking more information about copulas in general is encouraged to consult the books [23] by Harry Joe and [35] by Roger Nelson.

9.7   Exercises

In the exercises below, it is assumed, whenever applicable, that you can take positions corresponding to fractions of assets.

Exercise 9.1 (Risk minimization). Consider the value L of a liability and values $X_1, \ldots, X_d$ of assets at time $T > 0$ that may be used to hedge the liability. Suppose that L and the X_k have finite variances, and let ρ be a translation-invariant and positively homogeneous risk measure.

(a) Show that if $(X_1, \ldots, X_d, L)$ has an elliptical distribution, then the portfolio weights $h_0, h_1, \ldots, h_d$ minimizing

$$\mathrm{E}[(h_0 + h_1 X_1 + \cdots + h_d X_d - L)^2],$$

i.e., the optimal quadratic hedge, minimize $\rho(h_0 + h_1 X_1 + \cdots + h_d X_d - L)$.
(b) Show, by an explicit example, that the conclusion in (a) does not hold in general when $(X_1, \ldots, X_d, L)$ does not have an elliptical distribution.

Exercise 9.2 (Allocation invariance). Let $\mathbf{X} = (X_1, \ldots, X_d)^\mathrm{T}$ and $\mathbf{Y} = (Y_1, \ldots, Y_d)^\mathrm{T}$ be random vectors having normal variance mixture distributions with identical dispersion matrices and identical location vectors $R_0\mathbf{1}$, where R_0 is the return on a risk-free asset. Vectors $\mathbf{X}$ and $\mathbf{Y}$ represent returns on $2d$ risky assets. Let $V_\mathbf{X}(\mathbf{w})$ and $V_\mathbf{Y}(\mathbf{w})$ denote the values at the end of the investment horizon for an investment of the capital V_0 in positions in the risk-free asset and in the assets with return vectors $\mathbf{X}$ and $\mathbf{Y}$, respectively, where $\mathbf{w}$ is a vector of monetary portfolio weights corresponding to the positions in the risky assets.

(a) Show that if ρ is a translation-invariant and positively homogeneous risk measure, then
$$\frac{\rho(V_\mathbf{X}(\mathbf{w}) - V_0 R_0)}{\rho(V_\mathbf{Y}(\mathbf{w}) - V_0 R_0)} \tag{9.17}$$
does not depend on the allocation of the initial capital or on the common dispersion matrix of the return vectors.
(b) Suppose that $\mathbf{X}$ has a Student's t distribution with four degrees of freedom, that $\mathbf{Y}$ has a normal distribution, and that $\rho = \mathrm{VaR}_p$, and compute the expression in (9.17) as a function of p for $p \leq 0.05$.

Exercise 9.3 (Asymptotic dependence). Consider a random vector (X_1, X_2) whose components are equally distributed and use Propositions 9.5 and 9.6 to compute $\lim_{x\to\infty} \mathrm{P}(X_2 > x \mid X_1 > x)$ in the following two cases:

(a) X_1 and X_2 are Student's t-distributed with four degrees of freedom, and (X_1, X_2) has a Gaussian copula with linear correlation parameter 0.5.
(b) X_1 and X_2 are Student's t-distributed with four degrees of freedom, and (X_1, X_2) has a Student's t copula with linear correlation parameter 0.5 and degrees of freedom parameter 6.

Exercise 9.4 (Comonotonic additive risk). Show that if X_1 and X_2 are comonotone random variables, then $\mathrm{VaR}_p(X_1 + X_2) = \mathrm{VaR}_p(X_1) + \mathrm{VaR}_p(X_2)$ and $\rho_\phi(X_1 + X_2) = \rho_\phi(X_1) + \rho_\phi(X_2)$ for any spectral risk measure ρ_ϕ defined in (6.18).

Exercise 9.5 (Kendall's tau). Let Ψ be the Laplace transform of a strictly positive random variable, and consider the random pair (U_1, U_2) whose distribution function is the copula $C(u_1, u_2) = \Psi(\Psi^{-1}(u_1) + \Psi^{-1}(u_2))$.

(a) Show that $\tau(U_1, U_2) = 4\,\mathrm{E}[C(U_1, U_2)] - 1$.
(b) It can be shown that $\mathrm{P}(C(U_1, U_2) \leq v) = v - \Psi^{-1}(v)/(\Psi^{-1})'(v)$ for v in $(0, 1)$. Use this relation to show that

$$\tau(U_1, U_2) = 1 + 4 \int_0^1 \frac{\Psi^{-1}(v)}{(\Psi^{-1})'(v)} dv.$$

(c) Compute $\tau(U_1, U_2)$ when $C = C_\theta^{\mathrm{Cl}}$ is a Clayton copula.

Exercise 9.6 (Credit rating migration). Consider the two corporate bonds in Exercise 4.6. Let the credit ratings be numbered from 1 to 4 and correspond to the ratings Excellent, Good, Poor, and Default in Exercise 4.6. Let (X_1, X_2) denote the pair of credit ratings of the two issuers after 1 year with the distribution given in Table 4.1.

(a) Find a copula C such that $P(X_1 \le x_1, X_2 \le x_2) = C(P(X_1 \le x_1), P(X_2 \le x_2))$ for all (x_1, x_2).
(b) The copula C of (X_1, X_2) in (a) can be well approximated by a Gaussian copula. Investigate numerically what value of the correlation parameter in the Gaussian copula gives a good approximation of the copula of (X_1, X_2) in (a).

Exercise 9.7 (Portfolio default risk). Consider a latent variable model for a homogeneous portfolio of n risky loans. Let p be the default probability for each loan, let $Y, Y_1, \ldots, Y_n$ be independent and standard normally distributed, and let $\rho \in (0, 1)$ be a parameter. The default indicators are modeled as

$$X_k = \begin{cases} 1 \text{ if } \sqrt{\rho}Y + \sqrt{1-\rho}Y_k \le \Phi^{-1}(p), \\ 0 \text{ otherwise,} \end{cases} \tag{9.18}$$

where Φ denotes the standard normal distribution function.

(a) Determine the random variable $\Theta = g(Y)$ such that the default indicators are conditionally independent and $Be(\theta)$-distributed given $\Theta = \theta$.
(b) Show that the following formula holds for the q-quantile of Θ:

$$F_\Theta^{-1}(q) = \Phi\left(\Phi^{-1}(q)\frac{\sqrt{\rho}}{\sqrt{1-\rho}} + \Phi^{-1}(p)\frac{1}{\sqrt{1-\rho}}\right).$$

(c) Consider a loan portfolio of a bank consisting of one thousand loans, each of size one million dollars. Suppose that, for each of the loans, the probability of default within 1 year is 3%, and in case of default the bank makes a loss equal to 25% of the size of the loan. Suppose further that the bank makes a profit of $10,000 per year from interest payments on each loan that does not default and nothing on those that do. The bank decides to set aside an amount of buffer capital that equals its estimate of $ES_{0.01}(S)$, where S is the profit from interest income minus the loss from defaults over a 1-year period. Estimate the size of the buffer capital under the assumption that the default indicators are given by (9.18) with $\rho = 0.2$ and that the bank may invest in a risk-free, 1-year, zero-coupon bond with a zero rate of 3%.

Exercise 9.8 (Potential death spiral). Consider a life insurance company with a liability cash flow with long duration. The value of the liability 1 year from now is denoted by L and increases in value when interest rates decline. The premium received for insuring the liability is $V_0 = 1.1 E[L]$. The insurer invests its capital in a fixed-income portfolio with 1-year return R_1 and in a stock market portfolio with 1-year return R_2. The vector (R_1, R_2, L) is, for simplicity, assumed to have a multivariate Student's t distribution with four degrees of freedom. Its mean vector and correlation matrix are given by

$$E\begin{pmatrix} R_1 \\ R_2 \\ L \end{pmatrix} = \begin{pmatrix} 1.02 \\ 1.10 \\ 1.2 \cdot 10^7 \end{pmatrix} \quad \text{and} \quad \text{Cor}\begin{pmatrix} R_1 \\ R_2 \\ L \end{pmatrix} = \begin{pmatrix} 1 & 0.3 & 0.9 \\ 0.3 & 1 & 0.2 \\ 0.9 & 0.2 & 1 \end{pmatrix}.$$

The standard deviations of R_1, R_2, and L are given by 0.005, 0.05, and $1.2 \cdot 10^5$, respectively.

Let w_1, w_2 be the amount invested in the fixed-income portfolio and the stock market portfolio, respectively. The insurer invests the initial capital V_0 in the two portfolios so that the expected value of its asset portfolio has an expected return of 1.06.

(a) Determine w_1 and w_2.
(b) Is the insurer solvent in the sense that $\text{VaR}_{0.005}(A - L) \leq 0$?
(c) Suppose there is an instantaneous decline of 15% in the value of the stock market portfolio. Does the insurer remain solvent? If not, determine how the insurer must adjust the asset portfolio weights w_1 and w_2 simply to become solvent in the sense that $\text{VaR}_{0.005}(A - L) = 0$.
(d) Compute the expected return of the insurer's adjusted asset portfolio determined in (c).

Comment: A simultaneous decline in the value of stocks and in interest rates is particularly dangerous to an insurer with a liability having a long duration. The reduction in the value of the insurer's capital forces the insurer to adjust its asset allocation away from stocks to less risky fixed-income instruments to remain solvent. The adjusted allocation has a lower expected return, which makes it difficult for the insurer to make up for the suffered losses. Moreover, insurance companies often have large amounts of capital invested in the stock market, and a forced sale of large positions in stocks and an increase in the demand for safe bonds could reduce both the prices of stocks and the interest rates even more. This phenomenon, sometimes referred to as a death spiral, makes the insurer stuck in a near-insolvent state with an asset portfolio that is unlikely to generate good returns.

Project 10 (Scenario-based risk analysis). Consider a stylized model of a life insurer. The insurer faces a liability cash flow of 100 each year for the next 30 years. The current zero rates are given in Table 9.1, from which the current value of the liability can be computed. In the market there is a short supply of bonds with maturities longer than 10 years. Therefore, the insurer has purchased a bond portfolio with payments only within the next 10 years. The bond portfolio has the cash flow given in Table 9.1. The insurer has also invested in a stock portfolio. The initial capital of the insurer is 30% more than the current value of the liability. The insurer invests 70% of the initial capital in the bond portfolio and 30% of the initial capital in the stock portfolio. The objective in this project is to identify the most dangerous extreme scenario.

Suppose that there are two risk factors in the model, the log return Y_1 of the stock portfolio and the size Y_2 of a parallel shift of the zero-rate curve. The risk factors are assumed to have a bivariate normal distribution, means μ_1, μ_2, standard deviations

Table 9.1 Annual cash flow of bond portfolio and current zero rates

Time	1	2	3	4	5	6	7	8	9	10
Bond payment	4	2	3	1	4	2	3	1	5	5
Zero rate (%)	2.86	3.24	3.55	3.93	4.27	4.62	4.96	5.30	5.55	5.80
Time	11	12	13	14	15	16	17	18	19	20
Bond payment	0	0	0	0	0	0	0	0	0	0
Zero rate (%)	6.05	6.30	6.45	6.60	6.74	6.90	7.00	7.21	7.32	7.32
Time	21	22	23	24	25	26	27	28	29	30
Bond payment	0	0	0	0	0	0	0	0	0	0
Zero rate (%)	7.40	7.48	7.56	7.64	7.70	7.77	7.83	7.90	7.95	8.00

σ_1, σ_2, and linear correlation coefficient ρ given by

$$\mu_1 = 0.08, \quad \mu_2 = 0, \quad \sigma_1 = 0.2, \quad \sigma_2 = 0.01, \quad \rho = 0.1.$$

Consider equally likely extreme scenarios in the following sense. The risk factors can be represented via two independent standard normally distributed random variables Z_1 and Z_2 as

$$Y_1 = \mu_1 + \sigma_1 Z_1,$$

$$Y_2 = \sigma_2 \left(\rho Z_1 + \sqrt{1 - \rho^2} Z_2 \right).$$

All scenarios with $\sqrt{Z_1^2 + Z_2^2} = 3$ can be viewed as equally likely extreme scenarios corresponding to three-standard-deviation movements. The extreme scenarios for Z_1, Z_2 translate into extreme scenarios for the risk factors Y_1, Y_2 by the relation above.

(a) Plot the value of the insurer's portfolio, assets minus liabilities, in 1 year for all the equally likely extreme scenarios.
(b) Identify which scenario for Y_1, Y_2 leads to the worst outcome for the value of the insurer's assets minus that of the liabilities in 1 year.
(c) Repeat the analysis outlined above and find the most dangerous scenario when (Y_1, Y_2) has another bivariate elliptical distribution.

Project 11 (Tail dependence in large portfolios). Let $Z_1, \ldots, Z_{50}$ represent log returns from today until tomorrow for 50 hypothetical financial assets. Suppose that Z_k has a Student's t distribution with three degrees of freedom and standard deviation 0.01 for each k and that $\tau(Z_j, Z_k) = 0.4$ for $j \neq k$.

Consider an investment of $20,000 in long positions in each of the assets. Let V_0 and V_1 be the portfolio value today and tomorrow, respectively. Investigate the effect of tail dependence on the distribution of the portfolio value V_1 tomorrow and the distribution of the portfolio log return $\log(V_1/V_0)$ by simulating from the distribution of V_1. Simulate from the distribution of V_1 under the assumption that

(a) $(Z_1, \ldots, Z_{50})$ has a Gaussian copula.
(b) $(Z_1, \ldots, Z_{50})$ has a t_4-copula.
(c) $(Z_1, \ldots, Z_{50})$ has a Clayton copula.
(d) How large a sample size is needed to get stable estimates of $\text{VaR}_{0.01}(V_1 - V_0)$ and $\text{ES}_{0.01}(V_1 - V_0)$? Explain the differences in the estimates of $\text{VaR}_{0.01}(V_1 - V_0)$ and $\text{ES}_{0.01}(V_1 - V_0)$ in the three cases (a)–(c).
(e) Compare the results in (a)–(d) to the results when $1 million is invested in only one of the assets.
(f) Suppose that the Z_k are equally distributed and have a left-skewed polynomial normal distribution with zero mean and standard deviation 0.01. Study and explain the effect of the log-return distribution of the Z_k on the distribution of V_1 and the portfolio risk by simulating from the distribution of V_1 under assumptions (a)–(c).