In this section we describe two different approaches to approximating the loss operator for a bond portfolio, and we show how statistical analysis techniques from the area of factor modelling are used to calibrate the approximating functions.

The approach based on the Nelson and Siegel (1987) model. The Nelson–Siegel model is usually formulated in terms of instantaneous *forward interest rates*, which are defined from prices by

$$f(t, T) = -\frac{\partial}{\partial T} \ln p(t, T).$$

These can be interpreted as representing the rates that are offered at time t for borrowing at future times T. Yields are related to forward rates by

$$y(t, T) = \frac{1}{T-t} \int_t^T f(t, u)\, \mathrm{d}u. \tag{9.18}$$

In the Nelson–Siegel approach the forward curve is modelled by

$$f(\tau_t, T) = Z_{t,1} + Z_{t,2} \exp(-\eta_t(T-\tau_t)) + Z_{t,3}\eta_t(T-\tau_t)\exp(-\eta_t(T-\tau_t)),$$

where the factors are $(Z_{t,1}, Z_{t,2}, Z_{t,3})$ and η_t is a positive rate parameter, which is chosen to give the best fit to forward rate data. The relationship (9.18) between the forward and yield curves implies that

$$y(\tau_t, T) = Z_{t,1} + k_1(T-\tau_t, \eta_t)Z_{t,2} + k_2(T-\tau_t, \eta_t)Z_{t,3}, \tag{9.19}$$

where the functions k_1 and k_2 are given by

$$k_1(s, \eta) = \frac{1-\mathrm{e}^{-\eta s}}{\eta s}, \qquad k_2(s, \eta) = k_1(s, \eta) - \mathrm{e}^{-\eta s}.$$

These functions are illustrated in Figure 9.1. We now give an economic interpretation of the factors.

Clearly, $\lim_{s\to\infty} k_1(s, \eta) = \lim_{s\to\infty} k_2(s, \eta) = 0$, while $\lim_{s\to 0} k_1(s, \eta) = 1$ and $\lim_{s\to 0} k_2(s, \eta) = 0$. It follows that $\lim_{T\to\infty} y(\tau_t, T) = Z_{t,1}$, so that the first factor is usually interpreted as a long-term level factor. $Z_{t,2}$ is interpreted as a slope factor because the difference between short-term yield and long-term yield satisfies $\lim_{T\to\tau_t} y(\tau_t, T) - \lim_{T\to\infty} y(\tau_t, T) = Z_{t,2}$; $Z_{t,3}$ has an interpretation as a curvature factor.

Using the factor model (9.19), the mapping (9.11) for the bond portfolio becomes

$$V_t = g(\tau_t, \boldsymbol{Z}_t) = \sum_{i=1}^{d} \lambda_i \exp(-(T_i - \tau_t)\boldsymbol{k}_{t,i}'\boldsymbol{Z}_t).$$

where $\boldsymbol{k}_{t,i} = (1, k_1(T_i-\tau_t, \eta_t), k_2(T_i-\tau_t, \eta_t))'$. It is then straightforward to derive the loss operator $l_{[t]}(\boldsymbol{x})$ or its linear version $l^{\Delta}_{[t]}(\boldsymbol{x})$, which, in contrast to (9.13), are functions on $\mathbb{R}^3$ rather than $\mathbb{R}^d$ (d is the number of bonds in the portfolio).

To use this method to evaluate the linear loss operator at time t, in practice we require realized values $\boldsymbol{z}_t$ for the risk factors $\boldsymbol{Z}_t$. However, we have to overcome the

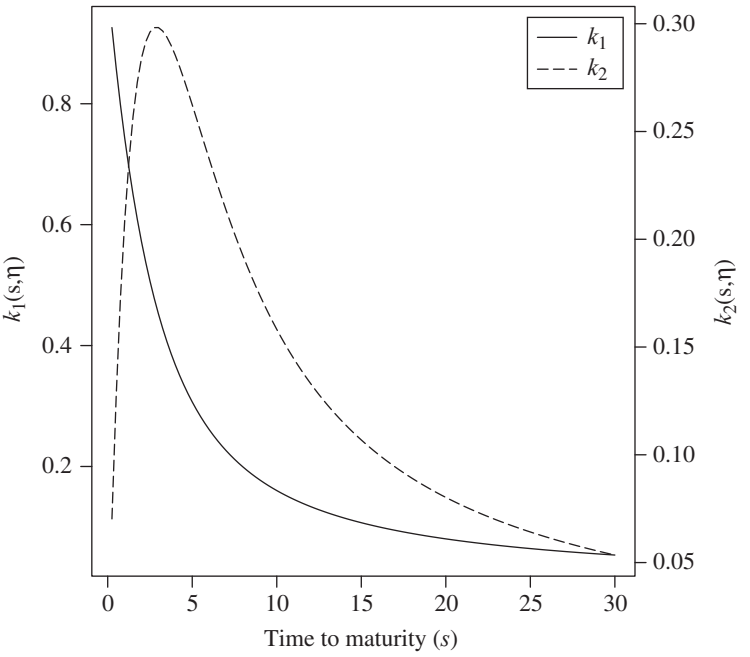

Figure 9.1. The Nelson–Siegel functions $k_1(s, \eta)$ and $k_2(s, \eta)$ for an η value of 0.623 (see also Example 9.2).

fact that the Nelson–Siegel factors $\boldsymbol{Z}_t$ are not directly observed at time t. Instead they have to be estimated from observable yield curve data.

Let us suppose that at time t we have the data vector

$$\boldsymbol{Y}_t = (y(\tau_t, \tau_t + s_1), \ldots, y(\tau_t, \tau_t + s_m))'$$

giving the yields for m different times to maturity, $s_1, \ldots, s_m$, where m is large. This is assumed to follow the factor model $\boldsymbol{Y}_t = B_t \boldsymbol{Z}_t + \boldsymbol{\varepsilon}_t$, where $B_t \in \mathbb{R}^{m\times 3}$ is the matrix with ith row consisting of $(1, k_1(s_i, \eta_t), k_2(s_i, \eta_t))$ and $\boldsymbol{\varepsilon}_t \in \mathbb{R}^m$ is an error vector. This model fits into the framework of the general factor model in (6.50).

For a given value of η_t the estimation of $\boldsymbol{Z}_t$ can be carried out as a cross-sectional regression using weighted least squares. To estimate η_t, a more complicated optimization is carried out (see Notes and Comments). We now show how the method works for real market yield data.

Example 9.2 (Nelson–Siegel factor model of yield curve). The data are daily Canadian zero-coupon bond yields for 120 different quarterly maturities ranging from 0.25 years to 30 years. They have been generated using pricing data for Government of Canada bonds and treasury bills. We model the yield curve on 8 August 2011. The estimated values are $z_{t,1} = 3.82$, $z_{t,2} = -2.75$, $z_{t,3} = -5.22$ and $\hat{\eta}_t = 0.623$. The curves $k_1(s, \eta)$ and $k_2(s, \eta)$ are therefore as shown in Figure 9.1.

Example 9.2 illustrates the estimation of the Nelson–Siegel factor model at a single time point t. We note that, to make statistical inferences about bond-portfolio

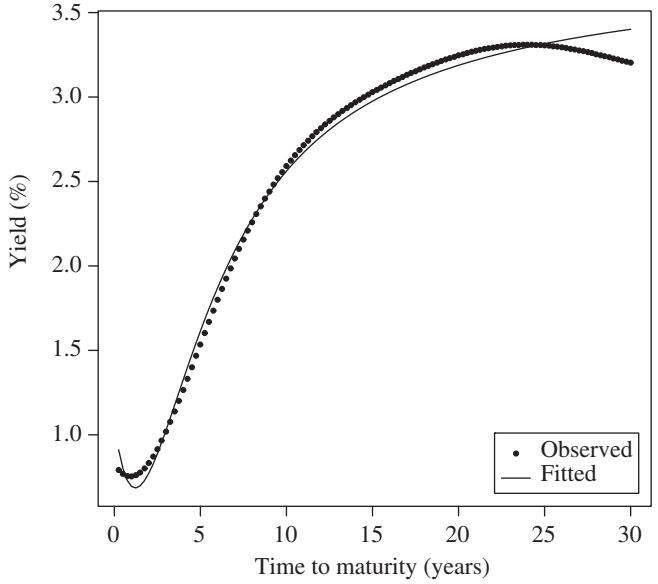

Figure 9.2. Canadian yield curve data (points) and fitted Nelson–Siegel curve for 8 August 2011 (see Example 9.2 for further details).

losses based on historical data, we would typically construct historical risk-factor time series $(\boldsymbol{Z}_u)$ from yield data $(\boldsymbol{Y}_u)$ by estimating cross-sectional regression models at all times u in some set $\{t-n+1, \ldots, t\}$. The estimated risk-factor time series $(\boldsymbol{Z}_u)$ and the corresponding risk-factor changes form the data for the statistical methods that are the subject of Section 9.2.

The approach based on principal component analysis. Another approach to building a factor model of the term structure involves the use of principal component analysis (PCA). We refer to Section 6.4 for an introduction to this method. The key difference to the Nelson–Siegel approach is that here the dimension reduction via factor modelling is applied at the level of the changes in yields rather than the yields themselves.

Let $\boldsymbol{R}_{t+1}$ denote the vector of yield changes for the bonds in the portfolio, so that $R_{t+1,i} = y(\tau_{t+1}, T_i) - y(\tau_t, T_i)$, $1 \leqslant i \leqslant d$. We recall from (6.62) that PCA can be used to construct approximate factor models of the form

$$\boldsymbol{R}_{t+1} = \boldsymbol{\mu} + \Gamma_1 \boldsymbol{X}_{t+1} + \boldsymbol{\varepsilon}_{t+1}, \tag{9.20}$$

where $\boldsymbol{X}_{t+1}$ is a p-dimensional vector of principal components ($p \ll d$), $\Gamma_1 \in \mathbb{R}^{d\times p}$ contains the corresponding loading matrix, $\boldsymbol{\mu}$ is the mean vector of $\boldsymbol{R}_{t+1}$ and $\boldsymbol{\varepsilon}_{t+1}$ is an error vector. The columns of the matrix Γ_1 consist of the first p eigenvectors (ordered by decreasing eigenvalue) of the covariance matrix of $\boldsymbol{R}_{t+1}$. The principal components $\boldsymbol{X}_{t+1}$ will form the risk-factor changes in our portfolio analysis, hence our choice of notation.

Typically, the error term is neglected and $\boldsymbol{\mu} \approx \boldsymbol{0}$, so that we make the approximation $\boldsymbol{R}_{t+1} \approx \Gamma_1 \boldsymbol{X}_{t+1}$. In the case of the linear loss operator for the bond portfolio

in (9.13), we use the approximation

$$l^{\Delta}_{[t]}(\boldsymbol{x}) = -\sum_{i=1}^{d} \lambda_i \, p(\tau_t, T_i)(y(\tau_t, T_i)\Delta t - (T_i - \tau_t)(\Gamma_1 \boldsymbol{x})_i), \tag{9.21}$$

so that a function of a p-dimensional argument $\boldsymbol{x}$ is substituted for a function of a d-dimensional argument.

To work with this function we require an estimate for the matrix Γ_1. This can be obtained from historical time-series data on yield changes by estimating sample principal components, as explained in Example 9.3.

Example 9.3 (PCA factor model of yield changes). We again analyse Canadian bond yield data as in Example 9.2. To estimate the Γ_1 matrix of principal component loadings we require longitudinal (time-series) data rather than the cross-sectional data that were used in the previous example.

We will assume for simplicity that the times to maturity $T_1 - \tau_t, \ldots, T_d - \tau_t$ of the bonds in the portfolio correspond exactly to the times to maturity $s_1, \ldots, s_d$ available in the historical data set (if not we would make an appropriate selection of the data) and that the risk-management horizon Δt is one day.

In the Canadian data set we have 2488 days of data spanning the period from 2 January 2002 to 30 December 2011 (ten full trading years); recall that each day gives rise to a data vector $\boldsymbol{Y}_u = (y(\tau_u, \tau_u + s_1), \ldots, y(\tau_u, \tau_u + s_d))'$ of yields for the different maturities. In line with (9.20) we analyse the daily returns (first differences) of these data $\boldsymbol{R}_u = \boldsymbol{Y}_u - \boldsymbol{Y}_{u-1}$ using PCA under the assumption that they form a stationary time series. (Note that a small error is incurred by analysing daily yield changes for yields with fixed times to maturity rather than fixed maturity date, but this will be neglected for the purposes of illustration.)

When we compute the variances of the sample principal components (using the same technique as for Figure 6.5), we find that the first component explains 87.0% of the variance of the data, the first two components explain 95.9%, and the first three components explain 97.5%. We choose to work with the first three principal components, meaning that we set $p = 3$. The matrix Γ_1 is estimated by G_1, a matrix whose columns are the first three eigenvectors of the sample covariance matrix. We recall from (6.63) that the complete eigenvector matrix for the sample covariance matrix is denoted by G.

The first three eigenvectors are shown graphically in Figure 9.3 and lend themselves to a standard interpretation. The first principal component has negative loadings for all maturities; the second has negative loadings up to ten years and positive loadings thereafter; the third has positive loadings for very short maturities (less than 2.5 years) and very long maturities (greater than 15 years) but negative loadings otherwise. This suggests that the first principal component can be thought of as inducing a change in the level of all yields, the second induces a change of slope and the third induces a change in the curvature of the yield curve.

We note that to make statistical inferences about bond-portfolio losses based on historical data we require a historical time series of risk-factor changes $(\boldsymbol{X}_u)$ for times

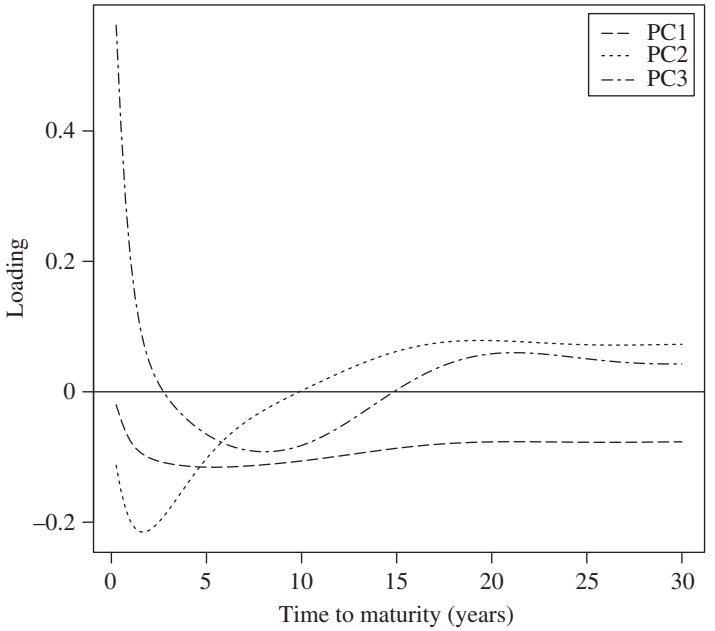

Figure 9.3. First three principal component loading vectors plotted against time to maturity. The data are daily changes in yield for Canadian zero-coupon bonds in the ten-year period 2002–11. The horizontal line at zero shows when loadings are positive or negative. See Example 9.3 for further details.

u in some set $\{t-n+1,\dots,t\}$. These data are not directly observed but are instead extracted from the time series of sample principal components $(G^{-1}(\boldsymbol{R}_u - \bar{\boldsymbol{R}}))$, where $\bar{\boldsymbol{R}}$ is the sample mean vector. The risk-factor change data $(\boldsymbol{X}_u)$ are taken to be the first p component series; these form the data for the statistical methods that are the subject of Section 9.2.

Notes and Comments

The mapping framework introduced in this section is similar to the approach pioneered by the RiskMetrics Group (see the RiskMetrics Technical Document (JPMorgan 1996) and Mina and Xiao (2001)). The mapping of positions is also discussed in Dowd (1998), Jorion (2007) and in Volume III of the Market Risk Analysis series: Alexander (2009). The latter series of four volumes is relevant to much of the material of this chapter.

The use of first- and second-order approximations to the portfolio value (the so-called delta–gamma approximation) may be found in Duffie and Pan (1997) and Rouvinez (1997) (see also Duffie and Pan 2001). Formulas for the Greeks in the Black–Scholes model may be found in a number of textbooks, including Haug (1998) and Wilmot (2000). Leoni (2014) is a very readable introduction.

Many standard finance textbooks treat interest-rate risk and bonds. For a detailed discussion of duration and its use in the management of interest-rate risk, the books

by Jarrow and Turnbull (1999) and Hull (1997) are good starting points. The construction of fixed-income portfolios with higher convexity (so-called barbell portfolios) is discussed in Tuckman and Serrat (2011).

More advanced mathematical textbooks on interest-rate modelling include Brigo and Mercurio (2006), Carmona and Tehranchi (2006) and Filipović (2009)

There are many different approaches to modelling the yield curve. We have concentrated on the parametric factor model proposed by Nelson and Siegel (1987) and further developed in Siegel and Nelson (1988) and Svensson (1994). One theoretical deficiency of the Nelson–Siegel model is that it is not consistent with no-arbitrage pricing theory (Filipović 1999). In more recent work, Christensen, Diebold and Rudebusch (2011) have proposed a model that approximates the Nelson–Siegel model but also fits into the class of three-factor, arbitrage-free affine models.

Useful information about estimating Nelson–Siegel models in practice can be found in Ferstl and Hayden (2011); we have used their R package `termstrc` to carry out the analysis in Example 9.2. Diebold and Li (2006) have developed an approach to forecasting the yield curve in which they fit cross-sectional Nelson–Siegel factor models to multivariate time series of yields for different times to maturity and then use vector autoregression (VAR) models to forecast the Nelson–Siegel factors and hence the entire yield curve. They report satisfactory results when the value of η is held constant over all cross-sectional regressions. The alternative factor-model approach based on PCA is discussed in Hull (1997) and Alexander (2001) (see examples in Section 6.2 of the latter in particular).

RiskMetrics takes a different approach to the problem of the dimension of bond portfolios. A few benchmark yields are taken for each country and a procedure is used to approximately map cash flows at days between benchmark points to the two nearest benchmark points; we refer to Section 6.2 of the RiskMetrics technical document (JPMorgan 1996) for details.

9.2 Market Risk Measurement

In this section we discuss methods used in the financial industry to estimate the loss distribution and associated risk measures for portfolios subject to market risk. In the formal framework of Section 9.1 this amounts to the problem of estimating the distribution of $L_{t+1} = l_{[t]}(\boldsymbol{X}_{t+1})$, or a linear or quadratic approximation thereof, where $\boldsymbol{X}_{t+1}$ is the vector of risk-factor changes from time t to time $t+1$ and $l_{[t]}$ is the known loss operator function at time t.

The problem comprises two tasks: on the one hand we have the statistical problem of estimating the distribution of $\boldsymbol{X}_{t+1}$; on the other hand we have the computational or numerical problem of evaluating the distribution of $L_{t+1} = l_{[t]}(\boldsymbol{X}_{t+1})$. To accomplish the first task, we first have to consider carefully the nature of the distribution we wish to estimate, in particular, whether we focus on the conditional or unconditional distribution of risk-factor changes.

9.2.1 Conditional and Unconditional Loss Distributions

Generally, in market-risk measurement, it is natural to compute conditional measures of risk based on the most recent available information about financial markets. In this case, the task is to estimate $F_{\boldsymbol{X}_{t+1}|\mathcal{F}_t}$, the conditional distribution of risk-factor changes, given $\mathcal{F}_t$, the sigma algebra representing the available information at time t. In most cases, $\mathcal{F}_t$ is given by $\mathcal{F}_t = \sigma(\{\boldsymbol{X}_s : s \leqslant t\})$, the sigma algebra generated by past and present risk-factor changes up to and including time t. The *conditional loss distribution* is the distribution of the loss operator $l_{[t]}(\cdot)$ under $F_{\boldsymbol{X}_{t+1}|\mathcal{F}_t}$, that is, the distribution with df $F_{L_{t+1}|\mathcal{F}_t}(l) = P(l_{[t]}(\boldsymbol{X}_{t+1}) \leqslant l \mid \mathcal{F}_t)$.

While the conditional approach is very natural in market-risk measurement, it may not always yield a prudent assessment of risk. If we are in the middle of a quiet period on financial markets, we may underestimate the possibility of extreme market losses, which can result in an overoptimistic view of a firm's capital adequacy. For this reason it can be informative to compute *unconditional* loss distributions based on assumptions of stationary behaviour over longer time windows that (ideally) contain previous episodes of market volatility.

In the unconditional approach we make the assumption that the process of risk-factor changes $(\boldsymbol{X}_s)_{s\leqslant t}$ forms a stationary multivariate time series. We recall the definition of a stationary univariate time series from Section 4.1.1; the multivariate definition is given in Section 14.1. We estimate the stationary distribution $F_{\boldsymbol{X}}$ of the time series and then evaluate the unconditional loss distribution of $l_{[t]}(\boldsymbol{X})$, where $\boldsymbol{X}$ represents a generic random vector in $\mathbb{R}^d$ with df $F_{\boldsymbol{X}}$. The unconditional loss distribution is thus the distribution of the loss operator $l_{[t]}(\cdot)$ under $F_{\boldsymbol{X}}$.

If the risk-factor changes form an independent and identically distributed (iid) series, we obviously have $F_{\boldsymbol{X}_{t+1}|\mathcal{F}_t} = F_{\boldsymbol{X}}$, so that the conditional and unconditional approaches coincide. However, in Section 3.1.1 we have argued that many types of risk-factor-change data show volatility clustering, which is inconsistent with iid behaviour. In the stationary models that are used to account for such behaviour, $F_{\boldsymbol{X}_{t+1}|\mathcal{F}_t}$ is not generally equal to the stationary distribution $F_{\boldsymbol{X}}$. An important example is provided by the popular models from the GARCH family. In Section 4.2.1 we observed that a simple stationary ARCH(1) model with a conditional normal distribution has a leptokurtic stationary distribution, i.e. a non-normal distribution with heavier tails. This is also true of more complicated univariate and multivariate GARCH models.

Since the financial crisis of 2007–9, regulators have called for regular estimates of VaR to be supplemented by *stressed VaR* estimates (Basel Committee on Banking Supervision 2013a). Firms are required to estimate VaR using historical data from stress periods in the financial markets (such as 2008). Stressed VaR calculations have more to do with the choice of historical data than with the choice of conditional or unconditional distribution. In principle, a stressed VaR estimate can be computed using either approach. The key point is that historical time-series data from stress periods are substituted for the up-to-date time-series data from which regular VaR estimates are made.

9.2.2 Variance–Covariance Method

This method was originally pioneered by JPMorgan's RiskMetrics group in the early 1990s. Although used by a minority of banks today, it remains an important contribution to the development of the methodology for market-risk measurement. As mentioned in Section 2.2.3, it is an example of an analytical method in which the linearized loss distribution has a known form and estimates of VaR and expected shortfall can be computed with simple formulas.

In the variance–covariance method we assume that the conditional distribution of risk-factor changes $F_{\boldsymbol{X}_{t+1}|\mathcal{F}_t}$ is a multivariate normal distribution with mean vector $\boldsymbol{\mu}_{t+1}$ and covariance matrix Σ_{t+1}. In other words, we assume that, given $\mathcal{F}_t$, $\boldsymbol{X}_{t+1} \sim N_d(\boldsymbol{\mu}_{t+1}, \Sigma_{t+1})$, where $\boldsymbol{\mu}_{t+1}$ and Σ_{t+1} are $\mathcal{F}_t$-measurable.

The estimation of $F_{\boldsymbol{X}_{t+1}|\mathcal{F}_t}$ can be carried out in a number of ways. We can fit a (multivariate) time-series model to historical data $\boldsymbol{X}_{t-n+1}, \ldots, \boldsymbol{X}_t$ and use the fitted model to derive estimates of $\boldsymbol{\mu}_{t+1}$ and Σ_{t+1}. In Section 4.2.5 we explained the procedure for univariate GARCH or ARMA–GARCH models (see Examples 4.25 and 4.26 in particular). The same idea carries over to the multivariate GARCH models of Section 14.2 (see Section 14.2.6 in particular).

Alternatively, and more straightforwardly, the model-free exponentially weighted moving-average (EWMA) procedure suggested by the RiskMetrics group can be used. The univariate version of this technique was presented in Section 4.2.5 and the multivariate version is a simple extension of the idea. Let us suppose that we work in the context of a multivariate model with conditional mean $\boldsymbol{\mu}_t = E(\boldsymbol{X}_t \mid \mathcal{F}_{t-1}) = \mathbf{0}$. The conditional covariance matrix Σ_{t+1} is estimated recursively by

$$\hat{\Sigma}_{t+1} = \theta \boldsymbol{X}_t \boldsymbol{X}_t' + (1-\theta)\hat{\Sigma}_t, \tag{9.22}$$

where θ is a small positive number (typically of the order $\theta \approx 0.04$). The estimator (9.22) takes the form of a weighted sum of the estimate of Σ_t calculated at time $t-1$ and a term $\boldsymbol{X}_t \boldsymbol{X}_t'$ that satisfies $E(\boldsymbol{X}_t \boldsymbol{X}_t' \mid \mathcal{F}_{t-1}) = \Sigma_t$. The interpretation is that the estimate at time t is obtained by perturbing the estimate at time $t-1$ by a term that reacts to the "latest information" about the joint variability of risk-factor changes.

For n large we can calculate that

$$\hat{\Sigma}_{t+1} \approx \theta \sum_{i=0}^{n-1} (1-\theta)^i \boldsymbol{X}_{t-i} \boldsymbol{X}_{t-i}'.$$

This means that, after the EWMA procedure has been running for a while, the influence of starting values for the conditional covariance matrix is negligible and estimates are effectively weighted sums of the matrices $\boldsymbol{X}_t \boldsymbol{X}_t'$ where the weights decay exponentially. These estimates are usually quite close to estimates derived by formal multivariate GARCH modelling. The method can be refined by relaxing the assumption that the conditional mean satisfies $\boldsymbol{\mu}_t = \mathbf{0}$ and including an estimate of $\boldsymbol{\mu}_t$ obtained by exponential smoothing to get the updating equation

$$\hat{\Sigma}_{t+1} = \theta(\boldsymbol{X}_t - \hat{\boldsymbol{\mu}}_t)(\boldsymbol{X}_t - \hat{\boldsymbol{\mu}}_t)' + (1-\theta)\hat{\Sigma}_t.$$

The second critical assumption in the variance–covariance method is that the linear loss operator (9.7) for the portfolio in question is a sufficiently accurate approximation of the actual loss operator (9.5). The linear loss operator is a function of the form

$$l^{\Delta}_{[t]}(\boldsymbol{x}) = -(c_t + \boldsymbol{b}_t'\boldsymbol{x}) \tag{9.23}$$

for some constant c_t and constant vector $\boldsymbol{b}_t$, which are known to us at time t. We have seen a number of examples including

- the stock portfolio of Example 2.1, where the loss operator takes the form $l^{\Delta}_{[t]}(\boldsymbol{x}) = -v_t \boldsymbol{w}_t'\boldsymbol{x}$ and $\boldsymbol{w}_t$ is the vector of portfolio weights at time t;
- the European call option of Example 2.2; and
- the zero-coupon bond portfolio with linear loss operator given by (9.13).

An important property of the multivariate normal is that a linear function (9.23) of a normal vector must have a univariate normal distribution, as discussed in Section 6.1.3. From (6.13) we infer that, conditional on $\mathcal{F}_t$,

$$L^{\Delta}_{t+1} = l^{\Delta}_{[t]}(\boldsymbol{X}_{t+1}) \sim N(-c_t - \boldsymbol{b}_t'\boldsymbol{\mu}_{t+1}, \boldsymbol{b}_t'\Sigma_{t+1}\boldsymbol{b}_t). \tag{9.24}$$

VaR and expected shortfall may be easily calculated for the normal loss distribution in (9.24). For VaR we use formula (2.18) in Example 2.11. For expected shortfall we use formula (2.24) in Example 2.14.

Weaknesses of the method and extensions. The variance–covariance method offers a simple analytical solution to the risk-measurement problem but this convenience is achieved at the cost of two crude simplifying assumptions. First, linearization may not always offer a good approximation of the relationship between the true loss distribution and the risk-factor changes, particularly for derivative portfolios and longer time intervals. Second, the assumption of normality is unlikely to be realistic for the conditional distribution of the risk-factor changes for shorter-interval data such as daily data and weekly data. Another way of putting this is to say that the *innovation distribution* in a suitable time-series model of such data is generally heavier tailed than normal (see Example 4.24).

The convenience of the variance–covariance method relies on the fact that a linear combination of a multivariate Gaussian vector has a univariate Gaussian distribution. However, we have seen in Chapter 6 that there are other multivariate distribution families that are *closed under linear operations*, and variance–covariance methods can also be developed for these. Examples include multivariate t distributions and multivariate *generalized hyperbolic* distributions (see, in particular, Proposition 6.13 and (6.45)).

For example, suppose we model risk-factor changes in such a way that the conditional distribution is a multivariate t distribution; in other words, assume that $\boldsymbol{X}_{t+1} \mid \mathcal{F}_t \sim t_d(\nu, \boldsymbol{\mu}, \Sigma)$, where this notation was explained in Section 6.4 (see Example 6.7). Then, conditional on $\mathcal{F}_t$, we get from (9.23) that

$$L^{\Delta}_{t+1} = l^{\Delta}_{[t]}(\boldsymbol{X}_{t+1}) \sim t(\nu, -c_t - \boldsymbol{b}_t'\boldsymbol{\mu}, \boldsymbol{b}_t'\tilde{\Sigma}\boldsymbol{b}_t), \tag{9.25}$$

and risk measures can be calculated using (2.19) and (2.25) in Example 2.14.

9.2.3 Historical Simulation

Historical simulation is by far the most popular method used by banks for the trading book; Pérignon and Smith (2010) report that 73% of US and international commercial banks that disclose their methodology use this method. Moreover, many of these firms use historical simulation in a simple unconditional manner, as will be explained in this section. In Section 9.2.4 we discuss different approaches to adapting historical simulation to give conditional measures of risk that take account of changing market volatility.

Instead of estimating the distribution of $l_{[t]}(\boldsymbol{X}_{t+1})$ under some explicit parametric model for $\boldsymbol{X}_{t+1}$, the historical-simulation method can be thought of as estimating the distribution of the loss operator under the *empirical distribution* of data $\boldsymbol{X}_{t-n+1}, \ldots, \boldsymbol{X}_t$. The method can be concisely described using the loss-operator notation; we construct a univariate data set by applying the operator to each of our historical observations of the risk-factor change vector to get a set of historically simulated losses:

$$\{\tilde{L}_s = l_{[t]}(\boldsymbol{X}_s) \colon s = t - n + 1, \ldots, t\}. \tag{9.26}$$

The values $\tilde{L}_s$ show what would happen to the current portfolio if the risk-factor changes on day s were to recur. We make inferences about the loss distribution and risk measures using these historically simulated loss data.

If we assume for a moment that the risk-factor changes are iid with df $F_{\boldsymbol{X}}$ and write $F_n(l)$ for the empirical df of the data $\tilde{L}_{t-n+1}, \ldots, \tilde{L}_t$, then we may use the strong law of large numbers to show that, as $n \to \infty$,

$$\begin{aligned} F_n(l) = \frac{1}{n} \sum_{s=t-n+1}^{t} I_{\{\tilde{L}_s \leqslant l\}} = \frac{1}{n} \sum_{s=t-n+1}^{t} I_{\{l_{[t]}(\boldsymbol{X}_s) \leqslant l\}} \\ \to P(l_{[t]}(\boldsymbol{X}) \leqslant l) = F_L(l), \end{aligned}$$

where $\boldsymbol{X}$ is a generic vector of risk-factor changes with distribution $F_{\boldsymbol{X}}$ and where $L = l_{[t]}(\boldsymbol{X})$. Thus $F_n(l)$ is a consistent estimator of the df of $l_{[t]}(\boldsymbol{X})$ under $F_{\boldsymbol{X}}$.

The same conclusion will also apply for many strictly stationary time-series models, such as GARCH processes, under a suitable adaptation of the strong law of large numbers. Since the empirical df of the historically simulated loss data estimates the distribution of $l_{[t]}(\boldsymbol{X})$ under $F_{\boldsymbol{X}}$, historical simulation in its basic form is an unconditional method.

In practice, there are various ways we can use the historically simulated loss data. It is common to estimate VaR using the method of *empirical quantile estimation*, whereby theoretical quantiles of the loss distribution are estimated by *sample quantiles* of the data. As an alternative, the EVT-based methods of Section 5.2 can also be used to derive parametric estimates of the tail of the loss distribution. Further discussion of these topics is deferred to Section 9.2.6.

Strengths and weaknesses of the method. The historical-simulation method has obvious attractions: it is easy to implement and reduces the risk-measure estimation

problem to a one-dimensional problem; no statistical estimation of the multivariate distribution of $\boldsymbol{X}$ is necessary, and no assumptions about the dependence structure of risk-factor changes are made; we usually work with the original loss operator and not a linearized approximation.

However, as we have observed, it is an unconditional method and is therefore unsuited to giving dynamic, conditional measures of risk that capture the volatile nature of risk in the trading book. Advocates of historical simulation often find a virtue in the fact that it gives more stable, less volatile estimates of risk than a conditional method. But it is more prudent to separate the statistical or econometric problem of accurately estimating risk measures from the regulatory problem of imposing more stable capital requirements; to some extent, stability is behind the sixty-day smoothing used in (2.20). For this reason we look at dynamic extensions of historical simulation in Section 9.2.4.

Another issue is that the success of the approach is dependent on our ability to collect sufficient quantities of relevant, synchronized data for all risk factors. Whenever there are gaps in the risk-factor history, or whenever new risk factors are introduced into the modelling, there may be problems filling the gaps and completing the historical record. These problems will tend to reduce the effective value of n and mean that empirical estimates of VaR and expected shortfall have very poor accuracy. Ideally we want n to be fairly large, since the method is an unconditional method and we want a number of extreme scenarios in the historical record to provide more informative estimates of the tail of the loss distribution.

The method has been described as being like "driving a car while looking through the rear view mirror", a deficiency that is shared to an extent by all purely statistical procedures. It is for this reason that the stressed VaR calculations mentioned in Section 9.2.1 were introduced by regulators.

Finally, although the historical-simulation method can be easily described, it may prove difficult to implement efficiently for large portfolios of derivative instruments. Computing the historically simulated losses in (9.26) involves what practitioners refer to as *full revaluation* of the portfolio under each of the historical scenarios $\boldsymbol{X}_s$, and this may be computationally costly. To get round the problem of full revaluation in producing the simulated losses in (9.26), we can consider substituting the quadratic loss operator $l_{[t]}^{\Delta\Gamma}$ for the loss operator $l_{[t]}$ and working with second-order approximations to the losses. This means that only the risk factor sensitivities are required, and it is these that are often routinely calculated for hedging purposes anyway.

9.2.4 Dynamic Historical Simulation

In this section we present two approaches to incorporating volatility forecasting into historical simulation: a univariate approach based on univariate volatility prediction using the kind of models presented in Chapter 4, and a multivariate approach.

A univariate approach to dynamic historical simulation. For a given loss operator $l_{[t]}$ at time t, we recall the construction of the historical simulation data $\{\tilde{L}_s = l_{[t]}(\boldsymbol{X}_s) \colon s = t-n+1, \ldots, t\}$ in (9.26). We assume that these are realizations from

a stationary univariate stochastic process $(\tilde{L}_s)$ obtained by applying the function $l_{[t]}\colon \mathbb{R}^d \to \mathbb{R}$ to a stationary multivariate process of risk-factor changes $(\boldsymbol{X}_s)$, and we also assume that $L_{t+1} = l_{[t]}(\boldsymbol{X}_{t+1})$ is the next random variable in this process.

Moreover, we assume that the stationary process $(\tilde{L}_s)$ satisfies, for all s, equations of the form $\tilde{L}_s = \mu_s + \sigma_s Z_s$, where μ_s is an $\mathcal{F}_{s-1}$-measurable conditional mean term, σ_s is an $\mathcal{F}_{s-1}$-measurable volatility, and the (Z_s) are SWN(0, 1) innovations with df F_Z. An example of a model satisfying these assumptions would be an ARMA process with GARCH errors as defined in Section 4.2.3. As shown in Section 4.2.5, we can derive simple formulas for the VaR and expected shortfall of the conditional loss distribution $F_{L_{t+1}|\mathcal{F}_t}$ under these assumptions.

Writing VaR_α^t for the α-quantile of $F_{L_{t+1}|\mathcal{F}_t}$ and ES_α^t for the corresponding expected shortfall, we obtain

$$\mathrm{VaR}_\alpha^t = \mu_{t+1} + \sigma_{t+1} q_\alpha(Z), \qquad \mathrm{ES}_\alpha^t = \mu_{t+1} + \sigma_{t+1}\,\mathrm{ES}_\alpha(Z), \tag{9.27}$$

where Z is a generic rv with the df F_Z.

To estimate the risk measures in (9.27), we require estimates of μ_{t+1} and σ_{t+1} and estimates of the quantile and expected shortfall of the innovation df F_Z. In a model with Gaussian innovations the latter need not be estimated and are simply $q_\alpha(Z) = \Phi^{-1}(\alpha)$ and $\mathrm{ES}_\alpha(Z) = \phi(\Phi^{-1}(\alpha))/(1-\alpha)$, where the latter formula was derived in Example 2.14. In a model with non-Gaussian innovations, $q_\alpha(Z)$ and $\mathrm{ES}_\alpha(Z)$ depend on any further parameters of the innovation distribution. For example, we might assume (scaled) t innovations; in this case, the quantile and expected shortfall of a standard univariate t distribution (the latter given in (2.25)) would have to be scaled by the factor $\sqrt{(\nu-2)/\nu}$ to take account of the fact that the innovation distribution is assumed to have variance 1.

We now give a number of possible estimation strategies. In all cases the data are the historical simulation data $\tilde{L}_{t-n+1}, \ldots, \tilde{L}_t$.

(1) Fit an ARMA–GARCH model with an appropriate innovation distribution to the data by the ML method and use the prediction methodology discussed in Section 4.2.5 to estimate σ_{t+1} and μ_{t+1}. Any further parameters of the innovation distribution can be estimated simultaneously in the model fitting.

For example, suppose we use an AR(1)–GARCH(1,1) model, which, according to Definition 4.22, takes the form

$$\begin{aligned}
\tilde{L}_s &= \mu_s + \sigma_s Z_s,\\
\mu_s &= \mu + \phi_1(\tilde{L}_{s-1} - \mu_{s-1}),\\
\sigma_s^2 &= \alpha_0 + \alpha_1(\tilde{L}_{s-1} - \mu_{s-1})^2 + \beta_1\sigma_{s-1}^2
\end{aligned}$$

at any time s. The conditional mean μ_{t+1} and standard deviation σ_{t+1} are then estimated recursively by

$$\begin{aligned}
\hat{\mu}_{t+1} &= \hat{\mu} + \hat{\phi}_1(\tilde{L}_t - \hat{\mu}_t),\\
\hat{\sigma}_{t+1} &= \sqrt{\hat{\alpha}_0 + \hat{\alpha}_1(\tilde{L}_t - \hat{\mu}_t)^2 + \hat{\beta}_1\hat{\sigma}_t^2},
\end{aligned}$$

where ML estimates of the parameters of the AR(1)–GARCH(1,1) model are denoted using hats.

(2) Fit an ARMA–GARCH model by QML (see Section 4.2.4) and use prediction methodology as in strategy (1) to estimate σ_{t+1} and μ_{t+1}. In a separate second step use the model residuals to find estimates of $q_\alpha(Z)$ and $\mathrm{ES}_\alpha(Z)$. As for the basic historical-simulation method, this can be achieved using simple empirical estimates of quantiles and expected shortfalls or semi-parametric estimators based on EVT (see Section 9.2.6).

(3) Use the univariate EWMA procedure (see Section 4.2.5) to estimate

$$\sigma_{t-n+1}, \ldots, \sigma_t, \sigma_{t+1}.$$

The conditional mean terms $\mu_{t-n+1}, \ldots, \mu_t, \mu_{t+1}$ could also be estimated by exponential smoothing but it is easier to set them equal to zero, as they are likely to be very small. Standardize each of the historical simulation losses $\tilde{L}_{t-n+1}, \ldots, \tilde{L}_t$ by dividing by the EWMA volatility estimates $\hat{\sigma}_{t-n+1}, \ldots, \hat{\sigma}_t$. This yields a set of residuals, from which the innovation distribution F_Z can be estimated as in strategy (2).

These procedures often work well in practice but there can be some loss of information involved with applying volatility modelling at the level of the historically simulated data rather than at the level of the risk-factor changes themselves. We now present a second method that incorporates volatility at the level of the individual risk factors. While the method is more computationally intensive, it can result in more accurate estimates of risk measures.

A multivariate approach to dynamic historical simulation. In this method we work with risk-factor change data $\boldsymbol{X}_{t-n+1}, \ldots, \boldsymbol{X}_t$ and assume that the data vectors are realizations from a multivariate time-series process $(\boldsymbol{X}_s)$ that satisfies equations of the form

$$\boldsymbol{X}_s = \boldsymbol{\mu}_s + \Delta_s \boldsymbol{Z}_s, \quad \Delta_s = \mathrm{diag}(\sigma_{s,1}, \ldots, \sigma_{s,d}),$$

where $(\boldsymbol{\mu}_s)$ is a process of vectors and (Δ_s) a process of diagonal matrices such that $\mu_{s,1}, \ldots, \mu_{s,d}, \sigma_{s,1}, \ldots, \sigma_{s,d}$ are all $\mathcal{F}_{s-1}$-measurable and $(\boldsymbol{Z}_s) \sim \mathrm{SWN}(\boldsymbol{0}, P)$ for some correlation matrix P (in other words, the $\boldsymbol{Z}_s$ are iid random vectors whose covariance matrix is the correlation matrix P). Under these assumptions, $E(X_{s,k} \mid \mathcal{F}_{s-1}) = \mu_{s,k}$ and $\mathrm{var}(X_{s,k} \mid \mathcal{F}_{s-1}) = \sigma^2_{s,k}$, so the vector $\boldsymbol{\mu}_s$ contains the conditional means and the matrix Δ_s contains the volatilities of the component series at time s. An example of a model that fits into this framework is the CCC–GARCH (constant conditional correlation) process (see Definition 14.11).

In this context we may use multivariate dynamic historical simulation. The key idea of the method is to apply historical simulation to the unobserved innovations $(\boldsymbol{Z}_s)$ rather than the observed data $(\boldsymbol{X}_s)$ (as in standard historical simulation). The first step is to compute estimates $\{\hat{\boldsymbol{\mu}}_s : s = t-n+1, \ldots, t\}$ and $\{\hat{\Delta}_s : s = t-n+1, \ldots, t\}$ of the conditional mean vectors and volatility matrices.

This can be achieved by fitting univariate time-series models of ARMA–GARCH type to each of the component series in turn; alternatively, we can use the univariate EWMA approach for each series. In either case we also use prediction methodology to obtain estimates of $\hat{\Delta}_{t+1}$, the volatility matrix in the next time period, and (if desired) $\hat{\boldsymbol{\mu}}_{t+1}$, the conditional mean vector.

In the second step we construct residuals

$$\{\hat{\boldsymbol{Z}}_s = \hat{\Delta}_s^{-1}(\boldsymbol{X}_s - \hat{\boldsymbol{\mu}}_s) \colon s = t-n+1, \ldots, t\}$$

and treat these as "observations" of the unobserved innovations. To make statistical inferences about the distribution of $L_{t+1} = l_{[t]}(\boldsymbol{X}_{t+1}) = l_{[t]}(\boldsymbol{\mu}_{t+1} + \Delta_{t+1}\boldsymbol{Z}_{t+1})$ given $\mathcal{F}_t$ we construct the data set

$$\{\tilde{L}_s = l_{[t]}(\hat{\boldsymbol{\mu}}_{t+1} + \hat{\Delta}_{t+1}\hat{\boldsymbol{Z}}_s) \colon s = t-n+1, \ldots, t\}. \tag{9.28}$$

To estimate VaR (or expected shortfall) we can apply simple empirical estimators or EVT-based methods directly to these data (see Section 9.2.6 for more details).

9.2.5 Monte Carlo

The Monte Carlo method is a rather general name for any approach to risk measurement that involves the simulation of an explicit parametric model for risk-factor changes. Many banks report that they use a Monte Carlo method to compute measures of market risk in the trading book (Pérignon and Smith 2010). However, the Monte Carlo method only offers a solution to the problem of evaluating the distribution of $L_{t+1} = l_{[t]}(\boldsymbol{X}_{t+1})$ under a given model for $\boldsymbol{X}_{t+1}$. It does not solve the statistical problem of finding a suitable model for $\boldsymbol{X}_{t+1}$.

In the market-risk context let us assume that we have estimated a time-series model for historical risk-factor change data $\boldsymbol{X}_{t-n+1}, \ldots, \boldsymbol{X}_t$ and that this is a model from which we can readily simulate. We use the model to generate m independent realizations $\tilde{\boldsymbol{X}}_{t+1}^{(1)}, \ldots, \tilde{\boldsymbol{X}}_{t+1}^{(m)}$ from the estimated conditional distribution of risk-factor changes $\hat{F}_{\boldsymbol{X}_{t+1}|\mathcal{F}_t}$.

In a similar fashion to the historical-simulation method, we apply the loss operator to these simulated vectors to obtain simulated realizations $\{\tilde{L}_{t+1}^{(i)} = l_{[t]}(\tilde{\boldsymbol{X}}_{t+1}^{(i)}) \colon i = 1, \ldots, m\}$ from the estimated conditional loss distribution $\hat{F}_{L_{t+1}|\mathcal{F}_t}$. As for the historical-simulation method, the simulated loss data from the Monte Carlo method are used to estimate VaR and expected shortfall, e.g. by using simple empirical estimators or EVT-based methods (see Section 9.2.6 for more details).

Note that the use of Monte Carlo means that we are free to choose the number of replications m ourselves, within the obvious constraints of computation time. Generally, m can be chosen to be much larger than n (the number of data) so we obtain more accuracy in empirical VaR and expected shortfall estimates than is possible in the case of historical simulation.

Weaknesses of the method. As we have already remarked, the method does not solve the problem of finding a multivariate model for $F_{\boldsymbol{X}_{t+1}|\mathcal{F}_t}$ and any results that are obtained will only be as good as the model that is used.

For large portfolios the computational cost of the Monte Carlo approach can be considerable, as every simulation ideally requires the full revaluation of the portfolio to compute the loss operator. This is particularly problematic if the portfolio contains many derivatives that cannot be priced in closed form. The problem of computational cost is even more relevant to the Monte Carlo method than it is to the historical-simulation method, because we typically choose larger numbers of scenarios m for risk-factor changes in the Monte Carlo method. If second-order sensitivities are available, the loss operator can be replaced by the quadratic loss operator $l_{[t]}^{\Delta\Gamma}$ to reduce the computational cost. Moreover, variance-reduction techniques for evaluating tail probabilities and quantiles, such as importance sampling, can also be of help (see Notes and Comments).

9.2.6 Estimating Risk Measures

In both the historical simulation and Monte Carlo methods we estimate risk measures using simulated loss data. In this section we discuss different methods for estimating VaR and expected shortfall from a data sample. Let us suppose that we have data $L_1, \dots, L_n$ from an underlying loss distribution F_L and the aim is to estimate $\mathrm{VaR}_\alpha = q_\alpha(F_L) = F_L^{\leftarrow}(\alpha)$ or $\mathrm{ES}_\alpha = (1-\alpha)^{-1}\int_\alpha^1 q_\theta(F_L)\,\mathrm{d}\theta$.

L-estimators. These estimators take the form of linear combinations of sample order statistics, and the "L" in their name refers to *linear*. In Chapter 5 we defined the upper-order statistics $L_{1,n} \geqslant \cdots \geqslant L_{n,n}$, as is standard in extreme value theory. Many of the results concerning L-estimators are given in terms of lower-order statistics $L_{(1)} \leqslant \cdots \leqslant L_{(n)}$. Note that we can easily move between the two conventions by observing that $L_{k,n} = L_{(n-k+1)}$ for $k = 1, \dots, n$.

The simplest L-estimator of VaR is the sample quantile obtained by inverting the empirical distribution function $F_n(x) = n^{-1}\sum_{i=1}^n 1_{\{L_i \leqslant x\}}$ of the data $L_1, \dots, L_n$. It may be easily verified that the inverse of the empirical df is given by

$$F_n^{\leftarrow}(\alpha) = L_{(k)} \quad \text{for } \frac{k-1}{n} < \alpha \leqslant \frac{k}{n}.$$

We may write this more compactly as $F_n^{\leftarrow}(\alpha) = L_{(\lceil n\alpha \rceil)}$, where $\lceil x \rceil = \min\{k \in \mathbb{Z}\colon k \geqslant x\}$. This is the *ceiling* function that gives the smallest integer not less than x.

In working with order statistics we often use both the ceiling function and the *floor* function, $\lfloor x \rfloor = \max\{k \in \mathbb{Z}\colon k \leqslant x\}$, the largest integer not greater than x. It is easy to see that they are related by $\lfloor -x \rfloor = -\lceil x \rceil$. This fact, together with the relation $L_{k,n} = L_{(n-k+1)}$, allows us to write the sample quantile in terms of upper-order statistics. We have that $L_{(\lceil n\alpha \rceil)} = L_{k,n}$, where $k = n - \lceil n\alpha \rceil + 1 = \lfloor n(1-\alpha) \rfloor + 1$, giving

$$\widehat{\mathrm{VaR}}_\alpha = L_{k,n}, \quad k = \lfloor n(1-\alpha) \rfloor + 1. \tag{9.29}$$

For example, if $n = 1000$ and $\alpha = 0.995$, the estimator would be $L_{6,1000}$, the sixth largest value. For the same data and $\alpha = 0.9945$ the estimator is also $L_{6,1000}$.

Inverting the empirical distribution function yields a sample quantile function that is discontinuous in α. To obtain a continuous function in α there are a number of

alternative definitions of sample quantiles that interpolate linearly between adjacent order statistics. For example, the default method in the statistical package R estimates the α-quantile to be

$$\widehat{\mathrm{VaR}}_\alpha = \lambda_{\alpha,k,n} L_{k+1,n} + (1 - \lambda_{\alpha,k,n}) L_{k,n}, \quad k = \lceil (n-1)(1-\alpha) \rceil, \tag{9.30}$$

where the weights are given by $\lambda_{\alpha,k,n} = (n-k) - (n-1)\alpha$. If $n = 1000$ and $\alpha = 0.995$, then $k = 5$ and the estimator is $0.995 L_{6,1000} + 0.005 L_{5,1000}$. If we want to estimate the 0.9945 quantile from the same data, then $k = 6$ and the estimator becomes $0.4945 L_{7,1000} + 0.5055 L_{6,1000}$.

The estimators (9.29) and (9.30), being based on only one or two order statistics, are subject to a large variance, particularly for quantiles in the tail of the distribution and for small sample sizes.

To obtain an L-estimator of expected shortfall we recall from Section 8.2.1 the general form of the distortion risk measures, of which expected shortfall is a special case. Distortion risk measures are given by

$$\varrho(L) = \int_0^1 F_L^{\leftarrow}(u) \, \mathrm{d}D(u)$$

for convex distortion functions D on $[0, 1]$; the distortion function for expected shortfall is $D_\alpha(u) = (1-\alpha)^{-1}(u-\alpha)^+$. L-estimators for distortion risk measures may be derived by inserting the inverse of the empirical df as an estimator of $F_L^{\leftarrow}$ to obtain

$$\hat{\varrho}(L) = \int_0^1 F_n^{\leftarrow}(u) \, \mathrm{d}D(u) = \sum_{k=1}^n L_{(k)} \left(D_\alpha\left(\frac{k}{n}\right) - D_\alpha\left(\frac{k-1}{n}\right) \right).$$

In the special case of expected shortfall the estimator is

$$\begin{aligned}
\widehat{\mathrm{ES}}_\alpha &= \frac{1}{n(1-\alpha)} \sum_{k=1}^n L_{(k)} ((k - n\alpha)^+ - ((k-1) - n\alpha)^+) \\
&= \frac{1}{n(1-\alpha)} \left(\left(\sum_{k=\lceil n\alpha \rceil + 1}^n L_{(k)} \right) + (\lceil n\alpha \rceil - n\alpha) L_{(\lceil n\alpha \rceil)} \right) \\
&= \frac{1}{n(1-\alpha)} \left(\left(\sum_{k=1}^{\lfloor n(1-\alpha) \rfloor} L_{k,n} \right) + (\lceil n\alpha \rceil - n\alpha) L_{\lfloor n(1-\alpha) \rfloor + 1, n} \right).
\end{aligned}$$

The final term involving $L_{\lfloor n(1-\alpha) \rfloor + 1, n}$ may sometimes be omitted for a simpler estimator.

EVT-based estimators. Simple empirical estimates of the VaR and, especially, the expected shortfall are likely to be inaccurate when n is of modest size (say only a few years of daily data). This is a problem for historical simulation in particular. A possible solution is to use the techniques of *extreme value theory* (EVT) to provide

estimates of the tail of the loss distribution that are as faithful as possible to the most extreme data and that use parametric forms that are supported by theory. In Section 5.2.3 we presented a standard EVT method based on the generalized Pareto distribution that is useful in this context.

To use this method to estimate VaR_α and ES_α we can set a high threshold $u = L_{k+1,n}$ at the $(k+1)$-upper-order statistic and fit a GPD distribution to excess losses over u. We thereby obtain ML estimates $\hat{\xi}$ and $\hat{\beta}$ based on k exceedances of the threshold. To form a quantile estimator, the value k must satisfy $k/n > 1-\alpha$; moreover, k should be sufficiently large to give reasonably accurate estimates of the GPD parameters.

We then form the risk-measure estimates

$$\widehat{\mathrm{VaR}}_\alpha = u + \frac{\hat{\beta}}{\hat{\xi}}\left(\left(\frac{1-\alpha}{k/n}\right)^{-\hat{\xi}} - 1\right),$$
$$\widehat{\mathrm{ES}}_\alpha = \frac{\widehat{\mathrm{VaR}}_\alpha}{1-\hat{\xi}} + \frac{\hat{\beta} - \hat{\xi}u}{1-\hat{\xi}}.$$

For more guidance on the choice of threshold, see Section 5.2.2; for a comparison of the EVT quantile estimates with simple empirical quantile estimates, see Section 5.2.5.

9.2.7 *Losses over Several Periods and Scaling*

In the banking context the methods we have described in previous sections are generally applied to daily risk-factor change data, and risk measures are routinely calculated for a one-day horizon. However, for regulatory capital purposes there is a requirement to calculate a 99% VaR estimate for a period of ten trading days (two weeks).

An obvious approach to this calculation is to model historical risk-factor changes over ten-day intervals using exactly the same methodology that has been discussed in this chapter. However, for a fixed amount n of historical time-series data, this results in a dramatic reduction in the precision of the statistical estimates of model parameters. For example, if we have $n = 1000$ days (just under four years) of historical data, this would give only 100 non-overlapping observations of ten-day risk-factor changes. To obtain similar accuracy to an analysis of the daily returns we would have to collect $n = 10\,000$ daily data (around thirty-eight years). It is possible to artificially preserve the value of n by the formation of overlapping risk-factor returns (a construction that is described in Section 3.1). However, this introduces new serial dependencies into the data, which complicates statistical modelling and does not lead to an obvious gain in statistical accuracy.

For these reasons most banks use a simple scaling rule, known as the *square-root-of-time rule*, to move between estimates of one-day VaR and estimates of ten-day VaR. We now look at the (limited) theoretical support for this rule and discuss an alternative Monte Carlo approach.

Scaling. For $h \in \mathbb{Z}$ and $h \geqslant 1$ suppose we denote the loss from time t over the next h periods by $L_{t+h}^{(h)}$. Arguing as in (9.3) and (9.4) we have

$$\begin{aligned} L_{t+h}^{(h)} &= -(V_{t+h} - V_t) \\ &= -(g(\tau_{t+h}, \boldsymbol{Z}_{t+h}) - g(\tau_t, \boldsymbol{Z}_t)) \\ &= -(g(\tau_{t+h}, \boldsymbol{Z}_t + \boldsymbol{X}_{t+1} + \cdots + \boldsymbol{X}_{t+h}) - g(\tau_t, \boldsymbol{Z}_t)) \\ &=: l_{[t]}^{(h)}\bigg(\sum_{i=1}^{h} \boldsymbol{X}_{t+i}\bigg), \end{aligned}$$

where $l_{[t]}^{(h)}$ represents a loss operator at time t for the h-period loss. The general question of interest is how risk measures applied to the conditional distribution of $L_{t+h}^{(h)}$ given $\mathcal{F}_t$ scale with h, and this has no simple answer except in special cases.

Note that the h-period loss operator differs from the one-period loss operator in situations where the mapping depends explicitly on time (such as derivative portfolios). For simplicity let us consider the case in which the mapping does not depend on calendar time, so that $l_{[t]}^{(h)}(\boldsymbol{x}) = l_{[t]}(\boldsymbol{x})$. The linearized form of this operator is of the form $l_{[t]}^{\Delta}(\boldsymbol{x}) = \boldsymbol{b}_t'\boldsymbol{x}$ for some vector $\boldsymbol{b}_t$ that is known at time t. We look at the simpler problem of scaling for risk measures applied to the linearized loss distribution:

$$L_{t+h}^{(h)\Delta} = l_{[t]}^{\Delta}\bigg(\sum_{i=1}^{h} \boldsymbol{X}_{t+i}\bigg) = \sum_{i=1}^{h} \boldsymbol{b}_t' \boldsymbol{X}_{t+i}. \tag{9.31}$$

The following example gives a justification for the square-root-of-time rule.

Example 9.4 (square-root-of-time scaling). Suppose the risk-factor change vectors are iid with distribution $N_d(\boldsymbol{0}, \Sigma)$. Then $\sum_{i=1}^{h} \boldsymbol{X}_{t+i} \sim N_d(\boldsymbol{0}, h\Sigma)$ and the distribution of $L_{t+h}^{(h)\Delta}$ in (9.31) satisfies $L_{t+h}^{(h)\Delta} \sim N(0, h\boldsymbol{b}_t'\Sigma\boldsymbol{b}_t)$. It then follows easily from (2.18) and (2.24) that both quantiles and expected shortfalls for this distribution scale according to the square root of time ($\sqrt{h}$). For example, writing $\mathrm{ES}_\alpha^{(h)}$ for the expected shortfall, we have

$$\mathrm{ES}_\alpha^{(h)} = \sqrt{h}\sigma \frac{\phi(\Phi^{-1}(\alpha))}{1-\alpha},$$

where $\sigma^2 = \boldsymbol{b}_t'\Sigma\boldsymbol{b}_t$. Clearly, $\mathrm{ES}_\alpha^{(h)} = \sqrt{h}\,\mathrm{ES}_\alpha^{(1)}$ and, with similar notation, $\mathrm{VaR}_\alpha^{(h)} = \sqrt{h}\,\mathrm{VaR}_\alpha^{(1)}$.

Although this scaling rule is quite commonly used in practice, empirical risk-factor change data generally support neither a Gaussian distributional assumption nor an iid assumption (see Section 3.1). Moreover, for the kinds of dynamic time-series model (such as GARCH) that are appropriate, very little is known about the scaling of risk measures for the conditional loss distribution of the h-period loss $L_{t+h}^{(h)}$ (or its linearized form).

Monte Carlo approach. It is possible to use a Monte Carlo approach to the problem of determining risk measures for the h-period conditional loss distribution. Suppose we have a time-series model for the risk-factor changes $(\boldsymbol{X}_s)_{s \leqslant t}$. We simulate future

paths of the process $\tilde{\boldsymbol{X}}_{t+1}^{(i)}, \ldots, \tilde{\boldsymbol{X}}_{t+h}^{(i)}$ for $i = 1, \ldots, m$, where m is a predetermined large number of replications. We then apply the h-period loss operator to these simulated data to obtain Monte Carlo simulated losses:

$$\{\tilde{L}_{t+h}^{(h)(i)} = l_{[t]}^{(h)}(\tilde{\boldsymbol{X}}_{t+1}^{(i)} + \cdots + \tilde{\boldsymbol{X}}_{t+h}^{(i)}) \colon i = 1, \ldots, m\}.$$

These are used to make statistical inferences about the loss distribution and associated risk measures, as described in Section 9.2.5. We can also use the Monte Carlo approach to examine the performance of square-root-of-time scaling and to experiment with alternative power laws (see Notes and Comments).

Notes and Comments

Standard methods for market risk are described in detail in Jorion (2007) and Crouhy, Galai and Mark (2001). For the variance–covariance approach using EWMA, see Mina and Xiao (2001). A useful overview of the popularity of different approaches in practice is given by Pérignon and Smith (2010). The multivariate approach to dynamic historical simulation is described by Hull and White (1998) and Barone-Adesi, Bourgoin and Giannopoulos (1998).

The book by Glasserman (2003) is an excellent general introduction to Monte Carlo simulation techniques in finance. Glasserman, Heidelberger and Shahabuddin (1999) present efficient numerical techniques (based on delta–gamma approximations and advanced simulation techniques) for estimating VaR for derivative portfolios in the presence of heavy-tailed risk factors.

For a reference on different definitions of empirical quantile estimates and their properties, see Hyndman and Fan (1996). Tsukahara (2009) describes L-estimators of distortion risk measures, which apply to the case of expected shortfall. The use of EVT to provide dynamic estimates of risk measures was introduced by McNeil and Frey (2000), who also highlight the differences between conditional and unconditional approaches. For risk-measure estimation applying EVT to a regime-switching model, see Chavez-Demoulin, Embrechts and Sardy (2014).

A useful summary of scaling results for market-risk measures may be found in Kaufmann (2004) (see also Brummelhuis and Kaufmann 2007; Embrechts, Kaufmann and Patie 2005). In these papers the message emerges that, for unconditional VaR scaling over longer time horizons, the square-root-of-time rule often works well. On the other hand, for conditional VaR scaling over short time horizons, McNeil and Frey (2000) use the Monte Carlo approach to present evidence against square-root-of-time scaling. For further comments on these and further scaling issues, see Diebold et al. (1998) and Daníelsson and de Vries (1997c).

For a practically oriented text on market-risk management see Daníelsson (2011).

9.3 Backtesting

Backtesting is the practice of evaluating risk measurement procedures by comparing out-of-sample estimates of risk measures with actual realized losses and gains. Backtesting allows us to address the question of whether a given estimation procedure

produces credible risk-measure estimates. In Section 9.2 we considered standard methods for estimating risk measures at a time t for the distribution of losses in the next period. At the end of the next period we have the opportunity to compare the risk-measure estimate with the actual realized loss. When this procedure is repeated over many time periods we can monitor the performance of methods and compare their relative performance.

In Section 9.3.1 we discuss the backtesting of VaR estimates, and in Section 9.3.2 we discuss the backtesting of expected shortfall. Section 9.3.3 examines the use of elicitability theory to construct natural scoring statistics for comparing the backtest results for different VaR estimation methods. An empirical example of backtesting is described in Section 9.3.4, and in Section 9.3.5 we briefly consider backtests of the whole estimated loss distribution.

9.3.1 Violation-Based Tests for VaR

At any time point t let VaR_α^t denote the α-quantile of the conditional loss distribution $F_{L_{t+1}|\mathcal{F}_t}$. We will refer to the event $\{L_{t+1} > \text{VaR}_\alpha^t\}$ as a VaR *violation* or *exception* and define the event indicator variable by $I_{t+1} = I_{\{L_{t+1} > \text{VaR}_\alpha^t\}}$. Assuming a continuous loss distribution, we have, by definition of the quantile, that

$$E(I_{t+1} \mid \mathcal{F}_t) = P(L_{t+1} > \text{VaR}_\alpha^t \mid \mathcal{F}_t) = 1 - \alpha, \tag{9.32}$$

so that I_{t+1} is a Bernoulli variable with event probability $1 - \alpha$. Moreover, the following lemma shows that the sequence of VaR violation indicators (I_t) forms a Bernoulli trials process, i.e. a process of iid Bernoulli random variables with event probability $1 - \alpha$.

Lemma 9.5. *Let $(Y_t)_{t\in\mathbb{Z}}$ be a sequence of Bernoulli indicator variables adapted to a filtration $(\mathcal{F}_t)_{t\in\mathbb{Z}}$ and satisfying $E(Y_{t+1} \mid \mathcal{F}_t) = p > 0$ for all t. Then (Y_t) is a process of iid Bernoulli variables.*

Proof. The process $(Y_t - p)_{t\in\mathbb{Z}}$ has the martingale-difference property (see Definition 4.6). Moreover, $\text{var}(Y_t - p) = E(E((Y_t - p)^2 \mid \mathcal{F}_{t-1})) = p(1 - p)$ for all t. As $(Y_t - p)$ is a martingale-difference sequence with a finite variance, it is a white noise process (see Section 4.1.1). Hence (Y_t) is a white noise processes of uncorrelated variables. If two Bernoulli variables Y_t and Y_s are uncorrelated, it follows that

$$\begin{aligned} 0 = \text{cov}(Y_t, Y_s) &= E(Y_t Y_s) - E(Y_t)E(Y_s) \\ &= P(Y_t = 1, Y_s = 1) - P(Y_t = 1)P(Y_s = 1), \end{aligned}$$

which shows that they are also independent. □

There are two important consequences of the independent Bernoulli behaviour for violations. First, if we sum the violation indicators over a number of different times, we obtain binomially distributed random variables. For example, $M = \sum_{t=1}^m I_{t+1} \sim B(m, 1 - \alpha)$. Second, the spacings between consecutive violations are independent and geometrically distributed. Suppose that the event

$\{L_{t+1} > \mathrm{VaR}_\alpha^t\}$ occurs for times $t \in \{T_1, \ldots, T_M\}$, and let $T_0 = 0$. Then the spacings $S_j = T_j - T_{j-1}$ will be independent geometrically distributed random variables with mean $1/(1-\alpha)$, so that $P(S_j = k) = \alpha^{k-1}(1-\alpha)$ for $k \in \mathbb{N}$. Both of these properties can be tested on empirical data.

Suppose that we now estimate VaR_α^t based on information available up to time t, and we denote our estimate by $\widehat{\mathrm{VaR}}_\alpha^t$. The empirical violation indicator variable

$$\hat{I}_{t+1} = I_{\{L_{t+1} > \widehat{\mathrm{VaR}}_\alpha^t\}}$$

represents a one-step, out-of-sample comparison, in which we compare the actual realized value L_{t+1} with our VaR estimate made at time t.

Under the null hypothesis that our estimation method is accurate, in the sense that $E(\hat{I}_{t+1} \mid \mathcal{F}_t) = 1 - \alpha$ at the time points $t = 1, \ldots, m$, the sequence of empirical indicator variables $(\hat{I}_{t+1})_{1 \leqslant t \leqslant m}$ will then form a realization from a Bernoulli trials process with event probability $1 - \alpha$. For example, the quantity $\sum_{t=1}^m \hat{I}_{t+1}$ should behave like a realization from a $B(m, 1-\alpha)$ distribution, and this hypothesis can be easily addressed with a binomial test. There are a number of varieties of binomial test; in a two-sided score test we compute the statistic

$$Z_m = \frac{\sum_{t=1}^m \hat{I}_{t+1} - m(1-\alpha)}{\sqrt{m\alpha(1-\alpha)}} \tag{9.33}$$

and reject the hypothesis of Bernoulli behaviour at the 5% level if $|Z_m| > \Phi^{-1}(0.975)$. Rejection would suggest either systematic underestimation or overestimation of VaR (see Notes and Comments for further references concerning binomial tests).

To check the independence of violations we can construct a test of the geometric hypothesis. Since violations should be rare events with probability $(1-\alpha) \leqslant 0.05$, it proves easier to use the fact that a discrete-time Bernoulli process for rare events can be approximated by a continuous-time Poisson process and that the discrete geometric distribution for the event spacings can be approximated by a continuous exponential distribution.

To be precise let us suppose that the time interval $[t, t+1]$ in discrete time has length Δt in the chosen unit of continuous time. For example, if $[t, t+1]$ represents a trading day, then $\Delta t = 1$ if time is measured in days and $\Delta t = 1/250$ if time is measured in years. If the Bernoulli rare event probability is $1 - \alpha$, then the approximating Poisson process has rate $\lambda = (1-\alpha)\Delta t$ and the approximating exponential distribution has parameter λ and mean $1/\lambda$.

The exponential hypothesis can be tested using a Q–Q plot of the spacings data against the quantiles of a standard exponential reference distribution, similar to the situation discussed in Section 5.3.2. Alternatively, Christoffersen and Pelletier (2004) have proposed a likelihood ratio test of the hypothesis of exponential spacings against a more general Weibull alternative. In the exponential model the so-called hazard function of an event is constant, but the Weibull distribution can model an event clustering phenomenon whereby the hazard function is initially high after an event takes place and then decreases (see Section 10.4.1 for more discussion of hazard rate models).

In Section 9.2.7 we also discussed VaR estimates for the h-period loss distribution. To use the tests described above on h-period estimates we would have to base our backtests on non-overlapping periods. For example, if we calculated two-week VaRs, we could make a comparison of the VaR estimate and the realized loss every two weeks, which would clearly lead to a relatively small amount of violation data with which to monitor the performance of the model. It is also possible to look at overlapping periods, e.g. by recording the violation indicator value every day for the loss incurred over the previous two weeks. However, this would create a series of dependent Bernoulli trials for which formal inference is difficult.

9.3.2 Violation-Based Tests for Expected Shortfall

It is also possible to use information about the magnitudes of VaR violations to backtest estimates of expected shortfall. Let ES_α^t denote the expected shortfall of the conditional loss distribution $F_{L_{t+1}|\mathcal{F}_t}$, and define a *violation residual* by

$$K_{t+1} = \left(\frac{L_{t+1} - \mathrm{ES}_\alpha^t}{\mathrm{ES}_\alpha^t - \mu_{t+1}}\right) I_{\{L_{t+1} > \mathrm{VaR}_\alpha^t\}}, \tag{9.34}$$

where $\mu_{t+1} = E(L_{t+1} \mid \mathcal{F}_t)$. In the event that there is a VaR violation $\{L_{t+1} > \mathrm{VaR}_\alpha^t\}$, the violation residual K_{t+1} compares the actual size of the violation L_{t+1} with its expected size conditional on information up to time t, given by ES_α^t; if there is no VaR violation, the violation residual is 0. The reason for scaling the residual by $(\mathrm{ES}_\alpha^t - \mu_{t+1})$ will become apparent below.

It follows from Lemma 2.13 that, for a continuous loss distribution, the identity

$$E(K_{t+1} \mid \mathcal{F}_t) = 0$$

is satisfied so that the series of violation residuals (K_t) forms a martingale-difference series. Under stronger assumptions we can use this as the basis for a backtest of expected shortfall estimates. Let us assume that the underlying process generating the losses (L_t) satisfies, for all t, equations of the form $L_t = \mu_t + \sigma_t Z_t$, where μ_t is an $\mathcal{F}_{t-1}$-measurable conditional mean term, σ_t is an $\mathcal{F}_{t-1}$-measurable volatility and the (Z_t) are SWN(0, 1) innovations. This assumption would be satisfied, for example, by an ARMA process with GARCH errors, which mimics many of the essential features of financial return data. Under this assumption we have that $\mathrm{ES}_\alpha^t = \mu_{t+1} + \sigma_{t+1}\,\mathrm{ES}_\alpha(Z)$, where $\mathrm{ES}_\alpha(Z)$ denotes the expected shortfall of the innovation distribution. We can then calculate that

$$K_{t+1} = \left(\frac{Z_{t+1} - \mathrm{ES}_\alpha(Z)}{\mathrm{ES}_\alpha(Z)}\right) I_{\{Z_{t+1} > q_\alpha(Z)\}},$$

so that the sequence of violation residuals (K_t) forms a process of iid variables with mean 0 and an atom of probability mass of size α at zero.

Empirical violation residuals are formed, in the obvious way, by calculating

$$\hat{K}_{t+1} = \left(\frac{L_{t+1} - \widehat{\mathrm{ES}}_\alpha^t}{\widehat{\mathrm{ES}}_\alpha^t - \hat{\mu}_{t+1}}\right) \hat{I}_{t+1}, \tag{9.35}$$

where $\widehat{\mathrm{ES}}_\alpha^t$ denotes the estimated value of ES_α^t at time t, $\hat{I}_{t+1}$ is the VaR violation indicator defined in Section 9.3.1, and $\hat{\mu}_{t+1}$ is an estimate of the conditional mean. In practice, the conditional mean μ_{t+1} is not always estimated (particularly in EWMA-based methods), and when it is estimated it is often close to zero; for this reason we might simplify the calculations by setting $\hat{\mu}_{t+1} = 0$.

We expect the empirical violation residuals to behave like realizations of iid variables from a distribution with mean 0. We can test the hypothesis that the non-zero violation residuals have a mean of zero. The simplest approach is to use a t-test, and this is the option we choose in Section 9.3.4. It is also possible to use a bootstrap test that makes no assumption about the underlying distribution of the violation residuals (see Notes and Comments).

9.3.3 Elicitability and Comparison of Risk Measure Estimates

We now present a more recent approach to backtesting that is useful for comparing sets of risk-measure estimates derived using different methodologies. This approach is founded on the observation that the problem of estimating financial risk measures for the next time period is a special case of the general statistical problem of estimating statistics of a predictive or forecasting distribution. It is therefore natural to use ideas from the forecasting literature in backtesting.

In forecasting it is common to make predictions based on the idea of minimizing a scoring function or prediction error function. For example, if we wish to minimize the squared prediction error, it is well known that we should use the mean of the predictive distribution as our forecast; if we wish to minimize the absolute prediction error, we use the median of the predictive distribution.

The mean and median of the predictive distribution are known as *elicitable* statistical functionals of the distribution because they provide optimal forecasts under particular choices of scoring function. When a statistic is elicitable, there are natural ways of comparing different sets of estimates of that statistic using empirical scores.

For example, let (X_t) denote a time series and suppose we use two procedures A and B to estimate the conditional mean $\mu_{t+1} = E(X_{t+1} \mid \mathcal{F}_t)$ at different time points, based on data up to time t, resulting in estimates $\hat{\mu}_{t+1}^{(j)}$, $j \in \{A, B\}$, $t = 1, \ldots, m$. The conditional mean μ_{t+1} is known to be the optimal prediction of X_{t+1} under a squared error scoring function. We therefore compare our estimates with the actual realized values X_{t+1} by computing squared differences. The superior estimation procedure j will tend to be the one that gives the lowest value of the total squared prediction error $\sum_{t=1}^m (X_{t+1} - \hat{\mu}_{t+1}^{(j)})^2$.

We now give a more formal treatment of basic concepts from elicitability theory and show in particular how the ideas relate to the problem of estimating value-at-risk.

Elicitability theory. A law-invariant risk measure ϱ defined on a space of random variables $\mathcal{M}$ can also be viewed as a statistical functional T defined on a space of distribution functions $\mathcal{X}$. If $L \in \mathcal{M}$ has df $F_L \in \mathcal{X}$, then ϱ and T are linked by $\varrho(L) = T(F_L)$; for example, $\mathrm{VaR}_\alpha(L) = F_L^{\leftarrow}(\alpha)$, so the functional in this case is the generalized inverse at α.

The theory of elicitability is usually presented as a theory of statistical functionals of distribution functions. Our account is based on the presentation in Bellini and Bignozzi (2013). Note that we follow Bellini and Bignozzi in restricting our attention to real-valued functionals; this is more natural for the application to financial risk measures but differs from the more common presentation in the statistical forecasting literature where set-valued functionals are allowed (see, for example, Gneiting 2011).

Elicitable statistical functionals are functionals that minimize expected scores where the expected scores are calculated using *scoring functions*. These quantify the discrepancy between a forecast and a realized value from the distribution. The formal definitions of scoring functions and elicitable functionals are as follows.

Definition 9.6. A scoring function is a function $S\colon \mathbb{R}\times\mathbb{R}\to[0,\infty)$ satisfying, for any $y, l\in\mathbb{R}$:

(i) $S(y,l)\geqslant 0$ and $S(y,l)=0$ if and only if $y=l$;

(ii) $S(y,l)$ is increasing for $y>l$ and decreasing for $y<l$;

(iii) $S(y,l)$ is continuous in y.

Definition 9.7. A real-valued statistical functional T defined on a space of distribution functions $\mathcal{X}$ is said to be elicitable on $\mathcal{X}_T\subseteq\mathcal{X}$ if there exists a scoring function S such that, for every $F\in\mathcal{X}_T$,

(1) $\int_{\mathbb{R}} S(y,l)\,\mathrm{d}F(l)<\infty, \forall y\in\mathbb{R}$,

(2) $T(F)=\arg\min_{y\in\mathbb{R}}\int_{\mathbb{R}} S(y,l)\,\mathrm{d}F(l)$.

In this case the scoring function S is said to be *strictly consistent* for T.

In the context of risk measures, if L is an rv with loss distribution function F_L, then an elicitable risk measure minimizes

$$E(S(y,L))=\int_{\mathbb{R}} S(y,l)\,\mathrm{d}F_L(l) \tag{9.36}$$

with respect to y for every $F_L\in\mathcal{X}_T$, where $\mathcal{X}_T$ is the set of dfs for which the integral in (9.36) is defined.

For example, let $\mathcal{X}$ be the set of dfs of integrable random variables. The mean $E(L)=\int_{\mathbb{R}} l\,\mathrm{d}F_L(l)$ is elicitable on the space $\mathcal{X}_T$ of distribution functions with finite variance. This is clear because it minimizes (9.36) for the strictly consistent scoring function $S(y,l)=(y-l)^2$, as may be easily verified.

Application to the VaR and expectile risk measures. We now consider the VaR and expectile risk measures. The former is elicitable for strictly increasing distribution functions and the latter is elicitable in general, subject of course to the moment conditions imposed by Definition 9.7. We summarize this information and give strictly consistent scoring functions in the following two propositions.

Proposition 9.8. *For any $0 < \alpha < 1$ the statistical functional $T(F_L) = F_L^{\leftarrow}(\alpha)$ is elicitable on the set of strictly increasing distribution functions with finite mean. The scoring function*

$$S_\alpha^q(y, l) = |1_{\{l \leqslant y\}} - \alpha|\,|l - y| \tag{9.37}$$

is strictly consistent for T.

Proof. For L with df F_L, the expected score $E(S_\alpha^q(y, L))$ is a continuous function that is differentiable at all the points of continuity of F_L. The derivative is

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}y} E(S_\alpha^q(y, L)) &= \frac{\mathrm{d}}{\mathrm{d}y} \int_{-\infty}^{\infty} |1_{\{y \geqslant x\}} - \alpha|\,|y - x|\,\mathrm{d}F_L(x) \\
&= \frac{\mathrm{d}}{\mathrm{d}y} \int_{-\infty}^{y} (1 - \alpha)(y - x)\,\mathrm{d}F_L(x) + \frac{\mathrm{d}}{\mathrm{d}y} \int_{y}^{\infty} \alpha(x - y)\,\mathrm{d}F_L(x) \\
&= (1 - \alpha) \int_{-\infty}^{y} \mathrm{d}F_L(x) - \alpha \int_{y}^{\infty} \mathrm{d}F_L(x) \\
&= F_L(y) - \alpha.
\end{aligned}$$

There are two cases to consider. If there exists a point y such that $F_L(y) = \alpha$, then $y = F_L^{\leftarrow}(\alpha)$ and y clearly minimizes $E(S_\alpha^q(y, L))$. If, on the other hand, the set $\{y\colon F_L(y) = \alpha\}$ is empty, then it must be the case that there is a point y at which the distribution function F_L jumps and $F_L(x) - \alpha < 0$ for $x < y$ and $F_L(x) - \alpha > 0$ for $x > y$. It follows again that $y = F_L^{\leftarrow}(\alpha)$ and that y is the unique minimizer of $E(S_\alpha^q(y, L))$. □

Proposition 9.9. *For any $0 < \alpha < 1$ the statistical functional T corresponding to the expectile risk measure e_α is elicitable for all loss distributions F_L with finite variance. The scoring function*

$$S_\alpha^e(y, l) = |1_{\{l \leqslant y\}} - \alpha|(l - y)^2 \tag{9.38}$$

is strictly consistent for T.

Proof. This follows easily from (8.31) and Definition 8.21 in Section 8.2.2. □

Characterizations of elicitable risk measures. The expectile is thus both an elicitable and a coherent risk measure (provided $\alpha \geqslant 0.5$); it is in fact the only risk measure to have both these properties, as shown by Ziegel (2015). Bellini and Bignozzi (2013) have provided an elegant result that characterizes all the elicitable risk measures that have the extra properties required for convexity, coherence and comonotone additivity.

Theorem 9.10 (Bellini and Bignozzi (2013)). *Let $T(F_L)$ be the statistical functional corresponding to a law-invariant risk measure that is both monotonic and translation invariant. Then*

(a) $T(F_L)$ is convex and elicitable if and only if it is a risk measure based on a (convex) loss function (see Example 8.8),

(b) *$T(F_L)$ is coherent and elicitable if and only if it is an expectile e_α with $\alpha \geqslant 0.5$, and*

(c) *$T(F_L)$ is coherent, comonotone additive and elicitable if and only if it coincides with the expected loss.*

In particular, we note that the expected shortfall risk measure cannot be elicitable according to this theorem. There are other ways of demonstrating this more directly (see, for example, Gneiting 2011).

Computing empirical scores. As indicated in the introduction to this section, if we want to compare the performance of different methods of estimating elicitable functionals, we can use the strictly consistent scoring functions suggested by elicitability theory. From Proposition 9.8 we know that VaR_α^t, the quantile of $F_{L_{t+1}|\mathcal{F}_t}$, minimizes $E(S_\alpha^q(y, L_{t+1}) \mid \mathcal{F}_t)$.

Suppose, as in Section 9.3.1, that we compute estimates $\widehat{\mathrm{VaR}}_\alpha^t$ of VaR_α^t on days $t = 1, \ldots, m$, based on information up to time t, and backtest each estimate on day $t + 1$. A natural score is given by

$$\sum_{t=1}^{m} S_\alpha^q(\widehat{\mathrm{VaR}}_\alpha^t, L_{t+1}).$$

If we compute this quantity for different estimation methods, then the methods that give the most accurate estimates of the conditional quantiles VaR_α^t will tend to give the smallest scores.

Unfortunately, since expected shortfall is not an elicitable functional, there is no natural empirical score for comparing sets of estimates of expected shortfall.

9.3.4 Empirical Comparison of Methods Using Backtesting Concepts

In this section we apply various VaR estimation methods to the portfolio of a hypothetical investor in international equity indices and backtest the resulting VaR estimates. The methods we compare belong to the general categories of variance–covariance and historical-simulation methods and are a mix of unconditional and conditional approaches.

The investor is assumed to have domestic currency sterling (GBP) and to invest in the FTSE100 Index, the S&P 500 Index and the SMI (Swiss Market Index). The investor thus has currency exposure to US dollars (USD) and Swiss francs (CHF), and the value of the portfolio is influenced by five risk factors (three log index values and two log exchange rates). The corresponding risk-factor time series for the period 2000–2012 are shown in Figure 9.4.

On any day t we standardize the total portfolio value V_t in sterling to be 1 and assume that the portfolio weights (the proportions of this total value invested in each of the FTSE 100, the S&P 500 and the SMI) are 30%, 40% and 30%, respectively. Using similar reasoning to that in Example 2.1, it may be verified that the loss operator is

$$l_{[t]}(\boldsymbol{x}) = 1 - (0.3\mathrm{e}^{x_1} + 0.4\mathrm{e}^{x_2+x_4} + 0.3\mathrm{e}^{x_3+x_5}),$$

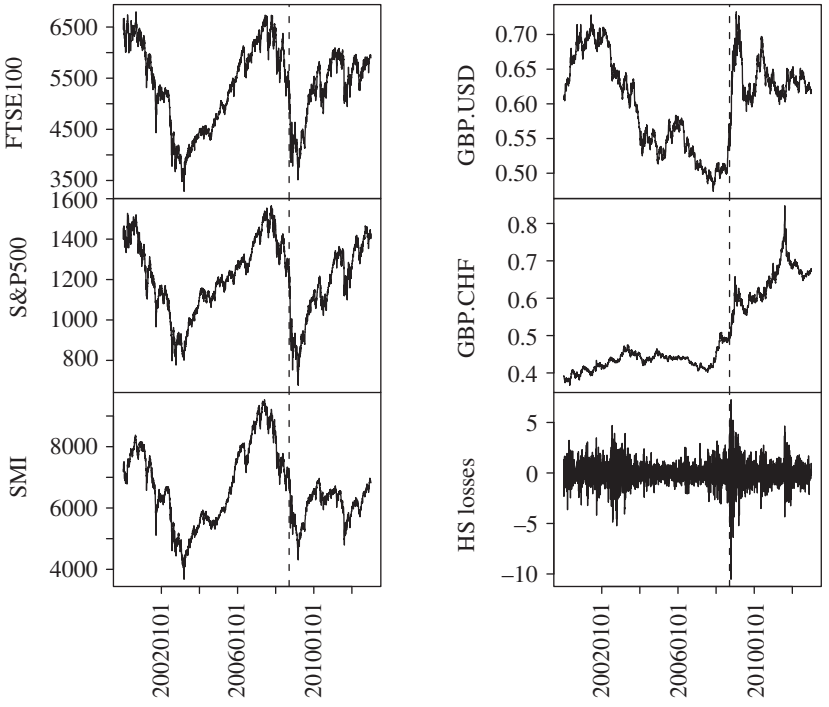

Figure 9.4. Time series of risk-factor values. These are (a) the FTSE 100, (b) the S&P 500, (c) the SMI, (d) the GBP/USD exchange rate and (e) the GBP/CHF exchange rate for the period 2000–2012. The final picture shows the corresponding historical simulation data (9.26) for the portfolio of Section 9.3.4. The vertical dashed line marks the date on which Lehman Brothers filed for bankruptcy.

and its linearized version is

$$l^{\Delta}_{[t]}(\boldsymbol{x}) = -(0.3x_1 + 0.4(x_2 + x_4) + 0.3(x_3 + x_5)),$$

where x_1, x_2 and x_3 represent log-returns on the three indices, and x_4 and x_5 are log-returns on the GBP/USD and GBP/CHF exchange rates.

Our objective is to calculate VaR estimates at the 95% and 99% levels for all trading days in the period 2005–12. Where local public holidays take place in individual markets (e.g. the Fourth of July in the US), we record artificial zero returns for the market in question, thus preserving around 258 days of risk-factor return data in each year. We use the last 1000 days of historical data $\boldsymbol{X}_{t-999}, \ldots, \boldsymbol{X}_t$ to make all VaR estimates for day $t + 1$ using the following methods.

VC. The variance–covariance method assuming multivariate Gaussian risk-factor changes and using the multivariate EWMA method to estimate the conditional covariance matrix of risk-factor changes as described in Section 9.2.2.

HS. The standard unconditional historical-simulation method as described in Section 9.2.3.

HS-GARCH. The univariate dynamic approach to historical simulation in which a GARCH(1, 1) model with a constant conditional mean term and Gaussian innovations is fitted to the historically simulated losses to estimate the volatility of the next day's loss (see Section 9.2.4).

HS-GARCH-*t*. A similar method to HS-GARCH but Student t innovations are assumed in the GARCH model.

HS-GARCH-EVT. A similar method to HS-GARCH and HS-GARCH-*t* but the model parameters are estimated by QML and EVT is applied to the model residuals; see strategy (2) in the univariate approach to dynamic historical simulation described in Section 9.2.4 and see also Section 5.2.6.

HS-MGARCH. The multivariate dynamic approach to historical simulation in which GARCH(1, 1) models with constant conditional mean terms are fitted to each time series of risk-factor changes to estimate volatilities (see Section 9.2.4).

HS-MGARCH-EVT. A similar method to HS-MGARCH but EVT estimators rather than simple empirical estimators are applied to the data constructed in (9.28) to calculate risk-measure estimates.

This collection of methods is of course far from complete and is merely meant as an indication of the kinds of strategies that are possible. In particular, we have confined our interest to rather simple GARCH models and not added, for example, asymmetric innovation distributions, leverage effects (see Section 4.2.3) or regime-switching models, which can often further improve the performance of such methods.

Table 9.1 contains the VaR violation counts for estimates of the 95% and 99% VaR for each of the methods. The violation counts have been broken down by year and the final column shows the total number of violations over the eight-year period. In each cell a binomial test has been carried out using the score statistic (9.33), and test results that are significant at the 5% level are indicated by italics.

At the 95% level the HS-MGARCH and HS-MGARCH-EVT methods clearly give the best overall results over the entire period; the former yields exactly the expected number of violations and the latter yields just one more; the third-best method, in terms of closeness of the number of violations to the expected number, is the HS-GARCH-EVT method. At the 99% level the HS-MGARCH, HS-MGARCH-EVT and HS-GARCH-EVT are again the best methods, although it is difficult to pick a favourite in terms of violation counts only. While HS-MGARCH gives insignificant results in every year period, the HS-MGARCH-EVT and HS-GARCH-EVT methods come closer to the expected number of overall violations; in fact, the HS-MGARCH method gives too few violations overall.

The volatile years 2007 and 2008 are problematic for most methods, with every method yielding more violations than expected. However, the VC method gives an insignificant result in both years at the 95% level, and the HS-MGARCH method gives an insignificant result in both years at the 99% level.

We now compare the results of the VaR backtests using elicitability theory. The results are contained in the first two columns of Table 9.2. According to this metric, the HS-MGARCH-EVT method gives the best VaR estimates at the 95% level, while

Table 9.1. Numbers of violations of the 95% and 99% VaR estimate calculated using various methods, as described in Section 9.3.4. Figures in italics show significant discrepancies between observed and expected violation counts according to a binomial score test at the 5% level.

Year	2005	2006	2007	2008	2009	2010	2011	2012	All
Trading days	258	257	258	259	258	259	258	258	2065
Results for 95% VaR									
Expected no. of violations	13	13	13	13	13	13	13	13	103
VC	8	16	17	19	13	15	14	14	116
HS	*0*	*6*	*28*	*49*	19	*6*	10	*1*	119
HS-GARCH	9	13	*22*	*22*	13	14	9	15	117
HS-GARCH-*t*	9	14	*23*	*22*	14	15	10	15	122
HS-GARCH-EVT	*5*	13	*22*	*21*	13	13	9	13	109
HS-MGARCH	*5*	14	*21*	19	12	9	11	12	103
HS-MGARCH-EVT	*5*	14	*22*	18	13	10	10	12	104
Results for 99% VaR									
Expected no. of violations	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.6	21
VC	2	*8*	*8*	*8*	2	4	5	*6*	*43*
HS	0	0	*10*	*22*	2	0	2	0	*36*
HS-GARCH	2	*8*	*8*	*10*	5	4	3	3	*43*
HS-GARCH-*t*	2	*8*	*6*	*8*	1	4	2	1	*32*
HS-GARCH-EVT	0	*6*	4	*7*	1	1	2	1	22
HS-MGARCH	0	4	4	5	0	1	2	1	17
HS-MGARCH-EVT	0	4	5	*6*	0	1	2	1	19

the HS-MGARCH method gives the best VaR results at the 99% level. At the 95% level the second lowest score is given by the HS-MGARCH method. At the 99% level the second lowest score is given by the HS-MGARCH-EVT method and the third lowest score is given by the HS-GARCH-EVT method.

It is also noticeable that the standard HS method gives very poor scores at both levels. As an unconditional method it is not well suited to giving estimates of quantiles of the conditional loss distribution; this is also evident from the violation counts in Table 9.1. In Figure 9.5 we show the second half of the year 2008: a period at the height of the 2007–9 credit crisis and one that includes the date on which Lehman Brothers filed for bankruptcy (15 September 2008). The plot shows actual losses as bars with risk-measure estimates for the HS and HS-MGARCH methods superimposed; violations are indicated by circles (HS) and crosses (HS-MGARCH).

Throughout the volatile year 2008, the standard historical-simulation method performs very poorly: there are forty-nine violations of the 95% VaR estimate and twenty-two violations of the 99% VaR estimate. The HS-MGARCH method, being a conditional method, is able to respond to the changes in volatility better and consequently gives nineteen and five violations. In the plot of the second half of the

Table 9.2. Scores based on elicitability theory for the VaR backtests (to four significant figures) and p-values (to two decimal places) for expected shortfall tests as described in Section 9.3.4; figures in italics indicate failure of the expected shortfall test.

	VaR score comparison		Violation residual test			
	95% VaR ($\times 10^6$)	99% VaR ($\times 10^6$)	95% ES	(n)	99% ES	(n)
VC	1081	308.1	*0.00*	116	0.05	43
HS	1399	466.4	*0.02*	119	0.25	36
HS-GARCH	1072	306.8	*0.00*	117	0.05	43
HS-GARCH-t	1074	299.6	0.12	122	0.68	32
HS-GARCH-EVT	1074	295.7	0.59	109	0.65	22
HS-MGARCH	1064	287.8	0.99	103	0.55	17
HS-MGARCH-EVT	1063	289.7	0.83	104	0.94	19

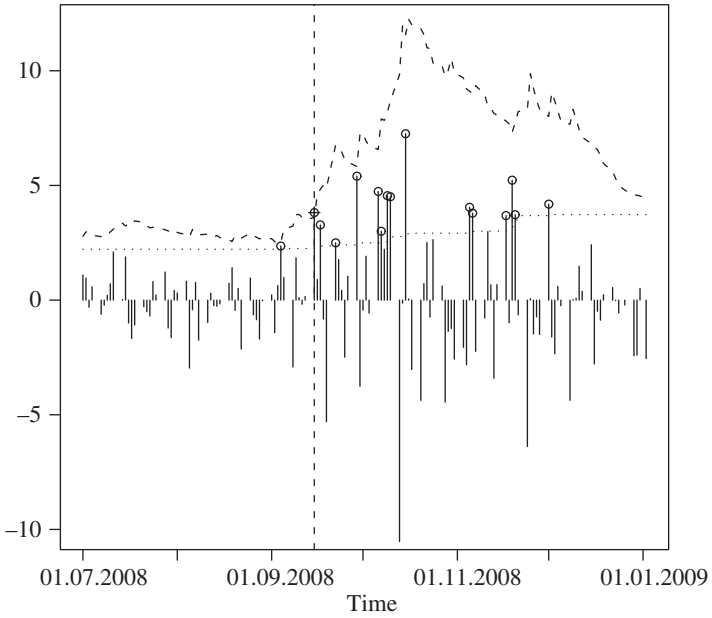

Figure 9.5. Daily losses for the second half of 2008 together with 99% VaR estimates and corresponding violations for the HS and HS-MGARCH methods. The HS VaR estimates are indicated by a dotted line and the corresponding violations are indicated by circles. The HS-MGARCH estimates are given by a dashed line; the only violation for this method in the time period occurred on 15 September 2008 (the day on which Lehman Brothers filed for bankruptcy) and is marked by a dashed vertical line and a crossed circle. For more information see Section 9.3.4.

year we see sixteen of the twenty-two violations of the 99% VaR estimate for the HS method and one of the five violations for the HS-MGARCH method.

In Figure 9.6 we address the hypothesis that VaR violations should form a Bernoulli trials process with geometrically distributed (or approximately exponentially distributed) spacings between violations. The figure shows a Q–Q plot of the

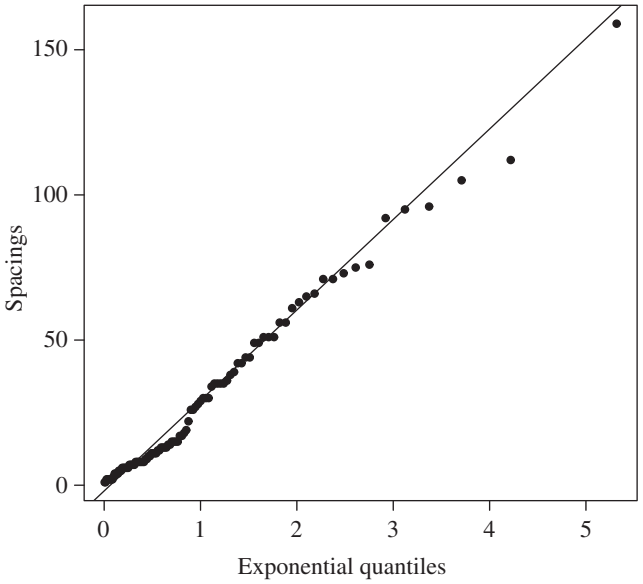

Figure 9.6. Q–Q plot of the spacings (in days) between violations of the estimates of the 99% VaR obtained by the HS-MGARCH method against the corresponding quantiles of a standard exponential distribution (see Section 9.3.4).

spacings (in days) between violations of the estimates of the 99% VaR obtained by the HS-MGARCH method against the corresponding quantiles of a standard exponential distribution; the exponential hypothesis is plausible on the basis of this picture.

In Table 9.2 we also give results for the backtest of expected shortfall based on violation residuals described in Section 9.3.2. For estimates of the 95% expected shortfall, three of the seven methods fail the zero-mean test for the violation residuals; the methods that do not fail are HS-GARCH-t, HS-GARCH-EVT, HS-MGARCH and HS-MGARCH-EVT. For estimates of the 99% expected shortfall, results appear to be even better; five of the methods give insignificant results (HS, HS-GARCH-t, HS-GARCH-EVT, HS-MGARCH, HS-MGARCH-EVT) while the other two give a marginally significant result ($p = 0.05$). However, at the 99% level we have only a small number of violation residuals to test, as indicated in the table by the column marked (n).

On the basis of all results, the HS-MGARCH and HS-MGARCH-EVT methods give the best overall performance in the example we have considered, with the HS-GARCH-EVT method also performing well.

9.3.5 Backtesting the Predictive Distribution

As well as backtesting VaR and expected shortfall we can also devise tests that assess the overall quality of the estimated conditional loss distributions from which the risk-measure estimates are derived. Of course, our primary interest focuses on the measures of tail risk, but it is still useful to backtest our estimates of the

whole predictive distribution to obtain additional confirmation of the risk-measure estimation procedure.

Suppose our objective at every time t is to estimate the conditional loss distribution $F_{L_{t+1}|\mathcal{F}_t}$ and let $U_{t+1} = F_{L_{t+1}|\mathcal{F}_t}(L_{t+1})$. For a given loss process (L_t), consider the formation of a process (U_t) by applying this transformation. As in Section 9.3.2, let us assume that the underlying process generating the losses (L_t) satisfies, for all t, equations of the form $L_t = \mu_t + \sigma_t Z_t$, where μ_t is an $\mathcal{F}_{t-1}$-measurable conditional mean term, σ_t is an $\mathcal{F}_{t-1}$-measurable volatility and the (Z_t) are SWN(0, 1) innovations.

Under this assumption it follows easily from the fact that

$$F_{L_{t+1}|\mathcal{F}_t}(l) = P(\mu_{t+1} + \sigma_{t+1} Z_{t+1} \leqslant l \mid \mathcal{F}_t) = F_Z((l - \mu_{t+1})/\sigma_{t+1})$$

that $U_{t+1} = G_Z(Z_{t+1})$, so (U_t) is a strict white noise process. Moreover, if G_Z is continuous, then the stationary or unconditional distribution of (U_t) must be standard uniform (see Proposition 7.2).

In actual applications we estimate $F_{L_{t+1}|\mathcal{F}_t}$ from data up to time t and we backtest our estimates by forming $\hat{U}_{t+1} := \hat{F}_{L_{t+1}|\mathcal{F}_t}(L_{t+1})$ on day $t + 1$. Suppose we estimate the predictive distribution on days $t = 0, \ldots, n - 1$ and form backtesting data $\hat{U}_1, \ldots, \hat{U}_n$; we expect these to behave like a sample of iid uniform data. The distributional assumption can be assessed by standard goodness-of-fit tests like the chi-squared test or the Kolmogorov–Smirnov test (see Section 15.1.2 for references). It is also possible to form the data $\Phi^{-1}(\hat{U}_1), \ldots, \Phi^{-1}(\hat{U}_n)$, where Φ is the standard normal df; these should behave like iid standard normal data (see again Proposition 7.2) and this can be tested as in Section 3.1.2. The strict white noise assumption can be tested using the approach described in Section 4.1.3.

Notes and Comments

The binomial test for numbers of VaR violations and the geometric test for the times between violations can be found in Kupiec (1995); in both cases, a likelihood ratio test is recommended. For the binomial test, alternatives are the Wald test and the score test (see, for example, Casella and Berger 2002, pp. 493–495). The score test in particular seems to give a test at about the right level for VaR probabilities $\alpha = 0.99$ and $\alpha = 0.95$ in samples of size $m = 250$ or $m = 500$ (i.e. a test with the right Type 1 error of falsely rejecting the null hypothesis of binomial behaviour). Further papers on testing VaR violations for independent Bernoulli behaviour include Christoffersen, Hahn and Inoue (2001) and Christoffersen and Pelletier (2004); the latter paper develops a number of tests of exponential behaviour for the durations between violations and finds that the likelihood ratio test against a Weibull alternative is generally most powerful for detecting clustering of violations.

There has been a large growth in papers on backtesting, and a good overview of regulatory implications is given in Embrechts et al. (2014). Our backtesting material is partly taken from McNeil and Frey (2000), where examples of the binomial test for violation counts and the test of expected shortfall using exceedance residuals can be found. In that paper a bootstrap test of the exceedance residuals is proposed as an

alternative to the simple t-test used in this chapter; see Efron and Tibshirani (1994, p. 224) for a description of the bootstrap hypothesis test. Kerkhof and Melenberg (2004) describe an econometric framework for backtesting risk-based regulatory capital. Two interesting papers on backtesting expected shortfall are Acerbi and Szekely (2014) and Costanzino and Curran (2014).

The relevance of elicitability theory to backtesting is discussed by Gneiting (2011), who also provides a proof that expected shortfall is not elicitable based on the work of Osband (1985). Further relevant papers on elicitability are Davis (2014), Ziegel (2015) and Bellini and Bignozzi (2013).

The idea of testing the estimate of the predictive distribution may be found in Berkowitz (2001, 2002). See also Berkowitz and O'Brien (2002) for a more general article on testing the accuracy of the VaR models of commercial banks. Finally, for a critical insider view on the use of VaR technology on Wall Street in the early days, see the relevant chapters in Brown (2012).

10
Credit Risk

Credit risk is the risk of a loss arising from the failure of a counterparty to honour its contractual obligations. This subsumes both default risk (the risk of losses due to the default of a borrower or a trading partner) and downgrade risk (the risk of losses caused by a deterioration in the credit quality of a counterparty that translates into a downgrading in some rating system). Credit risk is omnipresent in the portfolio of a typical financial institution. To begin with, the lending and corporate bond portfolios are obviously affected by credit risk. Perhaps less obviously, credit risk accompanies any over-the-counter (OTC, i.e. non-exchange-guaranteed) derivative transaction such as a swap, because the default of one of the parties involved may substantially affect the actual pay-off of the transaction. Moreover, there is a specialized market for credit derivatives, such as credit default swaps, in which financial institutions are active players. Credit risk therefore relates to the core activities of most banks. It is also highly relevant to insurance companies, who are exposed to substantial credit risk in their investment portfolios and counterparty default risk in their reinsurance treaties.

The management of credit risk at financial institutions involves a range of tasks. To begin with, an enterprise needs to determine the capital it should hold to absorb losses due to credit risk, for both regulatory and economic capital purposes. It also needs to manage the credit risk on its balance sheet. This involves ensuring that portfolios of credit-risky instruments are well diversified and that portfolios are optimized according to risk–return considerations. The risk profile of the portfolio can also be improved by hedging risk concentrations with credit derivatives or by transferring risk to investors through securitization. Moreover, institutions need to manage their portfolio of traded credit derivatives. This involves the tasks of pricing, hedging and managing collateral for such trades. Finally, financial institutions need to control the counterparty credit risk in their trades and contracts with other institutions. In fact, in the aftermath of the 2007–9 financial crisis, counterparty risk management became one of the most important issues for financial institutions.

With these tasks in mind we have split our treatment of credit risk into four chapters. In the present chapter we establish the foundations for the analysis of credit risk. We introduce the most common credit-risky instruments (Section 10.1), discuss various measures of credit quality (Section 10.2) and present models for the credit risk of a single firm (Sections 10.3–10.6). Moreover, we study basic single-name credit derivatives such as credit default swaps.

Chapters 11 and 12 are concerned with portfolio models, and the crucial issue of dependence between defaults comes to the fore. Chapter 11 treats one-period models with a view to capital adequacy and credit risk management issues for portfolios of largely non-traded assets. Chapter 12 deals with properties of portfolio credit derivatives such as collateralized debt obligations; moreover, we discuss the valuation of these products in standard copula models. Finally, Chapter 17 is concerned with more advanced fully dynamic models of portfolio credit risk. This chapter is also the natural place for a detailed discussion of counterparty credit risk because a proper analysis requires dynamic multivariate credit risk models.

Credit risk models can be divided into *structural* or *firm-value models* on the one hand and *reduced-form models* on the other. Broadly speaking, in a structural model default occurs when a stochastic variable (or, in dynamic models, a stochastic process), generally representing an asset value, falls below a threshold, generally representing liabilities. In reduced-form models the precise mechanism leading to default is left unspecified and the default time of a firm is modelled as a non-negative rv, whose distribution typically depends on economic covariables. In this chapter we treat structural models in Section 10.3, simple reduced-form models with deterministic hazard rates in Section 10.4 and more advanced reduced-form models in Sections 10.5 and 10.6.

10.1 Credit-Risky Instruments

In this section we give an overview of the universe of credit-risky instruments, starting with the simplest examples of loans and bonds. We include discussion of the counterparty credit risk in OTC derivatives trades and we also describe some of the more common modern credit derivative products. In what follows we often use the generic term *obligor* for the borrower, bond issuer, trading partner or counterparty to whom there is a credit exposure. The name stems from the fact that in all cases the obligor has a contractual obligation to make certain payments under certain conditions.

10.1.1 Loans

Loans are the oldest credit-risky "instruments" and come in a myriad of forms. It is common to categorize them according to the type of obligor into *retail* loans (to individuals and small or medium-sized companies), *corporate* loans (to larger companies), *interbank* loans and *sovereign* loans (to governments). In each of these categories there are likely to be a number of different lending products. For example, retail customers may borrow money from a bank using mortgages against property, credit cards and overdrafts.

The common feature of most loans is that a sum of money, known as the *principal*, is advanced to the borrower for a particular term in exchange for a series of defined interest payments, which may be at fixed or floating interest rates. At the end of the term the borrower is required to pay back the principal.

A useful distinction to make is between *secured* and *unsecured* lending. If a loan is secured, the borrower has pledged an asset as collateral for the loan. A prime

example is a mortgage, where the collateral is a property. In the event that the borrower is unable to fulfill its obligation to make interest payments or repay the principal, a situation that is termed *default*, the lender may take possession of the asset. In this way the loss in the event of default may be partly mitigated and money may be recovered by selling the asset. In an unsecured loan the lender has no such claim on a collateral asset and recoveries in the event of default may be a smaller fraction of the so-called *exposure*, which is the value of the outstanding principal and interest payments.

Unlike bonds, which are publicly traded securities, loans are private agreements between the borrower and the lender. Hence there is a wide variety of different loan contracts with different legal features. This makes loans difficult to value under fair-value principles. Book value is commonly used, and where fair-value approaches are applied these mostly fall under the heading of level 3 valuation (see Section 2.2.2).

10.1.2 Bonds

Bonds are publicly traded securities issued by companies and governments that allow the issuer to raise funding on financial markets. Bonds issued by companies are called *corporate bonds* and bonds issued by governments are known as *treasuries*, *sovereign bonds* or, particularly in the UK, *gilts* (gilt-edged securities).

The structure of the pay-offs is akin to that of a loan. The security commits the bond issuer (borrower) to make a series of interest payments to the bond buyer (lender) and pay back the principal at a fixed maturity. The interest payments, or coupons, may be *fixed* at the issuance of the bond (so-called fixed-coupon bonds). Alternatively, there are also bonds where the interest payments vary with market rates (so-called *floating-rate notes*). The reference for the floating rate is often LIBOR (the London Interbank Offered Rate). There are also *convertible bonds*, which allow the purchaser to convert them into shares of the issuing company at predetermined time points. These typically offer lower rates than conventional corporate bonds because the investor is being offered the option to participate in the future growth of the company.

A bondholder is subject to a number of risks, particularly interest-rate risk, spread risk and default risk. As for loans, default risk is the risk that promised coupon and principal payments are not made. Historically, government bonds issued by developed countries have been considered to be default free; for obvious reasons, after the European debt crisis of 2010–12, this notion was called into question.

Spread risk is a form of market risk that refers to changes in *credit spreads*. The credit spread of a defaultable bond measures the difference in the yield of the bond and the yield of an equivalent default-free bond (see Section 10.3.2 for a formal definition of credit spreads). An increase in the spread of a bond means that the market value of the bond falls, which is generally interpreted as indicating that the financial markets perceive an increased default risk for the bond.

10.1.3 Derivative Contracts Subject to Counterparty Risk

A significant proportion of all derivative transactions is carried out over the counter, and there is no central clearing counterparty such as an organized exchange that guarantees the fulfilment of the contractual obligations. These trades are therefore subject to the risk that one of the contracting parties defaults during the transaction, thus affecting the cash flows that are actually received by the other party. This risk, known as *counterparty credit risk*, received a lot of attention during the financial crisis of 2007–9, as some of the institutions heavily involved in derivative transactions experienced worsening credit quality or—in the case of Lehman Brothers—even a default event. Counterparty credit risk management is now a key issue for all financial institutions and is the focus of many new regulatory developments.

In order to illustrate the challenges in measuring and managing counterparty credit risk, we consider the example of an interest swap. This is a contract where two parties A and B agree to exchange a series of interest payments on a given nominal amount of money for a given period. For concreteness assume that A receives payments at a fixed interest rate and makes floating payments at a rate equal to the three-month LIBOR.

Suppose now that A defaults at time τ_A before the maturity of the contract, so that the contract is settled at that date. The consequences will depend on the value of the remaining interest payments at that point in time. If interest rates have risen relative to their value at inception of the contract, the fixed interest payments have decreased in value so that the value of the swap contract has increased for B. Since A is no longer able to fulfill its obligations, its default constitutes a loss for B; the exact size of the loss will depend on the term structure of interest rates at the default time τ_A. On the other hand, if interest rates have fallen relative to their value at $t = 0$, the fixed payments have increased in value so that the swap has a negative value for B. At settlement, B will still have to pay the value of the contract into the bankruptcy pool, so that there is no upside for B in A's default. If B defaults first, the situation is reversed: falling rates lead to a counterparty-risk-related loss for A. This simple example illustrates two important points: the size of the counterparty credit exposure is not known a priori, and it is not even clear who has the credit exposure.

The management of counterparty risk raises a number of issues. First, counterparty risk has to be taken into account in pricing and valuation. This has led to various forms of credit value adjustment (CVA). Second, counterparty risk needs to be controlled using risk-mitigation techniques such as netting and collateralization. Under a netting agreement, the value of all derivatives transactions between A and B is computed and only the aggregated value is subject to counterparty risk; since offsetting transactions cancel each other out, this has the potential to reduce counterparty risk substantially. Under a collateralization agreement, the parties exchange collateral (cash and securities) that serve as a pledge for the receiver. The value of the collateral is adjusted dynamically to reflect changes in the value of the underlying transactions.

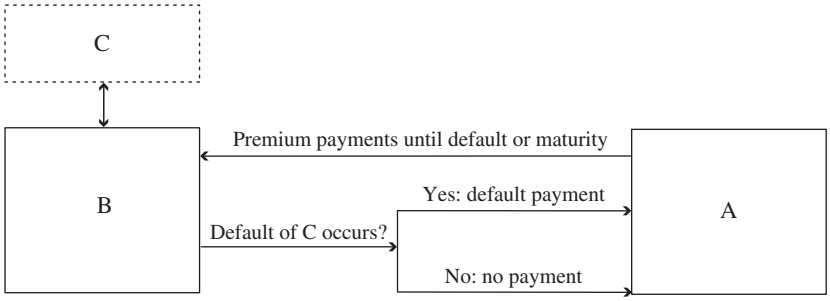

Figure 10.1. The basic structure of a CDS. Firm C is the reference entity, firm A is the protection buyer, and firm B is the protection seller.

The proper assessment of counterparty risk requires a joint modelling of the default times of the two counterparties (A and B) and of the price dynamics of the underlying derivative contract. For that reason we defer the detailed discussion of this topic to Chapter 17.

10.1.4 Credit Default Swaps and Related Credit Derivatives

Credit derivatives are securities that are primarily used for the hedging and trading of credit risk. In contrast to the products considered so far, the promised pay-off of a credit derivative is related to credit events affecting one or more firms. Major participants in the market for credit derivatives are banks, insurance companies and investment funds. Retail banks are typically net buyers of protection against credit events; other investors such as hedge funds and investment banks often act as both sellers and buyers of credit protection.

Credit default swaps. Credit default swaps (CDSs) are the workhorses of the credit derivatives market, and the market for CDSs written on larger corporations is fairly liquid; some numbers on the size of the market are given in Notes and Comments. The basic structure of a CDS is depicted in Figure 10.1. A CDS is a contract between two parties, the *protection buyer* and the *protection seller*. The pay-offs are related to the default of a reference entity (a financial firm or sovereign issuing bonds).

If the reference entity experiences a default event before the maturity date T of the contract, the protection seller makes a default payment to the protection buyer, which mimics the loss due to the default of a bond issued by the reference entity (the reference asset); this part of a CDS is called the *default payment leg*. In this way the protection buyer has acquired financial protection against the loss on the reference asset he would incur in case of a default. As compensation, the protection buyer makes periodic premium payments (typically quarterly or semiannually) to the protection seller (the *premium payment leg*); after the default of the reference entity, premium payments stop. There is no initial payment. The premium payments are quoted in the form of an annualized percentage x^* of the notional value of the reference asset; x^* is termed the (fair or market-quoted) *CDS spread*. For a mathematical description of the payments, see Section 10.4.4.

There are a number of technical and legal issues in the specification of a CDS. In particular, the parties have to agree on the precise definition of a default event and on a procedure to determine the size of the default payment in case a default event of the reference entity occurs. Due to the efforts of bodies such as the International Swaps and Derivatives Association (ISDA), some standardization of these issues has taken place.

Investors enter into CDS contracts for various reasons. To begin with, bond investors with a large credit exposure to the reference entity may buy CDS protection to insure themselves against losses due to the default of a bond. This may be easier than reducing the original bond position because CDS markets are often more liquid than bond markets. Moreover, CDS positions are quickly settled.

CDS contracts are also held for speculative reasons. In particular, so-called *naked* CDS positions, where the protection buyer does not own the bond, are often assumed by investors who are speculating on the widening of the credit spread of the reference entity. These positions are similar to short-selling bonds issued by the reference entity. Note that, in contrast to insurance, there is no requirement for the protection buyer to have *insurable interest*, that is, to actually own a bond issued by the reference entity. The speculative motive for holding CDSs is at least as important as the insurance motive.

There has been some debate about the risks of the CDS market, particularly with respect to the large volume of naked positions and whether or not these should be limited. By taking naked CDS positions speculators can depress the prices of the bonds issued by the reference entity so that default becomes a self-fulfilling prophecy. The debate about the pros and cons of limiting naked CDS positions is akin to the debate about the pros and cons of limiting short selling on equity markets.

A CDS is traded over the counter and is not guaranteed by a clearing house. A CDS position can therefore be subject to a substantial amount of counterparty risk, particularly if a trade is backed by insufficient collateral. A case in point arose during the credit crisis when AIG, which had sold many protection positions, had to be bailed out by the US government to prevent the systemic consequences of allowing it to default on its CDS contracts. There is concern that CDS markets have created a new form of dependency across financial institutions so that the default of one large (systemically important) institution could create a cascade of defaults across the financial sector due to counterparty risk.

On the other hand, CDSs are useful risk-management tools. Because of the liquidity of CDS markets, CDSs are the natural underlying security for many more complex credit derivatives. Models for pricing portfolio-related credit derivatives are usually calibrated to quoted CDS spreads. With improved collateral management in CDS markets it has been argued that the potential for CDS markets to create large-scale default contagion has been substantially reduced (see Notes and Comments).

Credit-linked notes. A credit-linked note is a combination of a credit derivative and a coupon bond that is sold as a fixed package. The coupon payments (and sometimes also the repayment of the principal) are reduced if a third party (the reference entity)

experiences a default event during the lifetime of the contract, so the buyer of a credit-linked note is providing credit protection to the issuer of the note.

Credit-linked notes are issued essentially for two reasons. First, from a legal point of view, a credit-linked note is treated as a fixed-income investment, so that investors who are unable to enter into a transaction involving credit derivatives directly (such as life insurance companies) may nonetheless sell credit protection by buying credit-linked notes. Second, an investor buying a credit-linked note pays the price up front, so that the protection buyer (the issuer of the credit-linked note) is protected against losses caused by the default of the protection seller.

10.1.5 PD, LGD and EAD

Regardless of whether we make a loan, buy a defaultable bond, engage in an OTC derivatives transaction, or act as protection seller in a CDS, the risk of a credit loss is affected by three, generally related, quantities: the *exposure at default* (EAD), the *probability of default* (PD) and the *loss given default* (LGD) or, equivalently, the size of the recovery in the event of default. They are key inputs to the Basel formula in the internal-ratings-based (IRB) approach to determining capital requirements for credit-risky portfolios, so it is important to consider them.

Exposure at default. If we make a loan or buy a bond, our exposure is relatively easy to determine, since it is mainly the principal that is at stake. However, there is some additional uncertainty about the value of the interest payments that could be lost. A further source of exposure uncertainty is due to the widespread use of credit lines. Essentially, a credit line is a ceiling up to which a corporate client can borrow money at given terms, and it is up to the borrower to decide which part of the credit line he actually wants to use. For OTC derivatives, the counterparty risk exposure is even more difficult to quantify, since it is a stochastic variable depending on the unknown time at which a counterparty defaults and the evolution of the value of the derivative up to that point; a case in point is the example of an interest rate swap discussed in Subsection 10.1.3.

In practice, the concept that is used to describe exposure is exposure at default or EAD, which recognizes that the exposure for many instruments will depend on the exact default time. In counterparty credit risk the use of collateral can also reduce the exposure and thus mitigate losses.

Probability of default. When measuring the risk of losses over a fixed time horizon, e.g. one year, we are particularly concerned with estimating the probability that obligors default by the time horizon, a quantity known to practitioners as the probability of default, or PD. The PD is related to the credit quality of an obligor, and Sections 10.2 and 10.3 discuss some of the models that are used to quantify default risk. For instruments where the loss is dependent on the exact timing of default, e.g. OTC derivatives with counterparty risk, the risk of default is described by the whole distribution of possible default times and not just the probability of default by a fixed horizon.

Loss given default. In the event of default, it is unlikely that the entire exposure is lost. For example, when a mortgage holder defaults on a residential mortgage, and there is no realistic possibility of restructuring the debt, the lender can sell the property (the collateral asset) and the proceeds from the sale will make good some of the lost principal. Similarly, when a bond issuer goes into administration, the bondholders join the group of creditors who will be partly recompensed for their losses by the sale of the firm's assets.

Practitioners use the term loss given default, or LGD, to describe the proportion of the exposure that is actually lost in the event of default, or its converse, the recovery, to describe the amount of the exposure that can be recovered through debt restructuring and asset sales.

Dependence of these quantities. It is important to realize that EAD, PD and LGD are dependent quantities. While it is common to attempt to model them in terms of independent random variables, it is unrealistic to do so. For example, in a period of financial distress, when PDs are high, the asset values of firms are depressed and firms are defaulting, recoveries are likely to be correspondingly low, so that there is positive dependence between PDs and LGDs. This will be discussed further in 11.2.3.

Notes and Comments

For further reading on loans and loan pricing we refer to Benzschawel, Dagraca and Fok (2010). For an overview of bonds see Sharpe, Alexander and Bailey (1999).

To get an idea of the size of the CDS market, note that the nominal value (gross notional amount) of the market stood at approximately $60 trillion by the end of 2007, before coming down to a still considerable amount of approximately $25 trillion by the end of 2012. In 2013 the net notional amount was of the order of $2 trillion. For comparison, by the end of 2012 world GDP stood at roughly $80 trillion. To give an example of the size of the speculative market in CDSs, Cont (2010) reports that "when it filed for bankruptcy on September 14, 2008, Lehman Brothers had $155 billion of outstanding debt, but more than $400 billion notional value of CDS contracts had been written with Lehman as reference entity". A good discussion of the role of such credit derivatives in the credit crisis is given in Stulz (2010). The effect of improved collateral management for CDSs on the risk of large-scale contagion in CDS markets is addressed in Brunnermeier, Clerc and Scheicher (2013).

In this brief introduction we have discussed a few essential features of credit derivatives but have omitted the rather involved regulatory, legal and accounting issues related to these instruments. Readers interested in these topics are referred to the paper collections edited by Gregory (2003) and Perraudin (2004), in which pricing issues are also discussed. An excellent treatment of credit derivatives at textbook level is Schönbucher (2003). For a discussion of credit derivatives from the viewpoint of financial engineering we refer to Neftci (2008).

10.2 Measuring Credit Quality

There are various ways of quantifying the credit quality or, equivalently, the default risk of obligors but, broadly speaking, these approaches may be divided into two philosophies. On the one hand, credit quality can be described by a credit rating or credit score that is based on empirical data describing the borrowing and repayment history of the obligor, or of similar obligors. On the other hand, for obligors whose equity is traded on financial markets, prices can be used to infer the market's view of the credit quality of the obligor. This section is devoted to the first philosophy, and market-implied measures of credit quality are treated in the context of structural models in Section 10.3.

Credit ratings and credit scores fulfill a similar function—they both allow us to order obligors according to their credit risk and map that risk to an estimate of default probability. Credit ratings tend to be expressed on an ordered categorical scale, whereas credit scores are often expressed in terms of points on a metric scale. The task of rating obligors, particularly large corporates or sovereigns, is often outsourced to a rating agency such as Moody's or Standard & Poor's (S&P); proprietary rating systems internal to a financial institution can also be used. In the S&P rating system there are seven pre-default rating categories, labelled AAA, AA, A, BBB, BB, B, CCC, with AAA being the highest rating and CCC the lowest rating; Moody's uses nine pre-default rating categories and these are labelled Aaa, Aa, A, Baa, Ba, B, Caa, Ca, C. A finer alpha-numeric system is also used by both agencies.

Credit scores are traditionally used for retail customers and are based on so-called scorecards that banks develop through extensive statistical analyses of historical data. The basic idea is that default risk is modelled as a function of demographic, behavioural and financial covariates that describe the obligor. Using techniques such as logistic regression these covariates are weighted and combined into a score.

10.2.1 Credit Rating Migration

In the credit-migration approach each firm is assigned to a credit-rating category at any given time point. The probability of moving from one credit rating to another over a given risk horizon (typically one year) is then specified. Transition probabilities are typically presented in the form of a matrix; an example from Moody's is presented in Table 10.1. Transition matrices are estimated from historical default data, and standard statistical methods used for this purpose are discussed in Section 10.2.2

In the credit-migration approach we assume that the current credit rating completely determines the default probability, so that this probability can be read from the transition matrix. For instance, if we use the transition matrix presented in Table 10.1, we obtain a one-year default probability for an A-rated company of 0.06%, whereas the default probability of a Caa-rated company is 13.3%. In practice, a correction to the figures in Table 10.1 would probably be undertaken to account for rating withdrawals: that is, transitions to the WR state. The simplest correction would be to divide the first nine probabilities in each row of the table by one minus the final probability in that row; this implicitly assumes that the act of rating withdrawal

Table 10.1. Probabilities of migrating from one rating quality to another within one year. "WR" represents the proportion of firms that were no longer rated at the end of the year, for various reasons including takeover by another company. Source: Ou (2013, Exhibit 26).

Initial rating	Rating at year-end (%)									
	Aaa	Aa	A	Baa	Ba	B	Caa	Ca–C	Default	WR
Aaa	87.20	8.20	0.63	0.00	0.03	0.00	0.00	0.00	0.00	3.93
Aa	0.91	84.57	8.43	0.49	0.06	0.02	0.01	0.00	0.02	5.48
A	0.06	2.48	86.07	5.47	0.57	0.11	0.03	0.00	0.06	5.13
Baa	0.039	0.17	4.11	84.84	4.05	7.55	1.63	0.02	0.17	5.65
Ba	0.01	0.05	0.35	5.52	75.75	7.22	0.58	0.07	1.06	9.39
B	0.01	0.03	0.11	0.32	4.58	73.53	5.81	0.59	3.85	11.16
Caa	0.01	0.02	0.02	0.12	0.38	8.70	61.71	3.72	13.34	12.00
Ca–C	0.00	0.00	0.00	0.00	0.40	2.03	9.38	35.46	37.93	14.80

Table 10.2. Average cumulative default rates (%). Source: Ou (2013, Exhibit 33).

Initial rating	Term						
	1	2	3	4	5	10	15
Aaa	0.00	0.01	0.01	0.04	0.11	0.50	0.93
Aa	0.02	0.07	0.14	0.26	0.38	0.92	1.75
A	0.06	0.20	0.41	0.63	0.87	2.48	4.26
Baa	0.18	0.50	0.89	1.37	1.88	4.70	8.62
Ba	1.11	3.08	5.42	7.93	10.18	19.70	29.17
B	4.05	9.60	15.22	20.13	24.61	41.94	52.22
Caa–C	16.45	27.87	36.91	44.13	50.37	69.48	79.18

contains no information about the likelihood of upgrade, downgrade or default of an obligor.

Rating agencies also produce cumulative default probabilities over larger time horizons. In Table 10.2 we reproduce Moody's cumulative default probabilities for companies with a given current credit rating. For instance, according to this table the probability that a company whose current credit rating is Baa defaults within the next four years is 1.37%. These cumulative default probabilities have been estimated directly from default data. Alternative estimates of multi-year default probabilities can be inferred from one-year transition matrices, as explained in more detail in the next section.

Remark 10.1 (accounting for business cycles). It is a well-established empirical fact that default rates tend to vary with the state of the economy, being high during recessions and low during periods of economic expansion (see Figure 10.2 for an illustration). Transition rates as estimated by S&P or Moody's, on the other hand, are historical averages over longer time horizons covering several business cycles. For instance, the transition rates in Table 10.1 have been estimated from rating-migration data over the period 1970–2012. Moreover, rating agencies focus on the

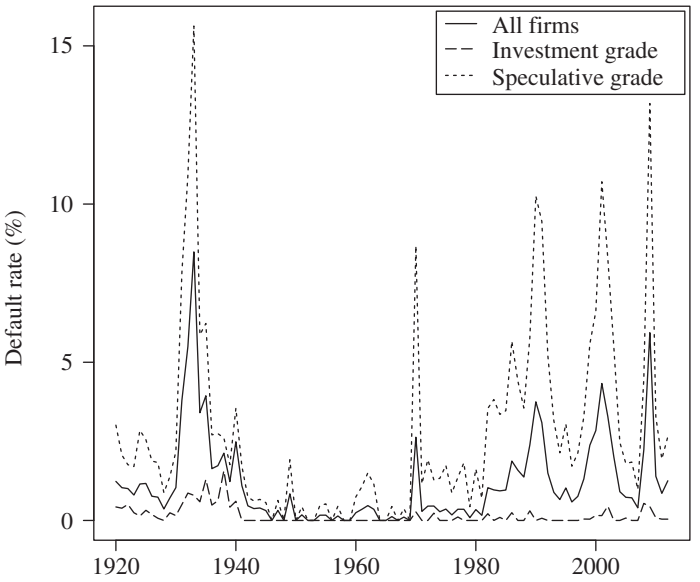

Figure 10.2. Moody's annual default rates from 1920 to 2012. Source for data: Ou (2013, Exhibit 30).

average credit quality "through the business cycle" when attributing a credit rating to a particular firm. The default probabilities from the credit-migration approach are therefore estimates for the average default probability, independent of the current economic environment. In some situations we are interested in "point-in-time" estimates of default probabilities reflecting the current macroeconomic environment, such as in the pricing of a short-term loan. In these situations adjustments to the long-term average default probabilities from the credit-migration approach can be made; for instance, we could use equity prices as an additional source of information, as is done in the public-firm EDF (expected default frequency) model discussed in Section 10.3.3.

10.2.2 Rating Transitions as a Markov Chain

Let (R_t) denote a discrete-time stochastic process defined at times $t = 0, 1, \ldots$ that takes values in $S = \{0, 1, \ldots, n\}$. The set S defines rating states of increasing creditworthiness, with 0 representing default. (R_t) models the evolution of an obligor's rating over time.

We will assume that (R_t) is a *Markov chain*. This means that conditional transition probabilities satisfy the Markov property

$$P(R_t = k \mid R_0 = r_0, R_1 = r_1, \ldots, R_{t-1} = j) = P(R_t = k \mid R_{t-1} = j)$$

for all $t \geqslant 1$ and all $j, r_0, r_1, \ldots, r_{t-2}, k \in S$. In words, the conditional probabilities of rating transitions, given an obligor's rating history, depend only on the previous rating $R_{t-1} = j$ at the last time point and not on the more distant history of how the obligor arrived at a rating state j at time $t - 1$.

The Markov assumption for rating migrations has been criticized; there is evidence for both *momentum* and *stickiness* in empirical rating histories (see Lando and Skodeberg 2002). Momentum is the phenomenon by which obligors who have been recently downgraded to a particular rating are more likely to experience further downgrades than obligors who have had the same rating for a long time. Stickiness is the converse phenomenon by which rating agencies are initially hesitant to downgrade obligors until the evidence for credit deterioration is overwhelming. But despite these issues, the Markov chain assumption is very widely made, because it leads to tractable models with a well-understood theory and to natural estimators for transition probabilities.

The Markov chain is *stationary* if

$$P(R_t = k \mid R_{t-1} = j) = P(R_1 = k \mid R_0 = j)$$

for all $t \geqslant 1$ and all rating states j and k. In this case we can define the transition matrix to be the $(n+1) \times (n+1)$ matrix $\boldsymbol{P} = (p_{jk})$ with elements $p_{jk} = P(R_t = k \mid R_{t-1} = j)$ for any $t \geqslant 1$. Simple conditional probability arguments can be used to derive the *Chapman–Kolmogorov equations*, which say that for any $t \geqslant 2$, and any $j, k \in S$,

$$\begin{aligned} P(R_t = k \mid R_{t-2} = j) &= \sum_{l \in S} P(R_t = k \mid R_{t-1} = l) P(R_{t-1} = l \mid R_{t-2} = j) \\ &= \sum_{l \in S} p_{lk} p_{jl}. \end{aligned}$$

An implication of this is that the matrix of transition probabilities over two time steps is given by $\boldsymbol{P}^2 = \boldsymbol{P} \times \boldsymbol{P}$. Similarly, the matrix of transition probabilities over T time periods is $\boldsymbol{P}^T$. It is, however, not clear how we would compute a matrix of transition probabilities for a fraction of a time period. In fact, this requires the notion of a Markov chain in *continuous time*, which is discussed below.

We now turn to the problem of estimating $\boldsymbol{P}$. Suppose we observe, or are given information about, the ratings of companies at the time points $0, 1, \ldots, T$. This information usually relates to a fluctuating population or cohort of companies, with only a few having complete rating histories throughout $[0, T]$: new companies may be added to the cohort at any time; some companies may default and leave the cohort; others may have their rating withdrawn. In the latter case we will assume that the withdrawal of rating occurs independently of the default or rating-migration risk of the company (which may not be true).

For $t = 0, \ldots, T-1$ and $j \in S \setminus \{0\}$, let N_{tj} denote the number of companies that are rated j at time t and for which a rating is available at time $t+1$; let N_{tjk} denote the subset of those companies that are rated k at time $t+1$. Under the discrete-time, homogeneous Markovian assumption, independent multinomial experiments effectively take place at each time t. In each experiment the N_{tj} companies rated j can be thought of as being randomly allocated to the ratings $k \in S$ according to probabilities p_{jk} that satisfy $\sum_{k=0}^{n} p_{jk} = 1$.

In this framework the likelihood is given by

$$L((p_{jk}); (N_{tj}), (N_{tjk})) = \prod_{t=0}^{T-1} \left(\prod_{j=1}^{n} \left(N_{tj}! \prod_{k=0}^{n} \frac{p_{jk}^{N_{tjk}}}{N_{tjk}!} \right) \right),$$

and if this is maximized subject to the constraints that $\sum_{k=0}^{n} p_{jk} = 1$ for $j = 1, \ldots, n$, we obtain the maximum likelihood estimator

$$\hat{p}_{jk} = \frac{\sum_{t=0}^{T-1} N_{tjk}}{\sum_{t=0}^{T-1} N_{tj}}. \tag{10.1}$$

There are a number of drawbacks to modelling rating transitions as a discrete-time Markov chain. In practice, rating changes tend to take place on irregularly spaced dates. While such data can be approximated by a regularly spaced time series (or panel) of, say, yearly, quarterly or monthly ratings, there will be a loss of information in doing so. The discrete-time model described above would ignore any information about intermediate transitions taking place between two times t and $t + 1$. For example, if an obligor is downgraded from A to BBB to BB over the course of the period $[t, t+1]$, this obligor will simply be recorded as migrating from A to BB and the information about transitions from A to BBB and BBB to BB will not be recorded. Moreover, the estimation procedure for a discrete-time chain tends to result in sparse estimates of transition matrices with quite a lot of zero entries. For example, if no transitions between AAA and default within a single time period are observed, then the probability of such a transition will be estimated to be zero. However, in reality such a transition is possible, if unlikely, and so its estimated probability of occurrence should not be zero.

It is thus more satisfactory to model rating transitions as a phenomenon in continuous time. In this case, transition probabilities are not modelled directly but are instead given in terms of transition rates. Intuitively, the relationship between transition rates and transition probabilities can be described as follows.

Assume that over any small time step of duration δt the probability of a transition from rating j to k is given approximately by $\lambda_{jk}\delta t$ for some constant $\lambda_{jk} > 0$, which is the *transition rate* between rating j and rating k. The probability of staying at rating j is given by $1 - \sum_{k \neq j} \lambda_{jk}\delta t$. If we define a matrix Λ to have off-diagonal entries λ_{jk} and diagonal entries $-\sum_{k \neq j} \lambda_{jk}$, we can summarize the implied transition probabilities for the small time step δt in the matrix $(I_{n+1} + \Lambda\delta t)$. We now consider transitions in the period $[0, t]$ and denote the corresponding matrix of transition probabilities by $\boldsymbol{P}(t)$. If we divide the time period into N small time steps of size $\delta t = t/N$ for N large, the matrix of transition probabilities can be approximated by

$$\boldsymbol{P}(t) \approx \left(I_{n+1} + \frac{\Lambda t}{N} \right)^N,$$

which converges, as $N \to \infty$, to the so-called matrix exponential of Λt:

$$\boldsymbol{P}(t) = \mathrm{e}^{\Lambda t}.$$

This formulation gives us a method of computing transition probabilities for any time horizon t in terms of the matrix Λ, the so-called *generator matrix*.

A Markov chain in continuous time with generator matrix Λ can be constructed in the following way. An obligor remains in rating state j for an exponentially distributed amount of time with parameter

$$\lambda_{jj} = \sum_{k \neq j} \lambda_{jk},$$

i.e. minus the diagonal element of the generator matrix. When a transition takes place the new rating is determined by a multinomial experiment in which the probability of a transition from state j to state k is given by $\lambda_{jk}/\lambda_{jj}$.

This construction also leads to natural estimators for the matrix Λ. Since λ_{jk} is the instantaneous rate of migrating from j to k, we can estimate it by

$$\hat{\lambda}_{jk} = \frac{N_{jk}(T)}{\int_0^T Y_j(t)\,\mathrm{d}t}, \tag{10.2}$$

where $N_{jk}(T)$ is the total number of observed transitions from j to k over the time period $[0, T]$ and $Y_i(t)$ is the number of obligors with rating j at time t; the denominator therefore represents the total time spent in state j by all the companies in the data set. Note that this is the continuous-time analogue of the maximum likelihood estimator in (10.1); it is not surprising, therefore, that (10.2) can be shown to be the maximum likelihood estimator for the transition intensities of a homogenous continuous-time Markov chain.

Notes and Comments

There is a large literature on credit scoring, and useful starter references are Thomas (2009) and Hand and Henley (1997). In addition to the well-known commercial rating agencies there are now open rating systems. One example is the Credit Research Initiative at the Risk Management Institute of the National University of Singapore (see www.rmicri.org).

An alternative discussion of models based on rating migration is given in Chapters 7 and 8 of Crouhy, Galai and Mark (2001). Statistical approaches to the estimation of rating-transition matrices are discussed in Hu, Kiesel and Perraudin (2002) and Lando and Skodeberg (2002). The latter paper also shows that there is momentum in rating-transition data, which contradicts the assumption that rating transitions form a Markov chain. An example of an industry model based on credit ratings is CreditMetrics: see RiskMetrics Group (1997).

The literature on the statistical properties of rating transitions is surveyed extensively in Chapter 4 of Duffie and Singleton (2003). The maximum likelihood estimator of the infinitesimal generator of a continuous-time Markov chain is formally derived in Albert (1962). For further information on Markov chains we refer to the standard textbook by Norris (1997).

10.3 Structural Models of Default

In structural or firm-value models of default one postulates a mechanism for the default of a firm in terms of the relationship between its assets and liabilities. Typically, default occurs whenever a stochastic variable (or in dynamic models a stochastic process) generally representing an asset value falls below a threshold representing liabilities. The kind of thinking embodied in these models has been very influential in the analysis of credit risk and in the development of industry solutions, so that this is a natural starting point for a discussion of credit risk models. We begin with a detailed analysis of the seminal model of Merton (1974) (in Sections 10.3.1 and 10.3.2). Industry implementations of structural models are discussed in Section 10.3.3.

From now on we denote a generic stochastic process in continuous time by (X_t); the value of the process at time $t \geqslant 0$ is given by the rv X_t.

10.3.1 The Merton Model

The model proposed in Merton (1974) is the prototype of all firm-value models. Consider a firm whose asset value follows some stochastic process (V_t). The firm finances itself by *equity* (i.e. by issuing shares) and by *debt*. In Merton's model, debt consists of zero-coupon bonds with common maturity T; the nominal value of debt at maturity is given by the constant B. Moreover, it is assumed that the firm cannot pay out dividends or issue new debt.

The values at time t of equity and debt are denoted by S_t and B_t. Default occurs if the firm misses a payment to its debtholders, which in the Merton model can occur only at the maturity T of the bonds. At T we have to distinguish between two cases.

(i) $V_T > B$: the value of the firm's assets exceeds the nominal value of the liabilities. In that case the debtholders (the owners of the zero-coupon bonds) receive B, the shareholders receive the residual value $S_T = V_T - B$, and there is no default.

(ii) $V_T \leqslant B$: the value of the firm's assets is less than its liabilities and the firm cannot meet its financial obligations. In that case shareholders have no interest in providing new equity capital, as these funds would go immediately to the bondholders. They therefore let the firm go into default. Control over the firm's assets is passed on to the bondholders, who liquidate the firm and distribute the proceeds among themselves. Shareholders pay and receive nothing, so that we have $B_T = V_T$, $S_T = 0$.

Summarizing, we have the relationships

$$S_T = \max(V_T - B, 0) = (V_T - B)^+, \tag{10.3}$$

$$B_T = \min(V_T, B) = B - (B - V_T)^+. \tag{10.4}$$

Equation (10.3) implies that the value of the firm's equity at time T equals the payoff of a European call option on V_T, while (10.4) implies that the value of the firm's

debt at maturity equals the nominal value of the liabilities minus the pay-off of a European put option on V_T with exercise price equal to B.

This model is of course a stylized description of default. In reality, the structure of a company's debt is much more complex, so that default can occur on many different dates. Moreover, under modern bankruptcy code, default does not automatically imply bankruptcy, i.e. liquidation of a firm. Nonetheless, Merton's model is a useful starting point for modelling credit risk and for pricing securities subject to default.

Remark 10.2. The option interpretation of equity and debt is useful in explaining potential conflicts of interest between the shareholders and debtholders of a company. It is well known that, all other things being equal, the value of an option increases if the volatility of the underlying security is increased. Shareholders therefore have an interest in the firm taking on risky projects. Bondholders, on the other hand, have a short position in a put option on the firm's assets and would therefore like to see the volatility of the asset value reduced.

In the Merton model it is assumed that under the real-world or physical probability measure P the process (V_t) follows a diffusion model (known as the Black–Scholes model or geometric Brownian motion) of the form

$$\mathrm{d}V_t = \mu_V V_t \,\mathrm{d}t + \sigma_V V_t \,\mathrm{d}W_t \tag{10.5}$$

for constants $\mu_V \in \mathbb{R}$ (the drift of the asset value process), $\sigma_V > 0$ (the volatility of the asset value process), and a standard Brownian motion (W_t). Equation (10.5) can be solved explicitly, and it can be shown that

$$V_T = V_0 \exp((\mu_V - \tfrac{1}{2}\sigma_V^2)T + \sigma_V W_T).$$

Since $W_T \sim N(0, T)$, it follows that $\ln V_T \sim N(\ln V_0 + (\mu_V - \frac{1}{2}\sigma_V^2)T, \sigma_V^2 T)$. Under the dynamics (10.5), the default probability of the firm is readily computed. We have

$$P(V_T \leqslant B) = P(\ln V_T \leqslant \ln B) = \Phi\left(\frac{\ln(B/V_0) - (\mu_V - \frac{1}{2}\sigma_V^2)T}{\sigma_V\sqrt{T}}\right). \tag{10.6}$$

It may be deduced from (10.6) that the default probability is increasing in B, decreasing in V_0 and μ_V and, for $V_0 > B$, increasing in σ_V. All these properties are perfectly in line with economic intuition.

Figure 10.3 shows two simulated trajectories for the asset value process (V_t) for values $V_0 = 1$, $\mu_V = 0.03$ and $\sigma_V = 0.25$. Assuming that $B = 0.85$ and $T = 1$, one path is a default path, terminating at a value $V_T < B$, while the other is a non-default path.

10.3.2 Pricing in Merton's Model

In the context of Merton's model one can price securities whose pay-off depends on the value V_T of the firm's assets at T. Prime examples are the firm's debt, or, equivalently, the zero-coupon bonds issued by the firm, and the firm's equity. In our analysis of pricing in the context of the Merton model we make use of a few basic

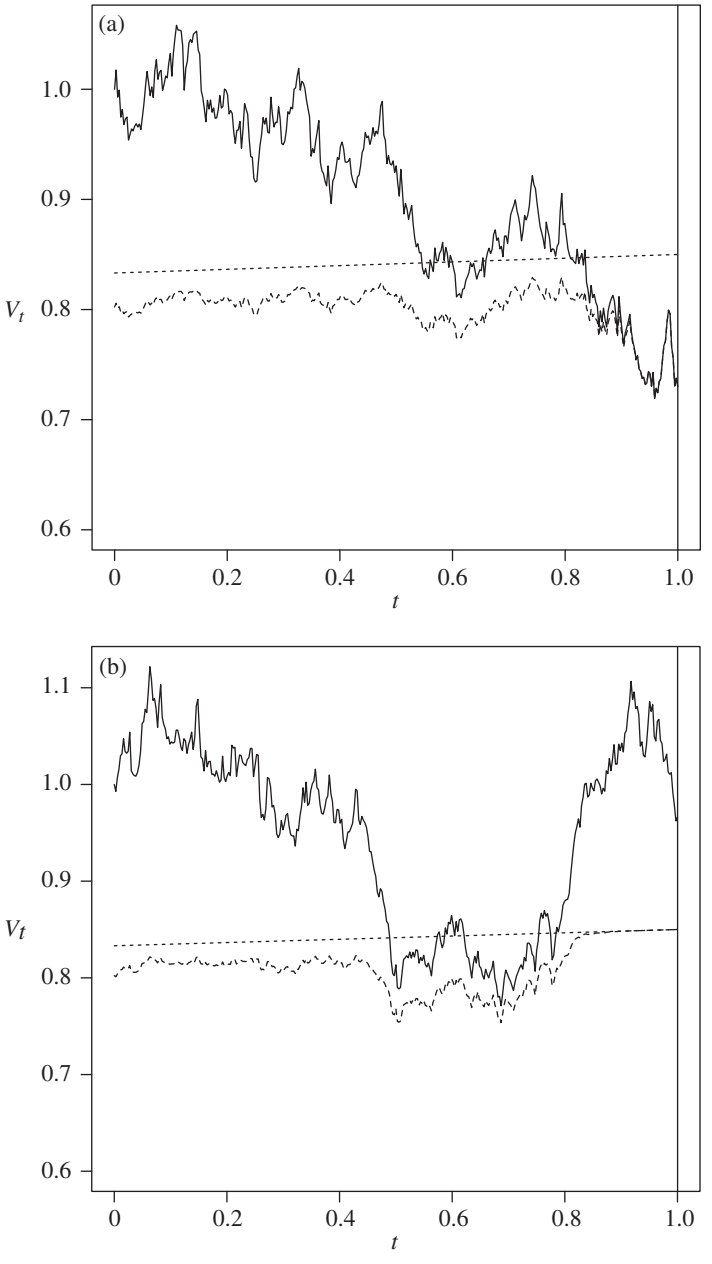

Figure 10.3. Illustration of (a) a default path and (b) a non-default path in Merton's model. The solid lines show simulated one-year trajectories for the asset value process (V_t) starting at $V_0 = 1$ with parameters $\mu_V = 0.03$ and $\sigma_V = 0.25$. Assuming that the debt has face value $B = 0.85$ and maturity $T = 1$ and that the interest rate is $r = 0.02$, the dotted curve shows the value of default-free debt $(Bp_0(t, T))$ while the dashed line shows the evolution of the company's debt B_t according to formula (10.12). The difference between the asset value V_t and the debt B_t is the value of equity S_t.

concepts from financial mathematics and stochastic calculus; references to useful texts in financial mathematics are given in Notes and Comments.

We make the following assumptions.

Assumption 10.3.

(i) *The risk-free interest rate is deterministic and equal to* $r \geqslant 0$.

(ii) *The firm's asset-value process* (V_t) *is independent of the way the firm is financed, and in particular it is independent of the debt level* B.

(iii) *The asset value* (V_t) *can be traded on a frictionless market, and the asset-value dynamics are given by the geometric Brownian motion (10.5).*

These assumptions merit some comments. First, the independence of (V_t) from the financial structure of the firm is questionable, because a very high debt level, and hence a high default probability, may adversely affect the ability of a firm to generate business, hence affecting the value of its assets. This is a special case of the indirect bankruptcy costs discussed in Section 1.4.2. Second, while there are many firms with traded equity, the value of the assets of a firm is usually neither completely observable nor traded. We come back to this issue in Section 10.3.3. For an example where (iii) holds, think of an investment company or trust that invests in liquidly traded securities and uses debt financing to leverage its position. In that case V_t corresponds to the value of the investment portfolio at time t, and this portfolio consists of traded securities by assumption.

Pricing of equity and debt. Consider a claim on the asset value of the firm with maturity T and pay-off $h(V_T)$, such as the firm's equity and debt in (10.3) and (10.4). Under Assumption 10.3, the fair value $f(t, V_t)$ of this claim at time $t \leqslant T$ can be computed using the risk-neutral pricing rule as the expectation of the discounted pay-off under the risk-neutral measure Q, that is,

$$f(t, V_t) = E^Q(\mathrm{e}^{-r(T-t)}h(V_T) \mid \mathcal{F}_t). \tag{10.7}$$

According to (10.3), the firm's equity corresponds to a European call on (V_t) with exercise price B and maturity T. The risk-neutral value of equity obtained from (10.7) is therefore given simply by the Black–Scholes price C^{BS} of a European call. This yields

$$S_t = C^{\mathrm{BS}}(t, V_t; r, \sigma_V, B, T) := V_t\Phi(d_{t,1}) - B\mathrm{e}^{-r(T-t)}\Phi(d_{t,2}), \tag{10.8}$$

where the arguments are given by

$$d_{t,1} = \frac{\ln V_t - \ln B + (r + \frac{1}{2}\sigma_V^2)(T-t)}{\sigma_V\sqrt{T-t}}, \qquad d_{t,2} = d_{t,1} - \sigma_V\sqrt{T-t}. \tag{10.9}$$

Next we turn to the valuation of the risky debt issued by the firm. Since we assumed a constant interest rate r, the price at $t \leqslant T$ of a default-free zero-coupon bond with maturity T and a face value of one equals $p_0(t, T) = \mathrm{e}^{-r(T-t)}$. According to (10.4) we have

$$B_t = Bp_0(t, T) - P^{\mathrm{BS}}(t, V_t; r, \sigma_V, B, T), \tag{10.10}$$

where $P^{\mathrm{BS}}(t, V; r, \sigma_V, B, T)$ denotes the Black–Scholes price of a European put with strike B, maturity T on (V_t) for given interest rate r, and volatility σ_V. It is well known that

$$P^{\mathrm{BS}}(t, V_t; r, \sigma_V, B, T) = B\mathrm{e}^{-r(T-t)}\Phi(-d_{t,2}) - V_t\Phi(-d_{t,1}), \tag{10.11}$$

with $d_{t,1}$ and $d_{t,2}$ as in (10.9). Combining (10.10) and (10.11) we get

$$B_t = p_0(t, T)B\Phi(d_{t,2}) + V_t\Phi(-d_{t,1}). \tag{10.12}$$

Lines showing the evolution of B_t as a function of the evolution of V_t under the assumption that $r = 0.02$ have been added to Figure 10.3. The difference between the asset value V_t and the debt B_t is the value of equity S_t; note how the value of equity is essentially negligible for $t > 0.8$ in the default path.

Volatility of the firm's equity. It is interesting to compute the volatility of the equity of the firm under Assumption 10.3. To this end we define the quantity

$$\nu(t, V_t) := \frac{V_t C_V^{\mathrm{BS}}(t, V_t)}{C^{\mathrm{BS}}(t, V_t)}.$$

In the context of option pricing this is known as the *elasticity* of a European call with respect to the price of the underlying security. In our context it measures the percentage change in the value of equity per percentage change in the value of the underlying assets.

If we apply Itô's formula to $S_t = C^{\mathrm{BS}}(t, V_t; r, \sigma_V, B, T)$ we obtain

$$\mathrm{d}S_t = \sigma_V C_V^{\mathrm{BS}}(t, V_t)V_t\,\mathrm{d}W_t + (C_t^{\mathrm{BS}}(t, V_t) + \mu_V C_V^{\mathrm{BS}}(t, V_t)V_t + \tfrac{1}{2}\sigma_V^2 V_t^2 C_{VV}^{\mathrm{BS}})\,\mathrm{d}t.$$

Using the definition of the elasticity ν, we may rewrite the $\mathrm{d}W_t$ term in the form

$$\sigma_V C_V^{\mathrm{BS}}(t, V_t)V_t\,\mathrm{d}W_t = \sigma_V \nu(t, V_t)C^{\mathrm{BS}}(t, V_t)\,\mathrm{d}W_t,$$

from which we conclude that the volatility of the firm's equity at time t is a function $\sigma_S(t, V_t)$ of time and of the current asset value V_t that takes the form

$$\sigma_S(t, V_t) = \nu(t, V_t)\sigma_V. \tag{10.13}$$

The volatility of the firm's equity is therefore greater than σ_V, since the elasticity of a European call is always greater than one.

Risk-neutral and physical default probabilities. Next we compare physical and risk-neutral default probabilities in Merton's model. It is a basic result from financial mathematics that under the risk-neutral measure Q the process (V_t) satisfies the stochastic differential equation (SDE) $\mathrm{d}V_t = rV_t\,\mathrm{d}t + \sigma_V V_t\,\mathrm{d}\tilde{W}_t$ for a standard Q-Brownian motion $\tilde{W}$. Note how the drift μ_V in (10.5) has been replaced by the risk-free interest rate r. The risk-neutral default probability is therefore given by the formula (10.6), evaluated with $\mu_V = r$:

$$q = Q(V_T \leqslant B) = \Phi\left(\frac{\ln B - \ln V_0 - (r - \frac{1}{2}\sigma_V^2)T}{\sigma_V\sqrt{T}}\right).$$

Comparing this with the physical default probability $p = P(V_T \leqslant B)$ as given in (10.6) we obtain the relationship

$$q = \Phi\left(\Phi^{-1}(p) + \frac{\mu_V - r}{\sigma_V}\sqrt{T}\right). \tag{10.14}$$

The correction term $(\mu_V - r)/\sigma_V$ equals the *Sharpe ratio* of V (a popular measure of the risk premium earned by the firm). The transition formula (10.14) is sometimes applied in practice to go from physical to risk-neutral default probabilities. Note, however, that (10.14) is supported by theoretical arguments only in the narrow context of the Merton model.

Credit spread. We may use (10.12) to infer the credit spread $c(t, T)$ implied by Merton's model. The credit spread measures the difference between the continuously compounded yield to maturity of a defaultable zero-coupon bond $p_1(t, T)$ and that of a default-free zero-coupon bond $p_0(t, T)$. It is defined by

$$c(t, T) = \frac{-1}{T - t}(\ln p_1(t, T) - \ln p_0(T, t)) = \frac{-1}{T - t} \ln \frac{p_1(t, T)}{p_0(t, T)}. \tag{10.15}$$

Throughout the book we use the convention that a zero-coupon bond has a nominal value equal to 1. In line with this convention we assume that the pay-off at T of a zero-coupon bond issued by the firm is given by $(1/B)B_T$, so that the price of such a bond at time $t \leqslant T$ is given by $p_1(t, T) = (1/B)B_t$. We therefore obtain

$$c(t, T) = \frac{-1}{T - t} \ln \left(\Phi(d_{t,2}) + \frac{V_t}{Bp_0(t, T)}\Phi(-d_{t,1})\right). \tag{10.16}$$

Since $d_{t,1}$ can be rewritten as

$$d_{t,1} = \frac{-\ln(Bp_0(t, T)/V_t) + \frac{1}{2}\sigma_V^2(T - t)}{\sigma_V\sqrt{T - t}},$$

and similarly for $d_{t,2}$, we conclude that, for a fixed time to maturity $T - t$, the spread $c(t, T)$ depends only on the volatility σ_V and on the ratio $d := Bp_0(t, T)/V_t$, which is the ratio of the discounted nominal value of the firm's debt to the value of the firm's assets and is hence a measure of the relative debt level or *leverage* of the firm. As the price of a European put (10.11) is increasing in the volatility, it follows from (10.10) that $c(t, T)$ is increasing in σ_V. In Figure 10.4 we plot the credit spread as a function of σ_V and of the time to maturity $\tau = T - t$.

Extensions. Merton's model is quite simplistic. Over the years this has given rise to a rich literature on firm-value models. We briefly comment on the most important research directions (bibliographic references are given in Notes and Comments). To begin with, the observation that, in reality, firms can default at essentially any time (and not only at a deterministic point in time T) has led to the development of so-called *first-passage-time models*. In this class of models default occurs when the asset-value process crosses a default threshold B for the first time; the threshold is usually interpreted as the average value of the liabilities. Formally, the default time τ is defined by $\tau = \inf\{t \geqslant 0 \colon V_t \leqslant B\}$. Further technical developments

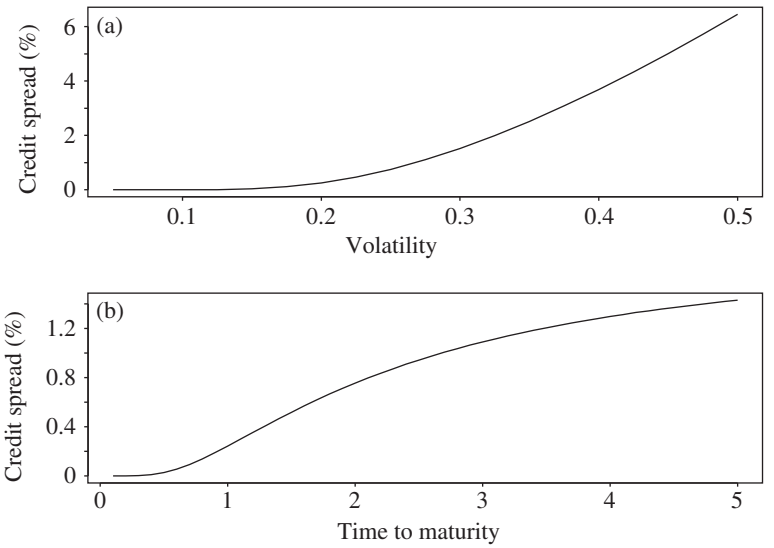

Figure 10.4. Credit spread $c(t, T)$ in per cent as a function of (a) the firm's volatility σ_V and (b) the time to maturity $\tau = T - t$ for fixed leverage measure $d = 0.6$ (in (a) $\tau = 2$ years; in (b) $\sigma_V = 0.25$). Note that, for a time to maturity smaller than approximately three months, the credit spread implied by Merton's model is basically equal to zero. This is not in line with most empirical studies of corporate bond spreads and has given rise to a number of extensions of Merton's model that are listed in Notes and Comments. We will see in Section 10.5.3 that reduced-form models lead to a more reasonable behaviour of short-term credit spreads.

include models with stochastic default-free interest rates and models where the asset-value process (V_t) is given by a diffusion with jumps.

Firm-value models with an *endogenous default threshold* are an interesting economic extension of Merton's model. Here the default boundary B is not fixed a priori by the modeller but is determined endogenously by strategic considerations of the shareholders. Finally, structural models with *incomplete information* on asset value and/or liabilities provide an important link between the structural and reduced-form approaches to credit risk modelling.

10.3.3 Structural Models in Practice: EDF and DD

There are a number of industry models that descend from the Merton model. An important example is the so-called *public-firm EDF model* that is maintained by Moody's Analytics. The acronym EDF stands for *expected default frequency*; this is a specific estimate of the physical default probability of a given firm over a one-year horizon. The methodology proposed by Moody's Analytics builds on earlier work by KMV (a private company named after its founders Kealhofer, McQuown and Vasicek) in the 1990s, and is also known as the KMV model. Our presentation of the public-firm EDF model is based on Crosbie and Bohn (2002) and Sun, Munves and Hamilton (2012). We concentrate on the main ideas, since detailed information about actual implementation and calibration procedures is proprietary and these procedures may change over time.

Overview. Recall that in the classic Merton model the one-year default probability of a given firm is given by the probability that the asset value in one year lies below the threshold B representing the overall liabilities of the firm. Under Assumption 10.3, the one-year default probability is a function of the current asset value V_0, the (annualized) drift μ_V and volatility σ_V of the asset-value process, and the threshold B; using (10.6) with $T = 1$ and recalling that $\Phi(d) = 1 - \Phi(-d)$ we infer that

$$\mathrm{EDF}_{\mathrm{Merton}} = 1 - \Phi\left(\frac{\ln V_0 - \ln B + (\mu_V - \frac{1}{2}\sigma_V^2)}{\sigma_V}\right). \tag{10.17}$$

In the public-firm EDF model a similar structure is assumed for the EDF; however, $1 - \Phi$ is replaced by some empirically estimated function, B is replaced by a new default threshold $\tilde{B}$ representing the structure of the firm's liabilities more closely, and the term $(\mu_V - \frac{1}{2}\sigma_V^2)$ in the numerator is sometimes omitted for expositional ease. Moreover, the current asset value V_0 and the asset volatility σ_V are inferred (or "backed out") from information about the firm's equity value.

Determination of the asset value and the asset volatility. Firm-value-based credit risk models are based on the *market value* V_0 of the firm's assets. This makes sense as the market value is a forward-looking measure that reflects investor expectations about the business prospects and future cash flows of the firm. Unfortunately, in contrast to the assumptions underlying Merton's model, in most cases there is no market for the assets of a firm, so that the asset value is not directly observable. Moreover, the market value can differ greatly from the value of a company as measured by accountancy rules (the so-called book value), so that accounting information and balance sheet data are of limited use in inferring the asset value V_0. For these reasons the public-firm EDF model relies on an indirect approach and infers values of V_t at different times t from the more easily observed values of a firm's equity S_t. This approach simultaneously provides estimates of V_0 and of the asset volatility σ_V. The latter estimate is needed since σ_V has a strong impact on default probabilities; all other things being equal, risky firms with a comparatively high asset volatility σ_V have a higher default probability than firms with a low asset volatility.

We explain the estimation approach in the context of the Merton model. Recall that under Assumption 10.3 we have that

$$S_t = C^{\mathrm{BS}}(t, V_t; r, \sigma_V, B, T). \tag{10.18}$$

Obviously, at a fixed point in time, $t = 0$ say, (10.18) is an equation with two unknowns, V_0 and σ_V. To overcome this difficulty, one may use an iterative procedure. In step (1), (10.18) with some initial estimate $\sigma_V^{(0)}$ is used to infer a time series of asset values $(V_t^{(0)})$ from equity values. Then a new volatility estimate $\sigma_V^{(1)}$ is estimated from this time series; a new time series $(V_t^{(1)})$ is then constructed using (10.18) with $\sigma_V^{(1)}$. This procedure is iterated n times, until the volatility estimates $\sigma_V^{(n-1)}$ and $\sigma_V^{(n)}$ generated in step $(n-1)$ and step (n) are sufficiently close.

In the public-firm EDF model, the capital structure of the firm is modelled in a more sophisticated manner than in Merton's model. There are several classes

of liabilities, such as long- and short-term debt and convertible bonds, the model allows for intermediate cash payouts corresponding to coupons and dividends, and default can occur at any time. Moreover, the default point (the threshold value $\tilde{B}$ such that the company defaults if (V_t) falls below $\tilde{B}$) is determined from a more detailed analysis of the term structure of the firm's debt. The equity value is thus no longer given by (10.18) but by some different function $f(t, V_t, \sigma_V)$, which has to be computed numerically. The general idea of the approach used to estimate V_0 and σ_V is, however, exactly as described above.

Calculation of EDFs. In the Merton model, default occurs if the value of a firm's assets falls below the value of its liabilities. With lognormally distributed asset values, as implied for instance by Assumption 10.3, this leads to default probabilities of the form $\mathrm{EDF}_{\mathrm{Merton}}$ as in (10.17). This relationship between asset value and default probability may be too simplistic to be an accurate description of actual default probabilities. For instance, asset values are not necessarily lognormal but might follow a distribution with heavy tails and there might be payments due at an intermediate point in time causing default at that date.

For these reasons, in the public-firm EDF model a new state variable is introduced in an intermediate step. This is the so-called *distance-to-default* (DD), given by

$$\mathrm{DD} := (\log V_0 - \log \tilde{B})/\sigma_V. \tag{10.19}$$

Here, $\tilde{B}$ represents the default threshold; in some versions of the model $\tilde{B}$ is modelled as the sum of the liabilities payable within one year and half of the longer-term debt. Sometimes practitioners call the distance-to-default the "number of standard deviations a company is away from its default threshold". Note that (10.19) is in fact an approximation of the argument of (10.17), since μ_V and σ_V^2 are usually small.

In the EDF methodology it is assumed that the distance-to-default *ranks* firms in the sense that firms with a higher DD exhibit a higher default probability. The functional relationship between DD and EDF is determined empirically; using a database of historical default events, the proportion of firms with DD in a given small range that default within a year is estimated. This proportion is the empirically estimated EDF. The DD-to-EDF mapping exhibits "heavy tails": for high-quality firms with a large DD the empirically estimated EDF is much higher than $\mathrm{EDF}_{\mathrm{Merton}}$ as given in (10.17). For instance, for a firm with a DD equal to 4 we find that $\mathrm{EDF}_{\mathrm{Merton}} \approx 0.003\%$, whereas the empirically estimated EDF equals 0.4%.

In Table 10.3 we illustrate the computation of the EDF for two different firms, Johnson & Johnson (a well-capitalized firm that operates in the relatively stable health care market) and RadioShack (a firm that is active in the highly volatile consumer electronics business). If we compare the numbers, we see that the EDF for Johnson & Johnson is close to zero whereas the EDF for RadioShack is quite high. This difference reflects the higher leverage of RadioShack and the riskiness of the underlying business, as reflected by the comparatively large asset volatility $\sigma_V = 24\%$. Indeed, on 11 September 2014, the *New York Times* reported that a bankruptcy filing for RadioShack could be near, suggesting that the EDF had good predictive power in this case.

Table 10.3. A summary of the public-firm EDF methodology. The example is taken from Sun, Munves and Hamilton (2012); it is concerned with the situation of Johnson & Johnson (J&J) and RadioShack as of April 2012. All quantities are in US dollars.

Variable	J&J	RadioShack	Notes
Market value of assets V_0	236 bn	1834 m	Determined from time series of equity prices
Asset volatility σ_V	11%	24%	
Default threshold $\tilde{B}$	39 bn	1042 m	Short-term liabilities and half of long-term liabilities
DD	16.4	2.3	Given by $(\log V_0 - \log \tilde{B})/\sigma_V$
EDF (one year)	0.01%	3.58%	Determined using empirical mapping between DD and EDF

10.3.4 Credit-Migration Models Revisited

Recall that in the credit-migration approach each firm is assigned to a credit-rating category at any given time point. There are a finite number of such ratings and they are ordered by credit quality and include the category of default. The probability of moving from one credit rating to another credit rating over the given risk horizon (typically one year) is then specified. In this section we explain how a migration model can be embedded in a firm-value model and thus be treated as a structural model. This will be useful in the discussion of portfolio versions of these models in Chapter 11. Moreover, we compare the public-firm EDF model and credit-migration approaches.

Credit-migration models as firm-value models. We consider a firm that has been assigned to some non-default rating category j at $t = 0$ and for which transition probabilities $p_{jk}, 0 \leqslant k \leqslant n$, over the period $[0, T]$ are available on the basis of that rating. These express the probability that the firm belongs to rating class k at the time horizon T, given that it is in class j at $t = 0$. In particular, $p_{j,0}$ is the default probability of the firm over $[0, T]$.

Suppose that the asset-value process (V_t) of the firm follows the model given in (10.5), so that

$$V_T = V_0 \exp((\mu_V - \tfrac{1}{2}\sigma_V^2)T + \sigma_V W_T) \tag{10.20}$$

is lognormally distributed. We can now choose thresholds

$$0 = \tilde{d}_0 < \tilde{d}_1 < \cdots < \tilde{d}_n < \tilde{d}_{n+1} = \infty \tag{10.21}$$

such that $P(\tilde{d}_k < V_T \leqslant \tilde{d}_{k+1}) = p_{jk}$ for $k \in \{0, \ldots, n\}$. We have therefore translated the transition probabilities into a series of thresholds for an assumed asset-value process. The threshold $\tilde{d}_1$ is the default threshold, which in the Merton model of Section 10.3.1 was interpreted as the value of the firm's liabilities. The higher thresholds are the asset-value levels that mark the boundaries of higher rating categories. The firm-value model in which we have embedded the migration model can be summarized by saying that the firm belongs to rating class k at the time horizon T if and only if $\tilde{d}_k < V_T \leqslant \tilde{d}_{k+1}$.

The migration probabilities in the firm-value model obviously remain invariant under simultaneous strictly increasing transformations of V_T and the thresholds $\tilde{d}_j$. If we define

$$X_T := \frac{\ln V_T - \ln V_0 - (\mu_V - \frac{1}{2}\sigma_V^2)T}{\sigma_V\sqrt{T}}, \tag{10.22}$$

$$d_k := \frac{\ln \tilde{d}_k - \ln V_0 - (\mu_V - \frac{1}{2}\sigma_V^2)T}{\sigma_V\sqrt{T}}, \tag{10.23}$$

then we can also say that the firm belongs to rating class k at the time horizon T if and only if $d_k < X_T \leqslant d_{k+1}$. Observe that X_T is a standardized version of the *asset-value log-return* $\ln V_T - \ln V_0$, and we can easily verify that $X_T = W_T/\sqrt{T}$ so that it has a standard normal distribution. In this case the formulas for the thresholds are easily obtained and are $d_k = \Phi^{-1}(\sum_{l=0}^{k-1} p_{jl})$ for $k = 1, \ldots, n$.

The public-firm EDF model and credit-migration approaches compared. The public-firm EDF model uses market data, most notably the current stock price, as inputs for the EDF computation. The EDF therefore reacts quickly to changes in the economic prospects of a firm, as these are reflected in the firm's share price and hence in the estimated distance-to-default. Moreover, EDFs are quite sensitive to the current macroeconomic environment. The distance-to-default is observed to rise in periods of economic expansion (essentially due to higher share prices reflecting better economic conditions) and to decrease in recession periods. Rating agencies, on the other hand, are typically slow in adjusting their credit ratings, so that the current rating does not always reflect the economic condition of a firm. This is particularly important if the credit quality of a firm deteriorates rapidly, as is typically the case with companies that are close to default. For instance, the investment bank Lehman Brothers had a fairly good rating (Aa or better) when it defaulted in September 2008. EDFs might therefore be better predictors of default probabilities over short time horizons.

On the other hand, the public-firm EDF model is quite sensitive to global over- and underreaction of equity markets. In particular, the bursting of a stock market bubble may lead to drastically increased EDFs, even if the economic outlook for a given corporation has not changed very much. This can lead to huge fluctuations in the amount of risk capital that is required to back a given credit portfolio. From this point of view the relative inertia of ratings-based models could be considered an advantage, as the ensuing risk capital requirements tend to be more stable over time.

Notes and Comments

There are many excellent texts, at varying technical levels, in which the basic mathematical finance results used in Section 10.3.2 can be found. Models in discrete time are discussed in Cox and Rubinstein (1985) and Jarrow and Turnbull (1999); excellent introductions to continuous-time models include Baxter and Rennie (1996), Björk (2004), Bingham and Kiesel (2004) and Shreve (2004b).

Lando (2004) gives a good overview of the rich literature on firm-value models. First-passage-time models have been considered by, among others, Black and Cox (1976) and, in a set-up with stochastic interest rates, Longstaff and Schwartz (1995). The problem of the unrealistically low credit spreads for small maturities $\tau = T - t$, which we pointed out in Figure 10.4, has also led to extensions of Merton's model. Partial remedies within the class of firm-value models include models with jumps in the firm value, as in Zhou (2001), time-varying default thresholds, as in Hull and White (2001), stochastic volatility models for the firm-value process with time-dependent dynamics, as in Overbeck and Schmidt (2005), and incomplete information on firm value or default threshold, as in Duffie and Lando (2001), Frey and Schmidt (2009) and Cetin (2012). Models with endogenous default thresholds have been considered by, among others, Leland (1994), Leland and Toft (1996) and Hilberink and Rogers (2002).

Duffie and Lando (2001) established a relationship between firm-value models and reduced-form models in continuous time. Essentially, they showed that, from the perspective of investors with *incomplete accounting information* (i.e. incomplete information about the assets or liabilities of a firm), a firm-value model becomes a reduced-form model. A less technical discussion of these issues can be found in Jarrow and Protter (2004).

The public-firm EDF model was first described in Crosbie and Bohn (2002); the model variant that is currently in use is described in Dwyer and Qu (2007) and Sun, Munves and Hamilton (2012).

10.4 Bond and CDS Pricing in Hazard Rate Models

Hazard rate models are the most basic reduced-form credit risk models and are therefore a natural starting point for our discussion of this model class. Moreover, hazard rate models are used as an input in the construction of the popular copula models for portfolio credit derivatives. For these reasons this section is devoted to bond and CDS pricing in hazard rate models. We begin by introducing the necessary mathematical background in Section 10.4.1. Since the pricing results that we present in this section rely on the concept of risk-neutral pricing and martingale modelling, we briefly review these notions in Section 10.4.2. The pricing of bonds and CDSs and some of the related empirical evidence is discussed in Sections 10.4.3, 10.4.4 and 10.4.5.

10.4.1 Hazard Rate Models

A hazard rate model is a model in which the distribution of the default time of an obligor is directly specified by a hazard function without modelling the mechanism by which default occurs.

To set up a hazard rate model we consider a probability space $(\Omega, \mathcal{F}, P)$ and a random time τ defined on this space, i.e. an $\mathcal{F}$-measurable rv taking values in $[0, \infty]$. In economic terms, τ can be interpreted as the default time of some company. We denote the df of τ by $F(t) = P(\tau \leqslant t)$ and the tail or survival function by

$\bar{F}(t) = 1 - F(t)$; we assume that $P(\tau = 0) = F(0) = 0$, and that $\bar{F}(t) > 0$ for all $t < \infty$. We define the *jump* or *default indicator process* (Y_t) associated with τ by

$$Y_t = I_{\{\tau \leqslant t\}}, \quad t \geqslant 0. \tag{10.24}$$

Note that (Y_t) is a right-continuous process that jumps from 0 to 1 at the default time τ and that $1 - Y_t = I_{\{\tau > t\}}$ is the *survival indicator* of the firm.

Definition 10.4 (cumulative hazard function and hazard function). The function $\Gamma(t) := -\ln(\bar{F}(t))$ is called the *cumulative hazard function* of the random time τ. If F is absolutely continuous with density f, the function $\gamma(t) := f(t)/(1 - F(t)) = f(t)/\bar{F}(t)$ is called the *hazard function* of τ.

By definition we have $\bar{F}(t) = \mathrm{e}^{-\Gamma(t)}$. If F has density f, we calculate that $\Gamma'(t) = f(t)/\bar{F}(t) = \gamma(t)$, so that we can represent the survival function of τ in terms of the hazard function by

$$\bar{F}(t) = \exp\left(-\int_0^t \gamma(s)\,\mathrm{d}s\right). \tag{10.25}$$

The hazard function $\gamma(t)$ at a fixed time t gives the *hazard rate* at t, which can be interpreted as a measure of the instantaneous risk of default at t, given survival up to time t. In fact, for $h > 0$ we have $P(\tau \leqslant t + h \mid \tau > t) = (F(t+h) - F(t))/\bar{F}(t)$. Hence we obtain

$$\lim_{h \to 0} \frac{1}{h} P(\tau \leqslant t + h \mid \tau > t) = \frac{1}{\bar{F}(t)} \lim_{h \to 0} \frac{F(t+h) - F(t)}{h} = \gamma(t).$$

For illustrative purposes we determine the hazard function for the Weibull distribution. This is a popular distribution for survival times with df $F(t) = 1 - \mathrm{e}^{-\lambda t^\alpha}$ for parameters $\lambda, \alpha > 0$. For $\alpha = 1$ the Weibull distribution reduces to the standard exponential distribution. Differentiation yields

$$f(t) = \lambda\alpha t^{\alpha-1}\mathrm{e}^{-\lambda t^\alpha} \quad \text{and} \quad \gamma(t) = \lambda\alpha t^{\alpha-1}.$$

In particular, γ is decreasing in t if $\alpha < 1$ and increasing if $\alpha > 1$. For $\alpha = 1$ (exponential distribution) the hazard rate is time independent and equal to the parameter λ.

Filtrations and conditional expectations. In financial models, filtrations are used to model the information available to investors at various points in time. Formally, a *filtration* $(\mathcal{F}_t)$ on $(\Omega, \mathcal{F})$ is an increasing family $\{\mathcal{F}_t : t \geqslant 0\}$ of sub-σ-algebras of $\mathcal{F}$: $\mathcal{F}_t \subset \mathcal{F}_s \subset \mathcal{F}$ for $0 \leqslant t \leqslant s < \infty$. The σ-algebra $\mathcal{F}_t$ represents the state of knowledge of an observer at time t, and $A \in \mathcal{F}_t$ is taken to mean that at time t the observer is able to determine if the event A has occurred.

In this section it is assumed that the only quantity that is observable for investors is the default event of the firm under consideration or, equivalently, the default indicator process (Y_t) associated with τ. The appropriate filtration is therefore given by $(\mathcal{H}_t)$ with

$$\mathcal{H}_t = \sigma(\{Y_u : u \leqslant t\}), \tag{10.26}$$

the *default history* up to and including time t. By definition, τ is an $(\mathcal{H}_t)$ stopping time, as $\{\tau \leqslant t\} = \{Y_t = 1\} \in \mathcal{H}_t$ for all $t \geqslant 0$; moreover, $(\mathcal{H}_t)$ is obviously the smallest filtration with this property.

In order to study bond and CDS pricing in hazard rate models we need to compute conditional expectations with respect to the σ-algebra $\mathcal{H}_t$. We begin our analysis of this issue with an auxiliary result on the structure of $\mathcal{H}_t$-measurable rvs. The result formalizes the fact that every $\mathcal{H}_t$-measurable rv can be expressed as a function of events related to the default history at t.

Lemma 10.5. *Every $\mathcal{H}_t$-measurable rv H is of the form $H = h(\tau)I_{\{\tau \leqslant t\}} + cI_{\{\tau > t\}}$ for a measurable function $h\colon [0,t] \to \mathbb{R}$ and some constant $c \in \mathbb{R}$.*

Proof. The σ-algebra $\mathcal{H}_t$ is generated by the events $\{Y_u = 1\} = \{\tau \leqslant u\}$, $u < t$, and $\{Y_t = 0\} = \{\tau > t\}$, and hence by the rvs $(\tau \wedge t) := \min\{\tau, t\}$ and $I_{\{\tau > t\}}$. This implies that any $\mathcal{H}_t$-measurable rv H can be written as $H = g(\tau \wedge t, I_{\{\tau > t\}})$ for some measurable function $g\colon [0,t] \times \{0,1\} \to \mathbb{R}$. The claim follows if we define $h(u) := g(u,0)$, $u \leqslant t$, and $c := g(t,1)$. □

Lemma 10.6. *Let τ be a random time with jump indicator process $Y_t = I_{\{\tau \leqslant t\}}$ and natural filtration $(\mathcal{H}_t)$. Then, for any integrable rv X and any $t \geqslant 0$, we have*

$$E(I_{\{\tau > t\}} X \mid \mathcal{H}_t) = I_{\{\tau > t\}} \frac{E(I_{\{\tau > t\}} X)}{P(\tau > t)}. \tag{10.27}$$

Proof. Since $E(I_{\{\tau > t\}} X \mid \mathcal{H}_t)$ is $\mathcal{H}_t$-measurable and zero on $\{\tau \leqslant t\}$, we obtain from Lemma 10.5 that $E(I_{\{\tau > t\}} X \mid \mathcal{H}_t) = I_{\{\tau > t\}} c$ for some constant c. Taking expectations yields $E(I_{\{\tau > t\}} X) = cP(\tau > t)$ and hence $c = E(I_{\{\tau > t\}} X)/P(\tau > t)$. □

Lemma 10.6 can be used to determine conditional survival probabilities. Fix $t < T$ and consider the quantity $P(\tau > T \mid \mathcal{H}_t)$. Applying (10.27) with $X := I_{\{\tau > T\}}$ yields

$$P(\tau > T \mid \mathcal{H}_t) = E(X \mid \mathcal{H}_t) = E(I_{\{\tau > t\}} X \mid \mathcal{H}_t) = I_{\{\tau > t\}} \frac{\bar{F}(T)}{\bar{F}(t)}. \tag{10.28}$$

If τ admits the hazard function $\gamma(t)$, we get the important formula

$$P(\tau > T \mid \mathcal{H}_t) = I_{\{\tau > t\}} \exp\left(- \int_t^T \gamma(s)\, \mathrm{d}s \right), \quad t < T. \tag{10.29}$$

The next proposition is concerned with stochastic process properties of the jump indicator process of a random time τ.

Proposition 10.7. *Let τ be a random time with absolutely continuous df F and hazard function γ. Then $M_t := Y_t - \int_0^t I_{\{\tau > u\}} \gamma(u)\, \mathrm{d}u$, $t \geqslant 0$, is an $(\mathcal{H}_t)$-martingale: that is, $E(M_s \mid \mathcal{H}_t) = M_t$ for all $0 \leqslant t \leqslant s < \infty$.*

In Section 10.5.1 we extend this result to doubly stochastic random times and discuss its financial and mathematical relevance.

Proof. Let $s > t$. We have to show that $E(M_s - M_t \mid \mathcal{H}_t) = 0$, i.e. that $E(Y_s - Y_t \mid \mathcal{H}_t) = E(\int_t^s \gamma(u) I_{\{u<\tau\}}\,\mathrm{d}u \mid \mathcal{H}_t)$. Using (10.28) we get

$$E(Y_s - Y_t \mid \mathcal{H}_t) = I_{\{\tau>t\}} P(\tau \leqslant s \mid \mathcal{H}_t) = I_{\{\tau>t\}}\left(1 - \frac{\bar{F}(s)}{\bar{F}(t)}\right)$$
$$= I_{\{\tau>t\}} \frac{\bar{F}(t) - \bar{F}(s)}{\bar{F}(t)}.$$

Note that $X := \int_t^s \gamma(u) I_{\{u<\tau\}}\,\mathrm{d}u$ is 0 on $\{\tau \leqslant t\}$, so $X = X I_{\{\tau>t\}}$. Hence we obtain from Lemma 10.6, the Fubini Theorem and the identity $\bar{F}'(t) = -f(t) = -\gamma(t)\bar{F}(t)$ that

$$E(X \mid \mathcal{H}_t) = I_{\{\tau>t\}} \frac{E(X)}{\bar{F}(t)} = I_{\{\tau>t\}} \frac{\int_t^s \gamma(u)\bar{F}(u)\,\mathrm{d}u}{\bar{F}(t)} = I_{\{\tau>t\}} \frac{\bar{F}(t) - \bar{F}(s)}{\bar{F}(t)},$$

and the result follows. □

10.4.2 Risk-Neutral Pricing Revisited

The remainder of Section 10.4 is devoted to an analysis of risk-neutral pricing results for credit products in hazard rate models. Risk-neutral pricing has become so popular that the conceptual underpinnings are often overlooked. A prime case in point is the mechanical use of the Gauss copula to price CDO tranches, a practice that led to well-documented problems during the 2007–9 financial crisis. It therefore seems appropriate to clarify the applicability and the limitations of risk-neutral pricing in the context of credit risk models.

Risk-neutral pricing. We build on the elementary discussion of risk-neutral valuation given in Section 2.2.2. In that section we considered a simple one-period default model for a defaultable zero-coupon bond with maturity T equal to one year and a deterministic recovery rate $1 - \delta$ equal to 60%. Moreover, we assumed that the real-world default probability was $p = 1\%$, the risk-free simple interest rate was $r_{0,1} = 5\%$, and the market price of the bond at $t = 0$ was $p_1(0, 1) = 0.941$.

Risk-neutral pricing is intimately linked to the notion of a risk-neutral measure. In general terms a risk-neutral measure is an artificial probability measure Q, equivalent to the historical measure P, such that the discounted prices of all traded securities are Q-martingales (fair bets). We have seen in Section 2.2.2 that in the simple one-period default model a risk-neutral measure Q is simply given by an artificial default probability q such that

$$p_1(0, 1) = (1.05)^{-1}((1 - q) \cdot 1 + q \cdot 0.6).$$

Obviously, q is uniquely determined by this equation and is given by $q = 0.03$. Note that in this example the risk-neutral default probability q is higher than the real-world default probability p. This reflects risk aversion on the part of investors and is typical for real markets; empirical evidence on the relationship between the physical and historical default probabilities will be presented in Section 10.4.5.

The *risk-neutral pricing rule* states that the price of a derivative security can be computed as the mathematical expectation of the discounted pay-off under a risk-neutral measure Q. In mathematical terms the price at time t of a derivative with pay-off H and maturity $T \geqslant t$ is thus given by

$$V_t^H = E^Q\left(\exp\left(-\int_t^T r_s \, \mathrm{d}s\right) H \;\middle|\; \mathcal{F}_t\right), \tag{10.30}$$

where r_s denotes the continuously compounded default-free short rate of interest at time s and where the σ-algebra $\mathcal{F}_t$ represents the information available to investors at time t (see the discussion of filtrations in Section 10.4.1). Note that in one-period models, (10.30) reduces to the simpler expression $V_0^H = E^Q(H/(1+r_{0,1}))$, where $r_{0,1}$ is the simple interest rate for the period.

There are two theoretical justifications for risk-neutral pricing. One argument is based on absence of arbitrage: according to the first fundamental theorem of asset pricing, a model for security prices is arbitrage free if and only if it admits at least one equivalent martingale measure Q. Hence, if a financial product is to be priced in accordance with no-arbitrage principles, its price must be given by the risk-neutral pricing formula for some risk-neutral measure Q. A second justification relies on hedging: in financial models it is often possible to replicate the pay-off of a financial product by (dynamic) trading in the available assets, and in a frictionless market the cost of carrying out such a hedge is given by the risk-neutral pricing rule.

Hedging and market completeness. Next we take a closer look at the concept of hedging. We work in the simple one-period default model that was introduced in the previous paragraph. Consider an investor, e.g. an investment bank, who plans to sell derivatives on the defaultable zero-coupon bond. For concreteness we consider a *default put option* with maturity date $T = 1$. This contract pays one unit if the bond defaults and zero otherwise; it can be thought of as a simplified version of a CDS. Obviously, the pay-off of the default put is unknown at date $t = 0$ and thus constitutes a risk for the investor. A possible strategy for dealing with this risk is to form a *hedging portfolio* in the defaultable bond and in cash that reduces the risk of selling the put: suppose that at time $t = 0$ we go short 2.5 units of the bond and hold $\frac{50}{21} \approx 2.38$ units of cash. At time $t = 1$ there are two possibilities for the value V_1 of this portfolio.

- Default occurs: in which case $V_1 = (-2.5) \cdot 0.6 + \frac{50}{21} \cdot 1.05 = 1$.
- No default: in which case $V_1 = (-2.5) \cdot 1 + \frac{50}{21} \cdot 1.05 = 0$.

In either case the value V_1 of the hedge portfolio equals the pay-off of the option and we have found a so-called *replicating strategy* for the option. In particular, by forming the replicating strategy the investor completely eliminates the risk from selling the option, and the *law of one price* dictates that the fair price at $t = 0$ of the option should equal the value of the hedge portfolio at $t = 0$ given by $V_0 = (-2.5) \cdot 0.941 + \frac{50}{21} \approx 0.0285$ (otherwise either the buyer or the seller could make some risk-free profit).

To construct the portfolio in this simple one-period, two-state setting we have to consider two linear equations. Denote by ξ_1 and ξ_2 the units of the defaultable bond and the amount of cash in our portfolio. At time $t = 1$ we must have

$$\xi_1 \cdot 0.6 + \xi_2 \cdot 1.05 = 1 \quad \text{(the default case)}, \tag{10.31}$$

$$\xi_1 \cdot 1.0 + \xi_2 \cdot 1.05 = 0 \quad \text{(the no-default case)}, \tag{10.32}$$

which leads to the above values of $\xi_1 = -2.5$ and $\xi_2 = \frac{50}{21}$. In mathematical finance a derivative security is called *attainable* if there is a replicating portfolio strategy in the underlying assets. The above argument shows that in the simple one-period default model with only two states every derivative security is attainable. Such models are termed *complete*.

The fair price of the default put (the initial value V_0 of the replicating portfolio) can alternatively be computed by the risk-neutral pricing rule. Recall that the risk-neutral default probability is given by $q = 0.03$. The risk-neutral pricing rule applied to the default put thus leads to a value of $(1.05)^{-1}(0.97 \cdot 0 + 0.03 \cdot 1) = 0.0285$, which is equal to V_0. This is, of course, not a lucky coincidence; a basic result from mathematical finance states that the fair price of any attainable claim can be computed as the expected value of the discounted pay-off under a risk-neutral measure. Armed with this result, we typically first compute the price (the expected value of the discounted pay-off under a risk-neutral measure) and then determine the replicating strategy. For this reason a lot of research focuses on the problem of computing prices. However, one should bear in mind that the economic justification for the risk-neutral pricing rule stems partially from the hedging argument, which applies only to attainable claims. This issue has, to a large extent, been neglected in the literature on the pricing of credit-risky securities. The next example illustrates some of the difficulties arising in *incomplete* markets, where most derivatives are not attainable.

Example 10.8 (a model with random recovery). As there is a substantial amount of randomness in real recovery rates, it is interesting to study the impact of random recovery rates on the validity of the above pricing arguments. We consider an extension of the basic one-period default model in which the loss given default may be either 30% or 50%. The price is assumed to be $p_1(0, 1) = 0.941$ and the risk-free simple interest rate is assumed to be $r_{0,1} = 5\%$ as before. The evolution of the price $p_1(\cdot, 1)$ is illustrated in Figure 10.5. We leave the physical measure unspecified—we assume only that all three possible outcomes have strictly positive probability.

We begin our analysis of this model by determining the equivalent martingale measures. Let q_1 be the risk-neutral probability that default occurs and the LGD is 0.5, let q_2 be the risk-neutral probability that default occurs and the LGD is 0.3, and let $q_3 = 1 - q_1 - q_2$. It follows that q_1 and q_2 satisfy the equation

$$p_1(0, 1) = 1.05^{-1}(q_1 \cdot 0.5 + q_2 \cdot 0.7 + (1 - q_1 - q_2) \cdot 1), \tag{10.33}$$

with the restrictions that $q_1 > 0$, $q_2 > 0$, $1 - q_1 - q_2 > 0$. Obviously, Q is no longer unique. It is easily seen from (10.33) that the set $\mathcal{Q}$ of equivalent martingale

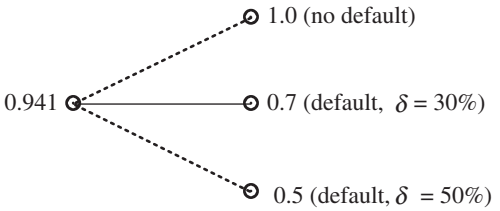

Figure 10.5. Evolution of the price $p_1(\cdot, 1)$ of the defaultable bond in Example 10.8.

measures is given by

$$\mathcal{Q} = \{\boldsymbol{q} \in \mathbb{R}^3 : q_1 \in (0, 0.024),\ q_2 = \tfrac{10}{3}(1 - 1.05 \cdot p_1(0,1) - 0.5 \cdot q_1),\ q_3 = 1 - (q_1 + q_2)\}. \tag{10.34}$$

It is interesting to look at the boundary cases. For $q_1 = 0$ we obtain $q_2 = 4\%$, $q_3 = 96\%$; this is the scenario where the risk-neutral default probability $q = q_1 + q_2$ is maximized. For $q_1 = 2.4\%$ we obtain $q_2 = 0$, $q_3 = 97.6\%$; this is the scenario where q is minimized. Note, however, that the measures $\boldsymbol{q}_0 := (0.024, 0, 0.976)$ and $\boldsymbol{q}_1 := (0, 0.04, 0.96)$ do not belong to $\mathcal{Q}$, as they are not equivalent to the physical measure P.

Consider a derivative security with pay-off H and maturity $T = 1$, such as the default put with pay-off $H = 0$ if $p_1(1, 1) = 1$ (no default) and $H = 1$ otherwise. Every price of the form $H_0 = E^Q(1.05^{-1}H)$ for some $Q \in \mathcal{Q}$ is consistent with no arbitrage and will therefore be called an *admissible value* for the derivative. If $\mathcal{Q}$ contains more than one element, as in our case, there is typically more than one admissible value. For instance, we obtain for the default put option that

$$\inf_{Q\in\mathcal{Q}} E^Q\left(\frac{H}{1.05}\right) \approx 0.023 \quad \text{and} \quad \sup_{Q\in\mathcal{Q}} E^Q\left(\frac{H}{1.05}\right) \approx 0.038; \tag{10.35}$$

obviously, the infimum and supremum in (10.35) correspond to the measures $\boldsymbol{q}_0$ and $\boldsymbol{q}_1$, where q is minimized and maximized, respectively. This non-uniqueness of admissible values reflects the fact that in our three-state model the put is no longer attainable. In fact, the hedging portfolio (ξ_1, ξ_2) now has to solve the following three equations:

$$\left.\begin{aligned} \xi_1 \cdot 0.5 + \xi_2 \cdot 1.05 &= 1 \quad \text{(default, low recovery)},\\ \xi_1 \cdot 0.7 + \xi_2 \cdot 1.05 &= 1 \quad \text{(default, high recovery)},\\ \xi_1 \cdot 1 + \xi_2 \cdot 1.05 &= 0 \quad \text{(no default)}. \end{aligned}\right\} \tag{10.36}$$

It is immediately seen that the system (10.36) of three equations and only two unknowns has no solution, so that the default put is not attainable. This illustrates two fundamental results from modern mathematical finance: a claim with bounded pay-off is attainable if and only if the set of admissible values consists of a single number; an arbitrage-free market is complete if and only if there is exactly one equivalent martingale measure Q. The latter result is known as the *second fundamental theorem of asset pricing*.

Example 10.8 shows that in an incomplete market new issues arise; in particular, it is not obvious how to choose the correct price of a derivative security from the range of admissible values or how to deal with the risk incurred by selling a derivative security. This is unfortunate, as realistic models, which capture the dynamics of financial time series, are typically incomplete. In recent years a number of interesting concepts for the risk management of derivative securities in incomplete markets have been developed. These approaches typically propose mitigating the risk by an appropriate trading strategy and often suggest a pricing formula for the remaining risk. However, a discussion of this work is outside the scope of this book. A brief overview of the existing literature on hedging in (incomplete) credit markets is given in Notes and Comments.

Advantages and limitations of risk-neutral pricing. The risk-neutral pricing approach is a *relative pricing theory*, which explains prices of credit products in terms of observable prices of other securities. If properly applied, it leads to arbitrage-free prices of credit-risky securities, which are consistent with prices quoted in the market. These features make the risk-neutral pricing approach to credit risk the method of choice in an environment where credit risk is actively traded and, in particular, for valuing credit instruments when the market for related products is relatively liquid. On the other hand, since pricing models have to be calibrated to prices of traded credit instruments, they are difficult to apply when we lack sufficient market information. Moreover, in such cases prices quoted using an ad hoc choice of some risk-neutral measure are more or less plucked out of thin air.

This can be contrasted with the more traditional pricing methodology for loans and related credit products, where a loan is taken on the balance sheet if the spread earned on the loan is deemed by the lender to be a sufficient compensation for bearing the default risk of the loan and where the default risk is measured using the real-world measure and historical (default) data. Such an approach is well suited to situations where the market for related credit instruments is relatively illiquid and little or no price information is available; loans to medium or small businesses are a prime example. On the other hand, the traditional pricing methodology does not necessarily lead to prices that are consistent (in the sense of absence of arbitrage) across products or compatible with quoted market prices for credit instruments, so it is less suitable in a trading environment.

Martingale modelling. Recall that, according to the first fundamental theorem of asset pricing, a model for security prices is arbitrage free if and (essentially) only if it admits at least one equivalent martingale measure Q. Moreover, in a complete market, the only thing that matters for the pricing of derivative securities is the Q-dynamics of the traded underlying assets. When building a model for pricing derivatives it is therefore a natural shortcut to model the objects of interest—such as interest rates, default times and the price processes of traded bonds—directly, under some exogenously specified martingale measure Q. In the literature this approach is termed *martingale modelling*.

Martingale modelling is particularly convenient if the value H of the underlying assets at some maturity date T is exogenously given, as in the case of zero-coupon bonds. In that case the price of the underlying asset at time $t < T$ can be computed as the conditional expectation under Q of the discounted value at maturity via the risk-neutral pricing rule (10.30). Model parameters are then determined using the requirement that at time $t = 0$ the price of traded securities, as computed from the model using (10.30), should coincide with the price of those securities as observed in the market; this is known as *calibration* of the model to market data.

Martingale modelling ensures that the resulting model is arbitrage free, which is advantageous if one has to model the prices of many different securities simultaneously. The approach is therefore frequently adopted in default-free term structure models and in reduced-form models for credit-risky securities. Martingale modelling has two drawbacks. First, historical information is, to a large extent, useless in estimating model parameters, as these may change in the transition from the real-world measure to the equivalent martingale measure. Second, as illustrated in Example 10.8, realistic models for pricing credit derivatives are typically incomplete, so that one cannot eliminate all risk by dynamic hedging. In those situations one is interested in the distribution of the remaining risk under the physical measure P, so martingale modelling alone is not sufficient. In summary, the martingale-modelling approach is most suitable in situations where the market for underlying securities is relatively liquid. In that case we have sufficient price information to calibrate our models, and issues of market completeness become less relevant.

10.4.3 Bond Pricing

In this section we discuss the pricing of defaultable zero-coupon bonds in hazard rate models. Note that coupon-paying corporate bonds can be represented as a portfolio of zero-coupon bonds, so our analysis applies to coupon-paying bonds as well.

Recovery models. We begin with a survey of different models for the recovery of defaultable zero-coupon bonds. As in previous sections we denote the price at time t of a defaultable zero-coupon bond with maturity $T \geqslant t$ by $p_1(t, T)$; $p_0(t, T)$ denotes the price of the corresponding default-free zero-coupon bond. The face value of these bonds is always taken to be one. The following recovery models are frequently used in the literature.

(i) *Recovery of Treasury (RT).* The RT model was proposed by Jarrow and Turnbull (1995). Under RT, if default occurs at some point in time $\tau \leqslant T$, the owner of the defaulted bond receives $(1 - \delta_\tau)$ units of the default-free zero-coupon bond $p_0(\cdot, T)$ at time τ, where $\delta_\tau \in [0, 1]$ models the percentage loss given default. At maturity T the holder of the defaultable bond therefore receives the payment $I_{\{\tau > T\}} + (1 - \delta_\tau) I_{\{\tau \leqslant T\}}$.

(ii) *Recovery of Face Value (RF).* Under RF, if default occurs at $\tau \leqslant T$, the holder of the bond receives a (possibly random) recovery payment of size $(1 - \delta_\tau)$

immediately at the default time τ. Note that even with deterministic loss given default $\delta_\tau \equiv \delta$ and deterministic interest rates, the value at maturity of the recovery payment is random as it depends on the exact timing of default.

A further recovery model, the so-called *recovery of market value*, is considered in Section 10.5.3. In real markets, recovery is a complex issue with many legal and institutional features, and all recovery models put forward in the literature are at best a crude approximation of reality. The RF assumption comes closest to legal practice, as debt with the same seniority is assigned the same (fractional) recovery, independent of the maturity. On the other hand, for "extreme" parameter values (long maturities and high risk-free interest rates), RF may lead to negative credit spreads, as we will see in Section 10.6.3. Moreover, the RF model leads to slightly more involved pricing formulas for defaultable bonds than the RT model. Empirical evidence on recovery rates for loans and bonds is discussed in Section 11.2.3.

Bond pricing. Next we turn to pricing formulas for defaultable bonds. We use martingale modelling and work directly under some martingale measure Q. We assume that under Q the default time τ is a random time with deterministic risk-neutral hazard function $\gamma^Q(t)$. The information available to investors at time t is given by $\mathcal{H}_t = \sigma(\{Y_u : u \leqslant t\})$. We take interest rates and recovery rates to be deterministic; the percentage loss given default is denoted by $\delta \in (0, 1)$, and the continuously compounded interest rate is denoted by $r(t) \geqslant 0$. Note that, in this setting, the price of the default-free zero-coupon bond with maturity $T \geqslant t$ equals

$$p_0(t, T) = \exp\left(-\int_t^T r(s)\,\mathrm{d}s\right).$$

This is the simplest type of model that can be calibrated to a given term structure of default-free interest rates and single-name credit spreads; generalizations allowing for stochastic interest rates, recovery rates and hazard rates will be discussed in Section 10.5.

The actual payments of a defaultable zero-coupon bond can be represented as a combination of a *survival claim* that pays one unit at the maturity date T and a recovery payment in case of default. The survival claim has pay-off $I_{\{\tau>T\}}$. Recall from (10.29) that

$$Q(\tau > T \mid \mathcal{H}_t) = I_{\{\tau>t\}} \exp\left(-\int_t^T \gamma^Q(s)\,\mathrm{d}s\right)$$

and define $R(t) = r(t) + \gamma^Q(t)$. The price of a survival claim at time t then equals

$$\begin{aligned} E^Q(p_0(t, T) I_{\{\tau>T\}} \mid \mathcal{H}_t) &= \exp\left(-\int_t^T r(s)\,\mathrm{d}s\right) Q(\tau > T \mid \mathcal{H}_t) \\ &= I_{\{\tau>t\}} \exp\left(-\int_t^T R(s)\,\mathrm{d}s\right). \end{aligned} \tag{10.37}$$

Note that for $\tau > t$, (10.37) can be viewed as the price of a default-free zero-coupon bond with adjusted interest rate $R(t) > r(t)$. A similar relationship between

defaultable and default-free bond prices can be established in many reduced-form credit risk models.

Under the RT model the value of the recovery payment at the maturity date T of the bond is $(1-\delta)I_{\{\tau\leqslant T\}} = (1-\delta) - (1-\delta)I_{\{\tau>T\}}$. Using (10.37), the value of the recovery payment at time $t < T$ is therefore

$$(1-\delta)p_0(t,T) - (1-\delta)I_{\{\tau>t\}}\exp\left(-\int_t^T (r(s)+\gamma^Q(s))\,\mathrm{d}s\right).$$

Under the RF hypothesis the recovery payment takes the form $(1-\delta)I_{\{\tau\leqslant T\}}$, where the payment occurs directly at time τ. Payments of this form will be referred to as *payment-at-default claims*). The value of the recovery payment at time $t \leqslant T$ therefore equals

$$E^Q\left((1-\delta)I_{\{t<\tau\leqslant T\}}\exp\left(-\int_t^\tau r(s)\,\mathrm{d}s\right)\,\middle|\,\mathcal{H}_t\right).$$

The evaluation of this expression is discussed in the following lemma.

Lemma 10.9. *Suppose that τ is a random time with hazard function $\gamma^Q(t)$, and let $R(t) = r(t) + \gamma^Q(t)$ as before. Then we have the identity*

$$E^Q\left(I_{\{t<\tau\leqslant T\}}\exp\left(-\int_t^\tau r(s)\,\mathrm{d}s\right)\,\middle|\,\mathcal{H}_t\right) = I_{\{\tau>t\}}\int_t^T \gamma^Q(s)\exp\left(-\int_t^s R(u)\,\mathrm{d}u\right)\mathrm{d}s.$$

Proof. Using Lemma 10.6 we get that

$$E^Q\left(I_{\{t<\tau\leqslant T\}}\exp\left(-\int_t^\tau r(s)\,\mathrm{d}s\right)\,\middle|\,\mathcal{H}_t\right) = I_{\{\tau>t\}}\frac{E^Q(I_{\{t<\tau\leqslant T\}}\exp(-\int_t^\tau r(s)\,\mathrm{d}s))}{\exp(-\int_0^t \gamma^Q(s)\,\mathrm{d}s)}. \quad (10.38)$$

Since τ has density

$$\gamma^Q(t)\exp\left(-\int_0^t \gamma^Q(s)\,\mathrm{d}s\right),$$

we have

$$E^Q\left(I_{\{t<\tau\leqslant T\}}\exp\left(-\int_t^\tau r(s)\,\mathrm{d}s\right)\right) = \int_t^T \exp\left(-\int_t^s r(u)\,\mathrm{d}u\right)\gamma^Q(s)\exp\left(-\int_0^s \gamma^Q(u)\,\mathrm{d}u\right)\mathrm{d}s.$$

Substitution of the right-hand side into equation (10.38) gives the result. □

10.4.4 CDS Pricing

The CDS market is among the most liquid markets for credit-risky securities, so the task of building a model using CDS spreads as input is frequently encountered in practice. In this section we therefore discuss CDS pricing and the calibration of hazard rate models to observed CDS spreads.

Pricing. We consider the following CDS contract. We take the notional to be one, so that percentage loss given default and absolute loss given default are the same. The premium payments are due at N points in time $0 < t_1 < \cdots < t_N$. If $\tau > t_k$, the protection buyer pays a premium of size $x^*(t_k - t_{k-1})$ at t_k, where x^* denotes the swap spread. After default, no further premium payments are made. If default occurs before the maturity date t_N of the swap, the protection seller makes a default payment of size δ to the buyer at the default time τ. In a standard CDS the protection buyer pays the protection seller at default the part of the premium that has accrued since the last regular premium payment date; here we ignore these accrued premium payments to simplify the exposition.

We use the same set-up as in the analysis of bond pricing in the previous section. As a first step we price the payments made by the protection buyer (the so-called premium payment leg of the swap) and the payments made by the protection seller (the default payment leg) separately, using a generic risk-neutral hazard function γ^Q and a generic spread x. The price of the premium payment leg at $t < t_N$ (the expected discounted value of the payments) is given by

$$\begin{aligned} V_t^{\text{prem}}(x;\gamma^Q) &= E^Q\bigg(\sum_{k:\,t_k>t} \exp\bigg(-\int_t^{t_k} r(u)\,\mathrm{d}u\bigg) x(t_k - t_{k-1}) I_{\{t_k<\tau\}} \;\bigg|\; \mathcal{H}_t\bigg) \\ &= x \sum_{k:\,t_k>t} p_0(t,t_k)(t_k - t_{k-1}) Q(\tau > t_k \mid \mathcal{H}_t), \end{aligned} \tag{10.39}$$

which is easily computed using the formula

$$Q(\tau > t_k \mid \mathcal{H}_t) = I_{\{\tau>t\}} \exp\bigg(-\int_t^{t_k} \gamma^Q(s)\,\mathrm{d}s\bigg).$$

The default payment leg is a typical payment-at-default claim. Using Lemma 10.9 we obtain

$$\begin{aligned} V_t^{\text{def}}(\gamma^Q) &= E^Q\bigg(\exp\bigg(-\int_t^{\tau} r(u)\,\mathrm{d}u\bigg)\delta I_{\{t<\tau\leqslant t_N\}} \;\bigg|\; \mathcal{H}_t\bigg) \\ &= I_{\{t<\tau\}}\delta \int_t^{t_N} \gamma^Q(s) \exp\bigg(-\int_t^s (r(u) + \gamma^Q(u))\,\mathrm{d}u\bigg)\,\mathrm{d}s. \end{aligned} \tag{10.40}$$

According to market convention the CDS spread x_t^* quoted for the contract at time t (the so-called *fair CDS spread* x^*) is chosen such that the value of the contract is equal to zero. Hence x_t^* is defined by the equation $V_t^{\text{prem}}(x_t^*;\gamma^Q) = V_t^{\text{def}}(\gamma^Q)$, which gives

$$x_t^* = I_{\{t<\tau\}} \frac{\delta \int_t^{t_N} \gamma^Q(s) \exp(-\int_t^s (r(u) + \gamma^Q(u))\,\mathrm{d}u)\,\mathrm{d}s}{\sum_{k:\,t_k>t} p_0(t,t_k)(t_k - t_{k-1}) \exp(-\int_t^{t_k} \gamma^Q(s)\,\mathrm{d}s)}. \tag{10.41}$$

Obviously, x_t^* depends on the hazard function γ^Q, as V_t^{prem} and V_t^{def} depend on γ^Q.

Note that in the pricing argument we have neglected the issue of counterparty risk and, in particular, the possibility that the protection seller might default before the maturity of the CDS. A discussion of counterparty risk for CDS contracts is given in Section 17.2.

Calibration. Assume now that we observe spreads quoted in the market for one or more CDSs on the same reference entity. Under the martingale-modelling approach we have to calibrate our model to the available market information: that is, we have to determine the implied risk-neutral hazard function γ^Q, which ensures that the fair CDS spreads implied by the model equal the spreads that are quoted in the market.

Suppose that the market information at time $t = 0$ consists of the fair spread x^* of one CDS with maturity t_N; the risk-neutral hazard function γ^Q is constant, so that, for all $s \geqslant 0$, $\gamma^Q(s) = \bar{\gamma}^Q$ for some $\bar{\gamma}^Q > 0$, which we refer to as the risk-neutral hazard rate. It follows from (10.39) and (10.40) that the implied risk-neutral hazard rate $\bar{\gamma}^Q$ satisfies the equation

$$x^* \sum_{k=1}^{N} p_0(0, t_k)(t_k - t_{k-1})\mathrm{e}^{-\bar{\gamma}^Q t_k} = \delta \bar{\gamma}^Q \int_0^{t_N} p_0(0, t)\mathrm{e}^{-\bar{\gamma}^Q t}\,\mathrm{d}t. \tag{10.42}$$

Here, the left-hand side equals $V_0^{\mathrm{prem}}(x^*, \bar{\gamma}^Q)$ and the right-hand side is obviously equal to $V_0^{\mathrm{def}}(\bar{\gamma}^Q)$. There is a unique implied risk-neutral hazard rate solving equation (10.42). This may be seen by first noting that $V_0^{\mathrm{prem}}(x^*, \bar{\gamma}^Q)$ is a decreasing function of $\bar{\gamma}^Q$ while $V_0^{\mathrm{def}}(\bar{\gamma}^Q)$ is an increasing function of $\bar{\gamma}^Q$. Moreover, $V_0^{\mathrm{def}}(0) = 0$, so the value of the premium payments exceeds the value of the default payment for small values of $\bar{\gamma}^Q$. On the other hand, as $\bar{\gamma}^Q$ tends to infinity, $V_0^{\mathrm{prem}}(x^*, \bar{\gamma}^Q)$ converges to zero, so $V_0^{\mathrm{prem}}(x^*, \bar{\gamma}^Q) < V_0^{\mathrm{def}}(\bar{\gamma}^Q)$ for large values of $\bar{\gamma}^Q$.

If one observes spreads for several CDSs on the same reference entity but with different maturities, a time-independent risk-neutral hazard function is generally not sufficient to calibrate the model to the observed swap spreads. Instead one typically uses hazard functions $\gamma^Q(t)$ that are piecewise constant or piecewise linear. An exception occurs in the special case where (1) the spread curve is *flat* (i.e. all CDSs on the reference entity have the same spread x^*, independent of the maturity), (2) the risk-free interest rate is constant, and (3) the time points t_k are equally spaced ($t_k - t_{k-1} = \Delta t$ for all k). In that case the implied risk-neutral hazard rate $\bar{\gamma}^Q$ is the solution of equation (10.42) in the case where $N = 1$, that is, the solution of

$$x^* \Delta t p_0(0, \Delta t)\mathrm{e}^{-\bar{\gamma}^Q \Delta t} = \delta \bar{\gamma}^Q \int_0^{\Delta t} \mathrm{e}^{-rt}\mathrm{e}^{-\bar{\gamma}^Q t}\,\mathrm{d}t. \tag{10.43}$$

For Δt relatively small (quarterly or semiannual spread payments), a good approximation to the solution of (10.43) is given by $\bar{\gamma}^Q \approx x^*/\delta$, i.e. by the ratio of the fair swap spread and the percentage loss given default. This approximation is frequently used in practice.

Note, finally, that for most issuers the implied hazard rate is relatively small (of the order of a few percentage points). We therefore have the following approximation for the one-year default probability:

$$Q(\tau \leqslant 1) = 1 - \mathrm{e}^{-\bar{\gamma}^Q} \approx \bar{\gamma}^Q \approx x^*/\delta, \tag{10.44}$$

so the quantity x^*/δ can be viewed as a proxy for the risk-neutral one-year default probability.

10.4.5 P versus Q: Empirical Results

We have now assembled the necessary technical tools to discuss some of the empirical work on the relationship between physical and risk-neutral default probabilities. Understanding this relationship is important; it enables market participants to use information about historical default probabilities in pricing credit-risky securities. Conversely, it allows the use of market quotes for CDSs or defaultable bonds as additional inputs in determining historical default probabilities.

In most empirical studies risk-neutral default probabilities are estimated from credit-spread data for CDSs. By comparing these estimates with estimates of the physical default probability—obtained, for instance, from the public-firm EDF methodology introduced in Section 10.3.3—it is possible to gain some empirical evidence on the relationship between physical and risk-neutral default probabilities in real markets. An extensive empirical study along these lines is found in Berndt et al. (2008). The authors carried out a very detailed regression analysis of the observed spreads for five-year CDSs against five-year EDFs for a large pool of firms. The five-year EDF of a firm with publicly traded stock is an annualized estimate of the physical five-year default probability. The computation of EDFs is described in detail in Section 10.3.3, and annualization is a way of expressing EDFs for different time horizons on a common yearly scale.

Berndt et al. (2008) begin by estimating a linear model for the relationship between the observed swap spread $x^*_{t,i}$ of firm i at date t and the five-year EDF of that firm on the same day, labelled $\mathrm{EDF}_{t,i}$. The model takes the form

$$x^*_{t,i} = \alpha + \beta\,\mathrm{EDF}_{t,i} + \varepsilon_{t,i}, \quad (t,i) \in S, \tag{10.45}$$

where S denotes the set of all time points/firms for which there is an observable EDF–CDS pair. The model was fitted to a sample of 33 912 EDF–CDS observations for a large set of publicly traded US firms in the period December 2000 to December 2004. The estimated coefficients were given by $\alpha = 33$ bp and $\beta = 1.6$; the R^2 was 0.73.

Berndt et al. (2008) propose the following interpretation of this regression result. Their model implies that the fair swap spread x^* of a firm increases by approximately 16 basis points for every 10 basis point increase in the five-year EDF of that firm; neglecting the intercept, we thus have that $x^*_{t,i}/\mathrm{EDF}_{t,i} \approx 1.6$. Assuming a fixed loss given default δ, we may use the quantity $q_{t,i} = x^*_{t,i}/\delta$ as a proxy for the risk-neutral default probability of firm i at time t; moreover, $\mathrm{EDF}_{t,i}$ can be viewed as a proxy for the physical default probability of firm i at time t. The ratio of risk-neutral to historical default probabilities is therefore given approximately by

$$\frac{q_{t,i}}{p_{t,i}} \approx \frac{x^*_{t,i}}{\delta\,\mathrm{EDF}_{t,i}} \approx 1.6\delta^{-1}.$$

With $\delta = 0.75$ we obtain $q_{t,i}/p_{t,i} \approx 2.13$; higher recovery rates, i.e. smaller values of δ, would lead to an even higher estimate for $q_{t,i}/p_{t,i}$. The analysis of Berndt et al. (2008) clearly shows that physical and risk-neutral default probabilities can differ substantially, and care must be taken to distinguish between the two concepts.

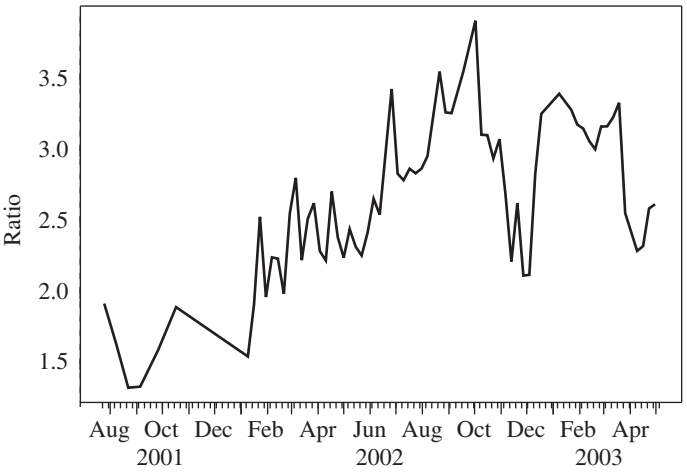

Figure 10.6. Ratio of one-year risk-neutral and historical default probabilities for Vintage Petroleum, as estimated by Berndt et al. (2008).

A careful inspection of the EDF–CDS relationship shows that the simple linear model (10.45) might not be appropriate for a number of reasons. First, the intercept of 33 basis points is implausible, as it would imply that even for a firm with historical default probability p close to zero the swap spread is still of the order of 30 basis points. Second, Berndt et al. (2008) found that the ratio x_i^*/EDF_i varies between industry sectors—reflecting different recovery rates for different industries—and over time, as is illustrated in Figure 10.6. Third, there seems to be some concavity in the relationship between swap spreads and EDFs; in particular, the ratio $x_{t,i}^*/\mathrm{EDF}_{t,i}$ is higher for high-quality firms with low EDF values than for low-quality firms. For these reasons the authors go on to consider more refined logarithmic regression models that fit the data significantly better.

Notes and Comments

Hazard rate models are a common tool in credit risk and survival analysis: see, for example, Bielecki and Rutkowski (2002) or, for a general introduction to survival analysis, the classical textbook by Cox and Oakes (1984). Further useful textbooks are Fleming and Harrington (2005), Marshall and Olkin (2007) and Aalen, Borgan and Gjessing (2010).

The fundamental theorems of asset pricing and the conceptual underpinnings of risk-neutral pricing are discussed in most textbooks on mathematical finance: see, for example, Duffie (2001), Björk (2004), Shreve (2004b) and Delbaen and Schachermayer (2006). The term martingale modelling was coined in Björk (2004) in the context of default-free short-rate models. In recent years a number of interesting approaches to the risk management of derivative securities in incomplete markets have been developed. *Quadratic hedging* approaches were first developed by Föllmer and Sondermann (1986) and Föllmer and Schweizer (1991); Schweizer (2001) is an excellent survey; *utility-based* approaches to pricing and hedging in incomplete markets are discussed in Delbaen et al. (2002) and Becherer (2004), and

the latter paper explicitly considers applications of utility-based hedging strategies to credit risk models. Papers dealing with dynamic hedging and market incompleteness in credit risk models include Bielecki, Jeanblanc and Rutkowski (2004), Bielecki, Jeanblanc and Rutkowski (2007), Frey and Backhaus (2010) and Cont and Kan (2011).

A detailed analysis of CDS pricing can be found in many sources; a good reference is Schönbucher (2003). Theoretical results on the relationship between physical and risk-neutral default probabilities were obtained by Artzner and Delbaen (1995) and Jarrow, Lando and Yu (2005). In their paper, Berndt et al. (2008) go beyond the regression analysis presented in our text and estimate a full time-series model for the joint evolution of risk-neutral and actual default intensities. Further empirical studies of the relationship between actual and risk-neutral default probabilities include Fons (1994), Bohn (2000), Driessen (2005) and Huang and Huang (2012). These results largely corroborate the findings of Berndt et al. (2008).

10.5 Pricing with Stochastic Hazard Rates

In the models with deterministic hazard functions discussed in Section 10.4, the only risk factor affecting a defaultable bond or a CDS is default risk. Hence in these models credit spreads evolve deterministically prior to default, which is clearly unrealistic. Moreover, it is not possible to price options on defaultable bonds or CDSs in such models. In this section we consider models where the hazard function is replaced by a stochastic *hazard process*. In mathematical terms this leads to the notion of doubly stochastic random times, which is discussed in Section 10.5.1. In Section 10.5.2 we derive pricing formulas for certain building blocks that can be used to value many important credit-risky securities. Applications of these formulas are studied in Section 10.5.3.

10.5.1 Doubly Stochastic Random Times

We now consider a situation where additional information affecting the distribution of the random time τ is available. Formally, we represent this additional information by some filtration $(\mathcal{F}_t)$ on the underlying probability space $(\Omega, \mathcal{F}, P)$. In credit risk models this information is typically generated by some background process $(\boldsymbol{\Psi}_t)$ representing, for instance, the risk-free interest rate or various measures of economic activity, so that $\mathcal{F}_t = \sigma(\{\boldsymbol{\Psi}_s : s \leqslant t\})$.

Consider some random time τ on $(\Omega, \mathcal{F}, P)$ with $P(\tau > 0) = 1$ and denote by $Y_t = I_{\{\tau \leqslant t\}}$ the associated jump indicator and by $(\mathcal{H}_t)$ the filtration generated by (Y_t) (see equation (10.26)). We introduce a new filtration $(\mathcal{G}_t)$ by

$$\mathcal{G}_t = \mathcal{F}_t \vee \mathcal{H}_t, \quad t \geqslant 0, \tag{10.46}$$

meaning that $\mathcal{G}_t$ is the smallest σ-algebra that contains $\mathcal{F}_t$ and $\mathcal{H}_t$. We will frequently use the notation $(\mathcal{G}_t) = (\mathcal{F}_t) \vee (\mathcal{H}_t)$ below. The filtration $(\mathcal{G}_t)$ contains information about the background processes and the occurrence or non-occurrence of default up to time t, and thus typically corresponds to the information available

to investors. Obviously, τ is an $(\mathcal{H}_t)$ stopping time and hence also a $(\mathcal{G}_t)$ stopping time. Note, however, that we do not assume that τ is a stopping time with respect to the background filtration $(\mathcal{F}_t)$.

Doubly stochastic random times are a straightforward extension of the models considered in Section 10.4 to the present set-up with additional information.

Definition 10.10 (doubly stochastic random time). A random time τ is said to be doubly stochastic if there exists a positive $(\mathcal{F}_t)$-adapted process (γ_t) such that $\Gamma_t = \int_0^t \gamma_s \, \mathrm{d}s$ is strictly increasing and finite for every $t > 0$ and such that, for all $t \geqslant 0$,

$$P(\tau > t \mid \mathcal{F}_\infty) = \exp\left(-\int_0^t \gamma_s \, \mathrm{d}s\right). \tag{10.47}$$

In that case (γ_t) is referred to as the $(\mathcal{F}_t)$-*conditional hazard process* of τ.

In (10.47) $\mathcal{F}_\infty$ denotes the smallest σ-algebra that contains $\mathcal{F}_t$ for all $t \geqslant 0$: that is, $\mathcal{F}_\infty = \sigma(\bigcup_{t \geqslant 0} \mathcal{F}_t)$. Conditioning on $\mathcal{F}_\infty$ thus means that we know the past and future economic environment and in particular the entire trajectory $(\gamma_s(\omega))_{s \geqslant 0}$ of the hazard rate process. Hence (10.47) implies that, given the economic environment, τ is a random time with deterministic hazard function given by the mapping $s \mapsto \gamma_s(\omega)$. The term *doubly stochastic* obviously refers to the fact that the hazard rate at any time is itself a realization of a stochastic process. In the literature, doubly stochastic random times are also known as *conditional Poisson* or *Cox* random times. Note, finally, that (10.47) implies that $P(\tau \leqslant t \mid \mathcal{F}_\infty)$ is $\mathcal{F}_t$-measurable, so we have the equality

$$P(\tau \leqslant t \mid \mathcal{F}_\infty) = P(\tau \leqslant t \mid \mathcal{F}_t). \tag{10.48}$$

In the next lemma we give an explicit construction of doubly stochastic random times. This construction is very useful for simulation purposes.

Lemma 10.11. *Let X be a standard exponentially distributed rv on $(\Omega, \mathcal{F}, P)$ independent of $\mathcal{F}_\infty$, i.e. $P(X \leqslant t \mid \mathcal{F}_\infty) = 1 - \mathrm{e}^{-t}$ for all $t \geqslant 0$. Let (γ_t) be a positive $(\mathcal{F}_t)$-adapted process such that $\Gamma_t = \int_0^t \gamma_s \, \mathrm{d}s$ is strictly increasing and finite for every $t > 0$. Define the random time τ by*

$$\tau := \Gamma^{\leftarrow}(X) = \inf\{t \geqslant 0 \colon \Gamma_t \geqslant X\}. \tag{10.49}$$

Then τ is doubly stochastic with $(\mathcal{F}_t)$-conditional hazard rate process (γ_t).

Proof. Note that by definition of τ it holds that $\{\tau > t\} = \{\Gamma_t < X\}$. Since Γ_t is $\mathcal{F}_\infty$-measurable and X is independent of $\mathcal{F}_\infty$, we obtain

$$P(\tau > t \mid \mathcal{F}_\infty) = P(\Gamma_t < X \mid \mathcal{F}_\infty) = \mathrm{e}^{-\Gamma_t},$$

which proves the claim. □

Lemma 10.11 has the following converse.

Lemma 10.12. *Let τ be a doubly stochastic random time with $(\mathcal{F}_t)$-conditional hazard process (γ_t). Denote by $\Gamma_t = \int_0^t \gamma_s \, \mathrm{d}s$ the $(\mathcal{F}_t)$-conditional cumulative hazard process of τ and set $X := \Gamma_\tau$. Then the rv X is standard exponentially distributed and independent of $\mathcal{F}_\infty$, and $\tau = \Gamma^{\leftarrow}(X)$ almost surely.*

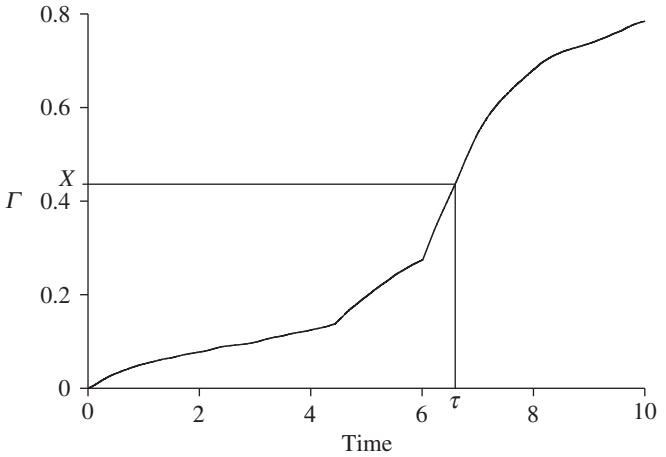

Figure 10.7. A graphical illustration of Algorithm 10.13: $X \approx 0.44$, $\tau \approx 6.59$.

Proof. Since (Γ_t) is strictly increasing by assumption, the relation $\tau = \Gamma^{\leftarrow}(X)$ is clear from the definition of X. To prove that X has the correct distribution we argue as follows:

$$P(X \leqslant t \mid \mathcal{F}_\infty) = P(\Gamma_\tau \leqslant t \mid \mathcal{F}_\infty) = P(\tau \leqslant \Gamma^{\leftarrow}(t) \mid \mathcal{F}_\infty).$$

Since τ is doubly stochastic, the last expression equals $1 - \exp(-\Gamma(\Gamma^{\leftarrow}(t))) = 1 - \mathrm{e}^{-t}$, as Γ is continuous and strictly increasing by assumption. This shows that X is independent of $\mathcal{F}_\infty$ and that it is standard exponentially distributed. □

Lemma 10.11 forms the basis for the following algorithm for the simulation of doubly stochastic random times.

Algorithm 10.13 (univariate threshold simulation).

(1) Generate a trajectory of the hazard process (γ_t). References for suitable simulation approaches are given in Notes and Comments.

(2) Generate a unit exponential rv X independent of (γ_t) (the threshold) and set $\tau = \Gamma^{\leftarrow}(X)$; this step is illustrated in Figure 10.7.

Moreover, Lemmas 10.11 and 10.12 provide an interesting interpretation of doubly stochastic random times in terms of *operational time*. For a given $(\mathcal{F}_t)$-adapted hazard process (γ_t), define a new timescale (operational time) by the associated cumulative hazard process $\Gamma_t = \int_0^t \gamma_s \, \mathrm{d}s$, so that c units of operational time correspond to $\Gamma^{\leftarrow}(c)$ units of real time. Take a standard exponential rv X independent of $\mathcal{F}_\infty$ and measure time in units of operational time. The associated calendar time $\tau := \Gamma^{\leftarrow}(X)$ is then doubly stochastic by Lemma 10.11. Conversely, by Lemma 10.12, if we take a doubly stochastic random time τ, the associated operational time $X := \Gamma_\tau$ is standard exponential, independent of $\mathcal{F}_\infty$. The notion of operational time plays an important role in insurance mathematics (see Section 13.2.7).

Intensity of doubly stochastic random times. We have seen in Proposition 10.7 that the jump indicator process (Y_t) can be turned into an $(\mathcal{H}_t)$-martingale if we subtract the process $\int_0^{t\wedge\tau} \gamma(s)\,\mathrm{d}s$, where $t \wedge \tau$ is a shorthand notation for $\min\{t, \tau\}$. We now generalize this result to doubly stochastic random times.

Proposition 10.14. *Let τ be a doubly stochastic random time with $(\mathcal{F}_t)$-conditional hazard process (γ_t). Then $M_t := Y_t - \int_0^{t\wedge\tau} \gamma_s\,\mathrm{d}s$ is a $(\mathcal{G}_t)$-martingale.*

Proof. Define a new artificial filtration $(\tilde{\mathcal{G}}_t)$ by $\tilde{\mathcal{G}}_t = \mathcal{F}_\infty \vee \mathcal{H}_t$, and note that $\tilde{\mathcal{G}}_0 = \mathcal{F}_\infty$ and $\mathcal{G}_t \subset \tilde{\mathcal{G}}_t$ for all t. As explained above, given $\mathcal{F}_\infty$, τ is a random time with deterministic hazard rate. Proposition 10.7 implies that $M_t := Y_t - \int_0^{t\wedge\tau} \gamma_s\,\mathrm{d}s$ is a martingale with respect to $(\tilde{\mathcal{G}}_t)$. Since (M_t) is $(\mathcal{G}_t)$-adapted and $\mathcal{G}_t \subset \tilde{\mathcal{G}}_t$, (M_t) is also a martingale with respect to $(\mathcal{G}_t)$. □

Finally, we relate Proposition 10.14 to the popular notion of the *intensity* of a random time.

Definition 10.15. Consider a filtration $(\mathcal{G}_t)$ and a random time τ with $(\mathcal{G}_t)$-adapted jump indicator process (Y_t). A non-negative $(\mathcal{G}_t)$-adapted process (λ_t) is called a $(\mathcal{G}_t)$-*intensity* process of the random time τ if $M_t := Y_t - \int_0^{t\wedge\tau} \lambda_s\,\mathrm{d}s$ is a $(\mathcal{G}_t)$-martingale.

In reduced-form credit risk models, (λ_t) is usually called the *default intensity* of the default time τ. It is well known that the intensity (λ_t) is uniquely defined on $\{t < \tau\}$. This is an immediate consequence of general results from stochastic calculus concerning the uniqueness of semimartingale decompositions (see, for example, Chapter 2 of Protter (2005)). Using the terminology of Definition 10.15, we may restate Proposition 10.14 in the following form: "the $(\mathcal{G}_t)$-intensity of a doubly stochastic random time τ is given by its $(\mathcal{F}_t)$-conditional hazard process (γ_t)". At this point a warning is in order: there are random times that admit an intensity in the sense of Definition 10.15 that are not doubly stochastic and for which the pricing formulas derived in Section 10.5.2 below do not hold.

Conditional expectations. Next we discuss the structure of conditional expectations with respect to the full-information σ-algebra $\mathcal{G}_t$; these results are crucial for the derivation of pricing formulas in models with doubly stochastic default times.

Proposition 10.16. *Let τ be an arbitrary random time (not necessarily doubly stochastic) such that $P(\tau > t \mid \mathcal{F}_t) > 0$ for all $t \geqslant 0$. We then have for every integrable rv X that*

$$E(I_{\{\tau>t\}}X \mid \mathcal{G}_t) = I_{\{\tau>t\}} \frac{E(I_{\{\tau>t\}}X \mid \mathcal{F}_t)}{P(\tau > t \mid \mathcal{F}_t)}.$$

Note that Proposition 10.16 allows us to replace certain conditional expectations with respect to $\mathcal{G}_t$ by conditional expectations with respect to the background information $\mathcal{F}_t$. The result is also known as the *Dellacherie formula*. In the special case where the background filtration is trivial, i.e. $\mathcal{F}_t = \{\emptyset, \Omega\}$ for all $t \geqslant 0$, Proposition 10.16 reduces to Lemma 10.6.

Proof. Standard measure-theoretic arguments show that for every $\mathcal{G}_t$-measurable rv X there is some $\mathcal{F}_t$-measurable rv $\tilde{X}$ such that $XI_{\{\tau>t\}} = \tilde{X}I_{\{\tau>t\}}$. This is quite intuitive since prior to default all information is generated by the background filtration $(\mathcal{F}_t)$; a formal proof is given in Section 5.1.1 of Bielecki and Rutkowski (2002). Now $E(I_{\{\tau>t\}}X \mid \mathcal{G}_t)$ is $\mathcal{G}_t$-measurable and zero on $\{\tau \leqslant t\}$. There is therefore an $\mathcal{F}_t$-measurable rv $\tilde{Z}$ such that $E(I_{\{\tau>t\}}X \mid \mathcal{G}_t) = I_{\{\tau>t\}}\tilde{Z}$. Taking conditional expectations with respect to $\mathcal{F}_t$ and noting that $\mathcal{F}_t \subset \mathcal{G}_t$ yields

$$E(I_{\{\tau>t\}}X \mid \mathcal{F}_t) = P(\tau > t \mid \mathcal{F}_t)\tilde{Z}.$$

Hence $\tilde{Z} = E(I_{\{\tau>t\}}X \mid \mathcal{F}_t)/P(\tau > t \mid \mathcal{F}_t)$, which proves the lemma. □

Corollary 10.17. *Let $T > t$ and assume that τ is doubly stochastic with hazard process (γ_t). If the rv $\tilde{X}$ is integrable and $\mathcal{F}_T$-measurable, we have*

$$E(I_{\{\tau>T\}}\tilde{X} \mid \mathcal{G}_t) = I_{\{\tau>t\}}E\left(\exp\left(-\int_t^T \gamma_s \,\mathrm{d}s\right)\tilde{X} \;\middle|\; \mathcal{F}_t\right).$$

Proof. Let $X := I_{\{\tau>T\}}\tilde{X}$. Since $X = I_{\{\tau>t\}}X$ (as $T > t$), Proposition 10.16 yields

$$E(I_{\{\tau>T\}}\tilde{X} \mid \mathcal{G}_t) = E(I_{\{\tau>t\}}X \mid \mathcal{G}_t) = I_{\{\tau>t\}}\mathrm{e}^{\int_0^t \gamma_s \,\mathrm{d}s}E(I_{\{\tau>T\}}\tilde{X} \mid \mathcal{F}_t),$$

where we have used the fact that

$$P(\tau > t \mid \mathcal{F}_t) = \exp\left(-\int_0^t \gamma_s \,\mathrm{d}s\right).$$

Since $\tilde{X}$ is $\mathcal{F}_T$-measurable,

$$E(I_{\{\tau>T\}}\tilde{X} \mid \mathcal{F}_t) = E(\tilde{X}P(\tau > T \mid \mathcal{F}_T) \mid \mathcal{F}_t) = E\left(\tilde{X}\exp\left(-\int_0^T \gamma_s \,\mathrm{d}s\right) \;\middle|\; \mathcal{F}_t\right),$$

and the result follows. □

Corollary 10.17 will be very useful for the pricing of various credit-risky securities in models with doubly stochastic default times. Moreover, the corollary implies that in the above setting γ_t gives a good approximation to the one-year default probability. This follows by setting $T = t + 1$ and $\tilde{X} = 1$ to obtain

$$P(\tau > t + 1 \mid \mathcal{G}_t) = I_{\{\tau>t\}}E\left(\exp\left(-\int_t^{t+1} \gamma_s \,\mathrm{d}s\right) \;\middle|\; \mathcal{F}_t\right). \tag{10.50}$$

Now assume that $\tau > t$ and that the hazard rate remains relatively stable over the time interval $[t, t + 1]$. Under these assumptions the right-hand side of (10.50) is approximated reasonably well by $\mathrm{e}^{-\gamma_t}$ and, if γ_t is small, the one-year default probability satisfies

$$P(\tau \leqslant t + 1 \mid \mathcal{G}_t) \approx 1 - \mathrm{e}^{-\gamma_t} \approx \gamma_t. \tag{10.51}$$

10.5.2 Pricing Formulas

The main result of this section concerns the pricing of three types of contingent claims that can be used as building blocks for constructing the pay-off of many important credit-risky securities. We will show that, for a default time that is doubly stochastic, the computation of prices for these claims can be reduced to a pricing problem for a corresponding default-free claim if we adjust the interest rate and replace the default-free interest rate r_t by the sum $R_t = r_t + \gamma_t$ of the default-free interest rate and the hazard rate of the default time.

The model. We consider a firm whose default time is given by a doubly stochastic random time as in Section 10.5.1. The economic background filtration represents the information generated by an arbitrage-free and complete model for non-defaultable security prices. More precisely, let $(\Omega, \mathcal{F}, (\mathcal{F}_t), Q)$ denote a filtered probability space, where Q is the equivalent martingale measure. Prices of default-free securities such as default-free bonds and the default-free rate of interest (r_t) are $(\mathcal{F}_t)$-adapted processes; $B_t = \exp(\int_0^t r_s \, \mathrm{d}s)$ models the default-free savings account.

Let τ be the default time of some company under consideration and let $Y_t = I_{\{\tau \leqslant t\}}$ be the associated default indicator process. As before we set $\mathcal{H}_t = \sigma(\{Y_s : s \leqslant t\})$ and $\mathcal{G}_t = \mathcal{F}_t \vee \mathcal{H}_t$; we assume that default is observable and that investors have access to the information contained in the background filtration $(\mathcal{F}_t)$, so that the information available to investors at time t is given by $\mathcal{G}_t$. We consider a market for credit products that is liquid enough that we may use the martingale-modelling approach, and we use Q as the pricing measure for defaultable securities. According to (10.30), the price at time t of an arbitrary, non-negative, $\mathcal{G}_T$-measurable contingent claim H is therefore given by

$$H_t = E^Q\left(\exp\left(-\int_t^T r_s \, \mathrm{d}s\right) H \;\middle|\; \mathcal{G}_t\right). \tag{10.52}$$

Finally, we assume that, under Q, the default time τ is a doubly stochastic random time with background filtration $(\mathcal{F}_t)$ and hazard process (γ_t). This latter assumption is crucial for the results that follow.

Definition 10.18. We introduce the following building blocks.

(i) A *survival claim*, i.e. an $\mathcal{F}_T$-measurable promised payment X that is made at time T if there is no default; the actual payment of the survival claim equals $X I_{\{\tau > T\}}$.

(ii) A *risky dividend stream*. Here, we consider a promised dividend stream given by the $(\mathcal{F}_t)$-adapted rate process ν_s, $0 \leqslant s \leqslant T$. The payments of a risky dividend stream stop when default occurs, so that the actual payments of this building block are given by the dividend stream with rate $\nu_t I_{\{\tau > t\}}$, $0 \leqslant t \leqslant T$.

(iii) A *payment-at-default claim* of the form $Z_\tau I_{\{\tau \leqslant T\}}$, where $Z = (Z_t)_{t \geqslant 0}$ is an $(\mathcal{F}_t)$-adapted stochastic process and where Z_τ is short for $Z_{\tau(\omega)}(\omega)$. Note that the payment is made directly at τ, provided that $\tau \leqslant T$, where T is the maturity date of the claim.

Recall from Section 10.4.3 that defaultable bonds can be viewed as portfolios of survival claims and payment-at-default claims. Credit default swaps can also be written as a combination of these claims, as will be shown in Section 10.5.3. A further example is provided by option contracts that are subject to counterparty risk. For concreteness we consider a call option on some default-free security (S_t). Denote the exercise price by K and the maturity date by T and suppose that if the writer defaults at time $\tau \leqslant T$, then the owner of the option receives a fraction $(1 - \delta_\tau)$ of the intrinsic value of the option at the time of default. This can be modelled as a combination of the survival claim $(S_T - K)^+ I_{\{\tau > T\}}$ and the payment-at-default claim $(1 - \delta_\tau)(S_\tau - K)^+ I_{\{\tau \leqslant T\}}$.

Pricing results. In the following theorem we show that the pricing of the building blocks introduced in Definition 10.18 can be reduced to a pricing problem in a default-free security market model with investor information given by the background filtration $(\mathcal{F}_t)$ and with adjusted default-free interest rate.

Theorem 10.19. *Suppose that, under Q, τ is doubly stochastic with background filtration $(\mathcal{F}_t)$ and hazard process (γ_t). Define $R_s := r_s + \gamma_s$. Assume that the rvs $\exp(-\int_t^T r_s \,\mathrm{d}s)|X|$, $\int_t^T |\nu_s| \exp(-\int_t^s r_u \,\mathrm{d}u)\,\mathrm{d}s$ and $\int_t^T |Z_s \gamma_s| \exp(-\int_t^s R_u \,\mathrm{d}u)\,\mathrm{d}s$ are all integrable with respect to Q. Then the following identities hold:*

$$E^Q\left(\exp\left(-\int_t^T r_s \,\mathrm{d}s\right) I_{\{\tau > T\}} X \;\middle|\; \mathcal{G}_t\right) = I_{\{\tau > t\}} E^Q\left(\exp\left(-\int_t^T R_s \,\mathrm{d}s\right) X \;\middle|\; \mathcal{F}_t\right), \tag{10.53}$$

$$E^Q\left(\int_t^T \nu_s I_{\{\tau > s\}} \exp\left(-\int_t^s r_u \,\mathrm{d}u\right) \mathrm{d}s \;\middle|\; \mathcal{G}_t\right) = I_{\{\tau > t\}} E^Q\left(\int_t^T \nu_s \exp\left(-\int_t^s R_u \,\mathrm{d}u\right) \mathrm{d}s \;\middle|\; \mathcal{F}_t\right), \tag{10.54}$$

$$E^Q\left(I_{\{t < \tau \leqslant T\}} \exp\left(-\int_t^\tau r_s \,\mathrm{d}s\right) Z_\tau \;\middle|\; \mathcal{G}_t\right) = I_{\{\tau > t\}} E^Q\left(\int_t^T Z_s \gamma_s \exp\left(-\int_t^s R_u \,\mathrm{d}u\right) \mathrm{d}s \;\middle|\; \mathcal{F}_t\right). \tag{10.55}$$

Proof. The integrability conditions ensure that all conditional expectations are well defined. We start with the pricing formula (10.53) for the vulnerable claim. Define the $\mathcal{F}_T$-measurable rv $\tilde{X} := \exp(-\int_t^T r_s \,\mathrm{d}s) X$. Using Corollary 10.17 with $s = T$ and $\Gamma_t = \int_0^t \gamma_s \,\mathrm{d}s$ we find that

$$E^Q(\tilde{X} I_{\{\tau > T\}} \mid \mathcal{G}_t) = I_{\{\tau > t\}} E^Q(\exp(-(\Gamma_T - \Gamma_t))\tilde{X} \mid \mathcal{F}_t).$$

Noting that $\Gamma_T - \Gamma_t = \int_t^T \gamma_s \,\mathrm{d}s$ and using the definition of $\tilde{X}$, it follows that the right-hand side equals $I_{\{\tau > t\}} E^Q(\exp(-\int_t^T R_s \,\mathrm{d}s) X \mid \mathcal{F}_t)$. The pricing formula (10.54) follows from (10.53) and the Fubini Theorem for conditional expectations. Finally,

we turn to (10.55). Lemma 10.16 implies that

$$E^Q\left(I_{\{\tau>t\}}\exp\left(-\int_t^\tau r_s\,\mathrm{d}s\right)Z_\tau I_{\{\tau\leqslant T\}}\;\middle|\;\mathcal{G}_t\right)$$
$$= I_{\{\tau>t\}}\frac{E^Q(I_{\{\tau>t\}}\exp(-\int_t^\tau r_s\,\mathrm{d}s)Z_\tau I_{\{\tau\leqslant T\}}\mid\mathcal{F}_t)}{P(\tau>t\mid\mathcal{F}_t)}. \quad (10.56)$$

Now note that

$$P(\tau\leqslant t\mid\mathcal{F}_T)=1-\exp\left(-\int_0^t\gamma_s\,\mathrm{d}s\right),$$

so the conditional density of τ given $\mathcal{F}_T$ equals $f_{\tau|\mathcal{F}_T}(t)=\gamma_t\exp(-\int_0^t\gamma_s\,\mathrm{d}s)$. Hence

$$E^Q\left(I_{\{\tau>t\}}\exp\left(-\int_t^\tau r_s\,\mathrm{d}s\right)Z_\tau I_{\{\tau\leqslant T\}}\;\middle|\;\mathcal{F}_T\right)$$
$$=\int_t^T\exp\left(-\int_t^s r_u\,\mathrm{d}u\right)Z_s\gamma_s\exp\left(-\int_0^s\gamma_u\,\mathrm{d}u\right)\mathrm{d}s$$
$$=\exp\left(-\int_0^t\gamma_u\,\mathrm{d}u\right)\int_t^T Z_s\gamma_s\exp\left(-\int_t^s R_u\,\mathrm{d}u\right)\mathrm{d}s.$$

Using iterated conditional expectations we obtain the formula

$$E^Q\left(I_{\{\tau>t\}}\exp\left(-\int_t^\tau r_s\,\mathrm{d}s\right)Z_\tau I_{\{\tau\leqslant T\}}\;\middle|\;\mathcal{F}_t\right)$$
$$=\exp\left(-\int_0^t\gamma_u\,\mathrm{d}u\right)E^Q\left(\int_t^T Z_s\gamma_s\exp\left(-\int_t^s R_u\,\mathrm{d}u\right)\mathrm{d}s\;\middle|\;\mathcal{F}_t\right),$$

and the identity (10.55) follows from (10.56). □

10.5.3 Applications

Credit default swaps. We extend our analysis of Section 10.4.4 and discuss the pricing of CDSs in models where the default time is doubly stochastic. This allows us to incorporate stochastic interest rates, recovery rates and hazard rates into the analysis.

We quickly recall the form of the payments of the CDS contract. As in our previous analysis, the premium payments are due at N points in time $0<t_1<\cdots<t_N$; at a pre-default date t_k, the protection buyer pays a premium of size $x(t_k-t_{k-1})$, where x denotes the swap spread in percentage points (again we take the nominal of the swap to be one). If $\tau\leqslant t_N$, the protection seller makes a default payment of size δ_τ to the buyer at the default time τ, where the percentage loss given default is now a general $(\mathcal{F}_t)$-adapted process. Using Theorem 10.19, both legs of the swap can be priced. The regular premium payments constitute a sequence of survival claims.

Using (10.53) the fair price of the premium leg at $t = 0$ is

$$V^{\text{prem},1} = \sum_{k=1}^{N} E^Q\left(\exp\left(-\int_0^{t_k} r_u \, \mathrm{d}u\right) x(t_k - t_{k-1}) I_{\{t_k < \tau\}}\right)$$
$$= x \sum_{k=1}^{N} (t_k - t_{k-1}) E^Q\left(\exp\left(-\int_0^{t_k} R_u \, \mathrm{d}u\right)\right).$$

The default payment leg is a payment-at-default claim with $Z_s = \delta_s$ and maturity t_N, so its value is given by $V^{\text{def}} = E^Q(\int_0^{t_N} \delta_s \gamma_s \exp(-\int_0^s R_u \, \mathrm{d}u) \, \mathrm{d}s)$. We have therefore reduced the pricing of credit default swaps to a pricing problem in the default-free world. Methods for solving this problem will be discussed in the next section.

Recovery of market value. Recovery of market value, abbreviated RM, is an alternative recovery model for defaultable bonds and other credit-risky securities that has been put forward by Duffie and Singleton (1999); its main virtue is the fact that it leads to particularly simple pricing formulas. Consider a claim whose payoff consists of the survival claim X and a recovery payment at the default time. Under the RM hypothesis it is assumed that this recovery payment is given by $(1 - \delta_\tau) V_\tau I_{\{\tau \leqslant T\}}$, where the $(\mathcal{F}_t)$-adapted process $(\delta_t) \in (0, 1)$ gives the percentage loss given default of the claim and where the $(\mathcal{F}_t)$-adapted process (V_t) gives the pre-default value of the claim. Note that this is a recursive definition, as the pre-default value at time t also depends on the form of the recovery payments in the time period $(t, T]$. Nonetheless, the following result can be established.

Proposition 10.20. *Suppose that, under Q, τ is doubly stochastic with hazard rate process (γ_t). Suppose, moreover, that X is integrable and that the RM assumption holds. Then the pre-default value process (V_t) is uniquely determined and is given by*

$$V_t = E^Q\left(\exp\left(-\int_t^T (r_s + \delta_s \gamma_s) \, \mathrm{d}s\right) X \;\middle|\; \mathcal{F}_t\right), \quad 0 \leqslant t \leqslant T. \tag{10.57}$$

Note that for $\delta_t \equiv 1$ the claim is a standard survival claim; in that case, (10.57) reduces to the formula (10.53). On the other hand, for $\delta_t \equiv 0$ the claim is essentially default free; in that case, (10.57) reduces to the standard pricing formula for the claim X in a default-free security market model. For a proof of Proposition 10.20 we refer to the references given in Notes and Comments.

Credit spreads and hazard rates. With doubly stochastic default times the risk-neutral hazard process (γ_t) and the credit spread

$$c(t, T) = -\frac{1}{T - t}(\ln p_1(t, T) - \ln p_0(t, T))$$

of defaultable bonds are closely related. Analytic results are most easily derived for the instantaneous credit spread given by

$$c(t, t) = \lim_{T \to t} c(t, T) = -\frac{\partial}{\partial T}\bigg|_{T=t} (\ln p_1(t, T) - \ln p_0(t, T)). \tag{10.58}$$

Assume that $\tau > t$, so that $p_1(t,t) = p_0(t,t) = 1$. We therefore obtain

$$\left.\frac{\partial}{\partial T}\right|_{T=t} \ln p_1(t,T) = \left.\frac{\partial}{\partial T}\right|_{T=t} p_1(t,T), \tag{10.59}$$

and similarly for $p_0(t,T)$. To compute the derivative in (10.59) we need to distinguish between the different recovery models. Under the RM hypothesis we can apply Proposition 10.20 with $X = 1$. Exchanging expectation and differentiation we obtain

$$\begin{aligned} -\left.\frac{\partial}{\partial T}\right|_{T=t} p_1(t,T) &= -E^Q\left(\left.\frac{\partial}{\partial T}\right|_{T=t} \exp\left(-\int_t^T (r_s + \delta_s\gamma_s)\,\mathrm{d}s\right) \,\middle|\, \mathcal{F}_t\right) \\ &= r_t + \delta_t\gamma_t. \end{aligned} \tag{10.60}$$

Applying (10.60) with $\delta_t \equiv 0$ yields

$$-\left.\frac{\partial}{\partial T}\right|_{T=t} p_0(t,T) = r_t,$$

so that $c(t,t) = \delta_t\gamma_t$, i.e. the instantaneous credit spread equals the product of the hazard rate and the percentage loss given default, which is quite intuitive from an economic point of view. Under the RF recovery model, $p_1(t,T)$ is given by the sum of the price of a survival claim $I_{\{\tau>T\}}$ and a payment at default of size $(1-\delta_\tau)$. Equation (10.60) with $\delta_t \equiv 1$ shows that the derivative with respect to T of the survival claim at $T = t$ is equal to $-(r_t + \gamma_t)$. For the recovery payment we get

$$\left.\frac{\partial}{\partial T}\right|_{T=t} E\left(\int_t^T \gamma_s(1-\delta_s)\exp\left(-\int_t^s R_u\,\mathrm{d}u\right)\mathrm{d}s \,\middle|\, \mathcal{F}_t\right) = (1-\delta_t)\gamma_t.$$

Hence

$$-\left.\frac{\partial}{\partial T}\right|_{T=t} p_1(t,T) = r_t + \gamma_t - (1-\delta_t)\gamma_t = r_t + \delta_t\gamma_t,$$

so that $c_1(t,t)$ is again equal to $\delta_t\gamma_t$. An analogous computation shows that we also have $c_1(t,t) = \delta_t\gamma_t$ under RT. However, for $T - t > 0$, the credit spread corresponding to the different recovery models differs, as is illustrated in Section 10.6.3.

Notes and Comments

The material discussed in this section is based on many sources. We mention in particular the books by Lando (2004) and Bielecki and Rutkowski (2002). The text by Bielecki and Rutkowski is more technical than our presentation; among other things the authors discuss various probabilistic characterizations of doubly stochastic random times. The threshold-simulation approach for doubly stochastic random times requires the simulation of trajectories of the hazard process. An excellent source for simulation techniques for stochastic processes is Glasserman (2003).

More general reduced-form models where the default time τ is not doubly stochastic are discussed, for example, in Kusuoka (1999), Elliott, Jeanblanc and Yor (2000), Bélanger, Shreve and Wong (2004), Collin-Dufresne, Goldstein and Hugonnier (2004) and Blanchet-Scalliet and Jeanblanc (2004).

Theorem 10.19 is originally due to Lando (1998); related results were obtained by Jarrow and Turnbull (1995) and Jarrow, Lando and Turnbull (1997). Proposition 10.20 is due to Duffie and Singleton (1999); extensions are discussed in Becherer and Schweizer (2005). An excellent text for the overall mathematical background is Jeanblanc, Yor and Chesney (2009).

The analogy with default-free term structure models makes the reduced-form models with doubly stochastic default times relatively easy to apply. However, some care is required in interpreting the results and applying the *linear pricing rules* for corporate debt that the models imply. In particular, one must bear in mind that in these models the default intensity does not explicitly take into account the structure of a firm's outstanding risky debt. A formal analysis of the effect of debt structure on bond values is best carried out in the context of firm-value models, where the default is explicitly modelled in terms of fundamental economic quantities. A good discussion of these issues can be found in Chapter 2 of Lando (2004).

10.6 Affine Models

In order to apply the pricing formulas for doubly stochastic random times obtained in Theorem 10.19 we need effective ways to evaluate the conditional expectations on the right-hand side of equations (10.53), (10.54) and (10.55). In most models, where default is modelled by a doubly stochastic random time, (r_t) and (γ_t) are modelled as functions of some p-dimensional Markovian state variable process $(\boldsymbol{\Psi}_t)$ with state space given by the domain $D \subset \mathbb{R}^p$, so that the natural background filtration is given by $(\mathcal{F}_t) = \sigma(\{\boldsymbol{\Psi}_s : s \leqslant t\})$. Moreover, $R_t := r_t + \gamma_t$ is of the form $R_t = R(\boldsymbol{\Psi}_t)$ for some function $R\colon D \subseteq \mathbb{R}^p \to \mathbb{R}_+$. We therefore have to compute conditional expectations of the form

$$E\left(\exp\left(-\int_t^T R(\boldsymbol{\Psi}_s)\,\mathrm{d}s\right)g(\boldsymbol{\Psi}_T) + \int_t^T h(\boldsymbol{\Psi}_s)\exp\left(-\int_t^s R(\boldsymbol{\Psi}_u)\,\mathrm{d}u\right)\mathrm{d}s \;\middle|\; \mathcal{F}_t\right) \tag{10.61}$$

for generic functions $g, h\colon D \to \mathbb{R}_+$. Since $(\boldsymbol{\Psi}_t)$ is a Markov process, this conditional expectation is given by some function $f(t, \boldsymbol{\Psi}_t)$ of time and the current value $\boldsymbol{\Psi}_t$ of the state variable process. It is well known that under some additional regularity assumptions the function f can be computed as solution of a parabolic partial differential equation (PDE)—this is the celebrated *Feynman–Kac formula*. The Feynman–Kac formula provides a way to determine f using analytical or numerical techniques for PDEs. In particular, it is known that in the case where $(\boldsymbol{\Psi}_t)$ belongs to the class of *affine jump diffusions* (see below), R is an affine function, $g(\boldsymbol{\psi}) = \mathrm{e}^{\boldsymbol{u}'\boldsymbol{\psi}}$ for some $\boldsymbol{u} \in \mathbb{R}^p$ and $h \equiv 0$, the function f takes the form

$$f(t, \boldsymbol{\psi}) = \exp(\alpha(t, T) + \boldsymbol{\beta}(t, T)'\boldsymbol{\psi}) \tag{10.62}$$

for deterministic functions $\alpha\colon [0, T] \to \mathbb{R}$ and $\boldsymbol{\beta}\colon [0, T] \to R^p$; moreover, α and $\boldsymbol{\beta}$ are determined by a $(p+1)$-dimensional ordinary differential equation (ODE) system that is easily solved numerically. Models based on affine jump diffusions and an affine specification of R are therefore relatively easy to implement, which

explains their popularity in practice. A relationship of the form (10.62) is often termed an *affine term structure*, as it implies that continuously compounded yields of bonds at time t are affine functions of $\boldsymbol{\Psi}_t$.

In this section we discuss these results. We concentrate on the case where the state variable process is given by a one-dimensional diffusion; extensions to processes with jumps will be considered briefly at the end.

10.6.1 Basic Results

The PDE characterization of f. We assume that the state variable process (Ψ_t) is the unique solution of the SDE

$$\mathrm{d}\Psi_t = \mu(\Psi_t)\,\mathrm{d}t + \sigma(\Psi_t)\,\mathrm{d}W_t, \quad \Psi_0 = \psi \in D, \tag{10.63}$$

with state space given by the domain $D \subseteq \mathbb{R}$. Here, (W_t) is a standard, one-dimensional Brownian motion on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), P)$, and μ and σ are continuous functions from D to $\mathbb{R}$ and D to $\mathbb{R}_+$, respectively. The next result shows that the conditional expectation (10.61) can be computed as the solution of a parabolic PDE.

Lemma 10.21 (Feynman–Kac). *Consider generic functions $R, g, h\colon D \to \mathbb{R}_+$. Suppose that the function $f\colon [0,T] \times D \to \mathbb{R}$ is continuous, once continuously differentiable in t and twice continuously differentiable in ψ on $[0,T) \times D$, and that f solves the terminal-value problem*

$$\left.\begin{aligned} f_t + \mu(\psi) f_\psi + \tfrac{1}{2}\sigma^2(\psi) f_{\psi\psi} + h(\psi) = R(\psi) f, \quad (t,\psi) \in [0,T) \times D, \\ f(T,\psi) = g(\psi), \quad \psi \in D. \end{aligned}\right\} \tag{10.64}$$

If f is bounded or, more generally, if $\max_{0\leqslant t\leqslant T} f(t,\psi) \leqslant C(1+\psi^2)$ for $\psi \in D$, then

$$E\left(\exp\left(-\int_t^T R(\boldsymbol{\Psi}_s)\,\mathrm{d}s\right) g(\boldsymbol{\Psi}_T) + \int_t^T h(\boldsymbol{\Psi}_s) \exp\left(-\int_t^s R(\boldsymbol{\Psi}_u)\,\mathrm{d}u\right)\mathrm{d}s \;\middle|\; \mathcal{F}_t\right) = f(t,\Psi_t). \tag{10.65}$$

The Feynman–Kac formula is a standard result of stochastic calculus and it is discussed in many textbooks on stochastic processes and financial mathematics, so we omit the proof (references are given in Notes and Comments).

Affine term structure. We begin with the case $h \equiv 0$; in financial terms this means that we concentrate on survival claims. The following assumption ensures that for $h \equiv 0$ the solution of the PDE (10.64), with terminal condition $g(\psi) = \mathrm{e}^{u\psi}$, $u\psi \leqslant 0$, for $\psi \in D$, is of the form (10.62), so that we have an affine term structure. Note that $g \equiv 1$ for $u = 0$; this is the appropriate terminal condition for pricing zero-coupon bonds.

Assumption 10.22. *R, μ and σ^2 are affine functions of ψ, i.e. there are constants ρ^0, ρ^1, k^0, k^1, h^0 and h^1 such that $R(\psi) = \rho^0 + \rho^1\psi$, $\mu(\psi) = k^0 + k^1\psi$ and $\sigma^2(\psi) = h^0 + h^1\psi$. Moreover, for all $\psi \in D$ we have $h^0 + h^1\psi \geqslant 0$ and $\rho_0 + \rho_1\psi \geqslant 0$.*

Fix some $T > 0$. We try to find a solution of (10.64) of the form $\tilde{f}(t,\psi) = \exp(\alpha(t,T) + \beta(t,T)\psi)$ for continuously differentiable functions $\alpha(\cdot,T)$ and $\beta(\cdot,T)$. As $\tilde{f}(T,\psi) = \mathrm{e}^{u\psi}$, we immediately obtain the terminal conditions $\alpha(T,T) = 0$, $\beta(T,T) = u$. Denote by $\alpha'(\cdot,T)$ and $\beta'(\cdot,T)$ the derivatives of α and β with respect to t. Using the special form of $\tilde{f}$ we obtain that

$$\tilde{f}_t = (\alpha' + \beta'\psi)\tilde{f}, \quad \tilde{f}_\psi = \beta\tilde{f} \quad \text{and} \quad \tilde{f}_{\psi\psi} = \beta^2\tilde{f}.$$

Hence, under Assumption 10.22 the PDE (10.64) takes the form

$$(\alpha' + \beta'\psi)\tilde{f} + (k^0 + k^1\psi)\beta\tilde{f} + \tfrac{1}{2}(h^0 + h^1\psi)\beta^2\tilde{f} = (\rho^0 + \rho^1\psi)\tilde{f}.$$

Dividing by $\tilde{f}$ and rearranging we obtain

$$\alpha' + k^0\beta + \tfrac{1}{2}h^0\beta^2 - \rho^0 + (\beta' + k^1\beta + \tfrac{1}{2}h^1\beta^2 - \rho^1)\psi = 0.$$

Since this equation must hold for all $\psi \in D$, we obtain the following ODE system:

$$\beta'(t,T) = \rho^1 - k^1\beta(t,T) - \tfrac{1}{2}h^1\beta^2(t,T), \quad \beta(T,T) = u, \tag{10.66}$$

$$\alpha'(t,T) = \rho^0 - k^0\beta(t,T) - \tfrac{1}{2}h^0\beta^2(t,T), \quad \alpha(T,T) = 0. \tag{10.67}$$

The ODE (10.66) for $\beta(\cdot,T)$ is a so-called *Ricatti equation*. While explicit solutions exist only in certain special cases, the ODE is easily solved numerically. The ODE (10.67) for $\alpha(\cdot,T)$ can be solved by simple (numerical) integration once β has been determined. Summing up, we have the following proposition.

Proposition 10.23. *Suppose that Assumption 10.22 holds, that the ODE system (10.66), (10.67) has a unique solution (α,β) on $[0,T]$, and that there is some C such that $\beta(t,T)\psi \leqslant C$ for all $t \in [0,T]$, $\psi \in D$. Then*

$$E\left(\exp\left(-\int_t^T R(\Psi_s)\,\mathrm{d}s\right)\mathrm{e}^{u\Psi_T} \;\middle|\; \mathcal{F}_t\right) = \exp(\alpha(t,T) + \beta(t,T)\Psi_t).$$

Proof. The result follows immediately from Lemma 10.21, as our assumption on β implies that $\tilde{f}(t,\psi) = \exp(\alpha(t,T) + \beta(t,T)\psi)$ is bounded. □

10.6.2 The CIR Square-Root Diffusion

A very popular affine model is the square-root diffusion model proposed by Cox, Ingersoll and Ross (1985) as a model for the short rate of interest. In this model (Ψ_t) is given by the solution of the SDE

$$\mathrm{d}\Psi_t = \kappa(\bar{\theta} - \Psi_t)\,\mathrm{d}t + \sigma\sqrt{\Psi_t}\,\mathrm{d}W_t, \quad \Psi_0 = \psi > 0, \tag{10.68}$$

for parameters $\kappa, \bar{\theta}, \sigma > 0$ and state space $D = [0,\infty)$. Clearly, (10.68) is an affine model in the sense of Assumption 10.22; the parameters are given by $k^0 = \kappa\bar{\theta}$, $k^1 = -\kappa$, $h^0 = 0$ and $h^1 = \sigma^2$.

It is well known that the SDE (10.68) admits a global solution (see Notes and Comments for a reference). This issue is non-trivial since the square-root function is not Lipschitz and since one has to ensure that the solution remains in D for all $t > 0$. Note that (10.68) implies that (Ψ_t) is a *mean-reverting process*: if Ψ_t deviates from the mean-reversion level $\bar{\theta}$, the process is pulled back towards $\bar{\theta}$. Moreover, if the mean reversion is sufficiently strong relative to the volatility, trajectories never reach zero. More precisely, let $\tau_0(\Psi) := \inf\{t \geqslant 0 \colon \Psi_t = 0\}$. It is well known that for $\kappa\bar{\theta} \geqslant \frac{1}{2}\sigma^2$ one has $P(\tau_0(\Psi) < \infty) = 0$, whereas for $\kappa\bar{\theta} < \frac{1}{2}\sigma^2$ one has $P(\tau_0(\Psi) < \infty) = 1$.

In the CIR square-root model the Ricatti equations (10.66) and (10.67) can be solved explicitly. Using Proposition 10.23 it can be computed that

$$E\left(\exp\left(-\int_t^T (\rho^0 + \rho^1 \Psi_s)\,\mathrm{d}s\right) \,\middle|\, \mathcal{F}_t\right) = \exp(\alpha(T-t) + \beta(T-t)\Psi_t),$$

with

$$\beta(\tau) = \frac{-2\rho^1(\mathrm{e}^{\gamma\tau} - 1)}{\gamma - \kappa + \mathrm{e}^{\gamma\tau}(\gamma + \kappa)}, \tag{10.69}$$

$$\alpha(\tau) = -\rho^0\tau + 2\frac{\kappa\bar{\theta}}{\sigma^2}\ln\left(\frac{2\gamma\mathrm{e}^{\tau(\gamma+\kappa)/2}}{\gamma - \kappa + \mathrm{e}^{\gamma\tau}(\gamma + \kappa)}\right), \tag{10.70}$$

and $\tau := T - t$, $\gamma := \sqrt{\kappa^2 + 2\sigma^2\rho^1}$. These formulas are the key to pricing bonds in models where the risk-free short rate and the default intensities are affine functions of independent square-root processes, as is shown in the next example.

Example 10.24 (a three-factor model). We now consider the pricing of zero-coupon bonds in a three-factor model similar to models that are frequently used in the literature. We assume that $\boldsymbol{\Psi}_t = (\Psi_{t,1}, \Psi_{t,2}, \Psi_{t,3})'$ is a vector of three independent square-root diffusions with dynamics $\mathrm{d}\Psi_{t,i} = \kappa_i(\bar{\theta}_i - \Psi_{t,i})\,\mathrm{d}t + \sigma_i\sqrt{\Psi_{t,i}}\,\mathrm{d}W_{t,i}$ for independent Brownian motions $(W_{t,i})$, $i = 1, 2, 3$. The risk-free short rate of interest is given by $r_t = r_0 + \Psi_{t,2} - \Psi_{t,1}$ for a constant $r_0 \geqslant 0$; the hazard rate of the counterparty under consideration is given by $\gamma_t = \gamma_1\Psi_{t,1} + \Psi_{t,3}$ for some constant $\gamma_1 > 0$. This parametrization allows for negative instantaneous correlation between (r_t) and (γ_t), which is in line with empirical evidence. Note, however, that this negative correlation comes at the expense of possibly negative risk-free interest rates. In this context the price of a default-free zero-coupon bond is given by

$$\begin{aligned} &p_0(t,T) \\ &\quad = E\left(\exp\left(-\int_t^T r_s\,\mathrm{d}s\right) \,\middle|\, \mathcal{F}_t\right) \\ &\quad = \mathrm{e}^{-r_0(T-t)} E\left(\exp\left(-\int_t^T \Psi_{s,2}\,\mathrm{d}s\right) \,\middle|\, \mathcal{F}_t\right) E\left(\exp\left(\int_t^T \Psi_{s,1}\,\mathrm{d}s\right) \,\middle|\, \mathcal{F}_t\right), \end{aligned} \tag{10.71}$$

where we have used the independence of $(\Psi_{t,1})$, $(\Psi_{t,2})$, $(\Psi_{t,3})$. Each of the terms in (10.71) can be evaluated using the above formulas for α and β (equations (10.69)

and (10.70)). Assuming that we have recovery of treasury in default (see Section 10.4.3) and a deterministic percentage loss given default δ, we find that the price of a defaultable zero-coupon bond is given by

$$p_1(t,T) = (1-\delta)\,p_0(t,T) + \delta E\left(\exp\left(-\int_t^T (r_s+\gamma_s)\,\mathrm{d}s\right)\,\middle|\,\mathcal{F}_t\right).$$

By definition of r_t and γ_t the last term on the right-hand side equals

$$\delta E\left(\exp\left(-\int_t^T (r_0 + (\gamma_1 - 1)\Psi_{s,1} + \Psi_{s,2} + \Psi_{s,3})\,\mathrm{d}s\right)\,\middle|\,\mathcal{F}_t\right),$$

which can be evaluated in a similar way to the evaluation of expression (10.71). In the next section we will show how to deal with more complicated recovery models, such as recovery of face value.

10.6.3 Extensions

A jump-diffusion model for (Ψ_t). We briefly discuss an extension of the basic model (10.63), where the economic factor process (Ψ_t) follows a diffusion with jumps. Adding jumps to the dynamics of (Ψ_t) provides more flexibility for modelling default correlations in models with conditionally independent defaults (see Section 17.3.2), and it also leads to more realistic credit spread dynamics.

In this section we assume that (Ψ_t) is the unique solution of the SDE

$$\mathrm{d}\Psi_t = \mu(\Psi_t)\,\mathrm{d}t + \sigma(\Psi_t)\,\mathrm{d}W_t + \mathrm{d}Z_t, \quad \Psi_0 = \psi \in D. \tag{10.72}$$

Here, (Z_t) is a pure jump process whose jump intensity at time t is equal to $\lambda^Z(\Psi_t)$ for some function $\lambda^Z\colon D \to \mathbb{R}_+$ and whose jump-size distribution has df ν on $\mathbb{R}$. Intuitively this means that, given the trajectory $(\Psi_t(\omega))_{t\geqslant 0}$ of the factor process, (Z_t) jumps at the jump times of an inhomogeneous Poisson process (see Section 13.2.7) with time-varying intensity $\lambda^Z(t,\Psi_t)$; the size of the jumps has df ν.

Suppose now that Assumption 10.22 holds, and that $\lambda^Z(\psi) = l^0 + l^1\psi$ for constants l^0, l^1 such that $\lambda^Z(\psi) > 0$ for all $\psi \in D$. In that case we say that (Ψ_t) follows an *affine jump diffusion*. For $x \in \mathbb{R}$ denote by $\hat{\nu}(x) = \int_{\mathbb{R}} \mathrm{e}^{-xy}\,\mathrm{d}\nu(y) \in (0,\infty]$ the extended Laplace–Stieltjes transform of ν (with domain $\mathbb{R}$ instead of the usual domain $[0,\infty)$). Consider the following extension of the ODE system (10.66), (10.67):

$$\beta'(t,T) = \rho^1 - k^1\beta(t,T) - \tfrac{1}{2}h^1\beta^2(t,T) - l^1(\hat{\nu}(-\beta(t,T)) - 1), \tag{10.73}$$

$$\alpha'(t,T) = \rho^0 - k^0\beta(t,T) - \tfrac{1}{2}h^0\beta^2(t,T) - l^0(\hat{\nu}(-\beta(t,T)) - 1), \tag{10.74}$$

with terminal condition $\beta(T,T) = u$ for some $u \leqslant 0$ and $\alpha(T,T) = 0$. Suppose that the system described by (10.74) and (10.73) has a unique solution α, β and that $\beta(t,T)\psi \leqslant C$ for all $t \in [0,T]$, $\psi \in D$ (for $l^0 \neq 0$ or $l^1 \neq 0$ this implicitly implies that $\hat{\nu}(-\beta(t,T)) < \infty$ for all t). Define $\tilde{f}(t,\psi) = \exp(\alpha(t,T) + \beta(t,T)\psi)$. Using similar arguments to those above it can then be shown that the conditional expectation $E(\exp(-\int_t^T R(\Psi_s)\,\mathrm{d}s)\mathrm{e}^{u\Psi_T} \mid \mathcal{F}_t)$ equals $\tilde{f}(t,\Psi_t)$.

Table 10.4. Parameters used in the model of Duffie and Gârleanu (2001). Recall that l^0 gives the intensity of jump in the factor process, μ gives the average jump size. With these parameters the average waiting time for a jump in the systematic factor process is $1/l^0 = 5$ years.

κ	$\bar{\theta}$	σ	l^0	μ
0.6	0.02	0.14	0.2	0.1

Example 10.25 (the model of Duffie and Gârleanu (2001)). The following jump-diffusion model has been used in the literature on CDO pricing. The dynamics of (Ψ_t) are given by

$$\mathrm{d}\Psi_t = \kappa(\bar{\theta} - \Psi_t)\,\mathrm{d}t + \sigma\sqrt{\Psi_t}\,\mathrm{d}W_t + \mathrm{d}Z_t \tag{10.75}$$

for parameters $\kappa, \bar{\theta}, \sigma > 0$ and a jump process (Z_t) with constant jump intensity $l^0 > 0$ and exponentially distributed jump sizes with parameter $1/\mu$. Following Duffie and Gârleanu, we will sometimes call the model (10.75) a *basic affine jump diffusion*. Note that these assumptions imply that the mean of ν is equal to μ and that ν has support $[0, \infty)$, so that (Ψ_t) has only upwards jumps. The existence of a solution to (10.75) therefore follows from the existence of solutions in the pure diffusion case. It is relatively easy to show that for $t \to \infty$ we obtain $E(\Psi_t) \to \bar{\theta} + l^0\mu/\kappa$. For illustrative purposes we present the parameter values used in Duffie and Gârleanu (2001) in Table 10.4; a typical trajectory of (Ψ_t) is simulated in Figure 10.8. Next we compute the Laplace–Stieltjes transform $\hat{\nu}$. We obtain for $u > -1/\mu$ that

$$\hat{\nu}(u) = \int_0^\infty \mathrm{e}^{-ux}(1/\mu)\mathrm{e}^{-x/\mu}\,\mathrm{d}x = \frac{1}{1+\mu u};$$

for $u \leqslant -1/\mu$ we get $\hat{\nu}(u) = \infty$. We therefore have all the necessary ingredients to set up the Ricatti equations (10.74) and (10.73). In the case of the model (10.75) it is in fact possible to solve these equations explicitly (see, for example, Chapter 11 of Duffie and Singleton (2003)). However, the explicit solution is given by a very lengthy expression so we omit the details.

Application to payment-at-default claims. According to Theorem 10.19, in a model with a doubly stochastic default time τ with risk-neutral hazard rate $\gamma(\Psi_t)$, the price at t of a payment-at-default claim of constant size $\delta > 0$ equals

$$\delta E\left(\int_t^T \gamma(\Psi_s)\exp\left(-\int_t^s R(\Psi_u)\,\mathrm{d}u\right)\mathrm{d}s \;\middle|\; \mathcal{F}_t\right), \tag{10.76}$$

where again $R(\psi) = r(\psi) + \gamma(\psi)$. Using the Fubini Theorem this equals

$$\delta\int_t^T E\left(\gamma(\Psi_s)\exp\left(-\int_t^s R(\Psi_u)\,\mathrm{d}u\right) \;\middle|\; \mathcal{F}_t\right)\mathrm{d}s. \tag{10.77}$$

Suppose now that $\gamma(\psi) = \gamma^0 + \gamma^1\psi$, that $R(\psi) = \rho^0 + \rho^1\psi$ and that (Ψ_t) is given by an affine jump diffusion as introduced above. In that case the inner expectation in (10.77) is given by a function $F(t, s, \Psi_t)$. This function can be computed using an extension of the basic affine methodology, so that (10.77) can be

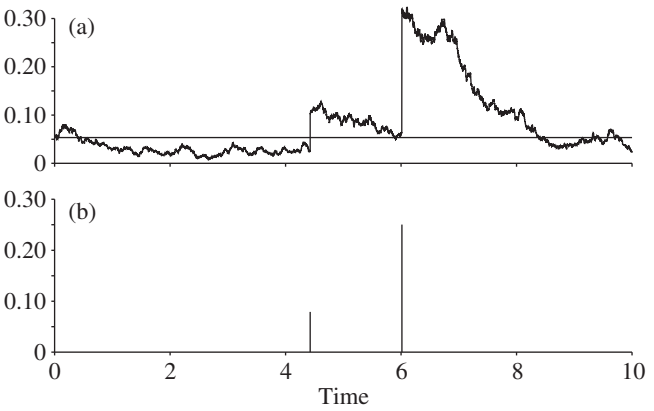

Figure 10.8. (a) A typical trajectory of the basic affine jump diffusion model (10.75) and (b) the corresponding jumps of (Z_t). The parameter values used are given in Table 10.4; the initial value Ψ_0 is equal to the long-run mean $\bar{\theta} + (l^0\mu)/\kappa$, which is marked by the horizontal line.

computed by one-dimensional numerical integration. Define for $0 \leqslant t \leqslant s$ the function $\tilde{f}(t, s, \psi) = \exp(\alpha(t, s) + \beta(t, s)\psi)$, where $\alpha(\cdot, s)$ and $\beta(\cdot, s)$ solve the ODEs (10.74), (10.73) with terminal condition $\alpha(s, s) = \beta(s, s) = 0$. Denote by $\hat{\nu}'(x)$ the derivative of the Laplace–Stieltjes transform of ν. It is then a straightforward application of standard calculus to show that, modulo some integrability conditions, $F(t, s, \psi) = \tilde{f}(t, s, \psi)(A(t, s) + B(t, s)\psi)$, where $A(\cdot, s)$ and $B(\cdot, s)$ solve the following ODE system:

$$B'(t, s) + k^1 B(t, s) + h^1 \beta B(t, s) - l^1 \hat{\nu}'(-\beta) B(t, s) = 0, \tag{10.78}$$

$$A'(t, s) + k^0 B(t, s) + h^0 \beta B(t, s) - l^0 \hat{\nu}'(-\beta) B(t, s) = 0, \tag{10.79}$$

with terminal conditions $A(s, s) = \gamma_0$, $B(s, s) = \gamma_1$. Again, (10.78) and (10.79) are straightforward to evaluate numerically.

It is of course possible to compute the conditional expectation (10.76) by using the Feynman–Kac formula (10.65) with $g \equiv 0$ and $h = \gamma$. However, in most cases the ensuing PDE needs to be solved numerically.

Example 10.26 (defaultable zero-coupon bonds and CDSs). We now have all the necessary ingredients to compute prices and credit spreads of defaultable zero-coupon bonds and CDS spreads in a model with a doubly stochastic default time with hazard rate $\gamma_t = \Psi_t$ for a one-dimensional affine jump diffusion (Ψ_t). In Figure 10.9 we plot the credit spread for defaultable bonds for the recovery assumptions discussed in Section 10.4.3. Note that, for $T \to t$, i.e. for time to maturity close to zero, the spread tends to $c(t, t) = \delta\Psi_t > 0$, as claimed in Section 10.5.3; in particular, the credit spread does not vanish as $T \to t$. This is in stark contrast to firm-value models, where typically $c(t, t) = 0$, as was shown in Section 10.3.1. Note further that, for $T - t$ large, under the RF assumption we obtain *negative* credit spreads, which is clearly unrealistic. These negative credit spreads are caused by the fact that under RF we obtain a payment of fixed size $1 - \delta$ immediately at default. If the

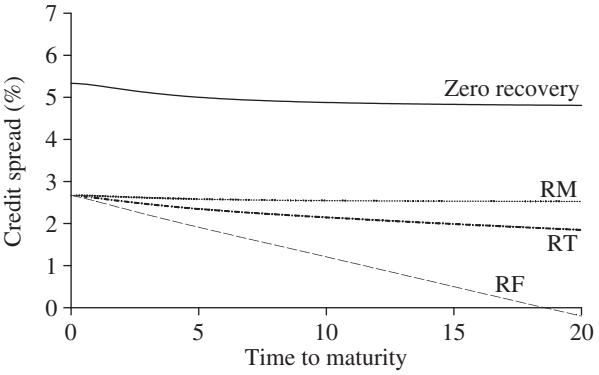

Figure 10.9. Spreads of defaultable zero-coupon bonds in the Duffie–Gârleanu model (10.75) for various recovery assumptions. The parameters of (Ψ_t) are given in Table 10.4; the initial value is $\Psi_0 \approx 0.0533$. The risk-free interest rate and the loss given default are deterministic and are given by $r = 6\%$ and $\delta = 0.5$. Note that under the RF recovery model, the spread becomes negative for large times to maturity.

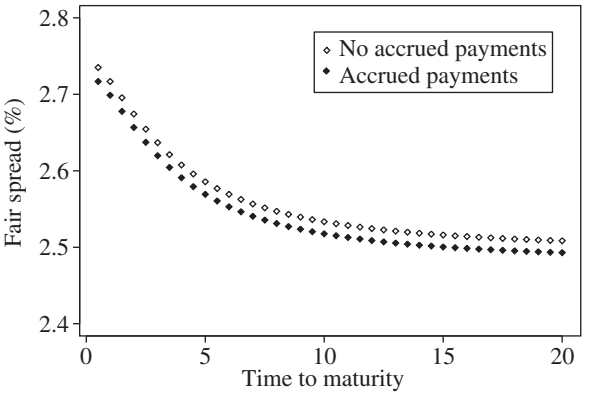

Figure 10.10. Fair CDS spreads in the Duffie–Gârleanu model (10.75) for a CDS contract with semiannual premium payments and varying time to maturity. The parameters of (Ψ_t) are given in Table 10.4; the initial value is $\Psi_0 \approx 0.0533$. The risk-free interest rate and the loss given default are deterministic and are given by $r = 6\%$ and $\delta = 0.5$. Note that, for small time to maturity, the fair swap spread is approximately equal to $\delta\Psi_0 \approx 2.7\%$.

default-free interest rate r is relatively large, it may happen that

$$E^Q\left(\exp\left(-\int_0^\tau r_s\,\mathrm{d}s\right)(1-\delta)I_{\{\tau\leqslant T\}}\right) > E^Q\left(\exp\left(-\int_0^T r_s\,\mathrm{d}s\right)I_{\{\tau\leqslant T\}}\right),$$

even if $\delta > 0$. This stems from the fact that on the right-hand side discounting is done over the whole period $[0, T]$ (as opposed to $[0, \tau]$), so that discounting has a large impact on the value of the right-hand side, compensating the higher terminal pay-off. In Figure 10.10 we have plotted the fair spreads for CDSs with and without accrued payments for varying maturities, assuming that the risk-neutral hazard process is a basic affine jump diffusion.

Notes and Comments

The Feynman–Kac formula is discussed, for example, in Section 5.5 of Björk (2004) and, at a slightly more technical level, in Section 5.7 of Karatzas and Shreve (1988).

Important original papers on affine models in term structure modelling are Duffie and Kan (1996) for diffusion models and Duffie, Pan and Singleton (2000) for jump diffusions. The latter paper also contains other applications of affine models, such as the pricing of equity options under stochastic volatility and econometric issues related to affine models. It should be mentioned that there is also a converse to Proposition 10.23; if the conditional expectations $E(\exp(-\int_t^T R(\Psi_s)\,\mathrm{d}s)\mathrm{e}^{u\Psi_T} \mid \mathcal{F}_t)$ are all exponentially affine functions of Ψ_t, the process (Ψ_t) is necessarily affine (see Duffie and Kan (1996), Duffie, Filipović and Schachermayer (2003) or Chapter 10 of Filipović (2009) for details).

The mathematical properties of the CIR model are discussed in, for example, Chapter 6.2 of Lamberton and Lapeyre (1996), where the explicit solution of the Ricatti equations in the CIR model (summarized by (10.69) and (10.70)) is also derived. The model studied in Example 10.24 is akin to models proposed by Duffie and Singleton (1999). Problems related to the modelling of negative correlation between state variable processes in an affine setting are discussed in Section 5.8 of Lando (2004). Empirical work on affine models for defaultable bonds includes the publications of Duffee (1999) and Driessen (2005).

11

Portfolio Credit Risk Management

This chapter is concerned with one-period models for credit portfolios with a view towards credit risk management issues for portfolios of largely non-traded credit products, such as the retail and commercial loans in the banking book of a typical bank.

The main theme in our analysis is the modelling of the dependence structure of the default events. In fact, default dependence has a profound impact on the upper tail of the credit loss distribution for a large portfolio. This is illustrated in Figure 11.1, where we compare the loss distribution for a portfolio of 1000 firms that default independently (portfolio 1) with a more realistic portfolio of the same size where defaults are dependent (portfolio 2). In portfolio 2 defaults are weakly dependent, in the sense that the correlation between default events is approximately 0.5%. In both cases the default probability is approximately 1%, so, on average, we expect ten defaults. As will be seen in Section 11.5, the model applied to portfolio 2 can be viewed as a realistic model for the loss distribution of a homogeneous portfolio of 1000 loans with a Standard & Poor's rating of BB. We see clearly from Figure 11.1 that the loss distribution of portfolio 2 is skewed and its right tail is substantially heavier than the right tail of the loss distribution of portfolio 1, illustrating the dramatic impact of default dependence.

Note that there are good economic reasons for expecting dependence between defaults of different obligors. Most importantly, the financial health of a firm varies with randomly fluctuating macroeconomic factors such as changes in economic growth. Since different firms are affected by common macroeconomic factors, this creates dependence between their defaults.

We begin our analysis with a discussion of threshold models in Section 11.1. These can be viewed as multivariate extensions of the Merton model considered in Chapter 10. In Section 11.2 we consider so-called mixture models in which defaults are assumed to be conditionally independent events given a set of common factors. The factors are usually interpreted as macroeconomic variables and are also modelled stochastically. Mixture models are commonly used in practice, essentially for tractability reasons; many threshold models also have convenient representations as mixture models.

Sections 11.3 and 11.4 are concerned with the calculation or approximation of the portfolio loss distribution and related measures of tail risk. We give asymptotic approximations for tail probabilities and quantiles that hold in large, relatively

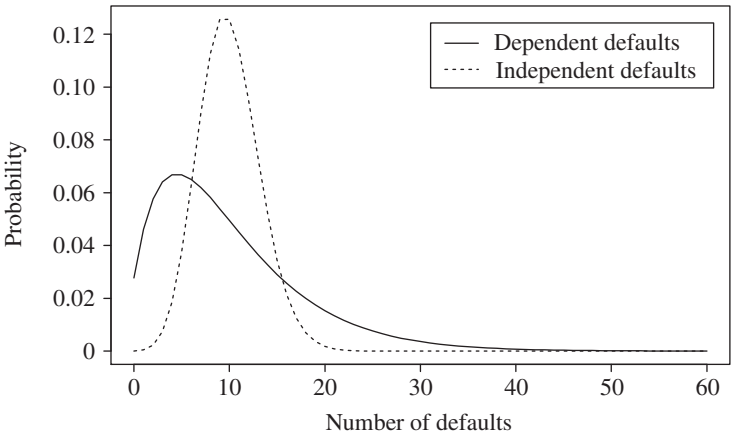

Figure 11.1. Comparison of loss distributions for two homogeneous portfolios of 1000 loans with a default probability of 1% and different dependence structures. In portfolio 1 defaults are assumed to be independent; in portfolio 2 we assume a default correlation of 0.5%. Portfolio 2 can be considered as representative for BB-rated loans. We clearly see that the default dependence generates a loss distribution with a heavier right tail and a shift of the mode to the left.

homogeneous portfolios (Section 11.3) and we discuss Monte Carlo methods for estimating tail probabilities, risk measures and capital allocations in mixture models (Section 11.4). Finally, in Section 11.5 we consider the important issue of statistical inference for credit models.

11.1 Threshold Models

The models of this section are one-period models for portfolio credit risk inspired by the firm-value models of Section 10.3. Their defining attribute is the idea that default occurs for a company i over the period $[0, T]$ if some critical rv X_i lies below some deterministic threshold d_i. The variable X_i can take on different interpretations. In a multivariate version of Merton's model, $X_i = X_{T,i}$ is a lognormally distributed asset value at the time horizon T and d_i represents liabilities to be repaid at T. More abstractly, X_i is frequently viewed as a latent variable representing the credit quality or "ability-to-pay" of obligor i.

The dependence among defaults stems from the dependence among the components of the vector $\boldsymbol{X} = (X_1, \ldots, X_m)'$. The distributions assumed for $\boldsymbol{X}$ can, in principle, be completely general, and indeed a major issue of this section will be the influence of the copula of $\boldsymbol{X}$ on the risk of the portfolio.

11.1.1 Notation for One-Period Portfolio Models

It is convenient to introduce some notation for one-period portfolio models that will be in force throughout the remainder of the chapter. We consider a portfolio of m obligors and fix a time horizon T. For $1 \leqslant i \leqslant m$, we let the rv R_i be a state indicator for obligor i at time T and assume that R_i takes integer values in the set $\{0, 1, \ldots, n\}$ representing, for example, rating classes; as in Section 10.2.1

we interpret the value 0 as default and non-zero values as states of increasing credit quality. At time $t = 0$ obligors are assumed to be in some non-default state.

Mostly we will concentrate on the binary outcomes of default and non-default and ignore the finer categorization of non-defaulted companies. In this case we write Y_i for the default indicator variables so that $Y_i = 1 \iff R_i = 0$ and $Y_i = 0 \iff R_i > 0$. The random vector $\boldsymbol{Y} = (Y_1, \ldots, Y_m)'$ is a vector of default indicators for the portfolio, and $p(\boldsymbol{y}) = P(Y_1 = y_1, \ldots, Y_m = y_m)$, $\boldsymbol{y} \in \{0, 1\}^m$, is its joint probability function; the marginal default probabilities are denoted by $p_i = P(Y_i = 1)$, $i = 1, \ldots, m$.

The *default* or *event correlations* will be of particular interest to us; they are defined to be the correlations of the default indicators. Because

$$\operatorname{var}(Y_i) = E(Y_i^2) - p_i^2 = E(Y_i) - p_i^2 = p_i - p_i^2,$$

we obtain, for firms i and j with $i \neq j$, the formula

$$\rho(Y_i, Y_j) = \frac{E(Y_i Y_j) - p_i p_j}{\sqrt{(p_i - p_i^2)(p_j - p_j^2)}}. \tag{11.1}$$

We count the number of defaulted obligors at time T with the rv $M := \sum_{i=1}^m Y_i$. The actual loss if company i defaults—termed the *loss given default* (LGD) in practice—is modelled by the random quantity $\delta_i e_i$, where e_i represents the overall exposure to company i and $0 \leqslant \delta_i \leqslant 1$ represents a random proportion of the exposure that is lost in the event of default. We will denote the overall loss by $L := \sum_{i=1}^m \delta_i e_i Y_i$ and make further assumptions about the e_i and δ_i variables as and when we need them.

It is possible to set up different credit risk models leading to the same multivariate distribution for $\boldsymbol{R}$ or $\boldsymbol{Y}$. Since this distribution is the main object of interest in the analysis of portfolio credit risk, we call two models with state vectors $\boldsymbol{R}$ and $\tilde{\boldsymbol{R}}$ (or $\boldsymbol{Y}$ and $\tilde{\boldsymbol{Y}}$) *equivalent* if $\boldsymbol{R} \overset{\mathrm{d}}{=} \tilde{\boldsymbol{R}}$ (or $\boldsymbol{Y} \overset{\mathrm{d}}{=} \tilde{\boldsymbol{Y}}$).

The exchangeable special case. To simplify the analysis we will often assume that the state indicator vector $\boldsymbol{R}$, and thus the default indicator vector $\boldsymbol{Y}$, are *exchangeable* random vectors. This is one way to mathematically formalize the notion of *homogeneous* groups that is used in practice. Recall that a random vector $\boldsymbol{R}$ is said to be exchangeable if $(R_1, \ldots, R_m) \overset{\mathrm{d}}{=} (R_{\Pi(1)}, \ldots, R_{\Pi(m)})$ for any permutation $(\Pi(1), \ldots, \Pi(m))$ of $(1, \ldots, m)$. Exchangeability implies in particular that, for any $k \in \{1, \ldots, m-1\}$, all of the $\binom{m}{k}$ possible k-dimensional marginal distributions of $\boldsymbol{R}$ are identical. In this situation we introduce a simple notation for default probabilities where $\pi := P(Y_i = 1)$, $i \in \{1, \ldots, m\}$, is the default probability of any firm and

$$\pi_k := P(Y_{i_1} = 1, \ldots, Y_{i_k} = 1), \quad \{i_1, \ldots, i_k\} \subset \{1, \ldots, m\},\ 2 \leqslant k \leqslant m, \tag{11.2}$$

is the joint default probability for any k firms. In other words, π_k is the probability that an arbitrarily selected subgroup of k companies defaults in $[0, T]$. When default

indicators are exchangeable, we get

$$E(Y_i) = E(Y_i^2) = P(Y_i = 1) = \pi, \quad \forall i,$$
$$E(Y_i Y_j) = P(Y_i = 1, Y_j = 1) = \pi_2, \quad \forall i \neq j,$$

so that $\operatorname{cov}(Y_i, Y_j) = \pi_2 - \pi^2$; this implies that the default correlation in (11.1) is given by

$$\rho_Y := \rho(Y_i, Y_j) = \frac{\pi_2 - \pi^2}{\pi - \pi^2}, \quad i \neq j, \tag{11.3}$$

which is a simple function of the first- and second-order default probabilities.

11.1.2 Threshold Models and Copulas

We start with a general definition of a threshold model before discussing the link to copulas.

Definition 11.1. Let $\boldsymbol{X} = (X_1, \ldots, X_m)'$ be an m-dimensional random vector and let $D \in \mathbb{R}^{m \times n}$ be a deterministic matrix with elements d_{ij} such that, for every i, the elements of the ith row form a set of increasing thresholds satisfying $d_{i1} < \cdots < d_{in}$. Augment these thresholds by setting $d_{i0} = -\infty$ and $d_{i(n+1)} = \infty$ for all obligors and then set

$$R_i = j \iff d_{ij} < X_i \leqslant d_{i(j+1)}, \quad j \in \{0, \ldots, n\},\ i \in \{1, \ldots, m\}.$$

Then $(\boldsymbol{X}, D)$ is said to define a threshold model for the state vector $\boldsymbol{R} = (R_1, \ldots, R_m)'$.

We refer to $\boldsymbol{X}$ as the vector of *critical variables* and denote its marginal dfs by $F_{X_i}(x) = P(X_i \leqslant x)$. The ith row of D contains the *critical thresholds* for firm i. By definition, default (corresponding to the event $R_i = 0$) occurs if $X_i \leqslant d_{i1}$, so the default probability of company i is given by $p_i = F_{X_i}(d_{i1})$. When working with a default-only model we simply write $d_i = d_{i1}$ and denote the threshold model by $(\boldsymbol{X}, \boldsymbol{d})$.

In the context of such models it is important to distinguish the default correlation $\rho(Y_i, Y_j)$ of two firms $i \neq j$ from the correlation of the critical variables X_i and X_j. Since the critical variables are often interpreted in terms of asset values, the latter correlation is often referred to as *asset correlation*. For given default probabilities, $\rho(Y_i, Y_j)$ is determined by $E(Y_i Y_j)$ according to (11.1). Moreover, in a threshold model, $E(Y_i Y_j) = P(X_i \leqslant d_{i1},\ X_j \leqslant d_{j1})$, which implies that default correlation depends on the joint distribution of X_i and X_j. If $\boldsymbol{X}$ is multivariate normal, as in many models used in practice, the correlation of X_i and X_j determines the copula of their joint distribution and hence the default correlation (see Lemma 11.2 below). For general critical variables outside the multivariate normal class, the correlation of the critical variables does not fully determine the default correlation; this can have serious implications for the tail of the distribution of $M = \sum_{i=1}^m Y_i$, as will be shown in Section 11.1.5.

We now give a simple criterion for the equivalence of two threshold models in terms of the marginal distributions of the state vector $\boldsymbol{R}$ and the copula of $\boldsymbol{X}$. This

result clarifies the central role of copulas in threshold models. For the necessary background information on copulas we refer to Chapter 7.

Lemma 11.2. *Let $(\boldsymbol{X}, D)$ and $(\tilde{\boldsymbol{X}}, \tilde{D})$ be a pair of threshold models with state vectors $\boldsymbol{R} = (R_1, \ldots, R_m)'$ and $\tilde{\boldsymbol{R}} = (\tilde{R}_1, \ldots, \tilde{R}_m)'$, respectively. The models are equivalent if the following conditions hold.*

(i) *The marginal distributions of the random vectors $\boldsymbol{R}$ and $\tilde{\boldsymbol{R}}$ coincide, i.e.*

$$P(R_i = j) = P(\tilde{R}_i = j), \quad j \in \{1, \ldots, n\},\ i \in \{1, \ldots, m\}.$$

(ii) *$\boldsymbol{X}$ and $\tilde{\boldsymbol{X}}$ admit the same copula C.*

Proof. According to Definition 11.1, $\boldsymbol{R} \overset{\mathrm{d}}{=} \tilde{\boldsymbol{R}}$ if and only if, for all $j_1, \ldots, j_m \in \{1, \ldots, n\}$,

$$\begin{aligned} &P(d_{1j_1} < X_1 \leqslant d_{1(j_1+1)}, \ldots, d_{mj_m} < X_m \leqslant d_{m(j_m+1)}) \\ &\qquad = P(\tilde{d}_{1j_1} < \tilde{X}_1 \leqslant \tilde{d}_{1(j_1+1)}, \ldots, \tilde{d}_{mj_m} < \tilde{X}_m \leqslant \tilde{d}_{m(j_m+1)}). \end{aligned}$$

By standard measure-theoretic arguments this holds if, for all $j_1, \ldots, j_m \in \{1, \ldots, n\}$,

$$P(X_1 \leqslant d_{1j_1}, \ldots, X_m \leqslant d_{mj_m}) = P(\tilde{X}_1 \leqslant \tilde{d}_{1j_1}, \ldots, \tilde{X}_m \leqslant \tilde{d}_{mj_m}).$$

By Sklar's Theorem (Theorem 7.3) this is equivalent to

$$C(F_{X_1}(d_{1j_1}), \ldots, F_{X_m}(d_{mj_m})) = C(F_{\tilde{X}_1}(\tilde{d}_{1j_1}), \ldots, F_{\tilde{X}_m}(\tilde{d}_{mj_m})),$$

where C is the copula of $\boldsymbol{X}$ and $\tilde{\boldsymbol{X}}$ (using condition (ii)). Condition (i) implies that $F_{X_i}(d_{ij}) = F_{\tilde{X}_i}(\tilde{d}_{ij})$ for all $j \in \{1, \ldots, n\}$, $i \in \{1, \ldots, m\}$, and the claim follows. □

The result shows that in a threshold model the copula of the critical variables determines the link between marginal probabilities of migration for individual firms and joint probabilities of migration for groups of firms. To illustrate this further, consider for simplicity a two-state model for default and non-default and a subgroup of k companies $\{i_1, \ldots, i_k\} \subset \{1, \ldots, m\}$ with individual default probabilities $p_{i_1}, \ldots, p_{i_k}$. Then

$$\begin{aligned} P(Y_{i_1} = 1, \ldots, Y_{i_k} = 1) &= P(X_{i_1} \leqslant d_{i_1 1}, \ldots, X_{i_k} \leqslant d_{i_k 1}) \\ &= C_{i_1 \cdots i_k}(p_{i_1}, \ldots, p_{i_k}), \end{aligned} \tag{11.4}$$

where $C_{i_1 \cdots i_k}$ denotes the corresponding k-dimensional margin of C. As a special case consider now a model for a single homogeneous group. We assume that $\boldsymbol{X}$ has an exchangeable copula (i.e. a copula of the form (7.20)) and that all individual default probabilities are equal to some constant π so that the default indicator vector $\boldsymbol{Y}$ is exchangeable. The formula (11.4) reduces to the useful formula

$$\pi_k = C_{1\cdots k}(\pi, \ldots, \pi), \quad 2 \leqslant k \leqslant m, \tag{11.5}$$

which will be used for the calibration of some copula models later on.

11.1.3 Gaussian Threshold Models

In this section we discuss the case where the critical variables have a Gauss copula.

Multivariate Merton model. It is straightforward to generalize the Merton model of Section 10.3.1 to a portfolio of m firms. We assume that the multivariate asset-value process $(\boldsymbol{V}_t)$ with $\boldsymbol{V}_t = (V_{t,1}, \dots, V_{t,m})'$ follows an m-dimensional geometric Brownian motion with drift vector $\boldsymbol{\mu}_V = (\mu_1, \dots, \mu_m)'$, vector of volatilities $\boldsymbol{\sigma}_V = (\sigma_1, \dots, \sigma_m)'$, and instantaneous correlation matrix P. This means that $(\boldsymbol{V}_t)$ solves the stochastic differential equations

$$\mathrm{d}V_{t,i} = \mu_i V_{t,i}\,\mathrm{d}t + \sigma_i V_{t,i}\,\mathrm{d}W_{t,i}, \quad i = 1, \dots, m,$$

for correlated Brownian motions with correlation $\rho(W_{t,i}, W_{t,j}) = \rho_{ij}$, $t \geqslant 0$. For all i the asset value $V_{T,i}$ is thus of the form

$$V_{T,i} = V_{0,i} \exp((\mu_i - \tfrac{1}{2}\sigma_i^2)T + \sigma_i W_{T,i}),$$

and $\boldsymbol{W}_T := (W_{T,1}, \dots, W_{T,m})'$ is a multivariate normal random vector satisfying $\boldsymbol{W}_T \sim N_m(\mathbf{0}, TP)$. In its basic form the Merton model is a default-only model in which the firm defaults if $V_{T,i} \leqslant B_i$ and B_i is the liability of firm i. Writing $\boldsymbol{B} = (B_1, \dots, B_m)'$, the threshold model representation is thus given by $(\boldsymbol{V}_T, \boldsymbol{B})$. Since in a threshold model the default event is invariant under strictly increasing transformations of critical variables and thresholds, this model is equivalent to a threshold model $(\boldsymbol{X}, \boldsymbol{d})$ with

$$X_i := \frac{\ln V_{T,i} - \ln V_{0,i} - (\mu_i - \frac{1}{2}\sigma_i^2)T}{\sigma_i\sqrt{T}},$$
$$d_i := \frac{\ln B_i - \ln V_{0,i} - (\mu_i - \frac{1}{2}\sigma_i^2)T}{\sigma_i\sqrt{T}}.$$

The transformed variables satisfy $\boldsymbol{X} \sim N_m(\mathbf{0}, P)$ and their copula is the Gauss copula C_P^{Ga}.

Gaussian threshold models in practice. In practice it is usual to start directly with threshold models of the form $(\boldsymbol{X}, \boldsymbol{d})$ with $\boldsymbol{X} \sim N_m(\mathbf{0}, P)$. There are two practical challenges: first, one has to calibrate the threshold vector $\boldsymbol{d}$ (or, in the case of a multi-state model, the threshold matrix D) in line with exogenously given default and transition probabilities; second, one needs to calibrate the correlation matrix P in a parsimonious way. The problem of calibrating the obligor-specific rows of the threshold matrix D to (rating) state transition probabilities was discussed in Section 10.3.4. In particular, in a default-only model we set $d_i = \Phi^{-1}(p_i)$ for given default probabilities p_i, for $i = 1, \dots, m$. Since $\boldsymbol{X}$ has standard normal margins, P is also the covariance matrix of $\boldsymbol{X}$.

In its most general form P has $m(m-1)/2$ distinct parameters. In portfolio credit risk applications m is typically large and it is important to use a more parsimonious parametrization of this matrix based on a factor model of the kind described in Section 6.4.1. Factor models also lend themselves to economic interpretation, and the

factors are commonly interpreted as country and industry effects. We now describe the mathematical form of typical factor models used in credit risk. The calibration of these models in practice is discussed in Section 11.5.1.

Factor models. We assume that

$$X_i = \sqrt{\beta_i}\tilde{F}_i + \sqrt{1-\beta_i}\varepsilon_i, \tag{11.6}$$

where $\tilde{F}_i$ and $\varepsilon_1, \ldots, \varepsilon_m$ are independent standard normal variables, and where $0 \leqslant \beta_i \leqslant 1$ for all i. In this formulation the $\tilde{F}_i$ are the *systematic* variables, which are correlated, and the ε_i are *idiosyncratic* variables. It follows that β_i can be viewed as a measure of the *systematic risk* of X_i: that is, the part of the variance of X_i that is explained by the systematic variable.

The systematic variables are assumed to be of the form $\tilde{F}_i = \boldsymbol{a}_i'\boldsymbol{F}$, where $\boldsymbol{F}$ is a vector of common factors satisfying $\boldsymbol{F} \sim N_p(\boldsymbol{0}, \Omega)$ with $p < m$, and where Ω is a correlation matrix. These factors typically represent country and industry effects. The assumption that $\mathrm{var}(\tilde{F}_i) = 1$ imposes the constraint that $\boldsymbol{a}_i'\Omega\boldsymbol{a}_i = 1$ for all i. Since $\mathrm{var}(X_i) = 1$ and since $\tilde{F}_i$ and $\varepsilon_1, \ldots, \varepsilon_m$ are independent and standard normal, the asset correlations in this model are given by

$$\rho(X_i, X_j) = \mathrm{cov}(X_i, X_j) = \sqrt{\beta_i\beta_j}\,\mathrm{cov}(\tilde{F}_i, \tilde{F}_j) = \sqrt{\beta_i\beta_j}\boldsymbol{a}_i'\Omega\boldsymbol{a}_j.$$

In order to set up the model we have to determine $\boldsymbol{a}_i$ and β_i for each obligor as well as Ω, with the additional constraint that $\boldsymbol{a}_i'\Omega\boldsymbol{a}_i = 1$ for all i. Since Ω has $p(p-1)/2$ parameters, the loading vectors $\boldsymbol{a}_i$ and coefficients β_i have a combined total of $mp + m$ parameters, and we are applying m constraints, the dimension of the calibration problem is $mp + p(p-1)/2$. In particular, the number of parameters grows linearly rather than quadratically in m.

Note that this factor model does fit into the general framework developed in Section 6.4.1. $\boldsymbol{X}$ can be written as

$$\boldsymbol{X} = B\boldsymbol{F} + \tilde{\boldsymbol{\varepsilon}}, \tag{11.7}$$

where $B = DA$, $D = \mathrm{diag}(\sqrt{\beta}_1, \ldots, \sqrt{\beta}_m)$, $A \in \mathbb{R}^{m\times p}$ is the matrix with ith row given by $\boldsymbol{a}_i'$, and $\tilde{\varepsilon}_i = \sqrt{1-\beta_i}\varepsilon_i$.

We often consider the special case of a one-factor model. This corresponds to a model where $\tilde{F}_i = F$ for a single common standard normal factor so that the equation in (11.6) takes the form

$$X_i = \sqrt{\beta_i}F + \sqrt{1-\beta_i}\varepsilon_i. \tag{11.8}$$

If, moreover, every obligor has the same systematic variance $\beta_i = \rho$, we get that $\rho(X_i, X_j) = \rho$ for all $i \neq j$. This model is often referred to as an equicorrelation model and was introduced previously in Example 6.32 and equation (6.53).

11.1.4 Models Based on Alternative Copulas

While most threshold models used in industry are based explicitly or implicitly on the Gauss copula, there is no reason why we have to assume a Gauss copula. In fact,

simulations presented in Section 11.1.5 show that the choice of copula may be very critical to the tail of the distribution of the number of defaults M. We now look at threshold models based on alternative copulas.

The first class of models attempts to preserve some of the flexibility of the Gaussian threshold models, which do have the appealing feature that they can accommodate a wide range of different correlation structures for the critical variables. This is clearly an advantage in modelling a portfolio where obligors are exposed to several risk factors and where the exposure to different risk factors differs markedly across obligors, such as a portfolio of loans to companies from different industry sectors or countries.

Example 11.3 (normal mean–variance mixtures). For the distribution of the critical variables we consider the kind of model described in Section 6.2.2. We start with an m-dimensional multivariate normal vector $\boldsymbol{Z} \sim N_m(\boldsymbol{0}, \Sigma)$ and a positive, scalar rv W, which is independent of $\boldsymbol{Z}$. The vector of critical variables $\boldsymbol{X}$ is assumed to have the structure

$$\boldsymbol{X} = \boldsymbol{m}(W) + \sqrt{W}\boldsymbol{Z}, \tag{11.9}$$

where $\boldsymbol{m}\colon [0, \infty) \to \mathbb{R}^m$ is a measurable function. In the special case where $\boldsymbol{m}(W)$ takes a constant value $\boldsymbol{\mu}$ not depending on W, the distribution is called a normal variance mixture. An important example of a normal variance mixture is the multivariate t distribution, as discussed in Example 6.7, which is obtained when W has an inverse gamma distribution, $W \sim \mathrm{Ig}(\frac{1}{2}\nu, \frac{1}{2}\nu)$, or equivalently when $\nu/W \sim \chi^2_\nu$. An example of a general mean–variance mixture is the GH distribution discussed in Section 6.2.3.

In a normal mean–variance mixture model the default condition may be written in the form

$$X_i \leqslant d_{i1} \iff Z_i \leqslant \frac{d_{i1}}{\sqrt{W}} - \frac{m_i(W)}{\sqrt{W}} =: \tilde{D}_i, \tag{11.10}$$

where $m_i(W)$ is the ith component of $\boldsymbol{m}(W)$. A possible economic interpretation of the model (11.9) is to consider Z_i as the asset value of company i and d_{i1} as an a priori estimate of the corresponding default threshold. The actual default threshold is *stochastic* and is represented by $\tilde{D}_i$, which is obtained by applying a multiplicative shock and an additive shock to the estimate d_{i1}. If we interpret this shock as a stylized representation of global factors such as the overall liquidity and risk appetite in the banking system, it makes sense to assume that the shocks to the default thresholds of different obligors are driven by the same rv W.

Normal variance mixtures, such as the multivariate t, provide the most tractable examples of normal mean–variance mixtures; they admit a calibration approach using linear factor models that is similar to the approach used for models based on the Gauss copula. In normal variance mixture models the correlation matrices of $\boldsymbol{X}$ (when defined) and $\boldsymbol{Z}$ coincide. Moreover, if $\boldsymbol{Z}$ follows a linear factor model (11.7), then $\boldsymbol{X}$ inherits the linear factor structure from $\boldsymbol{Z}$. Note, however, that the systematic factors $\sqrt{W}\boldsymbol{F}$ and the idiosyncratic factors $\sqrt{W}\boldsymbol{\varepsilon}$ are no longer independent but merely uncorrelated.

The class of threshold models based on the t copula can be thought of as containing the Gaussian threshold models as limiting cases when $\nu \to \infty$. However, the additional parameter ν adds a great deal of flexibility. We will come back to this point in Section 11.1.5.

Another class of parametric copulas that could be used in threshold models is the Archimedean family of Section 7.4.

Example 11.4 (Archimedean copulas). Recall that an Archimedean copula is the distribution function of a uniform random vector of the form

$$C(u_1, \ldots, u_m) = \psi(\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_m)), \tag{11.11}$$

where $\psi : [0, \infty) \to [0, 1]$ is a continuous, decreasing function, known as the copula generator, satisfying $\psi(0) = 1$ and $\lim_{t\to\infty} \psi(t) = 0$, and ψ^{-1} is its inverse. We assume that ψ is completely monotonic (see equation (7.47) and surrounding discussion). As explained in Section 7.4, these conditions ensure that (11.11) defines a copula for any portfolio size m. Our main example in this chapter is the Clayton copula. Recall from Section 7.4 that this copula has generator $\psi_\theta(t) = (1 + \theta t)^{-1/\theta}$, where $\theta > 0$, leading to the expression

$$C_\theta^{\mathrm{Cl}}(u_1, \ldots, u_m) = (u_1^{-\theta} + \cdots + u_m^{-\theta} + 1 - m)^{-1/\theta}. \tag{11.12}$$

As discussed in Section 7.4, exchangeable Archimedean copulas suffer from the deficiency that they are not rich in parameters and can model only exchangeable dependence and not a fully flexible dependence structure for the critical variables. Nonetheless, they yield useful parsimonious models for relatively small homogeneous portfolios, which are easy to calibrate and simulate, as we discuss in more detail in Section 11.2.4.

Suppose that $\boldsymbol{X}$ is a random vector with an Archimedean copula and marginal distributions F_{X_i}, $1 \leqslant i \leqslant m$, so that $(\boldsymbol{X}, \boldsymbol{d})$ specifies a threshold model with individual default probabilities $F_{X_i}(d_i)$. As a particular example consider the Clayton copula and assume a homogeneous situation where all individual default probabilities are identical to π. Using equations (11.5) and (11.12) we can calculate that the probability that an arbitrarily selected group of k obligors from a portfolio of m such obligors defaults over the time horizon is given by $\pi_k = (k\pi^{-\theta} - k + 1)^{-1/\theta}$. Essentially, the dependent default mechanism of the homogeneous group is now determined by this equation and the parameters π and θ. We study this Clayton copula model further in Example 11.13.

11.1.5 Model Risk Issues

Model risk is the risk associated with working with misspecified models—in our case, models that are a poor representation of the true mechanism governing defaults and migrations in a credit portfolio. For example, if we intend to use our models to estimate measures of tail risk, like VaR and expected shortfall, then we should be particularly concerned with the possibility that they might underestimate the tail of the portfolio loss distribution.

Table 11.1. Results of simulation study. We tabulate the estimated 95th and 99th percentiles of the distribution of M in an exchangeable model with 10 000 firms. The values for the default probability π and the asset correlation ρ corresponding to the three groups A, B and C are given in the text.

	$q_{0.95}(M)$			$q_{0.99}(M)$		
Group	$\nu = \infty$	$\nu = 50$	$\nu = 10$	$\nu = \infty$	$\nu = 50$	$\nu = 10$
A	14	23	24	21	49	118
B	109	153	239	157	261	589
C	1618	1723	2085	2206	2400	3067

As we have seen, a threshold model essentially consists of a collection of default (and migration) probabilities for individual firms and a copula that describes the dependence of certain critical variables. In discussing model risk in this context we will concentrate on models for default only and assume that individual default probabilities have been satisfactorily determined. It is much more difficult to determine the copula describing default dependence and we will look at model risk associated with the misspecification of this component of the threshold model. See also Section 8.4.4 for a discussion of the issue of dependence uncertainty.

The impact of the choice of copula. Since most threshold models used in practice use the Gauss copula, we are particularly interested in the sensitivity of the distribution of the number of defaults M with respect to the assumption of Gaussian dependence. Our interest is motivated by the observation made in Section 7.3.1 that, by assuming a Gaussian dependence structure, we may underestimate the probability of joint large movements of risk factors, with potentially drastic implications for the performance of risk-management models.

We compare a simple exchangeable model with multivariate normal critical variables and a model where the critical variables are multivariate t. Given a standard normal rv F, an iid sequence $\varepsilon_1, \ldots, \varepsilon_m$ of standard normal variates independent of F, and an asset correlation parameter $\rho \in [0, 1]$, we define a random vector $\boldsymbol{Z}$ by $Z_i = \sqrt{\rho}F + \sqrt{1-\rho}\varepsilon_i$. Observe that this is the equicorrelation special case of the factor model (11.8).

In the t copula case we define the critical variables $X_i := \sqrt{W}Z_i$, where $W \sim \mathrm{Ig}(\frac{1}{2}\nu, \frac{1}{2}\nu)$ is independent of $\boldsymbol{Z}$, so that $\boldsymbol{X}$ has a multivariate t distribution. In the Gauss copula case we simply set $\boldsymbol{X} := \boldsymbol{Z}$. In both cases we choose thresholds so that $P(Y_i = 1) = \pi$ for all i and for some $\pi \in (0, 1)$. Note that the correlation matrix P of $\boldsymbol{X}$ (the asset correlation matrix) is identical in both models and is given by an equicorrelation matrix with off-diagonal element ρ. However, the copula of $\boldsymbol{X}$ differs, and we expect more joint defaults in the t model due to the higher level of dependence in the joint tail of the t copula.

We consider three portfolios of decreasing credit quality, labelled A, B and C. In group A we set $\pi = 0.06\%$ and $\rho = 2.58\%$; in group B we set $\pi = 0.50\%$ and $\rho = 3.80\%$; in group C we set $\pi = 7.50\%$ and $\rho = 9.21\%$. We consider a portfolio of size $m = 10\,000$. For each group we vary the degrees-of-freedom

Table 11.2. Results of simulation study. Estimated 95th and 99th percentiles of the distribution of M in an exchangeable model for varying values of asset correlation ρ.

Quantile	$\rho = 2.58\%$	$\rho = 3.80\%$	$\rho = 9.21\%$
$q_{0.95}(M)$	98	109	148
$q_{0.99}(M)$	133	157	250

parameter ν. In order to represent the tail of the number of defaults M, we use simulations to determine (approximately) the 95% and 99% quantiles, $q_{0.95}(M)$ and $q_{0.99}(M)$, and tabulate them in Table 11.1. The actual simulation was performed using a representation of threshold models as Bernoulli mixture models that is discussed later in Section 11.2.4.

Table 11.1 shows that ν clearly has a massive influence on the high quantiles. For the important 99% quantile the impact is most pronounced for group A, where $q_{0.99}(M)$ is increased by a factor of almost six when we go from a Gaussian model to a model with $\nu = 10$.

The impact of changing asset correlation. Here we retain the assumption that $\boldsymbol{X}$ has a Gauss copula and study the impact of the factor structure of the asset returns on joint default events and hence on the tail of M. More specifically, we increase the systematic risk component of the critical variables for the obligors in our portfolio and analyse how this affects the tail of M. We use the exchangeable model introduced above. We fix the default probability at $\pi = 0.50\%$ (the value for group B above) and vary the asset correlation ρ using the values $\rho = 2.58\%$, $\rho = 3.80\%$ and $\rho = 9.21\%$. In Table 11.2 we tabulate $q_{0.95}(M)$ and $q_{0.99}(M)$ for a portfolio with 10 000 counterparties. Clearly, varying ρ also has a sizeable effect on the quantiles of M. However, this effect is less dramatic and, in particular, less surprising than the impact of varying the copula in our previous experiment.

Both simulation experiments suggest that the loss distributions implied by threshold models are very sensitive to the copula of the critical variables. For this reason a substantial effort should be devoted to the calibration of the dependence model for the critical variables. Moreover, it is important to conduct sensitivity analyses to understand the implications of model risk for risk capital calculations.

Notes and Comments

Our presentation of threshold models is based, to a large extent, on Frey and McNeil (2001, 2003). In those papers we referred to the models as "latent variable" models, because of structural similarities with statistical models of that name (see Joe 1997). However, whereas in statistical latent variable models the critical variables are treated as unobserved, in credit models they are often formally identified, e.g. as asset values or asset-value returns.

To see how the factor modelling approaches used in industry correspond to our presentation in Section 11.1.3 readers should consult the CreditMetrics technical document (RiskMetrics Group 1997) and the description of Moody's GCorr model in Huang et al. (2012). The latter model is used to model correlations between

changes in credit quality for many different kinds of obligor including publicly traded firms, private firms, small and medium-sized enterprises (SMEs) and retail borrowers. For public firms, weekly asset returns, calculated as part of the public-firm EDF methodology described in Section 10.3.3, are used as the measure of changing credit quality or "ability-to-pay".

The first systematic study of model risk for credit portfolio models is Gordy (2000). Our analysis of the impact of the copula of $\boldsymbol{X}$ on the tail of M follows Frey, McNeil and Nyfeler (2001). For an excellent discussion of various aspects of model risk in risk management in general, we refer to Gibson (2000).

11.2 Mixture Models

In a mixture model the default risk of an obligor is assumed to depend on a set of common factors, usually interpreted as macroeconomic variables, which are also modelled stochastically. Given a realization of the factors, defaults of individual firms are assumed to be independent. Dependence between defaults stems from the dependence of individual default probabilities on the set of common factors. We start with a general definition of a Bernoulli mixture model in Section 11.2.1 before looking in detail at the important special case of one-factor Bernoulli mixture models in Section 11.2.2. In Section 11.2.4 we show that many threshold models can be represented as Bernoulli mixtures, and in Section 11.2.5 we discuss the approximation of Bernoulli mixture models through Poisson mixture models and the important example of CreditRisk$^+$.

11.2.1 Bernoulli Mixture Models

Definition 11.5 (Bernoulli mixture model). Given some $p < m$ and a p-dimensional random vector $\boldsymbol{\Psi} = (\Psi_1, \ldots, \Psi_p)'$, the random vector $\boldsymbol{Y} = (Y_1, \ldots, Y_m)'$ follows a Bernoulli mixture model with factor vector $\boldsymbol{\Psi}$ if there are functions $p_i : \mathbb{R}^p \to [0, 1]$, $1 \leqslant i \leqslant m$, such that, conditional on $\boldsymbol{\Psi}$, the components of $\boldsymbol{Y}$ are independent Bernoulli rvs satisfying $P(Y_i = 1 \mid \boldsymbol{\Psi} = \boldsymbol{\psi}) = p_i(\boldsymbol{\psi})$.

For $\boldsymbol{y} = (y_1, \ldots, y_m)'$ in $\{0, 1\}^m$ we have that

$$P(\boldsymbol{Y} = \boldsymbol{y} \mid \boldsymbol{\Psi} = \boldsymbol{\psi}) = \prod_{i=1}^{m} p_i(\boldsymbol{\psi})^{y_i}(1 - p_i(\boldsymbol{\psi}))^{1-y_i}, \tag{11.13}$$

and the unconditional distribution of the default indicator vector $\boldsymbol{Y}$ is obtained by integrating over the distribution of the factor vector $\boldsymbol{\Psi}$. In particular, the default probability of company i is given by $p_i = P(Y_i = 1) = E(p_i(\boldsymbol{\Psi}))$.

Note that the two-stage hierarchical structure of a Bernoulli mixture model facilitates sampling from the model: first we generate the economic factor realizations, then we generate the pattern of defaults conditional on those realizations. The second step is easy because of the conditional independence assumption.

In general, Bernoulli mixture models have a number of computational advantages. Consider the portfolio loss $L = \sum_{i=1}^{m} e_i \delta_i Y_i$ in the case where the exposures e_i and LGDs δ_i are deterministic. While it is difficult to compute the df F_L of L, it is easy to

use the conditional independence of the defaults to show that the Laplace–Stieltjes transform of F_L for $t \in \mathbb{R}$ is given by

$$\begin{aligned}\hat{F}_L(t) = E(\mathrm{e}^{-tL}) = E(E(\mathrm{e}^{-tL} \mid \boldsymbol{\Psi})) &= E\left(E\left(\exp\left(-t\sum_{i=1}^{m} e_i\delta_i Y_i\right) \,\middle|\, \boldsymbol{\Psi}\right)\right) \\ &= E\left(\prod_{i=1}^{m} E(\mathrm{e}^{-te_i\delta_i Y_i} \mid \boldsymbol{\Psi})\right) \\ &= E\left(\prod_{i=1}^{m} (p_i(\boldsymbol{\Psi})\mathrm{e}^{-te_i\delta_i} + 1 - p_i(\boldsymbol{\Psi}))\right),\end{aligned}$$

which can also be obtained by integrating over the distribution of the factors $\boldsymbol{\Psi}$. The Laplace–Stieltjes transform is useful in a number of practical tasks relating to Bernoulli mixture models, as follows.

- To implement an efficient Monte Carlo scheme for sampling losses from a Bernoulli mixture model we often use importance sampling. For this we need the moment-generating function of L, which can be calculated from the Laplace–Stieltjes transform according to $M_L(t) = E(\mathrm{e}^{Lt}) = \hat{F}_L(-t)$ (see Section 11.4 for more details).
- The probability mass function of L may be calculated by using the inverse Fourier transform to invert the characteristic function of L given by $\phi_L(t) = E(\mathrm{e}^{\mathrm{i}Lt})$. The characteristic function has the same functional form as the Laplace–Stieltjes transform $\hat{F}_L(t)$ but with the imaginary argument $-\mathrm{i}t$ (see also the discussion after Theorem 11.16).

11.2.2 One-Factor Bernoulli Mixture Models

One-factor models, i.e. models where Ψ is one dimensional, are particularly important special cases because of their tractability. Their behaviour for large portfolios is particularly easy to understand, as will be shown in Section 11.3, and this has had an influence on the Basel capital framework. Moreover, they have relatively few parameters and are thus easier to estimate from data. Throughout this section we consider an rv Ψ with values in $\mathbb{R}$ and functions $p_i(\Psi)\colon \mathbb{R} \to [0, 1]$ such that, conditional on Ψ, the default indicator $\boldsymbol{Y}$ is a vector of independent Bernoulli rvs with $P(Y_i = 1 \mid \Psi = \psi) = p_i(\psi)$. We now consider a variety of special cases.

Exchangeable Bernoulli mixture models. A further simplification occurs if the functions p_i are all identical. In this case the Bernoulli mixture model is termed *exchangeable*, since the random vector $\boldsymbol{Y}$ is exchangeable. It is convenient to introduce the rv $Q := p_1(\Psi)$ and to denote the distribution function of this mixing variable by $G(q)$. Conditional on $Q = q$, the number of defaults M is the sum of m independent Bernoulli variables with parameter q and it therefore has a binomial distribution with parameters q and m, i.e. $P(M = k \mid Q = q) = \binom{m}{k} q^k (1-q)^{m-k}$. The unconditional distribution of M is obtained by integrating over q. We have

$$P(M = k) = \binom{m}{k} \int_0^1 q^k (1-q)^{m-k} \, \mathrm{d}G(q). \tag{11.14}$$

Using the notation of Section 11.1.1 we can calculate default probabilities and joint default probabilities for the exchangeable group. Simple calculations give $\pi = E(Y_1) = E(E(Y_1 \mid Q)) = E(Q)$ and, more generally,

$$\pi_k = P(Y_1 = 1, \ldots, Y_k = 1) = E(E(Y_1 \cdots Y_k \mid Q)) = E(Q^k), \qquad (11.15)$$

so that unconditional default probabilities of first and higher order are seen to be moments of the mixing distribution. Moreover, for $i \neq j$, $\operatorname{cov}(Y_i, Y_j) = \pi_2 - \pi^2 = \operatorname{var}(Q) \geqslant 0$, which means that in an exchangeable Bernoulli mixture model the default correlation ρ_Y defined in (11.3) is always non-negative. Any value of ρ_Y in $[0, 1]$ can be obtained by an appropriate choice of the mixing distribution G. In particular, if $\rho_Y = \operatorname{var}(Q) = 0$, the rv Q has a degenerate distribution with all mass concentrated on the point π and the default indicators are independent. The case $\rho_Y = 1$ corresponds to a model where $\pi = \pi_2$ and the distribution of Q is concentrated on the points 0 and 1.

Example 11.6 (beta, probit-normal and logit-normal mixtures). The following mixing distributions are frequently encountered in Bernoulli mixture models.

Beta mixing distribution. In this model $Q \sim \text{Beta}(a, b)$ for some parameters $a > 0$ and $b > 0$. See Section A.2.1 for more details concerning the beta distribution.

Probit-normal mixing distribution. Here, $Q = \Phi(\mu + \sigma\Psi)$ for $\Psi \sim N(0, 1)$, $\mu \in \mathbb{R}$ and $\sigma > 0$, where Φ is the standard normal distribution function. We show later, in Section 11.2.4, that this model is equivalent to an exchangeable version of the one-factor Gaussian threshold model in (11.8).

Logit-normal mixing distribution. Here, $Q = F(\mu + \sigma\Psi)$ for $\Psi \sim N(0, 1)$, $\mu \in \mathbb{R}$ and $\sigma > 0$, where $F(x) = (1 + \mathrm{e}^{-x})^{-1}$ is the df of a so-called logistic distribution.

In the model with beta mixing distribution, the higher-order default probabilities π_k and the distribution of M can be computed explicitly (see Example 11.7 below). Calculations for the logit-normal, probit-normal and other models generally require numerical evaluation of the integrals in (11.14) and (11.15). If we fix any two of π, π_2 and ρ_Y in a beta, logit-normal or probit-normal model, then this fixes the parameters a and b or μ and σ of the mixing distribution, and higher-order joint default probabilities are automatically determined.

Example 11.7 (beta mixing distribution). By definition, the density of a beta distribution is given by

$$g(q) = \frac{1}{\beta(a, b)} q^{a-1}(1 - q)^{b-1}, \quad a, b > 0,\ 0 < q < 1,$$

where $\beta(a, b)$ denotes the beta function. Below we use the fact that the beta function satisfies the recursion formula $\beta(a + 1, b) = (a/(a + b))\beta(a, b)$; this is easily established from the representation of the beta function in terms of the gamma

function in Section A.2.1. Using (11.15), for the higher-order default probabilities we obtain

$$\pi_k = \frac{1}{\beta(a,b)} \int_0^1 q^k q^{a-1}(1-q)^{b-1}\,\mathrm{d}q = \frac{\beta(a+k,b)}{\beta(a,b)}, \quad k = 1, 2, \ldots.$$

The recursion formula for the beta function yields $\pi_k = \prod_{j=0}^{k-1}(a+j)/(a+b+j)$; in particular, $\pi = a/(a+b)$, $\pi_2 = \pi(a+1)/(a+b+1)$ and $\rho_Y = (a+b+1)^{-1}$. The rv M has a so-called beta-binomial distribution. From (11.14) we obtain

$$\begin{aligned} P(M=k) &= \binom{m}{k} \frac{1}{\beta(a,b)} \int_0^1 q^{k+a-1}(1-q)^{m-k+b-1}\,\mathrm{d}q \\ &= \binom{m}{k} \frac{\beta(a+k, b+m-k)}{\beta(a,b)}. \end{aligned} \tag{11.16}$$

One-factor models with covariates. It is straightforward to extend the one-factor probit-normal and logit-normal mixture models to include covariates that influence default probability and default correlation; these covariates might be indicators for group membership, such as a rating class or industry sector, or key ratios taken from a company's balance sheet.

Writing $\boldsymbol{x}_i \in \mathbb{R}^k$ for a vector of deterministic covariates, a general model for the conditional default probabilities $p_i(\Psi)$ in (11.13) would be to assume that

$$\left.\begin{aligned} p_i(\Psi) &= h(\mu_i + \sigma_i \Psi), \\ \mu_i &= \mu + \boldsymbol{\beta}' \boldsymbol{x}_i, \\ \sigma_i &= \exp(\delta + \boldsymbol{\gamma}' \boldsymbol{x}_i), \end{aligned}\right\} \tag{11.17}$$

where $\Psi \sim N(0,1)$, $h(x) = \Phi(x)$ or $h(x) = (1+\mathrm{e}^{-x})^{-1}$, the vectors $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)'$ and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_k)'$ contain regression parameters, and $\mu \in \mathbb{R}$ and $\delta \in \mathbb{R}$ are intercept parameters. Similar specifications are commonly used in the class of generalized linear models in statistics (see Section 11.5.3).

Clearly, if $\boldsymbol{x}_i = \boldsymbol{x}$ for all i, so that all risks have the same covariates, then we are back in the situation of full exchangeability. Note also that, since the function $p_i(\Psi)$ is increasing in Ψ, the conditional default probabilities $(p_1(\Psi), \ldots, p_m(\Psi))$ form a comonotonic random vector; hence, in a state of the world where the default probability is comparatively high for one counterparty, it is high for all counterparties. For a discussion of comonotonicity we refer to Section 7.2.1.

Example 11.8 (model for several exchangeable groups). The regression structure in (11.17) includes partially exchangeable models, where we define a number of groups within which risks are exchangeable. These groups might represent rating classes according to some internal or rating-agency classification.

If the covariates $\boldsymbol{x}_i$ are simply k-dimensional unit vectors of the form $\boldsymbol{x}_i = \boldsymbol{e}_{r(i)}$, where $r(i) \in \{1, \ldots, k\}$ indicates, say, the rating class of firm i, then the model (11.17) can be written in the form

$$p_i(\Psi) = h(\mu_{r(i)} + \sigma_{r(i)} \Psi) \tag{11.18}$$

for parameters $\mu_r := \mu + \beta_r$ and $\sigma_r := \mathrm{e}^{\delta+\gamma_r}$ for $r = 1, \ldots, k$.

Inserting this specification into (11.13) allows us to find the conditional distribution of the default indicator vector. Suppose there are m_r obligors in rating category r for $r = 1, \ldots, k$, and write M_r for the number of defaults. The conditional distribution of the vector $\boldsymbol{M} = (M_1, \ldots, M_k)'$ is given by

$$P(\boldsymbol{M} = \boldsymbol{l} \mid \Psi = \psi) = \prod_{r=1}^{k} \binom{m_r}{l_r} (h(\mu_r + \sigma_r \psi))^{l_r} (1 - h(\mu_r + \sigma_r \psi))^{m_r - l_r}, \tag{11.19}$$

where $\boldsymbol{l} = (l_1, \ldots, l_k)'$. A model of the form (11.19) with $\sigma_1 = \cdots = \sigma_k$ will be fitted to Standard & Poor's default data in Section 11.5.4. The asymptotic behaviour of such a model (when m is large) is investigated in Example 11.20.

11.2.3 Recovery Risk in Mixture Models

In standard portfolio risk models it is assumed that the loss given default is independent of the default event. This is likely to be an oversimplification, as economic intuition suggests that recovery rates depend on risk factors similar to those for default probabilities; in that case one speaks of *systematic recovery risk*. Consider, for instance, the market for mortgages. During a property crisis many mortgages default. At the same time property prices are low, so that real estate can be sold only for very low prices in a foreclosure (a forced sale in which a bank liquidates a property it holds as collateral), so that recovery rates are low.

The presence of systematic recovery risk is confirmed in a number of empirical studies. Among others, Frye (2000) has carried out a formal empirical analysis using recovery data collected by Moody's on rated corporate bonds. He found that recovery rates are substantially lower than average in times of economic recession. To quote from his paper:

> Using that data [the Moody's data] to estimate an appropriate credit model, we can extrapolate that in a severe economic downturn recoveries might decline 20–25 percentage points from the normal-year average. This could cause loss given default to increase by nearly 100% and to have a similar effect on economic capital. Such systematic recovery risk is absent from first-generation credit risk models. Therefore these models may significantly understate the capital required at banking institutions.

In a similar vein, Hamilton et al. (2005) estimated formal models for the relationship between one-year default rate q and recovery rate R for corporate bonds; according to their analysis the best-fitting relationship is $R(q) \approx (0.52 - 6.9q)^+$.

Clearly, these findings call for the inclusion of systematic recovery risk in standard credit risk models. This is easily accomplished in the mixture-model framework—we replace the constant δ_i with some function $\delta_i(\boldsymbol{\psi})$—but the challenge lies in the estimation of the function $\delta_i(\cdot)$ describing the relationship between loss given default and the systematic factors.

11.2.4 Threshold Models as Mixture Models

Although the mixture models of this section seem, at first glance, to be different in structure from the threshold models of Section 11.1, it is important to realize that the majority of useful threshold models, including all the examples we have given, can be represented as Bernoulli mixture models. This is a very useful insight, because the Bernoulli mixture format has a number of advantages over the threshold format.

- Bernoulli mixture models lend themselves to Monte Carlo risk studies. From the analyses of this section we obtain methods for sampling from many of the models we have discussed, such as the t copula threshold model used in Section 11.1.5.
- Mixture models are arguably more convenient for statistical fitting purposes. We show in Section 11.5.3 that statistical techniques for generalized linear mixed models can be used to fit mixture models to empirical default data gathered over several time periods.
- The large-portfolio behaviour of Bernoulli mixtures can be analysed and understood in terms of the behaviour of the distribution of the common economic factors, as will be shown in Section 11.3.

To motivate the subsequent analysis we begin by computing the mixture model representation of the simple one-factor Gaussian threshold model in (11.8). It is convenient to identify the variable Ψ in the mixture representation with minus the factor F in the threshold representation; this yields conditional default probabilities that are increasing in Ψ and leads to formulas that are in line with the Basel IRB formula. With $F = -\Psi$ the one-factor model takes the form

$$X_i = -\sqrt{\beta_i}\Psi + \sqrt{1-\beta_i}\varepsilon_i.$$

By definition, company i defaults if and only if $X_i \leqslant d_i$ and hence if and only if $\sqrt{1-\beta_i}\varepsilon_i \leqslant d_i + \sqrt{\beta_i}\Psi$. Since the variables $\varepsilon_1, \ldots, \varepsilon_m$ and Ψ are independent, default events are independent conditional on Ψ and we can compute

$$\begin{aligned} p_i(\psi) = P(Y_i = 1 \mid \Psi = \psi) &= P(\sqrt{1-\beta_i}\varepsilon_i \leqslant d_i + \sqrt{\beta_i}\Psi \mid \Psi = \psi) \\ &= \Phi\left(\frac{d_i + \sqrt{\beta_i}\psi}{\sqrt{1-\beta_i}}\right), \end{aligned} \tag{11.20}$$

where we have used the fact that ε_i is standard normally distributed. The threshold is typically set so that the default probability matches an exogenously chosen value p_i, so that $d_i = \Phi^{-1}(p_i)$. In that case we obtain

$$p_i(\psi) = \Phi\left(\frac{\Phi^{-1}(p_i) + \sqrt{\beta_i}\psi}{\sqrt{1-\beta_i}}\right). \tag{11.21}$$

In the following we want to extend this idea to more general threshold models with a factor structure for the critical variables. We give a condition that ensures that a threshold model can be written as a Bernoulli mixture model.

Definition 11.9. A random vector $\boldsymbol{X}$ has a p-dimensional *conditional independence structure* with conditioning variable $\boldsymbol{\Psi}$ if there is some $p < m$ and a p-dimensional random vector $\boldsymbol{\Psi} = (\Psi_1, \ldots, \Psi_p)'$ such that, conditional on $\boldsymbol{\Psi}$, the rvs $X_1, \ldots, X_m$ are independent.

In the motivating example the conditioning variable was taken to be $\Psi = -F$. The next lemma generalizes the computations in (11.20) to any threshold model with a conditional independence structure.

Lemma 11.10. *Let $(\boldsymbol{X}, \boldsymbol{d})$ be a threshold model for an m-dimensional random vector $\boldsymbol{X}$. If $\boldsymbol{X}$ has a p-dimensional conditional independence structure with conditioning variable $\boldsymbol{\Psi}$, then the default indicators $Y_i = I_{\{X_i \leqslant d_i\}}$ follow a Bernoulli mixture model with factor $\boldsymbol{\Psi}$, where the conditional default probabilities are given by $p_i(\boldsymbol{\psi}) = P(X_i \leqslant d_i \mid \boldsymbol{\Psi} = \boldsymbol{\psi})$.*

Proof. For $\boldsymbol{y} \in \{0,1\}^m$ define the set $B := \{1 \leqslant i \leqslant m \colon y_i = 1\}$ and let $B^{\mathrm{c}} = \{1, \ldots, m\} \backslash B$. We have

$$\begin{aligned} P(\boldsymbol{Y} = \boldsymbol{y} \mid \boldsymbol{\Psi} = \boldsymbol{\psi}) &= P\Bigg(\bigcap_{i \in B} \{X_i \leqslant d_i\} \bigcap_{i \in B^{\mathrm{c}}} \{X_i > d_i\} \;\Bigg|\; \boldsymbol{\Psi} = \boldsymbol{\psi} \Bigg) \\ &= \prod_{i \in B} P(X_i \leqslant d_i \mid \boldsymbol{\Psi} = \boldsymbol{\psi}) \prod_{i \in B^{\mathrm{c}}} (1 - P(X_i \leqslant d_i \mid \boldsymbol{\Psi} = \boldsymbol{\psi})). \end{aligned}$$

Hence, conditional on $\boldsymbol{\Psi} = \boldsymbol{\psi}$, the Y_i are independent Bernoulli variables with success probability $p_i(\boldsymbol{\psi}) := P(X_i \leqslant d_i \mid \boldsymbol{\Psi} = \boldsymbol{\psi})$. □

We now consider a number of examples.

Example 11.11 (Gaussian threshold model). Consider the general Gaussian threshold model with the factor structure in (11.6), which takes the form

$$X_i = \sqrt{\beta_i}\boldsymbol{a}_i'\boldsymbol{F} + \sqrt{1 - \beta_i}\varepsilon_i, \tag{11.22}$$

where $\varepsilon_1, \ldots, \varepsilon_m$ are iid standard normal and where $\mathrm{var}(\boldsymbol{a}_i'\boldsymbol{F}) = 1$ for all i. Conditional on $\boldsymbol{\Psi} = -\boldsymbol{F}$, the vector $\boldsymbol{X}$ is normally distributed with diagonal covariance matrix and thus has conditional independence structure. With $d_i = \Phi^{-1}(p_i)$ the conditional default probabilities are given by

$$\begin{aligned} p_i(\boldsymbol{\psi}) &= P(Y_i = 1 \mid \boldsymbol{\Psi} = \boldsymbol{\psi}) = P(\sqrt{1 - \beta_i}\varepsilon_i \leqslant d_i + \sqrt{\beta_i}\boldsymbol{a}_i'\boldsymbol{\psi}) \\ &= \Phi\left(\frac{\Phi^{-1}(p_i) + \sqrt{\beta_i}\boldsymbol{a}_i'\boldsymbol{\psi}}{\sqrt{1 - \beta_i}} \right). \end{aligned} \tag{11.23}$$

By comparison with Example 11.6 we see that the individual stochastic default probabilities $p_i(\boldsymbol{\Psi})$ have a probit-normal distribution with parameters μ_i and σ_i given by

$$\mu_i = \Phi^{-1}(p_i)/\sqrt{1 - \beta_i} \quad \text{and} \quad \sigma_i^2 = \beta_i/(1 - \beta_i).$$

Example 11.12 (Student t threshold model). Now consider the case where the critical variables are of the form $\boldsymbol{X} = \sqrt{W}\boldsymbol{Z}$, where $\boldsymbol{Z}$ follows the Gaussian factor model in (11.22) and $W \sim \mathrm{Ig}(\frac{1}{2}\nu, \frac{1}{2}\nu)$. The vector $\boldsymbol{X}$ has a multivariate t distribution with ν degrees of freedom and standard univariate t margins with ν degrees of freedom.

This time we condition on $\boldsymbol{\Psi} = (-\boldsymbol{F}', W)'$. Given $\boldsymbol{\Psi} = (\tilde{\boldsymbol{\psi}}, w)$, the vector $\boldsymbol{X}$ has a multivariate normal distribution with independent components, and a computation similar to that in the previous example gives

$$p_i(\boldsymbol{\psi}) = p_i(\tilde{\boldsymbol{\psi}}, w) = \Phi\left(\frac{t_\nu^{-1}(p_i)w^{-1/2} + \sqrt{\beta_i}\boldsymbol{a}_i'\tilde{\boldsymbol{\psi}}}{\sqrt{1-\beta_i}}\right). \tag{11.24}$$

The formulas (11.23) and (11.24) are useful for Monte Carlo simulation of the corresponding threshold models. For example, rather than simulating an m-dimensional t distribution to implement the t model, one only needs to simulate a p-dimensional normal vector $\tilde{\boldsymbol{\Psi}}$ with $p \ll m$ and an independent gamma-distributed variate $V = W^{-1}$. In the second step of the simulation one simply conducts a series of independent Bernoulli experiments with default probabilities $p_i(\boldsymbol{\Psi})$ to decide whether individual companies default.

Application to Archimedean copula models. Another class of threshold models with an equivalent mixture representation is provided by models where the critical variables have an exchangeable LT-Archimedean copula in the sense of Definition 7.52. Consider a threshold model $(\boldsymbol{X}, \boldsymbol{d})$, where $\boldsymbol{X}$ has an exchangeable LT-Archimedean copula C with generator given by the Laplace transform $\hat{G}$ of some df G on $[0, \infty)$ with $G(0) = 0$. Let $\boldsymbol{p} = (p_1, \ldots, p_m)'$ denote the vector of default probabilities.

Consider now a non-negative rv $\Psi \sim G$ and rvs $U_1, \ldots, U_m$ that are conditionally independent given Ψ with conditional distribution function $P(U_i \leqslant u \mid \Psi = \psi) = \exp(-\psi \hat{G}^{-1}(u))$ for $u \in [0, 1]$. Proposition 7.51 then shows that $\boldsymbol{U}$ has df C. Moreover, by Lemma 11.2, $(\boldsymbol{X}, \boldsymbol{d})$ and $(\boldsymbol{U}, \boldsymbol{p})$ are two equivalent threshold models for default. By construction, $\boldsymbol{U}$ has a one-dimensional conditional independence structure with conditioning variable Ψ, and the conditional default probabilities are given by

$$p_i(\psi) = P(U_i \leqslant p_i \mid \Psi = \psi) = \exp(-\psi \hat{G}^{-1}(p_i)). \tag{11.25}$$

In order to simulate from a threshold model based on an LT-Archimedean copula we may therefore use the following efficient and simple approach. In a first step we simulate a realization ψ of Ψ and then we conduct m independent Bernoulli experiments with default probabilities $p_i(\psi)$ as in (11.25) to simulate a realization of the defaulting counterparties.

Example 11.13 (Clayton copula). As an example consider the Clayton copula with parameter $\theta > 0$. Suppose we wish to construct an exchangeable Bernoulli mixture model with default probability π and joint default probability π_2 that is equivalent to a threshold model with the Clayton copula for the critical variables. As mentioned

in Algorithm 7.53, a gamma-distributed rv $\Psi \sim \mathrm{Ga}(1/\theta, 1)$ (see Section A.2.4 for a definition) has Laplace transform $\hat{G}(t) = (1+t)^{-1/\theta}$. Using (11.25), the mixing variable of the equivalent Bernoulli mixture model can be defined by setting $Q = p_1(\Psi) = \exp(-\Psi(\pi^{-\theta} - 1))$.

Using (11.4), the required value of θ to give the desired joint default probabilities is the solution to the equation $\pi_2 = C_\theta(\pi, \pi) = (2\pi^{-\theta} - 1)^{-1/\theta}$, $\theta > 0$. It is easily seen that π_2 and, hence, the default correlation in our exchangeable Bernoulli mixture model are increasing in θ; for $\theta \to 0$ we obtain independent defaults and for $\theta \to \infty$ defaults become comonotonic and default correlation tends to one.

11.2.5 Poisson Mixture Models and CreditRisk$^+$

Since default is typically a rare event, it is possible to approximate Bernoulli indicator rvs for default with Poisson rvs and to approximate Bernoulli mixture models with Poisson mixture models. By choosing independent gamma distributions for the economic factors $\boldsymbol{\Psi}$ and using the Poisson approximation, we obtain a particularly tractable model for portfolio losses, known as CreditRisk$^+$.

Poisson approximation and Poisson mixture models. To be more precise, assume that, given the factors $\boldsymbol{\Psi}$, the default indicator variables $Y_1, \ldots, Y_m$ for a particular time horizon are conditionally independent Bernoulli variables satisfying $P(Y_i = 1 \mid \boldsymbol{\Psi} = \boldsymbol{\psi}) = p_i(\boldsymbol{\psi})$. Moreover, assume that the distribution of $\boldsymbol{\Psi}$ is such that the conditional default probabilities $p_i(\boldsymbol{\psi})$ tend to be very small. In this case the Y_i variables can be approximated by conditionally independent Poisson variables $\tilde{Y}_i$ satisfying $\tilde{Y}_i \mid \boldsymbol{\Psi} = \boldsymbol{\psi} \sim \mathrm{Poi}(p_i(\boldsymbol{\psi}))$, since

$$P(\tilde{Y}_i = 0 \mid \boldsymbol{\Psi} = \boldsymbol{\psi}) = \mathrm{e}^{-p_i(\boldsymbol{\psi})} \approx 1 - p_i(\boldsymbol{\psi}),$$
$$P(\tilde{Y}_i = 1 \mid \boldsymbol{\Psi} = \boldsymbol{\psi}) = p_i(\boldsymbol{\psi})\mathrm{e}^{-p_i(\boldsymbol{\psi})} \approx p_i(\boldsymbol{\psi}).$$

Moreover, the portfolio loss $L = \sum_{i=1}^m e_i \delta_i Y_i$ can be approximated by $\tilde{L} = \sum_{i=1}^m e_i \delta_i \tilde{Y}_i$. Of course, it is possible for a company to "default more than once" in the approximating Poisson model, albeit with a very low probability.

We now give a formal definition of a Poisson mixture model for counting variables that parallels the definition of a Bernoulli mixture model in Section 11.2.1.

Definition 11.14 (Poisson mixture model). Given some $p < m$ and a p-dimensional random vector $\boldsymbol{\Psi} = (\Psi_1, \ldots, \Psi_p)'$, the random vector $\tilde{\boldsymbol{Y}} = (\tilde{Y}_1, \ldots, \tilde{Y}_m)'$ follows a Poisson mixture model with factors $\boldsymbol{\Psi}$ if there are functions $\lambda_i : \mathbb{R}^p \to (0, \infty)$, $1 \leqslant i \leqslant m$, such that, conditional on $\boldsymbol{\Psi} = \boldsymbol{\psi}$, the random vector $\tilde{\boldsymbol{Y}}$ is a vector of independent Poisson distributed rvs with rate parameter $\lambda_i(\boldsymbol{\psi})$.

If $\tilde{\boldsymbol{Y}}$ follows a Poisson mixture model and if we define the indicators $Y_i = I_{\{\tilde{Y}_i \geqslant 1\}}$, then $\boldsymbol{Y}$ follows a Bernoulli mixture model and the mixing variables are related by $p_i(\cdot) = 1 - \mathrm{e}^{-\lambda_i(\cdot)}$.

The CreditRisk$^+$ model. The CreditRisk$^+$ model for credit risk was proposed by Credit Suisse Financial Products in 1997 (see Credit Suisse Financial Products 1997). It has the structure of the Poisson mixture model in Definition 11.14, where

the factor vector $\boldsymbol{\Psi}$ consists of p independent gamma-distributed rvs. The distributional assumptions and functional forms imposed in CreditRisk$^+$ make it possible to compute the distribution of the number of defaults and the aggregate portfolio loss fairly explicitly using techniques for compound distributions and mixture distributions that are well known in actuarial mathematics and which are also discussed in Chapter 13 (see Sections 13.2.2 and 13.2.4 in particular).

The (stochastic) parameter $\lambda_i(\boldsymbol{\Psi})$ of the conditional Poisson distribution for firm i is assumed to take the form

$$\lambda_i(\boldsymbol{\Psi}) = k_i \boldsymbol{w}_i' \boldsymbol{\Psi} \tag{11.26}$$

for a constant $k_i > 0$, for non-negative factor weights $\boldsymbol{w}_i = (w_{i1}, \ldots, w_{ip})'$ satisfying $\sum_j w_{ij} = 1$, and for p independent Ga(α_j, β_j)-distributed factors $\Psi_1, \ldots, \Psi_p$ with parameters set to be $\alpha_j = \beta_j = \sigma_j^{-2}$ for $\sigma_j > 0$ and $j = 1, \ldots, p$. This parametrization of the gamma variables ensures that we have $E(\Psi_j) = 1$ and $\operatorname{var}(\Psi_j) = \sigma_j^2$.

It is easy to verify that

$$E(\tilde{Y}_i) = E(E(\tilde{Y}_i \mid \boldsymbol{\Psi})) = E(\lambda_i(\boldsymbol{\Psi})) = k_i E(\boldsymbol{w}_i' \boldsymbol{\Psi}) = k_i,$$

so that k_i is the expected number of defaults for obligor i over the time period. Setting $Y_i = I_{\{\tilde{Y}_i \geqslant 1\}}$ we also observe that

$$P(Y_i = 1) = E(P(\tilde{Y}_i > 0 \mid \boldsymbol{\Psi})) = E(1 - \exp(-k_i \boldsymbol{w}_i' \boldsymbol{\Psi})) \approx k_i E(\boldsymbol{w}_i' \boldsymbol{\Psi}) = k_i,$$

for k_i small, so that k_i is approximately equal to the default probability.

Remark 11.15. The exchangeable version of CreditRisk$^+$ is extremely close to an exchangeable Bernoulli mixture model with beta mixing distribution. To see this, observe that in the exchangeable case the implied Bernoulli mixture model for $\boldsymbol{Y}$ has mixing variable Q given by $Q = 1 - \mathrm{e}^{-k\Psi}$ for some $\Psi \sim \text{Ga}(\alpha, \beta)$, $k > 0$ and $\alpha = \beta$. For $q \in (0, 1)$ we therefore obtain

$$P(Q \leqslant q) = P(1 - \mathrm{e}^{-k\Psi} \leqslant q) = P\left(\Psi \leqslant -\frac{\ln(1-q)}{k}\right),$$

so that the densities g_Q and g_Ψ are related by $g_Q(q) = g_\Psi(-\ln(1-q)/k)/(k(1-q))$. Using the form of the density of the Ga(α, β) distribution we obtain

$$\begin{aligned} g_Q(q) &= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{1}{k(1-q)} \left(\frac{-\ln(1-q)}{k}\right)^{\alpha-1} \exp\left(\frac{\beta \ln(1-q)}{k}\right) \\ &= \left(\frac{\beta}{k}\right)^\alpha \frac{1}{\Gamma(\alpha)} (-\ln(1-q))^{\alpha-1} (1-q)^{(\beta/k)-1}. \end{aligned}$$

In a realistic credit risk model the parameters are chosen in such a way that the mass of the distribution of Q is concentrated on values of q close to zero, since default is typically a rare event. For small q we may use the approximation $-\ln(1-q) \approx q$ to observe that the functional form of g_Q is extremely close to that of a beta distribution with parameters α and β/k, where we again recall that the model is parametrized to have $\alpha = \beta$.

Distribution of the number of defaults. In CreditRisk$^+$ we have that, given $\boldsymbol{\Psi} = \boldsymbol{\psi}$, $\tilde{Y}_i \sim \text{Poi}(k_i \boldsymbol{w}_i' \boldsymbol{\psi})$, which implies that the distribution of the number of defaults $\tilde{M} := \sum_{i=1}^m \tilde{Y}_i$ satisfies

$$\tilde{M} \mid \boldsymbol{\Psi} = \boldsymbol{\psi} \sim \text{Poi}\left(\sum_{i=1}^m k_i \boldsymbol{w}_i' \boldsymbol{\psi} \right), \tag{11.27}$$

since the sum of independent Poisson variables is also a Poisson variable with a rate parameter given by the sum of the rate parameters of the independent variables.

To compute the unconditional distribution of $\tilde{M}$ we require a well-known result on mixed Poisson distributions, which appears as Proposition 13.21 in a discussion of relevant actuarial methodology for quantitative risk management in Chapter 13. This result says that if the rv N is conditionally Poisson with a gamma-distributed rate parameter $\Lambda \sim \text{Ga}(\alpha, \beta)$, then N has a negative binomial distribution, $N \sim \text{NB}(\alpha, \beta/(\beta + 1))$.

In the case when $p = 1$ we may apply this result directly to (11.27) to deduce that $\tilde{M}$ has a negative binomial distribution (since a constant times a gamma variable remains gamma distributed). For arbitrary p we now show that $\tilde{M}$ is equal in distribution to a sum of p independent negative binomial rvs. This follows by observing that

$$\sum_{i=1}^m k_i \boldsymbol{w}_i' \boldsymbol{\Psi} = \sum_{i=1}^m k_i \sum_{j=1}^p w_{ij} \Psi_j = \sum_{j=1}^p \Psi_j \left(\sum_{i=1}^m k_i w_{ij} \right).$$

Now consider rvs $\tilde{M}_1, \ldots, \tilde{M}_p$ such that $\tilde{M}_j$ is conditionally Poisson with mean $(\sum_{i=1}^m k_i w_{ij}) \psi_j$ conditional on $\Psi_j = \psi_j$. The independence of the components $\Psi_1, \ldots, \Psi_p$ implies that the $\tilde{M}_j$ are independent, and by construction we have $\tilde{M} \stackrel{\text{d}}{=} \sum_{j=1}^p \tilde{M}_j$. Moreover, the rvs $(\sum_{i=1}^m k_i w_{ij}) \Psi_j$ are gamma distributed, so that each of the $\tilde{M}_j$ has a negative binomial distribution by Proposition 13.21.

Distribution of the aggregate loss. To obtain a tractable model, exposures are discretized in CreditRisk$^+$ using the concept of exposure bands. The CreditRisk$^+$ documentation (see Credit Suisse Financial Products 1997) suggests that the LGD can be subsumed in the exposure by multiplying the actual exposure by a value for the LGD that is typical for an obligor with the same credit rating. We will adopt this approach and assume that the losses arising from the individual obligors are of the form $\tilde{L}_i = e_i \tilde{Y}_i$, where the e_i are known (LGD-adjusted) exposures.

For all i, we discretize e_i in units of an amount ϵ; we replace e_i by a value $\ell_i \epsilon \geqslant e_i$, where ℓ_i is a positive integer multiplier. We now define exposure bands $b = 1, \ldots, n$ corresponding to the distinct values $\ell^{(1)}, \ldots, \ell^{(n)}$ for the multipliers. In other words, we group obligors in exposure bands according to the values of their discretized exposures.

Let s_b denote the set of indices for the obligors in exposure band b; this means that $i \in s_b \iff \ell_i = \ell^{(b)}$. Let $\tilde{L}^{(b)} = \sum_{i \in s_b} \epsilon \ell_i \tilde{Y}_i$ denote the aggregate loss in exposure band b. We have $\tilde{L}^{(b)} = \epsilon \ell^{(b)} \tilde{M}^{(b)}$, where $\tilde{M}^{(b)} = \sum_{i \in s_b} \tilde{Y}_i$ denotes the number of defaults in exposure band b. Let $\tilde{L} = \sum_{b=1}^n \tilde{L}^{(b)}$ denote the aggregate portfolio loss.

We now want to determine the distribution of $\tilde{L}$. The following theorem gives the necessary information for achieving this with Fourier inversion.

Theorem 11.16. *Let $\tilde{L}$ represent the aggregate loss in the general p-factor CreditRisk$^+$ model with exposures discretized into exposure bands as described above. The following then hold.*

(i) *The Laplace–Stieltjes transform of the df of $\tilde{L}$ is given by*

$$\hat{F}_{\tilde{L}}(s) = \prod_{j=1}^{p} \left(1 + \sigma_j^2 \sum_{i=1}^{m} k_i w_{ij} \left(1 - \sum_{b=1}^{n} \mathrm{e}^{-s\epsilon\ell^{(b)}} q_{jb}\right)\right)^{-\sigma_j^{-2}}, \qquad (11.28)$$

where $q_{jb} = \sum_{i \in s_b} k_i w_{ij} / \sum_{i=1}^{m} k_i w_{ij}$ for $b = 1, \ldots, n$.

(ii) *The distribution of $\tilde{L}$ has the structure $\tilde{L} \stackrel{d}{=} \sum_{j=1}^{p} Z_j$, where the Z_j are independent variables that follow a compound negative binomial distribution. More precisely, it holds that $Z_j \sim \mathrm{CNB}(\sigma_j^{-2}, \theta_j, G_{X_j})$ with $\theta_j = (1 + \sigma_j^2 \sum_{i=1}^{m} k_i w_{ij})^{-1}$ and G_{X_j} the df of a multinomial random variable X_j taking the value $\epsilon\ell^{(b)}$ with probability q_{jb}.*

Proof. The proof requires the mathematics of compound distributions as described in Section 13.2.2.

(i) Conditional on $\boldsymbol{\Psi} = \boldsymbol{\psi}$, $M^{(b)} = \sum_{i \in s_b} \tilde{Y}_i$ has a Poisson distribution with parameter

$$\lambda^{(b)}(\boldsymbol{\psi}) := \sum_{i \in s_b} \lambda_i(\boldsymbol{\psi}) = \sum_{i \in s_b} k_i \boldsymbol{w}_i' \boldsymbol{\psi},$$

and the loss $\tilde{L}^{(b)} = \epsilon\ell^{(b)} \tilde{M}^{(b)}$ in exposure band b has a compound Poisson distribution given by

$$\tilde{L}^{(b)} \mid \boldsymbol{\Psi} = \boldsymbol{\psi} \sim \mathrm{CPoi}(\lambda^{(b)}(\boldsymbol{\psi}), G^{(b)}),$$

where $G^{(b)}$ is the df of point mass at $\epsilon\ell^{(b)}$. It follows from Proposition 13.10 on sums of compound Poisson variables that

$$\tilde{L} \mid \boldsymbol{\Psi} = \boldsymbol{\psi} \sim \mathrm{CPoi}\left(\lambda(\boldsymbol{\psi}), \sum_{b=1}^{n} \frac{\lambda^{(b)}(\boldsymbol{\psi})}{\lambda(\boldsymbol{\psi})} G^{(b)}\right), \qquad (11.29)$$

where $\lambda(\boldsymbol{\psi}) := \sum_{b=1}^{n} \lambda^{(b)}(\boldsymbol{\psi})$. The (conditional) severity distribution in (11.29) is the distribution function of a multinomial random variable that takes the values $\epsilon\ell^{(b)}$ with probabilities $\lambda^{(b)}(\boldsymbol{\psi})/\lambda(\boldsymbol{\psi})$ for $b = 1, \ldots, n$.

Writing $\hat{F}_{L|\boldsymbol{\Psi}}(s \mid \boldsymbol{\psi})$ for the Laplace–Stieltjes transform of the conditional distribution function of $\tilde{L}$ given $\boldsymbol{\Psi}$, we can use equation (13.11) and Example 13.5 to

infer that

$$\begin{aligned}\hat{F}_{\tilde{L}|\boldsymbol{\Psi}}(s \mid \boldsymbol{\psi}) &= \exp\left(-\lambda(\boldsymbol{\psi})\left(1 - \sum_{b=1}^{n} \frac{\lambda^{(b)}(\boldsymbol{\psi})}{\lambda(\boldsymbol{\psi})} \mathrm{e}^{-s\epsilon\ell^{(b)}}\right)\right) \\ &= \exp\left(-\left(\sum_{i=1}^{m} k_i \sum_{j=1}^{p} w_{ij}\psi_j - \sum_{b=1}^{n} \mathrm{e}^{-s\epsilon\ell^{(b)}} \sum_{i\in s_b} k_i \sum_{j=1}^{p} w_{ij}\psi_j\right)\right) \\ &= \exp\left(-\sum_{j=1}^{p} \psi_j \left(\sum_{i=1}^{m} k_i w_{ij} - \sum_{b=1}^{n} \mathrm{e}^{-s\epsilon\ell^{(b)}} \sum_{i\in s_b} k_i w_{ij}\right)\right) \\ &= \prod_{j=1}^{p} \exp\left(-\psi_j \sum_{i=1}^{m} k_i w_{ij}\left(1 - \sum_{b=1}^{n} \mathrm{e}^{-s\epsilon\ell^{(b)}} q_{jb}\right)\right).\end{aligned}$$

Writing g_{Ψ_j} for the density of the factor Ψ_j and using the independence of the factors, it follows that

$$\hat{F}_{\tilde{L}}(s) = \prod_{j=1}^{p} \int_0^\infty \exp\left(-\psi_j \sum_{i=1}^{m} k_i w_{ij}\left(1 - \sum_{b=1}^{n} \mathrm{e}^{-s\epsilon\ell^{(b)}} q_{jb}\right)\right) g_{\Psi_j}(\psi_j)\,\mathrm{d}\psi_j,$$

and equation (11.28) is derived by evaluating the integrals and recalling that the parameters of the gamma distribution are chosen to be $\alpha_j = \beta_j = \sigma_j^{-2}$.

(ii) To see that this is the Laplace–Stieltjes transform of the df of a sum of independent compound negative binomial distributions, observe that (11.28) may be written as

$$\hat{F}_{\tilde{L}}(s) = \prod_{j=1}^{p} \left(1 + \sigma_j^2 \sum_{i=1}^{m} k_i w_{ij}(1 - \hat{G}_{X_j}(s))\right)^{-\sigma_j^{-2}}, \tag{11.30}$$

where $\hat{G}_{X_j}$ is the Laplace–Stieltjes transform of G_{X_j}. Substituting $\theta_j = (1 + \sigma_j^2 \sum_{i=1}^{m} k_i w_{ij})^{-1}$ we obtain

$$\hat{F}_{\tilde{L}}(s) = \prod_{j=1}^{p} \left(\frac{\theta_j}{1 - \hat{G}_{X_j}(s)(1-\theta_j)}\right)^{\sigma_j^{-2}},$$

and, by comparing with Example 13.6, this can be seen to be the product of Laplace–Stieltjes transforms of the dfs of variables $Z_j \sim \mathrm{CNB}(\sigma_j^{-2}, \theta_j, G_{X_j})$. □

This theorem gives the key information required to evaluate the distribution of $\tilde{L}$ by applying Fourier inversion to the characteristic function of $\tilde{L}$. Indeed, we have

$$\phi_{\tilde{L}}(s) = \prod_{j=1}^{p} \left(\frac{\theta_j}{1 - \phi_{X_j}(s)(1-\theta_j)}\right)^{\sigma_j^{-2}},$$

where ϕ_{X_j} is the characteristic function of the multinomial severity distribution. It is now straightforward to compute ϕ_{X_j} using the fast Fourier transform and hence to compute the cf of $\tilde{L}$. The probability mass function of $\tilde{L}$ can then be computed by inverting $\phi_{\tilde{L}}$ using the inverse fast Fourier transform.

Notes and Comments

The logit-normal mixture model can be thought of as a one-factor version of the CreditPortfolioView model of Wilson (1997a,b). Details of this model can be found in Section 5 of Crouhy, Galai and Mark (2000). Further details of the beta-binomial distribution can be found in Joe (1997).

The rating agency Moody's uses a so-called binomial expansion technique to model default dependence in a simplistic way. The method, which is very popular with practitioners, is not based on a formal default risk model but is related to binomial distributions. The basic idea is to approximate a portfolio of m dependent counterparties by a homogeneous portfolio of $d < m$ independent counterparties with adjusted exposures and identical default probabilities; the index d is called the *diversity score* and is chosen according to rules defined by Moody's. For further information we refer to Davis and Lo (2001) and to Section 9.2.7 of Lando (2004).

The equivalence between threshold models and mixture models has been observed by Koyluoglu and Hickman (1998) and Gordy (2000) for the special case of CreditMetrics and CreditRisk$^+$. Applications of Proposition 7.51 to credit risk modelling are also discussed in Schönbucher (2005). The study of mixture representations for sequences of exchangeable Bernoulli rvs is related to a well-known result of de Finetti, which states that any *infinite* sequence $Y_1, Y_2, \ldots$ of exchangeable Bernoulli rvs has a representation as an exchangeable Bernoulli mixture; see, for instance, Theorem 35.10 in Billingsley (1995) for a precise statement. Any exchangeable model for $\boldsymbol{Y}$ that can be extended to arbitrary portfolio size m therefore has a representation as an exchangeable Bernoulli mixture model

A comprehensive description of CreditRisk$^+$ is given in its original documentation (Credit Suisse Financial Products 1997). An excellent discussion of the model structure from a more academic viewpoint is provided in Gordy (2000). Both sources also provide further information concerning the calibration of the factor variances σ_i and factor weights w_{ij}. The derivation of recursion formulas for the probabilities $P(\tilde{M} = k)$, $k = 0, 1, \ldots$, via Panjer recursion is given in Appendix A10 of the CreditRisk$^+$ documentation. In Gordy (2002) an alternative approach to the computation of the loss distribution in CreditRisk$^+$ is proposed using the saddlepoint approximation (see, for example, Jensen 1995). Further numerical work for CreditRisk$^+$ can be found in papers by Kurth and Tasche (2003), Glasserman (2004) and Haaf, Reiss and Schoenmakers (2004). Importance-sampling techniques for CreditRisk$^+$ are discussed in Glasserman and Li (2005).

11.3 Asymptotics for Large Portfolios

We now provide some asymptotic results for large portfolios in Bernoulli mixture models. These results can be used to approximate the credit loss distribution and associated risk measures in a large portfolio. Moreover, they are useful for identifying the key sources of model risk in a Bernoulli mixture model. In particular, we will see that in one-factor models the tail of the loss distribution is essentially determined by the tail of the mixing distribution, which has direct consequences for

the analysis of model risk in mixture models and for the setting of capital adequacy rules for loan books.

11.3.1 Exchangeable Models

We begin our discussion of the asymptotic properties of Bernoulli mixture models with the special case of an exchangeable model. We consider an infinite sequence of obligors indexed by $i \in \mathbb{N}$ with identical exposures $e_i = e$ and LGD equal to 100%. We assume that, given a mixing variable $Q \in [0, 1]$, the default indicators Y_i are independent Bernoulli random variables with conditional default probability $P(Y_i = 1 \mid Q = q) = q$. This simple model can be viewed an as idealization of a large pool of homogeneous obligors.

We are interested in the asymptotic behaviour of the relative loss (the loss expressed as a proportion of total exposure). Writing $L^{(m)} = \sum_{i=1}^{m} eY_i$ for the total loss of the first m companies, the corresponding relative loss is given by

$$\frac{L^{(m)}}{me} = \frac{1}{m} \sum_{i=1}^{m} Y_i.$$

Conditioning on $Q = q$, the Y_i are independent with mean q and the strong law of large numbers implies that, given $Q = q$, $\lim_{m\to\infty} L^{(m)}/(me) = q$ almost surely. This shows that, for large m, the behaviour of the relative loss is essentially governed by the mixing distribution $G(q)$ of Q. In particular, it can be shown that, for G strictly increasing,

$$\lim_{m\to\infty} \mathrm{VaR}_\alpha \left(\frac{L^{(m)}}{me} \right) = q_\alpha(Q) \tag{11.31}$$

(see Proposition 11.18 below).

These results can be used to analyse model risk in exchangeable Bernoulli mixture models. We consider the risk related to the choice of mixing distribution under the constraint that the default probability π and the default correlation ρ_Y (or, equivalently, π and π_2) are known and fixed.

According to (11.31), for large m the tail of $L^{(m)}$ is essentially determined by the tail of the mixing variable Q. In Figure 11.2 we plot the tail function of the probit-normal distribution (corresponding to the Gaussian threshold model), the logit-normal distribution, the beta distribution (close to CreditRisk$^+$; see Remark 11.15) and the mixture distribution corresponding to the Clayton copula (see Example 11.13). The plots are shown on a logarithmic scale and in all cases the first two moments have the values $\pi = 0.049$ and $\pi_2 = 0.003\,13$, which correspond roughly to Standard & Poor's rating category B; the parameter values for each of the models can be found in Table 11.3.

Inspection of Figure 11.2 shows that the tail functions differ significantly only after the 99% quantile, the logit-normal distribution being the one with the heaviest tail. From a practical point of view this means that the particular parametric form of the mixing distribution in a Bernoulli mixture model is of lesser importance once π and ρ_Y have been fixed. Of course this does not mean that Bernoulli mixtures are

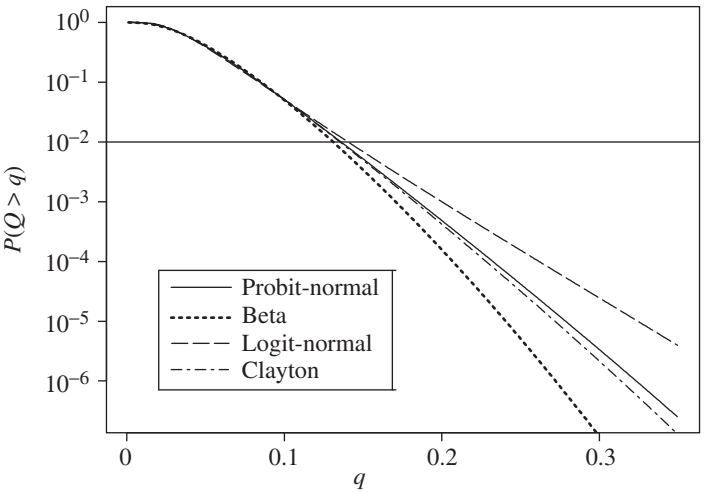

Figure 11.2. The tail of the mixing distribution of Q in four different exchangeable Bernoulli-mixture models: beta, probit-normal, logit-normal and Clayton. In all cases the first two moments have the values $\pi = 0.049$ and $\pi_2 = 0.003\,13$, which correspond roughly to Standard & Poor's rating category B; the actual parameter values can be found in Table 11.3. The horizontal line at 10^{-2} shows that the models only really start to differ around the 99th percentile of the mixing distribution.

Table 11.3. Parameter values for various exchangeable Bernoulli mixture models with identical values of π and π_2 (and ρ_Y). The values of π and π_2 correspond roughly to Standard & Poor's ratings CCC, B and BB (in fact, they have been estimated from 20 years of Standard & Poor's default data using the simple moment estimator in (11.50)). This table is used in the model-risk study of Section 11.1.5. and the simulation study of Section 11.5.2.

Model	Parameter	CCC	B	BB
All models	π	0.188	0.049	0.011 2
	π_2	0.042	0.003 13	0.000 197
	ρ_Y	0.044 6	0.015 7	0.006 43
Beta	a	4.02	3.08	1.73
	b	17.4	59.8	153
Probit-normal	μ	−0.93	−1.71	−2.37
	σ	0.316	0.264	0.272
Logit-normal	μ	−1.56	−3.1	−4.71
	σ	0.553	0.556	0.691
Clayton	π	0.188	0.049	0.011 2
	θ	0.070 4	0.032	0.024 7

immune to model risk; the tail of $L^{(m)}$ is quite sensitive to π and, in particular, to ρ_Y, and these parameters are not easily estimated (see Section 11.5.4 for a discussion of statistical inference for mixture models).

11.3.2 General Results

Since we are interested in asymptotic properties of the overall loss distribution, we also consider exposures and losses given default. Let $(e_i)_{i\in\mathbb{N}}$ be an infinite sequence of positive deterministic exposures, let $(Y_i)_{i\in\mathbb{N}}$ be the corresponding sequence of default indicators, and let $(\delta_i)_{i\in\mathbb{N}}$ be a sequence of rvs with values in $(0, 1]$ representing percentage losses given that default occurs. In this setting the loss for a portfolio of size m is given by $L^{(m)} = \sum_{i=1}^m L_i$, where $L_i = e_i\delta_i Y_i$ are the individual losses. We are interested in results for the relative loss $L^{(m)}/\sum_{i=1}^m e_i$, which expresses the loss as a proportion of total exposure. We introduce the notation $a_m = \sum_{i=1}^m e_i$ for the aggregate exposure to the first m obligors.

We now make some technical assumptions for our model.

(A1) There is a p-dimensional random vector $\boldsymbol{\Psi}$ such that, conditional on $\boldsymbol{\Psi}$, the $(L_i)_{i\in\mathbb{N}}$ form a sequence of independent rvs.

In this assumption the conditional independence structure is extended from the default indicators to the losses. Note that (A1) allows for the situation where δ_i depends on the systematic factors $\boldsymbol{\Psi}$. This extension is relevant from an empirical viewpoint since evidence suggests that losses given default tend to depend on the state of the underlying economy (see Section 11.2.3).

(A2) There is a function $\bar{\ell}\colon \mathbb{R}^p \to [0, 1]$ such that

$$\lim_{m\to\infty} \frac{1}{a_m} E(L^{(m)} \mid \boldsymbol{\Psi} = \boldsymbol{\psi}) = \bar{\ell}(\boldsymbol{\psi}) \tag{11.32}$$

for all $\boldsymbol{\psi} \in \mathbb{R}^p$. We call $\bar{\ell}(\boldsymbol{\psi})$ the asymptotic relative loss function.

Assumption (A2) implies that we preserve the essential composition of the portfolio as we allow it to grow (see, for instance, Example 11.20).

(A3) The sequence of exposures satisfies $\lim_{m\to\infty} a_m = \infty$ and $\sum_{i=1}^\infty (e_i/a_i)^2 < \infty$.

This is a very weak technical assumption that would be satisfied by any realistic sequence of exposures. For example, if $e_i = e$ for all i, then $\sum_{i=1}^\infty (e_i/a_i)^2 = \sum_{i=1}^\infty i^{-2} < \infty$. Even in the case where $e_i = i$, so that exposures grow linearly with portfolio size, we have $\sum_{i=1}^\infty (e_i/a_i)^2 = \sum_{i=1}^\infty (2/(i+1))^2 < \infty$, since $a_i = \sum_{k=1}^i k = i(i+1)/2$. To find a counterexample we would need to take a sequence of exposures where the cumulative sum grows at the same rate as the maximum of the first m exposures. In intuitive terms this means that the portfolio is dominated by a few large exposures (name concentration).

The following result shows that under these assumptions the average portfolio loss is essentially determined by the asymptotic relative loss function $\bar{\ell}$ and by the realization of the factor random vector $\boldsymbol{\Psi}$.

Proposition 11.17. *Consider a sequence $L^{(m)} = \sum_{i=1}^m L_i$ satisfying Assumptions (A1)–(A3) above. Denote by $P(\cdot \mid \boldsymbol{\Psi} = \boldsymbol{\psi})$ the conditional distribution of*

the sequence $(L_i)_{i\in\mathbb{N}}$ *given* $\boldsymbol{\Psi} = \boldsymbol{\psi}$. *Then*

$$\lim_{m\to\infty} \frac{1}{a_m} L^{(m)} = \bar{\ell}(\boldsymbol{\psi}), \quad P(\cdot \mid \boldsymbol{\Psi} = \boldsymbol{\psi}) \text{ a.s.}$$

Proof. The proof is based on the following version of the law of large numbers for non-identically distributed random variables given by Petrov (1995, Theorem 6.7). If $(Z_i)_{i\in\mathbb{N}}$ is a sequence of independent random variables and $(a_i)_{i\in\mathbb{N}}$ is a sequence of positive constants satisfying $\lim_{m\to\infty} a_m = \infty$ and $\sum_{i=1}^{\infty}(\operatorname{var}(Z_i)/a_i^2) < \infty$, then, as $m \to \infty$,

$$\frac{1}{a_m}\left(\sum_{i=1}^{m} Z_i - E\left(\sum_{i=1}^{m} Z_i\right)\right) \to 0 \text{ a.s.}$$

We set $Z_i = L_i = e_i\delta_i Y_i$ and $a_i = \sum_{k=1}^{i} e_k$ as before. We apply this result conditional on $\boldsymbol{\Psi} = \boldsymbol{\psi}$; that is, we work under the measure $P(\cdot \mid \boldsymbol{\Psi} = \boldsymbol{\psi})$. Under this measure the L_i are independent by Assumption (A1). Note that $\operatorname{var}(L_i) = e_i^2 \operatorname{var}(\delta_i Y_i) \leqslant e_i^2$, since $\delta_i Y_i$ is an rv on [0, 1]. Using Assumption (A3) we verify that

$$\sum_{i=1}^{\infty} \frac{\operatorname{var}(Z_i)}{a_i^2} \leqslant \sum_{i=1}^{\infty} \left(\frac{e_i}{a_i}\right)^2 < \infty.$$

Applying Petrov's result and Assumption (A2) we get

$$\lim_{m\to\infty} \frac{1}{a_m} L^{(m)} - \bar{\ell}(\boldsymbol{\psi}) = \lim_{m\to\infty} \left(\frac{1}{a_m}\sum_{i=1}^{m} L_i - \frac{1}{a_m} E\left(\sum_{i=1}^{m} L_i \mid \boldsymbol{\Psi} = \boldsymbol{\psi}\right)\right) = 0.$$

□

For one-factor Bernoulli mixture models a stronger result can be obtained that links the quantiles of the relative portfolio loss $L^{(m)}/a_m$ to quantiles of the mixing distribution.

Proposition 11.18. *Consider a sequence* $L^{(m)} = \sum_{i=1}^{m} L_i$ *satisfying Assumptions (A1)–(A3) with a one-dimensional mixing variable* Ψ *with df* G. *Assume that the conditional asymptotic loss function* $\bar{\ell}(\psi)$ *is strictly increasing and continuous and that* G *is strictly increasing at* $q_\alpha(\Psi)$, *i.e. that* $G(q_\alpha(\Psi)+\delta) > \alpha$ *for every* $\delta > 0$. *Then*

$$\lim_{m\to\infty} \operatorname{VaR}_\alpha\left(\frac{1}{a_m} L^{(m)}\right) = \bar{\ell}(q_\alpha(\Psi)). \tag{11.33}$$

The assumption that $\bar{\ell}$ is strictly increasing makes sense if it is assumed that low values of Ψ correspond to good states of the world with lower conditional default probabilities and lower losses given default than average, while high values of Ψ correspond to bad states with correspondingly higher losses given default.

Proof. The proof is based on the following simple intuition. Since $L^{(m)}/a_m$ converges to $\bar{\ell}(\Psi)$ and since $\bar{\ell}$ is strictly increasing by assumption, we have for large m

$$q_\alpha\left(\frac{L^{(m)}}{a_m}\right) \approx q_\alpha(\bar{\ell}(\Psi)) = \bar{\ell}(q_\alpha(\Psi)).$$

To turn this into a formal argument we use the following continuity result for quantiles, a proof of which may be found in Fristedt and Gray (1997, Proposition 5, p. 250) or Resnick (2008, Proposition 0.1).

Lemma 11.19. *Consider a sequence of random variables $(Z_m)_{m\in\mathbb{N}}$ that converges in distribution to a random variable Z. Then $\lim_{m\to\infty} q_u(Z_m) = q_u(Z)$ at all points of continuity u of the quantile function $u \mapsto q_u(Z)$.*

In our case, $L^{(m)}/a_m$ converges to $\bar{\ell}(\Psi)$ almost surely, and hence in distribution. The assumption that the function $\bar{\ell}$ is strictly increasing and that G is strictly increasing at $q_\alpha(\Psi)$ ensures that the distribution function of $\bar{\ell}(\Psi)$ is strictly increasing at that point so that the quantile function of the rv $\bar{\ell}(\Psi)$ is continuous at α. This shows that

$$\lim_{m\to\infty} q_\alpha\left(\frac{1}{a_m}L^{(m)}\right) = q_\alpha(\bar{\ell}(\Psi)).$$

Finally, the equality $q_\alpha(\bar{\ell}(\Psi)) = \bar{\ell}(q_\alpha(\Psi))$ follows from Proposition A.3. □

Example 11.20. Consider the one-factor Bernoulli mixture model for k exchangeable groups defined by (11.18). Denote by $r(i)$ the group of obligor i and assume that, within each group r, the exposures, LGDs and conditional default probabilities are identical and are given by e_r, δ_r and $p_r(\psi)$, respectively.

Suppose that we allow the portfolio to grow and that we write $m_r^{(m)}$ for the number of obligors in group r when the portfolio size is m. The relative exposure to group r is given by $\lambda_r^{(m)} = e_r m_r^{(m)} / \sum_{r=1}^k e_r m_r^{(m)}$, and we assume that $\lambda_r^{(m)} \to \lambda_r$ as $m \to \infty$. In this case the asymptotic relative loss function in equation (11.32) is

$$\begin{aligned}
\bar{\ell}(\psi) &= \lim_{m\to\infty} \sum_{i=1}^m \frac{e_{r(i)}}{\sum_{i=1}^m e_{r(i)}} \delta_{r(i)} p_{r(i)}(\psi) \\
&= \lim_{m\to\infty} \sum_{r=1}^k \frac{e_r m_r^{(m)}}{\sum_{r=1}^k e_r m_r^{(m)}} \delta_r h(\mu_r + \sigma\psi) \\
&= \sum_{r=1}^k \lambda_r \delta_r h(\mu_r + \sigma\psi).
\end{aligned}$$

Since Ψ is assumed to have a standard normal distribution, (11.33) implies that

$$\lim_{m\to\infty} q_\alpha\left(\frac{L^{(m)}}{\sum_{i=1}^m e_i}\right) = \sum_{r=1}^k \lambda_r \delta_r h(\mu_r + \sigma\Phi^{-1}(\alpha)). \tag{11.34}$$

For large m, since $\lambda_r \sum_{1=1}^m e_i \approx m_r^{(m)} e_r$, we get that

$$\mathrm{VaR}_\alpha(L^{(m)}) \approx \sum_{r=1}^k m_r^{(m)} e_r \delta_r h(\mu_r + \sigma_r\Phi^{-1}(\alpha)). \tag{11.35}$$

11.3.3 The Basel IRB Formula

In this section we examine how the considerations of Sections 11.3.1 and 11.3.2 have influenced the Basel capital adequacy framework, which was discussed in more general terms in Section 1.3. Under this framework a bank is required to hold 8% of the so-called *risk-weighted assets* (RWA) of its credit portfolio as risk capital. The RWA of a portfolio is given by the sum of the RWA of the individual risks in the portfolio, i.e. $\text{RWA}^{\text{portfolio}} = \sum_{i=1}^{m} \text{RWA}_i$. The quantity RWA_i reflects the exposure size and the riskiness of obligor i and takes the form $\text{RWA}_i = w_i e_i$, where w_i is a risk weight and e_i denotes exposure size.

Banks may choose between two options for determining the risk weight w_i, which must then be implemented for the entire portfolio. Under the simpler *standardized approach*, the risk weight w_i is determined by the type (sovereign, bank or corporate) and the credit rating of counterparty i. For instance, $w_i = 50\%$ for a corporation with a Moody's rating in the range A+ to A−. Under the more advanced *internal-ratings-based* (IRB) approach, the risk weight takes the form

$$w_i = (0.08)^{-1} c \delta_i \Phi\left(\frac{\Phi^{-1}(p_i) + \sqrt{\beta_i}\Phi^{-1}(0.999)}{\sqrt{1-\beta_i}} \right). \tag{11.36}$$

Here, c is a technical adjustment factor that is of minor interest to us, p_i represents the default probability, and δ_i is the percentage loss given default of obligor i. The parameter $\beta_i \in (0.12, 0.24)$ measures the systematic risk of obligor i. Estimates for p_i and (under the so-called advanced IRB approach) for δ_i and e_i are provided by the individual bank; the adjustment factor c and, most importantly, the value of β_i are determined by fixed rules within the Basel II Accord independently of the structure of the specific portfolio under consideration. The risk capital to be held for counterparty i is thus given by

$$\text{RC}_i = 0.08\text{RWA}_i = c \delta_i e_i \Phi\left(\frac{\Phi^{-1}(p_i) + \sqrt{\beta_i}\Phi^{-1}(0.999)}{\sqrt{1-\beta_i}} \right). \tag{11.37}$$

The interesting part of equation (11.37) is, of course, the expression involving the standard normal df, and we now give a derivation.

Consider a one-factor Gaussian threshold with default probabilities $p_1, \ldots, p_m$ and critical variables given by

$$X_i = \sqrt{\beta_i} F + \sqrt{1-\beta_i}\varepsilon_i \tag{11.38}$$

for iid standard normal rvs $F, \varepsilon_1, \ldots, \varepsilon_m$. By taking $\Psi = -F$, the equivalent Bernoulli mixture model was shown in Section 11.2.4 to have conditional default probabilities $p_i(\psi) = \Phi((\Phi^{-1}(p_i) + \sqrt{\beta_i}\psi)/\sqrt{1-\beta_i})$. Note that this is of the form $p_i(\psi) = h(\mu_i + \sigma_i \psi)$ for $h = \Phi$, $\mu_i = \Phi^{-1}(p_i)/\sqrt{1-\beta_i}$ and $\sigma_i = \sqrt{\beta_i/(1-\beta_i)}$. Assume, moreover, that the portfolio has a homogeneous group structure consisting of a few large groups with (approximately) identical exposures, PDs, LGDs and factor weights within the groups, as in Example 11.20.

Applying the analysis of that example and, in particular, equation (11.35), it follows that

$$\begin{aligned}\mathrm{VaR}_\alpha(L) &\approx \sum_{i=1}^{m} \delta_i e_i \, p_i(q_\alpha(\Psi)) \\ &= \sum_{i=1}^{m} \delta_i e_i \Phi\left(\frac{\Phi^{-1}(p_i) + \sqrt{\beta_i}\Phi^{-1}(\alpha)}{\sqrt{1-\beta_i}}\right).\end{aligned}$$

For $c = 1$ the risk capital RC_i in (11.37) can thus be considered as the asymptotic contribution of risk i to the 99.9% VaR of the overall portfolio in a one-factor Gaussian threshold model with homogeneous group structure. Note, further, that β_i can be viewed as the asset correlation for firms i within the same group.

While formula (11.36) is influenced by portfolio-theoretic considerations, the Basel framework falls short of reflecting the true dependence structure of a bank's credit portfolio for a number of reasons. First, in the Basel framework the parameters β_i are specified ad hoc by regulatory rules irrespective of the composition of the portfolio at hand. Second, the homogeneous group structure and the simple one-factor model (11.38) are typically oversimplified representations of the factor structure underlying default dependence, particularly for internationally active banks. Third, the rule is based on an asymptotic result. Moreover, historical default experience for the portfolio under consideration has no formal role to play in setting capital adequacy standards. These deficiencies should be weighed against the relative simplicity of the IRB approach, which makes it suitable for use in a supervisory setting. For economic capital purposes, on the other hand, most banks develop fully internal models with more sophisticated factor models to describe dependencies.

Notes and Comments

The results in Section 11.3 are an amalgamation of results from Frey and McNeil (2003) and Gordy (2003). The first limit result for large portfolios was obtained in Vasicek (1997) for a probit-normal mixture model equivalent to the KMV model. Asymptotic results for credit portfolios related to the theory of large deviations are discussed in Dembo, Deuschel and Duffie (2004). For details of the IRB approach, and the Basel II Capital Accord in general, we refer to the website of the Basel Committee: www.bis.org/bcbs. Our discussion in Section 11.3.3 is related to the analysis by Gordy (2003).

There have been a number of papers on second-order corrections or so-called *granularity adjustments* to the large-portfolio results in Propositions 11.17 and 11.18. While the results assume that idiosyncratic risk diversifies away in sufficiently large portfolios, these corrections take into account the fact that a certain amount of idiosyncratic risk and name concentration will remain in real portfolios. References include Martin and Wilde (2002), Gordy (2004), Gordy and Marrone (2012), Gordy and Lütkebohmert (2013) and Gagliardini and Gouriéroux (2013). See also the book by Lütkebohmert (2009) on concentration risk in credit portfolios.

11.4 Monte Carlo Methods

In this section we consider a Bernoulli mixture model for a loan portfolio and assume that the overall loss is of the form $L = \sum_{i=1}^{m} L_i$, where the L_i are conditionally independent given some economic factor vector $\boldsymbol{\Psi}$. A possible method for calculating risk measures and related quantities such as capital allocations is to use Monte Carlo (MC) simulation, although the problem of *rare-event simulation* arises. Suppose, for example, that we wish to compute expected shortfall and expected shortfall contributions at the confidence level α for our portfolio. We need to evaluate the conditional expectations

$$E(L \mid L \geqslant q_\alpha(L)) \quad \text{and} \quad E(L_i \mid L \geqslant q_\alpha(L)). \tag{11.39}$$

If $\alpha = 0.99$, say, then only 1% of our standard Monte Carlo draws will lead to a portfolio loss higher than $q_{0.99}(L)$. The standard MC estimator of (11.39), which consists of averaging the simulated values of L or L_i over all draws, leading to a simulated portfolio loss $L \geqslant q_\alpha(L)$, will be unstable and subject to high variability, unless the number of simulations is very large. The problem is of course that most simulations are "wasted", in that they lead to a value of L that is smaller than $q_\alpha(L)$. Fortunately, there exists a variance-reduction technique known as *importance sampling* (IS), which is well suited to such problems.

11.4.1 Basics of Importance Sampling

Consider an rv X on some probability space $(\Omega, \mathcal{F}, P)$ and assume that it has an absolutely continuous df with density f. A generalization to general probability spaces is discussed below. The problem we consider is the computation of the expected value

$$\theta = E(h(X)) = \int_{-\infty}^{\infty} h(x) f(x)\, \mathrm{d}x \tag{11.40}$$

for some known function h. To calculate the probability of an event we consider a function of the form $h(x) = I_{\{x \in A\}}$ for some set $A \subset \mathbb{R}$; for expected shortfall computation we consider functions of the form $h(x) = x I_{\{x \geqslant c\}}$ for some $c \in \mathbb{R}$. Where the analytical evaluation of (11.40) is difficult, due to the complexity of the distribution of X, we can resort to an MC approach, for which we only have to be able to simulate variates from the distribution with density f.

Algorithm 11.21 (Monte Carlo integration).

(1) Generate $X_1, \ldots, X_n$ independently from density f.

(2) Compute the standard MC estimate $\hat{\theta}_n^{\mathrm{MC}} = (1/n) \sum_{i=1}^{n} h(X_i)$.

The MC estimator converges to θ by the strong law of large numbers, but the speed of convergence may not be particularly fast, particularly when we are dealing with rare-event simulation.

Importance sampling is based on an alternative representation of the integral in (11.40). Consider a second probability density g (whose support should contain that of f) and define the *likelihood ratio* $r(x)$ by $r(x) := f(x)/g(x)$ whenever

$g(x) > 0$, and $r(x) = 0$ otherwise. The integral (11.40) may be written in terms of the likelihood ratio as

$$\theta = \int_{-\infty}^{\infty} h(x)r(x)g(x)\,\mathrm{d}x = E_g(h(X)r(X)), \tag{11.41}$$

where E_g denotes expectation with respect to the density g. We can therefore approximate the integral with the following algorithm.

Algorithm 11.22 (importance sampling).

(1) Generate $X_1, \ldots, X_n$ independently from density g.

(2) Compute the IS estimate $\hat{\theta}_n^{\mathrm{IS}} = (1/n)\sum_{i=1}^n h(X_i)\,r(X_i)$.

The density g is often termed the *importance-sampling density*. The art (or science) of importance sampling is in choosing an importance-sampling density such that, for fixed n, the variance of the IS estimator is considerably smaller than that of the standard Monte Carlo estimator. In this way we can hope to obtain a prescribed accuracy in evaluating the integral of interest using far fewer random draws than are required in standard Monte Carlo simulation. The variances of the estimators are given by

$$\begin{aligned}\mathrm{var}_g(\hat{\theta}_n^{\mathrm{IS}}) &= (1/n)(E_g(h(X)^2 r(X)^2) - \theta^2),\\ \mathrm{var}(\hat{\theta}_n^{\mathrm{MC}}) &= (1/n)(E(h(X)^2) - \theta^2),\end{aligned}$$

so the aim is to make $E_g(h(X)^2 r(X)^2)$ small compared with $E(h(X)^2)$. In theory, the variance of $\hat{\theta}^{\mathrm{IS}}$ can be reduced to zero by choosing an optimal g. To see this, suppose for the moment that h is non-negative and set

$$g^*(x) = f(x)h(x)/E(h(X)). \tag{11.42}$$

With this choice, the likelihood ratio becomes $r(x) = E(h(X))/h(x)$. Hence $\hat{\theta}_1^{\mathrm{IS}} = h(X_1)r(X_1) = E(h(X))$, and the IS estimator gives the correct answer in a single draw. In practice, it is of course impossible to choose an IS density of the form (11.42), as this requires knowledge of the quantity $E(h(X))$ that one wants to compute; nonetheless, (11.42) can provide useful guidance in choosing an IS density, as we will see in the next section.

Consider the case of estimating a rare-event probability corresponding to $h(x) = I_{\{x \geqslant c\}}$ for c significantly larger than the mean of X. Then we have that $E(h(X)^2) = P(X \geqslant c)$ and, using (11.41), that

$$E_g(h(X)^2 r(X)^2) = E_g(r(X)^2; X \geqslant c) = E(r(X); X \geqslant c). \tag{11.43}$$

Clearly, we should try to choose g such that the likelihood ratio $r(x) = f(x)/g(x)$ is small for $x \geqslant c$; in other words, we should make the event $\{X \geqslant c\}$ more likely under the IS density g than it is under the original density f.

Exponential tilting. We now describe a useful way of finding IS densities when X is light tailed. For $t \in \mathbb{R}$ we write $M_X(t) = E(\mathrm{e}^{tX}) = \int_{-\infty}^{\infty} \mathrm{e}^{tx} f(x)\,\mathrm{d}x$ for the moment-generating function of X, which we assume is finite for $t \in \mathbb{R}$. If $M_X(t)$ is finite, we can define an IS density by $g_t(x) := \mathrm{e}^{tx} f(x)/M_X(t)$. The likelihood ratio is $r_t(x) = f(x)/g_t(x) = M_X(t)\mathrm{e}^{-tx}$. Define μ_t to be the mean of X with respect to the density g_t, i.e.

$$\mu_t := E_{g_t}(X) = E(X\mathrm{e}^{tX})/M_X(t). \tag{11.44}$$

How can we choose t optimally for a particular IS problem? We consider the case of tail probability estimation and recall from (11.43) that the objective is to make

$$E(r(X); X \geqslant c) = E(I_{\{X \geqslant c\}} M_X(t)\mathrm{e}^{-tX}) \tag{11.45}$$

small. Now observe that $\mathrm{e}^{-tx} \leqslant \mathrm{e}^{-tc}$ for $x \geqslant c$ and $t \geqslant 0$, so

$$E(I_{\{X \geqslant c\}} M_X(t)\mathrm{e}^{-tX}) \leqslant M_X(t)\mathrm{e}^{-tc}.$$

Instead of solving the (difficult) problem of minimizing (11.45) over t, we choose t such that this bound becomes minimal. Equivalently, we try to find t minimizing $\ln M_X(t) - tc$. Using (11.44) we obtain that

$$\frac{\mathrm{d}}{\mathrm{d}t} \ln M_X(t) - tc = \frac{E(X\mathrm{e}^{tX})}{M_X(t)} - c = \mu_t - c,$$

which suggests choosing $t = t(c)$ as the solution of the equation $\mu_t = c$, so that the rare event $\{X \geqslant c\}$ becomes a normal event if we compute probabilities using the density $g_{t(c)}$. A unique solution of the equation $\mu_t = c$ exists for all relevant values of c. In the cases that are of interest to us this is immediately obvious from the form of the exponentially tilted distributions, so we omit a formal proof.

Example 11.23 (exponential tilting for normal distribution). We illustrate the concept of exponential tilting in the simple case of a standard normal rv. Suppose that $X \sim N(0, 1)$ with density $\phi(x)$. Using exponential tilting we obtain the new density $g_t(x) = \mathrm{e}^{tx}\phi(x)/M_X(t)$. The moment-generating function of X is known to be $M_X(t) = \mathrm{e}^{t^2/2}$. Hence

$$g_t(x) = \frac{1}{\sqrt{2\pi}} \exp(tx - \tfrac{1}{2}(t^2 + x^2)) = \frac{1}{\sqrt{2\pi}} \exp(-\tfrac{1}{2}(x - t)^2),$$

so that, under the tilted distribution, $X \sim N(t, 1)$. Note that in this case exponential tilting corresponds to changing the mean of X.

An abstract view of importance sampling. To handle the more complex application to portfolio credit risk in the next section it helps to consider importance sampling from a slightly more general viewpoint. Given densities f and g as above, define probability measures P and Q by

$$P(A) = \int_A f(x)\,\mathrm{d}x \quad \text{and} \quad Q(A) = \int_A g(x)\,\mathrm{d}x, \quad A \subset \mathbb{R}.$$

With this notation, (11.41) becomes $\theta = E^P(h(X)) = E^Q(h(X)r(X))$, so that $r(X)$ equals $\mathrm{d}P/\mathrm{d}Q$, the (measure-theoretic) density of P with respect to Q. Using this more abstract view, exponential tilting can be applied in more general situations: given an rv X on $(\Omega, \mathcal{F}, P)$ such that $M_X(t) = E^P(\mathrm{e}^{tX}) < \infty$, define the measure Q_t on $(\Omega, \mathcal{F})$ by

$$\frac{\mathrm{d}Q_t}{\mathrm{d}P} = \frac{\mathrm{e}^{tX}}{M_X(t)}, \quad \text{i.e. } Q_t(A) = E^P\left(\frac{\mathrm{e}^{tX}}{M_X(t)}; A\right),$$

and note that $(\mathrm{d}Q_t/\mathrm{d}P)^{-1} = M_X(t)\mathrm{e}^{-tX} = r_t(X)$. The IS algorithm remains essentially unchanged: simulate independent realizations X_i under the measure Q_t and set $\hat{\theta}^{\mathrm{IS}} = (1/n)\sum_{i=1}^n X_i r_t(X_i)$ as before.

11.4.2 Application to Bernoulli Mixture Models

In this section we return to the subject of credit losses and consider a portfolio loss of the form $L = \sum_{i=1}^m e_i Y_i$, where the e_i are deterministic, positive exposures and the Y_i are default indicators with default probabilities p_i. We assume that $\boldsymbol{Y}$ follows a Bernoulli mixture model in the sense of Definition 11.5 with factor vector $\boldsymbol{\Psi}$ and conditional default probabilities $p_i(\boldsymbol{\Psi})$. We study the problem of estimating exceedance probabilities $\theta = P(L \geqslant c)$ for c substantially larger than $E(L)$ using importance sampling. This is useful for risk-management purposes, as, for $c \approx q_\alpha(L)$, a good IS distribution for the computation of $P(L \geqslant c)$ also yields a substantial variance reduction for computing expected shortfall or expected shortfall contributions.

We consider first the situation where the default indicators $Y_1, \ldots, Y_m$ are independent, and then we discuss the extension to the case of conditionally independent default indicators. Our exposition is based on Glasserman and Li (2005).

Independent default indicators. Here we use the more general IS approach outlined at the end of the previous section. Set $\Omega = \{0, 1\}^m$, the state space of $\boldsymbol{Y}$. The probability measure P is given by

$$P(\{\boldsymbol{y}\}) = \prod_{i=1}^m p_i^{y_i}(1 - p_i)^{1-y_i}, \quad \boldsymbol{y} \in \{0, 1\}^m.$$

We need to understand how this measure changes under exponential tilting using L. The moment-generating function of L is easily calculated to be

$$M_L(t) = E\left(\exp\left(t\sum_{i=1}^m e_i Y_i\right)\right) = \prod_{i=1}^m E(\mathrm{e}^{te_i Y_i}) = \prod_{i=1}^m (\mathrm{e}^{te_i} p_i + 1 - p_i).$$

The measure Q_t is given by $Q_t(\{\boldsymbol{y}\}) = E^P(\mathrm{e}^{tL}/M_L(t); \boldsymbol{Y} = \boldsymbol{y})$ and hence

$$Q_t(\{\boldsymbol{y}\}) = \frac{\exp(t\sum_{i=1}^m e_i y_i)}{M_L(t)} P(\{\boldsymbol{y}\}) = \prod_{i=1}^m \frac{\mathrm{e}^{te_i y_i}}{\mathrm{e}^{te_i} p_i + 1 - p_i} p_i^{y_i}(1 - p_i)^{1-y_i}.$$

Define new default probabilities by $\bar{q}_{t,i} := \mathrm{e}^{te_i} p_i/(\mathrm{e}^{te_i} p_i + 1 - p_i)$. It follows that $Q_t(\{\boldsymbol{y}\}) = \prod_{i=1}^m \bar{q}_{t,i}^{y_i}(1 - \bar{q}_{t,i})^{1-y_i}$, so that after exponential tilting the default indicators remain independent but with new default probability $\bar{q}_{t,i}$. Note that $\bar{q}_{t,i}$ tends

to 1 for $t \to \infty$ and to 0 for $t \to -\infty$, so that we can shift the mean of L to any point in $(0, \sum_{i=1}^m e_i)$.

In analogy with our previous discussion, for IS purposes, the optimal value of t is chosen such that $E^{Q_t}(L) = c$, leading to the equation $\sum_{i=1}^m e_i \bar{q}_{t,i} = c$.

Conditionally independent default indicators. The first step in the extension of the importance-sampling approach to conditionally independent defaults is obvious: given a realization $\boldsymbol{\psi}$ of the economic factors, the conditional exceedance probability $\theta(\boldsymbol{\psi}) := P(L \geqslant c \mid \boldsymbol{\Psi} = \boldsymbol{\psi})$ is estimated using the approach for independent default indicators described above. We have the following algorithm.

Algorithm 11.24 (IS for conditional loss distribution).

(1) Given $\boldsymbol{\psi}$, calculate the conditional default probabilities $p_i(\boldsymbol{\psi})$ according to the particular model, and solve the equation

$$\sum_{i=1}^m e_i \frac{\mathrm{e}^{te_i} p_i(\boldsymbol{\psi})}{\mathrm{e}^{te_i} p_i(\boldsymbol{\psi}) + 1 - p_i(\boldsymbol{\psi})} = c;$$

the solution $t = t(c, \boldsymbol{\psi})$ gives the optimal degree of tilting.

(2) Generate n_1 conditional realizations of the default vector $\boldsymbol{Y} = (Y_1, \ldots, Y_m)'$. The defaults of the companies are simulated independently, with the default probability of the ith company given by

$$\frac{\exp(t(c, \boldsymbol{\psi})e_i)\, p_i(\boldsymbol{\psi})}{\exp(t(c, \boldsymbol{\psi})e_i)\, p_i(\boldsymbol{\psi}) + 1 - p_i(\boldsymbol{\psi})}.$$

(3) Denote by $M_L(t, \boldsymbol{\psi}) := \prod_{i=1}^m \{\mathrm{e}^{te_i} p_i(\boldsymbol{\psi}) + 1 - p_i(\boldsymbol{\psi})\}$ the conditional moment-generating function of L. From the simulated default data construct n_1 conditional realizations of $L = \sum_{i=1}^m e_i Y_i$ and label these $L^{(1)}, \ldots, L^{(n_1)}$. Determine the IS estimator for the conditional loss distribution:

$$\hat{\theta}_{n_1}^{\mathrm{IS},1}(\boldsymbol{\psi}) = M_L(t(c, \boldsymbol{\psi}), \boldsymbol{\psi}) \frac{1}{n_1} \sum_{j=1}^{n_1} I_{\{L^{(j)} \geqslant c\}} \exp(-t(c, \boldsymbol{\psi}) L^{(j)}).$$

In principle, the approach discussed above also applies in the more general situation where the loss given default is random; all we need to assume is that the L_i are conditionally independent given $\boldsymbol{\Psi}$, as in Assumption (A1) of Section 11.3. However, the actual implementation can become quite involved.

IS for the distribution of the factor variables. Suppose we now want to estimate the unconditional probability $\theta = P(L \geqslant c)$. A naive approach would be to generate realizations of the factor vector $\boldsymbol{\Psi}$ and to estimate θ by averaging the IS estimator of Algorithm 11.24 over these realizations. As is shown in Glasserman and Li (2005), this is not the best solution for large portfolios of dependent credit risks. Intuitively, this is due to the fact that for such portfolios most of the variation in L is caused by fluctuations of the economic factors, and we have not yet applied IS to the distribution

of $\boldsymbol{\Psi}$. For this reason we now discuss a full IS algorithm that combines IS for the economic factor variables with Algorithm 11.24.

We consider the important case of a Bernoulli mixture model with multivariate Gaussian factors and conditional default probabilities $p_i(\boldsymbol{\Psi})$ for $\boldsymbol{\Psi} \sim N_p(\mathbf{0}, \Omega)$, such as the probit-normal Bernoulli mixture model described in Example 11.11. In this context it is natural to choose an importance-sampling density such that $\boldsymbol{\Psi} \sim N_p(\boldsymbol{\mu}, \Omega)$ for a new mean vector $\boldsymbol{\mu} \in \mathbb{R}^p$, i.e. we take g as the density of $N_p(\boldsymbol{\mu}, \Omega)$. For a good choice of $\boldsymbol{\mu}$ we expect to generate realizations of $\boldsymbol{\Psi}$ leading to high conditional default probabilities more frequently. The corresponding likelihood ratio $r_{\boldsymbol{\mu}}(\boldsymbol{\Psi})$ is given by the ratio of the respective multivariate normal densities, so that

$$r_{\boldsymbol{\mu}}(\boldsymbol{\Psi}) = \frac{\exp(-\frac{1}{2}\boldsymbol{\Psi}'\Omega^{-1}\boldsymbol{\Psi})}{\exp(-\frac{1}{2}(\boldsymbol{\Psi}-\boldsymbol{\mu})'\Omega^{-1}(\boldsymbol{\Psi}-\boldsymbol{\mu}))} = \exp(-\boldsymbol{\mu}'\Omega^{-1}\boldsymbol{\Psi} + \tfrac{1}{2}\boldsymbol{\mu}'\Omega^{-1}\boldsymbol{\mu}).$$

Essentially, this is a multivariate analogue of the exponential tilting applied to a univariate normal distribution in Example 11.23.

Now we can describe the algorithm for full IS. At the outset we have to choose the overall number of simulation rounds, n, the number of repetitions of conditional IS per simulation round, n_1, and the mean of the IS distribution for the factors, $\boldsymbol{\mu}$. Whereas the value of n depends on the desired degree of precision and is best determined in a simulation study, n_1 should be taken to be fairly small. An approach to determine a sensible value of $\boldsymbol{\mu}$ is discussed below.

Algorithm 11.25 (full IS for mixture models with Gaussian factors).

(1) Generate $\boldsymbol{\Psi}_1, \ldots, \boldsymbol{\Psi}_n \sim N(\boldsymbol{\mu}, I_p)$.

(2) For each $\boldsymbol{\Psi}_i$ calculate $\hat{\theta}_{n_1}^{\mathrm{IS},1}(\boldsymbol{\Psi}_i)$ as in Algorithm 11.24.

(3) Determine the full IS estimator:

$$\hat{\theta}_n^{\mathrm{IS}} = \frac{1}{n}\sum_{i=1}^{n} r_{\boldsymbol{\mu}}(\boldsymbol{\Psi}_i)\hat{\theta}_{n_1}^{\mathrm{IS},1}(\boldsymbol{\Psi}_i).$$

Choosing $\boldsymbol{\mu}$. A key point in the full IS approach is the determination of a value for $\boldsymbol{\mu}$ that gives a low variance for the IS estimator. Here we sketch the solution proposed by Glasserman and Li (2005). Since $\hat{\theta}_{n_1}^{\mathrm{IS},1}(\boldsymbol{\psi}) \approx P(L \geqslant c \mid \boldsymbol{\Psi} = \boldsymbol{\psi})$, applying IS to the factors essentially amounts to finding a good IS density for the function $\boldsymbol{\psi} \to P(L \geqslant c \mid \boldsymbol{\Psi} = \boldsymbol{\psi})$. Now recall from our discussion in the previous section that the optimal IS density g^* satisfies

$$g^*(\boldsymbol{\psi}) \propto P(L \geqslant c \mid \boldsymbol{\Psi} = \boldsymbol{\psi}) \exp(-\tfrac{1}{2}\boldsymbol{\psi}'\Omega^{-1}\boldsymbol{\psi}), \tag{11.46}$$

where "$\propto$" stands for "proportional to". Sampling from that density is obviously not feasible, as the normalizing constant involves the exceedance probability $P(L \geqslant c)$ that we are interested in. In this situation the authors suggest using a multivariate normal density with the same mode as g^* as an approximation to the optimal IS

density. Since a normal density attains its mode at the mean $\boldsymbol{\mu}$, this amounts to choosing $\boldsymbol{\mu}$ as the solution to the optimization problem

$$\max_{\boldsymbol{\psi}} P(L \geqslant c \mid \boldsymbol{\Psi} = \boldsymbol{\psi}) \exp(-\tfrac{1}{2}\boldsymbol{\psi}' \Omega^{-1} \boldsymbol{\psi}). \tag{11.47}$$

An exact (numerical) solution of (11.47) is difficult because the function $P(L \geqslant c \mid \boldsymbol{\Psi} = \boldsymbol{\psi})$ is usually not available in closed form. Glasserman and Li (2005) discuss several approaches to overcoming this difficulty (see their paper for details).

Example 11.26. We give a very simple example of IS to show the gains that can be obtained by applying IS at both the level of the factor variables and the level of the conditional loss distribution.

Consider an exchangeable one-factor Bernoulli mixture model in which the factor Ψ is standard normally distributed and the conditional probability of default is given by

$$p_i(\psi) = P(Y_i = 1 \mid \Psi = \psi) = \Phi\left(\frac{\Phi^{-1}(p) + \sqrt{\rho}\psi}{\sqrt{1-\rho}}\right)$$

for all obligors i. Let the unconditional default probability be $p = 0.05$, let the asset correlation $\rho = 0.05$ and consider $m = 100$ obligors, each with an identical exposure $e_i = 1$. We are interested in the probability $\theta = P(L \geqslant 20)$, where $L = \sum_{i=1}^{m} Y_i$. In this set-up we can calculate, using numerical integration, that $\theta \approx 0.00112$, so $\{L \geqslant 20\}$ is a relatively rare event. In the first panel of Figure 11.3 we apply naive Monte Carlo estimation of θ and plot $\hat{\theta}_n^{\mathrm{MC}}$ against n; the true value is shown by a horizontal line.

In the second panel we apply importance sampling at the level of the factor Ψ using the value $\mu = -2.8$ for the mean of the distribution of Ψ under Q and plot the resulting estimate for different values of n, the number of random draws of the factor. In the third panel we apply IS at the level of the conditional loss distribution, using Algorithm 11.24 with $n_1 = 50$, but we apply naive Monte Carlo to the distribution of the factor.

In the final panel we apply full IS using Algorithm 11.25, and we plot $\hat{\theta}_n^{\mathrm{IS}}$ against n using $n_1 = 50$ as before. This is clearly the only estimate that appears to have converged to the true value by the time we have sampled $n = 10\,000$ values of the factor.

Notes and Comments

Our discussion of IS for credit portfolios follows Glasserman and Li (2005) closely. Theoretical results on the asymptotics of the IS estimator for large portfolios and numerical case studies contained in Glasserman and Li (2005) indicate that full IS is a very useful tool for dealing with large Bernoulli mixture models. Merino and Nyfeler (2004) and Kalkbrener, Lotter and Overbeck (2004) undertook related work—the latter paper gives an interesting alternative solution to finding a reasonable IS mean $\boldsymbol{\mu}$ for the factors.

For a general introduction to importance sampling we refer to the excellent textbook by Glasserman (2003) (see also Asmussen and Glynn 2007; Robert and Casella

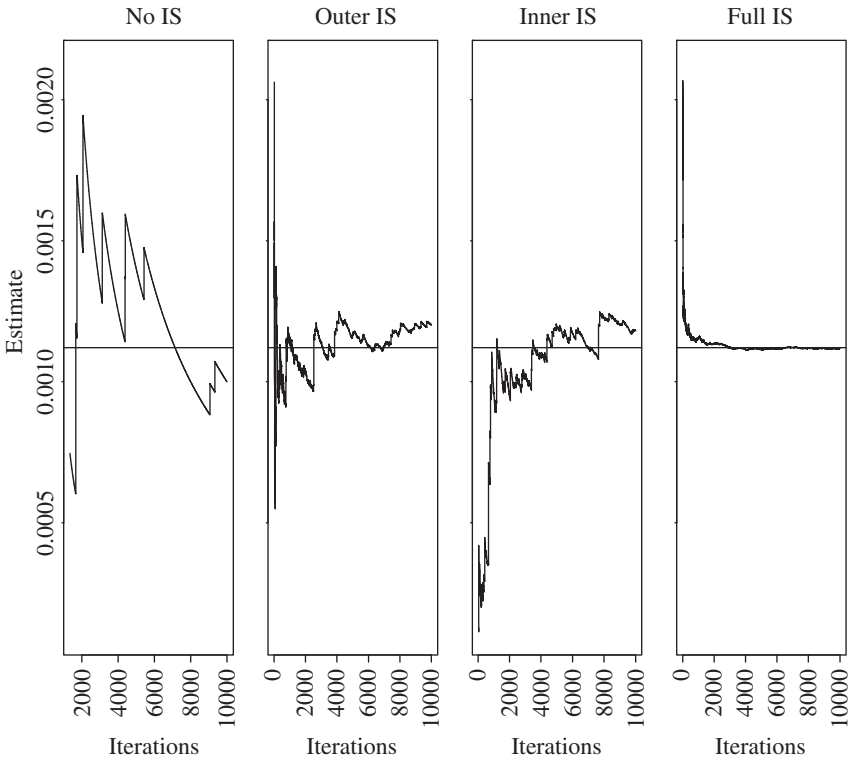

Figure 11.3. Illustration of the significant improvements that can be made in the estimation of rare-event probabilities for Bernoulli mixture models when importance sampling is applied at the level of both the factors and the conditional loss distribution given the factors (see Example 11.26 for details).

1999). For applications of importance sampling to heavy-tailed distributions, where exponential families cannot be applied directly, see Asmussen, Binswanger and Højgaard (2000) and Glasserman, Heidelberger and Shahabuddin (1999).

An alternative to simulation is the use of analytic approximations to the portfolio loss distribution. Applications of the saddle-point approximation (see Jensen 1995) are discussed in Martin, Thompson and Browne (2001) and Gordy (2002).

11.5 Statistical Inference in Portfolio Credit Models

In the remainder of this chapter we consider two different approaches to the estimation of portfolio credit risk models. In Section 11.5.1 we discuss the calibration of industry threshold models, such as the CreditMetrics model and the portfolio version of the Moody's public-firm EDF model; we focus on the estimation of the factor model describing the dependence structure of the critical variables using proxy data on equity or asset returns.

In Sections 11.5.2–11.5.4 we discuss the direct estimation of Bernoulli or Poisson mixture models from historical default data. This approach has been less widely applied in industry due to the relative scarcity of data on defaults, particularly for

higher-rated firms. However, it has become increasingly feasible with the availability of larger databases of historical defaults and rating migrations.

11.5.1 Factor Modelling in Industry Threshold Models

Recall from Section 11.1.3 that many industry models take the form of a Gaussian threshold model $(\boldsymbol{X}, \boldsymbol{d})$ with $\boldsymbol{X} \sim N_m(\mathbf{0}, P)$, where the random vector $\boldsymbol{X}$ contains the critical variables representing the credit quality or "ability-to-pay" of the obligors in the portfolio, the deterministic vector $\boldsymbol{d}$ contains the critical default thresholds, and P is the so-called asset correlation matrix, which is estimated with the help of a factor model for $\boldsymbol{X}$.

Industry models generally separate the calibration of the vector $\boldsymbol{d}$ (or the threshold matrix D in a multi-state model) and the calibration of the factor model for $\boldsymbol{X}$. As discussed in Section 11.1.3, in a default-only model the threshold d_i is usually set at $d_i = \Phi^{-1}(p_i)$, where p_i is an estimate of the default probability for obligor i for the time period in question (generally one year). Depending on the type of obligor, the default probability may be estimated in different ways: for larger corporates it may be estimated using credit ratings or using a firm-value approach, such as the Moody's public-firm EDF model; for retail obligors it may be estimated on the basis of credit scores.

We concentrate in this section on the estimation of the factor model for $\boldsymbol{X}$. We recall that this takes the form

$$X_i = \sqrt{\beta_i}\tilde{F}_i + \sqrt{1-\beta_i}\varepsilon_i, \quad i = 1, \ldots, m, \tag{11.48}$$

where $\tilde{F}_i$ and $\varepsilon_1, \ldots, \varepsilon_m$ are independent standard normal variables, and where $0 \leqslant \beta_i \leqslant 1$ for all i. The systematic variables $\tilde{F}_i$ are assumed to be of the form $\tilde{F}_i = \boldsymbol{a}_i'\boldsymbol{F}$, where $\boldsymbol{F}$ is a vector of common factors satisfying $\boldsymbol{F} \sim N_p(\mathbf{0}, \Omega)$ with $p < m$, and where Ω is a correlation matrix. The factors typically represent country and industry effects and the assumption that $\operatorname{var}(\tilde{F}_i) = 1$ imposes the constraint that $\boldsymbol{a}_i'\Omega\boldsymbol{a}_i = 1$ for all i.

Different industry models use different data for $\boldsymbol{X}$ to calibrate the factor model (11.48). The Moody's Analytics Global Correlation, or GCorr, model has sub-models for many different kinds of obligor, including public corporate firms, private firms, small and medium-sized enterprises (SMEs), retail customers and sovereigns (Huang et al. 2012). The sub-model for public firms (GCorr Corporate) is calibrated using data on weekly asset value returns, where asset values are determined as part of the public-firm EDF methodology described in Section 10.3.3. In the CreditMetrics framework, weekly equity returns are viewed as a proxy for asset returns and used to estimate the factor model (RiskMetrics Group 1997). In both cases the data contain information about changing credit quality.

We now provide a sketch of a generic procedure for estimating a factor model for corporates where the factors have country- and industry-sector interpretations. Specific industry models follow this procedure in outline but may differ in the details of the calculations at certain steps. We assume that we have a high-dimensional multivariate time series $(\boldsymbol{X}_t)_{1 \leqslant t \leqslant n}$ of asset returns (or other proxy data for changing

credit quality) over a period of time in which stationarity can be assumed; we also assume that each component time series has been scaled to have mean 0 and variance 1.

(1) We first fix the structure of the factor vector $\boldsymbol{F}$. For example, the first block of components might represent country factors and the second block of components might represent industry factors. We then assign vectors of factor weights $\boldsymbol{a}_i$ to each obligor based on our knowledge of the companies. The elements of $\boldsymbol{a}_i$ may simply consist of ones and zeros if the company can be clearly identified with a single country and industry, but may also consist of weights if the company has significant activity in more than one country or more than one industry sector. For example, a firm that does 60% of its business in one country and 40% in another would be coded with weights of 0.6 and 0.4 in the relevant positions of $\boldsymbol{a}_i$.

(2) We then use cross-sectional estimation techniques to estimate the factor values $\boldsymbol{F}_t$ at each time point t. Effectively, the factor estimates $\hat{\boldsymbol{F}}_t$ are constructed as weighted sums of the $X_{t,i}$ data for obligors i that are exposed to each factor. One way of achieving this is to construct a matrix A with rows $\boldsymbol{a}_i$ and then to estimate a fundamental factor model of the form $\boldsymbol{X}_t = A\boldsymbol{F}_t + \boldsymbol{\varepsilon}_t$ at each time point t, as described in Section 6.4.4.

(3) The raw factor estimates form a multivariate time series of dimension p. We standardize each component series to have mean 0 and variance 1 to obtain $(\hat{\boldsymbol{F}}_t)_{1\leqslant t\leqslant n}$ and calculate the sample covariance matrix of the standardized factor estimates, which serves as our estimate of Ω.

(4) We then scale the vectors of factor weights $\boldsymbol{a}_i$ such that the conditions $\boldsymbol{a}_i'\hat{\Omega}\boldsymbol{a}_i = 1$ are met for each obligor.

(5) Time series of estimated systematic variables for each obligor are then constructed by calculating $\tilde{F}_{t,i} = \boldsymbol{a}_i'\hat{\boldsymbol{F}}_t$ for $t = 1, \ldots, n$.

(6) Finally, we estimate the β_i parameters by performing a time-series regression of $X_{t,i}$ on $\tilde{\tilde{F}}_{t,i}$ for each obligor.

Note that the accurate estimation of the β_i in the last step is particularly important. In Section 11.1.5 we showed that there is considerable model risk associated with the size of the specific risk component, particularly when the tail of a credit loss distribution is of central importance. The estimate of β_i is the so-called R-squared of the time-series regression model in step (6) and will be largest for the firms whose credit-quality changes are best explained by systematic factors.

11.5.2 Estimation of Bernoulli Mixture Models

We now turn our attention to the estimation of Bernoulli mixture models of portfolio credit risk from historical default data. The models we describe are motivated by the format of the data we consider, which can be described as *repeated cross-sectional data*. This kind of data, comprising observations of the default or non-default of groups of monitored companies in a number of time periods, can be readily extracted

from the rating-migration and default databases of rating agencies. Since the group of companies may differ from period to period, as new companies are rated and others default or cease to be rated, we have a cross-section of companies in each period, but the cross-section may change from period to period.

In this section we discuss the estimation of default probabilities and default correlations for homogeneous groups, e.g. groups with the same credit rating. In Sections 11.5.3 and 11.5.4 we consider more complicated one-factor models allowing more heterogeneity and make a link to the important class of generalized linear mixed models (GLMMs) used in many statistical applications.

Suppose that we observe historical default numbers over n periods of time for a homogeneous group; typically these might be yearly data. For $t = 1, \ldots, n$, let m_t denote the number of observed companies at the start of period t and let M_t denote the number that defaulted during the period; the former will be treated as fixed at the outset of the period and the latter as an rv. Suppose further that within a time period these defaults are generated by an exchangeable Bernoulli mixture model of the kind described in Section 11.2.2. In other words, assume that, given some mixing variable Q_t taking values in $(0, 1)$ and the cohort size m_t, the number of defaults M_t is conditionally binomially distributed and satisfies $M_t \mid Q_t = q \sim B(m_t, q)$. Further assume that the mixing variables $Q_1, \ldots, Q_n$ are identically distributed. We consider two methods for estimating the fundamental parameters of the mixing distribution $\pi = \pi_1, \pi_2$ and ρ_Y (default correlation); these are the method of moments and the maximum likelihood method.

A simple moment estimator. For $1 \leqslant t \leqslant n$, let $Y_{t,1}, \ldots, Y_{t,m_t}$ be default indicators for the m_t companies in the cohort. Suppose we define the rv

$$\binom{M_t}{k} := \sum_{\{i_1,\ldots,i_k\}\subset\{1,\ldots,m_t\}} Y_{t,i_1} \cdots Y_{t,i_k}; \tag{11.49}$$

this represents the number of possible subgroups of k obligors among the defaulting obligors in period t (and takes the value zero when $k > M_t$). By taking expectations in (11.49) we get

$$E\left(\binom{M_t}{k}\right) = \binom{m_t}{k}\pi_k$$

and hence

$$\pi_k = E\left(\binom{M_t}{k}\right) \Big/ \binom{m_t}{k}.$$

We estimate the unknown theoretical moment π_k by taking a natural empirical average (11.50) constructed from the n years of data:

$$\hat{\pi}_k = \frac{1}{n}\sum_{t=1}^{n} \frac{\binom{M_t}{k}}{\binom{m_t}{k}} = \frac{1}{n}\sum_{t=1}^{n} \frac{M_t(M_t-1)\cdots(M_t-k+1)}{m_t(m_t-1)\cdots(m_t-k+1)}. \tag{11.50}$$

For $k = 1$ we get the standard estimator of default probability

$$\hat{\pi} = \frac{1}{n}\sum_{t=1}^{n} \frac{M_t}{m_t},$$

and ρ_Y can obviously be estimated by taking $\hat{\rho}_Y = (\hat{\pi}_2 - \hat{\pi}^2)/(\hat{\pi} - \hat{\pi}^2)$. The estimator is unbiased for π_k and consistent as $n \to \infty$ (for more details see Frey and McNeil (2001)). Note that, for Q_t random, consistency requires observations for a large number of years; it is not sufficient to observe a large pool in a single year.

Maximum likelihood estimators. To implement a maximum likelihood (ML) procedure we assume a simple parametric form for the density of the Q_t (such as beta, logit-normal or probit-normal). The joint probability function of the default counts $M_1, \ldots, M_n$ given the cohort sizes $m_1, \ldots, m_n$ can then be calculated using (11.14), under the assumption that the Q_t variables in different years are independent. This expression is then maximized with respect to the natural parameters of the mixing distribution (i.e. a and b in the case of beta and μ and σ for the logit-normal and probit-normal). Of course, independence may be an unrealistic assumption for the mixing variables, due to the phenomenon of economic cycles, but the method could then be regarded as a quasi-maximum likelihood (QML) procedure, which misspecifies the serial dependence structure but correctly specifies the marginal distribution of defaults in each year and still gives reasonable parameter estimates.

In practice, it is easiest to use the beta mixing distribution, since, in this case, given the group size m_t in period t, the rv M_t has a beta-binomial distribution with probability function given in (11.16). The likelihood to be maximized therefore takes the form

$$L(a, b; \text{data}) = \prod_{t=1}^{n} \binom{m_t}{M_t} \frac{\beta(a + M_t, b + m_t - M_t)}{\beta(a, b)},$$

and maximization can be performed numerically with respect to a and b. For further information about the ML method consult Section A.3. The ML estimates of $\pi = \pi_1, \pi_2$ and ρ_Y are calculated by evaluating moments of the fitted distribution using (11.15); the formulas are given in Example 11.7.

A comparison of moment estimation and ML estimation. To compare these two approaches we conduct a simulation study summarized in Table 11.4. To generate data in the simulation study we consider the beta, probit-normal and logit-normal mixture models of Section 11.2.2. In any single experiment we generate 20 years of data using parameter values that roughly correspond to one of the Standard & Poor's credit ratings CCC, B or BB (see Table 11.3 for the parameter values). The number of firms m_t in each of the years is generated randomly using a binomial-beta model to give a spread of values typical of real data; the defaults are then generated using one of the Bernoulli mixture models, and estimates of π, π_2 and ρ_Y are calculated. The experiment is repeated 5000 times and a relative root mean square error (RRMSE) is estimated for each parameter and each method: that is, we take the square root of the estimated MSE and divide by the true parameter value. Methods are compared by calculating the percentage increase of the estimated RRMSE with respect to the better method (i.e. the RRMSE-minimizing method) for each parameter.

It may be concluded from Table 11.4 that the ML method is better in all but one experiment. Surprisingly, it is better even in the experiments when it is misspecified

Table 11.4. Each part of the table relates to a block of 5000 simulations using a particular exchangeable Bernoulli mixture model with parameter values roughly corresponding to a particular S&P rating class. For each parameter of interest, an estimated RRMSE is tabulated for both estimation methods: moment estimation using (11.50) and ML estimation based on the beta model. Methods can be compared by using Δ, the percentage increase of the estimated RRMSE with respect to the better method (i.e. the RRMSE-minimizing method) for each parameter. For each parameter the better method therefore has $\Delta = 0$. The table clearly shows that MLE is at least as good as the moment estimator in all but one case.

			Moment		MLE-beta	
Group	True model	Parameter	RRMSE	Δ	RRMSE	Δ
CCC	Beta	π	0.101	0	0.101	0
CCC	Beta	π_2	0.202	0	0.201	0
CCC	Beta	ρ_Y	0.332	5	0.317	0
CCC	Probit-normal	π	0.100	0	0.100	0
CCC	Probit-normal	π_2	0.205	1	0.204	0
CCC	Probit-normal	ρ_Y	0.347	11	0.314	0
CCC	Logit-normal	π	0.101	0	0.101	0
CCC	Logit-normal	π_2	0.209	1	0.208	0
CCC	Logit-normal	ρ_Y	0.357	11	0.320	0
B	Beta	π	0.130	0	0.130	0
B	Beta	π_2	0.270	0	0.269	0
B	Beta	ρ_Y	0.396	8	0.367	0
B	Probit-normal	π	0.130	0	0.130	0
B	Probit-normal	π_2	0.286	3	0.277	0
B	Probit-normal	ρ_Y	0.434	19	0.364	0
B	Logit-normal	π	0.131	0	0.132	0
B	Logit-normal	π_2	0.308	7	0.289	0
B	Logit-normal	ρ_Y	0.493	26	0.392	0
BB	Beta	π	0.199	0	0.199	0
BB	Beta	π_2	0.435	0	0.438	1
BB	Beta	ρ_Y	0.508	7	0.476	0
BB	Probit-normal	π	0.197	0	0.197	0
BB	Probit-normal	π_2	0.492	10	0.446	0
BB	Probit-normal	ρ_Y	0.607	27	0.480	0
BB	Logit-normal	π	0.196	0	0.196	0
BB	Logit-normal	π_2	0.572	24	0.462	0
BB	Logit-normal	ρ_Y	0.752	45	0.517	0

and the true mixing distribution is either probit-normal or logit-normal; in fact, in these cases, it offers more of an improvement than in the beta case. This can partly be explained by the fact that when we constrain well-behaved, unimodal mixing distributions with densities to have the same first and second moments, these distributions are very similar (see Figure 11.2). Finally, we observe that the ML method tends to outperform the moment method more as we increase the credit quality, so that defaults become rarer.

11.5.3 Mixture Models as GLMMs

A one-factor Bernoulli mixture model. Recall the simple one-factor model (11.17), which generalizes the exchangeable model in Section 11.2.2, and consider the case where $\sigma_i = \sigma$ for all obligors i. Rewriting slightly, this model has the form

$$p_i(\Psi) = h(\mu + \boldsymbol{\beta}' \boldsymbol{x}_i + \Psi), \tag{11.51}$$

where h is a link function, the vector $\boldsymbol{x}_i$ contains covariates for the ith firm, such as indicators for group membership or key balance sheet ratios, and $\boldsymbol{\beta}$ and μ are model parameters. Examples of link functions include the standard normal df $\Phi(x)$ and the logistic df $(1 + \mathrm{e}^{-x})^{-1}$. The scale parameter σ has been subsumed in the normally distributed random variable $\Psi \sim N(0, \sigma^2)$, representing a common or systematic factor.

This model can be turned into a multi-period model for default counts in different periods by assuming that a series of mixing variables $\Psi_1, \ldots, \Psi_n$ generates default dependence in each time period $t = 1, \ldots, n$. The default indicator $Y_{t,i}$ for the ith company in time period t is assumed to be Bernoulli with default probability $p_{t,i}(\Psi_t)$ depending on Ψ_t according to

$$p_{t,i}(\Psi_t) = h(\mu + \boldsymbol{x}_{t,i}' \boldsymbol{\beta} + \Psi_t), \tag{11.52}$$

where $\Psi_t \sim N(0, \sigma^2)$ and $\boldsymbol{x}_{t,i}$ are covariates for the ith company in time period t. Moreover, the default indicators $Y_{t,1}, \ldots, Y_{t,m_t}$ in period t are assumed to be conditionally independent given Ψ_t.

To complete the model we need to specify the joint distribution of $\Psi_1, \ldots, \Psi_n$, and it is easiest to assume that these are iid mixing variables. To capture possible economic cycle effects causing dependence between numbers of defaults in successive time periods, one could either enter covariates at the level of $\boldsymbol{x}_{ti}$ that are known to be good proxies for "the state of the economy", such as changes in GDP over the time period, or an index like the Chicago Fed National Activity Index (CFNAI) in the US, or one could consider a serially dependent time-series structure for the systematic factors (Ψ_t).

A one-factor Poisson mixture model. When considering higher-grade portfolios of companies with relatively low default risk, there may sometimes be advantages (particularly in the stability of fitting procedures) in formulating Poisson mixture models instead of Bernoulli mixture models. A multi-period mixture model based on Definition 11.14 can be constructed by assuming that the default count variable $\tilde{Y}_{t,i}$ for the ith company in time period t is conditionally Poisson with rate parameter $\lambda_{t,i}(\Psi_t)$ depending on Ψ_t according to

$$\lambda_{t,i}(\Psi_t) = \exp(\mu + \boldsymbol{x}_{t,i}' \boldsymbol{\beta} + \Psi_t), \tag{11.53}$$

with all other elements of the model as in (11.52). Again the variables $\tilde{Y}_{t,1}, \ldots, \tilde{Y}_{t,m_t}$ are assumed to be conditionally independent given Ψ_t.

GLMMs. Both the multi-period Bernoulli and Poisson mixture models in (11.52) and (11.53) belong to a family of widely used statistical models known as *generalized linear mixed models* (GLMMs). The three basic elements of such a model are as follows.

(1) The vector of *random effects*. In our examples this is the vector $(\Psi_1, \ldots, \Psi_n)$ containing the systematic factors for each time period.

(2) A distribution from the *exponential family* for the conditional distribution of the responses ($Y_{t,i}$ or $\tilde{Y}_{t,i}$) given the random effects. Responses are assumed to be conditionally independent given the random effects. The Bernoulli, binomial and Poisson distributions all belong to the exponential family (see, for example, McCullagh and Nelder 1989, p. 28).

(3) A *link function* relating $E(Y_{t,i} \mid \Psi_t)$, the mean response conditional on the random effects, to the so-called *linear predictor*. In our examples the linear predictor for $Y_{t,i}$ is

$$\eta_{t,i}(\Psi_t) = \mu + \boldsymbol{x}_{t,i}'\boldsymbol{\beta} + \Psi_t. \tag{11.54}$$

We have considered the so-called probit and logit link functions in the Bernoulli case and the log-link function in the Poisson case. (Note that it is usual in GLMMs to write the model as $g(E(Y_{t,i} \mid \Psi_t)) = \eta_{t,i}(\Psi_t)$ and to refer to g as the link function; hence the probit link function is the quantile function of the standard normal, and the link in the Poisson case (11.53) is referred to as "log" rather than "exponential".)

When no random effects are modelled in a GLMM, the model is simply known as a generalized linear model, or GLM. The role of the random effects in the GLMM is, in a sense, to capture patterns of variability in the responses that cannot be explained by the observed covariates alone, but which might be explained by additional unobserved factors. In our case, these unobserved factors are bundled into a time-period effect that we loosely describe as the state of the economy in that time period; alternatively, we refer to it as the systematic risk.

The GLMM framework allows models of much greater complexity. We can add further random effects to obtain multi-factor mixture models. For example, we might know the industry sector of each firm and wish to include random effects for sectors that are nested within the year effect; in this way we might capture additional variability associated with economic effects in different sectors over and above the global variability associated with the year effect. Such models can be considered in the GLMM framework by allowing the linear predictor in (11.54) to take the form $\eta_{t,i}(\Psi_t) = \mu + \boldsymbol{x}_{t,i}'\boldsymbol{\beta} + \boldsymbol{z}_{t,i}'\boldsymbol{\Psi}_t$ for some vector of random effects $\boldsymbol{\Psi}_t = (\Psi_{t,1}, \ldots, \Psi_{t,p})'$; the vector $\boldsymbol{z}_{t,i}$ is a known *design element* of the model that selects the random effects that are relevant to the response $Y_{t,i}$. We would then have a total of $p \times n$ random effects in the model. We may or may not want to model serial dependence in the time series $\boldsymbol{\Psi}_1, \ldots, \boldsymbol{\Psi}_n$.

Inference for GLMMs. Full ML inference for a GLMM is an option for the simplest models. Consider the form of the likelihood for the one-factor models in (11.52) and (11.53). If we write $p_{Y_{t,i}|\Psi_t}(y \mid \psi)$ for the conditional probability mass function of the response $Y_{t,i}$ (or $\tilde{Y}_{t,i}$) given Ψ_t, we have, for data $\{Y_{t,i} : t = 1, \ldots, n,\ i = 1, \ldots, m_t\}$,

$$L(\boldsymbol{\beta}, \sigma; \text{data}) = \int \cdots \int \left(\prod_{t=1}^{n} \prod_{i=1}^{m_t} p_{Y_{t,i}|\Psi_t}(Y_{t,i} \mid \psi_t) \right) f(\psi_1, \ldots, \psi_n)\, \mathrm{d}\psi_1 \cdots \mathrm{d}\psi_n, \tag{11.55}$$

where f denotes the assumed joint density of the random effects. If we do not assume independent random effects from time period to time period, then we are faced with an n-dimensional integral (or an $(n \times p)$-dimensional integral in multi-factor models). Assuming iid Gaussian random effects with marginal Gaussian density f_Ψ, the likelihood (11.55) becomes

$$L(\boldsymbol{\beta}, \sigma; \text{data}) = \prod_{t=1}^{n} \left(\int \prod_{i=1}^{m_t} p_{Y_{t,i}|\Psi_t}(Y_{t,i} \mid \psi_t) f_\Psi(\psi_t)\, \mathrm{d}\psi_t \right), \tag{11.56}$$

so we have a product of one-dimensional integrals and this can be easily evaluated numerically and maximized over the unknown parameters. Alternatively, faster approximate likelihood methods, such as penalized quasi-likelihood (PQL) and marginal quasi-likelihood (MQL), can be used (see Notes and Comments).

Another attractive possibility is to treat inference for these models from a Bayesian point of view and to use Markov chain Monte Carlo (MCMC) methods to make inferences about parameters (McNeil and Wendin 2006, 2007, see, for example,). The Bayesian approach has two main advantages. First, a Bayesian MCMC approach allows us to work with much more complex models than can be handled in the likelihood framework, such as a model with serially dependent random effects. Second, the Bayesian approach is ideal for handling the considerable parameter uncertainty in portfolio credit risk, particularly in models for higher-rated counterparties, where default data are scarce.

11.5.4 A One-Factor Model with Rating Effect

In this section we fit a Bernoulli mixture model to annual default count data from Standard & Poor's for the period 1981–2000; these data have been reconstructed from published default rates in Brand and Bahr (2001, Table 13, pp. 18–21). Standard & Poor's uses the ratings AAA, AA, A, BBB, BB, B, CCC, but because the observed one-year default rates for AAA-rated and AA-rated firms are mostly zero, we concentrate on the rating categories A–CCC.

In our model we assume a single yearly random effect representing the state of the economy and treat the rating category as an observed covariate for each firm in each time period. Our model is a particular instance of the one-factor Bernoulli mixture model in (11.52) and a multi-period extension of the model described in Example 11.8. We assume for simplicity that random effects in each year are iid normal, which allows us to use the likelihood (11.56).

Since we are able to pool companies into groups by year and rating category, we note that it is possible to reformulate the model as a binomial mixture model. Let $r = 1, \ldots, 5$ index the five rating categories in our study, and write $m_{t,r}$ for the number of followed companies in year t with rating r, and $M_{t,r}$ for the number of these that default. Our model assumption is that, conditional on Ψ_t (and the group sizes), the default counts $M_{t,1}, \ldots, M_{t,5}$ are independent and are distributed in such a way that $M_{t,r} \mid \Psi_t = \psi \sim B(m_{t,r}, p_r(\psi))$. Using the probit link, the conditional default probability of an r-rated company in year t is given by

$$p_r(\Psi_t) = \Phi(\mu_r + \Psi_t). \tag{11.57}$$

The model may be fitted under the assumption of iid random effects in each year by straightforward maximization of the likelihood in (11.56). The parameter estimates and obtained standard errors are given in Table 11.5, together with the estimated default probabilities $\hat{\pi}^{(r)}$ for each rating category and estimated default correlations $\hat{\rho}_Y^{(r_1,r_2)}$ implied by the parameter estimates. Writing Ψ for a generic random effect variable, the default probability for rating category r is given by

$$\hat{\pi}^{(r)} = E(\hat{p}_r(\Psi)) = \int_{-\infty}^{\infty} \Phi(\hat{\mu}_r + \hat{\sigma} z)\phi(z)\,\mathrm{d}z, \quad 1 \leqslant r \leqslant 5,$$

where ϕ is the standard normal density. The default correlation for two firms with ratings r_1 and r_2 in the same year is calculated easily from the joint default probability for these two firms, which is

$$\hat{\pi}_2^{(r_1,r_2)} = E(\hat{p}_{r_1}(\Psi)\hat{p}_{r_2}(\Psi)) = \int_{-\infty}^{\infty} \Phi(\hat{\mu}_{r_1} + \hat{\sigma} z)\Phi(\hat{\mu}_{r_2} + \hat{\sigma})\phi(z)\,\mathrm{d}z.$$

The default correlation is then

$$\hat{\rho}_Y^{(r_1,r_2)} = \frac{\hat{\pi}_2^{(r_1,r_2)} - \hat{\pi}^{(r_1)}\hat{\pi}^{(r_2)}}{\sqrt{(\hat{\pi}^{(r_1)} - (\hat{\pi}^{(r_1)})^2)(\hat{\pi}^{(r_2)} - (\hat{\pi}^{(r_2)})^2)}}.$$

Note that the default correlations are correlations between event indicators for very low probability events and are necessarily very small.

The model in (11.57) assumes that the variance of the systematic factor Ψ_t is the same for all firms in all years. When compared with the very general Bernoulli mixture model (11.23) we might be concerned that the simple model considered in this section does not allow for enough heterogeneity in the variance of the systematic risk. A simple extension of the model is to allow the variance to be different for different rating categories: that is, to fit a model where $p_r(\Psi_t) = \Phi(\mu_r + \sigma_r \Psi_t)$ and where Ψ_t is a standard normally distributed random effect. This increases the number of parameters in the model by four but is no more difficult to fit than the basic model. The maximized value of the log-likelihood in the model with heterogeneous scaling is -2557.4, and the value in the model with homogeneous scaling is -2557.7; a likelihood ratio test suggests that no significant improvement results from allowing heterogeneous scaling. If rating is the only categorical variable, the simple model seems adequate, but if we had more information on the industrial and geographical sectors to which the companies belonged, it would be natural to introduce further

Table 11.5. Maximum likelihood parameter estimates and standard errors (se) for a one-factor Bernoulli mixture model fitted to historical Standard & Poor's one-year default data, together with the implied estimates of default probabilities $\hat{\pi}^{(r)}$ and default correlations $\hat{\rho}_Y^{(r_1,r_2)}$. The MLE of the scaling parameter σ is 0.24 with standard error 0.05. Note that we have tabulated default correlation in absolute terms and not in percentage terms.

Parameter	A	BBB	BB	B	CCC	
μ_r	−3.43	−2.92	−2.40	−1.69	−0.84	
se (μ_r)	0.13	0.09	0.07	0.06	0.08	
$\pi^{(r)}$	0.000 4	0.002 3	0.009 7	0.050 3	0.207 8	
$\rho_Y^{(r_1,r_2)}$	0.000 40	0.000 77	0.001 30	0.002 19	0.003 04	A
	0.000 77	0.001 49	0.002 55	0.004 35	0.006 15	BBB
	0.001 30	0.002 55	0.004 40	0.007 63	0.010 81	BB
	0.002 19	0.004 35	0.007 63	0.013 28	0.019 06	B
	0.003 04	0.006 15	0.010 81	0.019 06	0.027 88	CCC

random effects for these sectors and to allow more heterogeneity in the model in this way.

The implied default probability and default correlation estimates in Table 11.5 can be a useful resource for calibrating simple credit models to homogeneous groups defined by rating. For example, to calibrate a Clayton copula to group BB we use the inputs $\pi^{(3)} = 0.0097$ and $\rho_Y^{(3,3)} = 0.0044$ to determine the parameter θ of the Clayton copula (see Example 11.13). Note also that we can now immediately use the scaling results of Section 11.3 to calculate approximate risk measures for large portfolios of companies that have been rated with the Standard & Poor's system (see Example 11.20).

Notes and Comments

The main references for our account of industry factor models are Huang et al. (2012) and RiskMetrics Group (1997).

The estimator (11.50) for joint default probabilities is also used in Lucas (1995) and Nagpal and Bahar (2001), although de Servigny and Renault (2002) suggest there may be problems with this estimator for groups with low default rates. A related moment-style estimator has been suggested by Gordy (2000) and appears to have a similar performance to (11.50) (see Frey and McNeil 2003). A further paper on default correlation estimation is Gordy and Heitfield (2002).

A good overview article on generalized linear mixed models is Clayton (1996). For generalized linear models a standard reference is McCullagh and Nelder (1989) (see also Fahrmeir and Tutz 1994).

The analysis of Section 11.5.4 is very similar to the analysis in Frey and McNeil (2003) (where heterogeneous variances for each rating category were assumed). The results reported in this book were obtained by full maximization of the likelihood using our own R code. Very similar results are obtained with the `glmer` function

in the `lme4` R package, which maximizes an adaptive Gauss–Hermite approximation to the log-likelihood. GLMMS may also be estimated using the approximate penalized quasi-likelihood (PQL) and marginal quasi-likelihood (MQL) methods (see Breslow and Clayton 1993). For a Bayesian approach to fitting the model using Markov chain Monte Carlo techniques, see McNeil and Wendin (2007), who also incorporate an autoregressive time-series structure for the random effects.

Although we have only described default models it is also possible to analyse rating migrations in the generalized linear model framework (with or without random effects). A standard model is the ordered probit model, which is used without random effects in Nickell, Perraudin and Varotto (2000) to provide evidence of time variation in default rates attributable to macroeconomic factors; a similar message is found in Bangia et al. (2002). McNeil and Wendin (2006) show how random effects and unobserved factors may be included in such models and carry out Bayesian inference. See also Gagliardini and Gouriéroux (2005), in which a variety of rating-migration models with serially dependent unobserved factors are studied.

There is a large literature on models with latent structure designed to capture the dynamics of systematic risk, and there is quite a lot of variation in the types of model considered. Crowder, Davis and Giampieri (2005) use a two-state hidden Markov structure to capture periods of high and low default risk, Koopman, Lucas and Klaassen (2005) use an unobserved components time-series model to describe US company failure rates, Koopman, Lucas and Monteiro (2008) develop a latent factor intensity model for rating transitions, and Koopman, Lucas and Schwaab (2012) combine macroeconomic factors, unobserved frailties and industry effects in a model of US defaults through the crisis of 2008.

12
Portfolio Credit Derivatives

In this chapter we study portfolio credit derivatives such as collateralized debt obligations (CDOs) and related products. The primary use of portfolio credit derivatives is in the securitization of credit risk: that is, the transformation of credit risk into securities that may be bought and sold by investors. The market for portfolio credit derivatives peaked in the period leading up to the 2007–9 credit crisis (as discussed in Section 1.2.1) and has only partly recovered since.

In Section 12.1 we describe the most important portfolio credit derivatives and their properties. We also provide some more discussion of the role that these products played in the credit crisis. Section 12.2 introduces copula models for portfolio credit risk, which have become the market standard in pricing CDOs and related credit derivatives. In Section 12.3 we discuss pricing and model calibration in factor copula models. More advanced dynamic portfolio credit risk models are studied in Chapter 17.

This chapter makes extensive use of the analysis of single-name credit risk models in Sections 10.1–10.4 and of basic notions in copula theory. We restrict our attention to models for random default times with deterministic hazard functions without adding the extra complexity of doubly stochastic default times as in Sections 10.5 and 10.6. The simpler models are sufficient to understand the key features of portfolio credit derivative pricing.

12.1 Credit Portfolio Products

In this section we describe the pay-off and qualitative properties of certain important credit portfolio products such as CDOs. We begin by introducing the necessary notation.

We consider a portfolio of m firms with default times $\tau_1, \dots, \tau_m$. In keeping with the notation introduced in the earlier credit chapters, the random vector $\boldsymbol{Y}_t = (Y_{t,1}, \dots, Y_{t,m})'$ with $Y_{t,i} = I_{\{\tau_i \leqslant t\}}$ describes the default state of the portfolio at some point in time $t \geqslant 0$. Note that $Y_{t,i} = 1$ if firm i has defaulted by time t, and $Y_{t,i} = 0$ otherwise. We assume throughout that there are no simultaneous defaults, so we may define the *ordered default times* $T_0 < T_1 < \cdots < T_m$ by setting $T_0 = 0$ and recursively setting

$$T_n := \min\{\tau_1, \dots, \tau_m : \tau_i > T_{n-1}\}, \quad 1 \leqslant n \leqslant m.$$

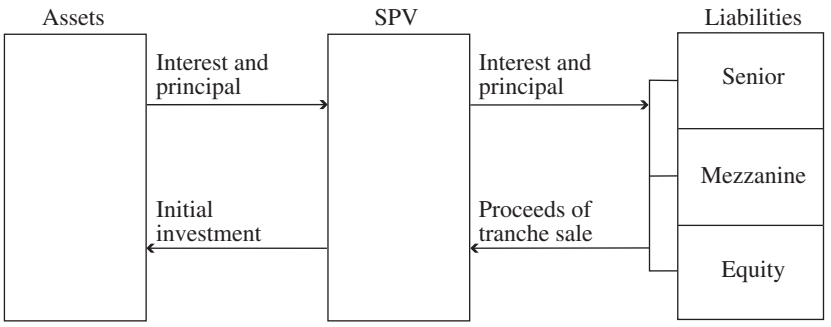

Figure 12.1. Schematic representation of the payments in a CDO structure.

As in Chapter 11, the exposure to firm (or so-called reference entity) i is denoted by e_i and the percentage loss given default (LGD) of firm i is denoted by $\delta_i \in [0, 1]$. The *cumulative loss* of the portfolio up to time t is therefore given by $L_t = \sum_{i=1}^{m} \delta_i e_i Y_{t,i}$. While δ_i and e_i may in principle be random, we mostly work with deterministic exposures and LGDs; further assumptions about these quantities are introduced as and when needed.

12.1.1 Collateralized Debt Obligations

Before the credit crisis of 2007–9 CDO markets were a fast-growing segment of the credit market. Although activity on CDO markets has slowed down since the crisis, CDOs and credit products with a similar structure remain an important asset class for risk managers to study.

A CDO is a financial instrument for the securitization of a portfolio of credit products such as bonds, loans or mortgages. This portfolio forms the so-called *asset pool* underlying the contract. The CDOs that are traded in practice come in many different varieties, but the basic structure is the same. The assets are sold to a *special-purpose vehicle* (SPV): a company that has been set up with the single purpose of carrying out the securitization deal. To finance the acquisition of the assets, the SPV issues securities in tranches of differing seniority, which form the *liability side* of the structure. The tranches of the liability side are called (in order of increasing seniority) equity, mezzanine and senior tranches (sometimes there are also super-senior tranches). The rules that determine the exact cash flow of the tranches are known as the *waterfall structure* of the CDO. These rules can be quite complex. Roughly speaking, the waterfall structure ensures that losses due to credit events on the asset side are borne first by the equity tranche; if the equity tranche is exhausted, losses are borne by the mezzanine tranches and only thereafter by the senior tranches. The credit quality of the more senior tranches is therefore usually higher than the average credit quality of the asset pool. The payments associated with a typical CDO are depicted schematically in Figure 12.1.

CDOs where the asset pool consists mainly of bonds are known as collateralized bond obligations (CBOs); if the asset side consists mainly of loans, a CDO is termed a collateralized loan obligation (CLO); CDOs for the securitization of mortgages are also known as mortgage-backed securities (MBSs) or asset-backed securities

(ABSs). There is also a liquid market for *synthetic* CDO tranches. In these contracts payments are triggered by default events in a pool of reference names (typically major corporations), but there are no actual assets underlying the contract. A precise pay-off description for synthetic CDO tranches is given in Section 12.1.2.

There are a number of economic motivations for arranging a CDO transaction. To begin with, in a typical CDO structure a large part of the asset pool is allocated to the senior tranches with a fairly high credit quality, even if the quality of the underlying assets is substantially lower. For instance, according to Hull and White (2010), for a typical ABS created from residential mortgages, about 75–80% of the underlying mortgage principal is allocated to senior tranches with a AAA rating. Many institutional investors prefer an investment in highly rated securities because of legal or institutional constraints. Securitization can therefore be a way to sell a large part of the underlying assets to investors who are unable to invest directly in the asset pool. Another incentive to set up a CDO transaction is related to capital adequacy; CDOs are often issued by banks who want to sell some of the credit-risky securities on their balance sheet in order to reduce their regulatory capital requirements.

Securitization via CDOs or ABSs is an important tool in credit markets. It allows lenders to reduce concentration risk and leverage and to refinance themselves more efficiently. Securitization can therefore increase the lending capacity of the financial sector. On the other hand, the credit crisis of 2007–9 clearly exposed a number of problems related to securitization and the use of asset-backed CDOs.

- Securitization can create *incentive problems*: if a mortgage originator knows that most of the mortgages he sells to homeowners will be securitized later on, he has little interest in evaluating the credit quality of the borrowers carefully. This can lead to a deterioration of lending standards. There is a lot of evidence that this actually happened in the years preceding the subprime crisis (see, for example, Crouhy, Jarrow and Turnbull 2008). This problem could be addressed by better aligning the interests of loan originators and of ABS investors. For instance, originators could be forced to keep a certain percentage of all the tranches they sell on the securitization market.
- The exact cash-flow structure of most asset-backed CDOs is very complicated. In fact, the legal documentation defining the payments of an asset-backed CDO can run to several hundred pages. This makes it difficult for investors to form an opinion of the value and the riskiness of any given tranche, thus contributing to the low trading volume in securitization markets during the credit crisis. At the height of the subprime crisis CDO products were on offer that crossed ethical and even legal boundaries. An example is the infamous ABACUS 2700-AC-1 CDO of Goldman Sachs (see Duffie (2010) for details).
- CDOs and ABSs were clearly misused by banks before the credit crisis to exploit *regulatory arbitrage*. They allowed loan-related credit risk to be transferred from the banking book to the trading book, where it enjoyed a more lenient capital treatment. By holding tranches that had been overoptimistically

rated as AAA in the trading book, banks were able to substantially lower their regulatory capital.

- The pay-off distribution associated with CDO tranches is very sensitive to the characteristics of the underlying asset pool, which makes it difficult to properly assess the risk associated with these instruments. In particular, a small increase in default correlation often leads to a substantial increase in the likelihood of losses for senior tranches. An intuitive explanation of this correlation sensitivity is given below.

Stylized CDOs and correlation sensitivity. In order to gain a better understanding of the main qualitative features of CDOs without getting bogged down in the details of the waterfall structure, we introduce a hypothetical contract that we label a *stylized CDO*. We consider a portfolio of m firms with cumulative loss $L_t = \sum_{i=1}^{m} \delta_i e_i Y_{t,i}$ and deterministic exposures. The stylized CDO has k tranches, indexed by $\kappa \in \{1, \dots, k\}$ and characterized by attachment points $0 = K_0 < K_1 < \cdots < K_k \leqslant \sum_{i=1}^{m} e_i$. The value of the *notional* corresponding to tranche κ can be described as follows. Initially, the notional is equal to $K_\kappa - K_{\kappa-1}$; it is reduced whenever there is a default event such that the cumulative loss falls in the layer $[K_{\kappa-1}, K_\kappa]$. In mathematical terms, $N_{t,\kappa}$, the notional of tranche κ at time t, is given by

$$N_{t,\kappa} = N_\kappa(L_t) \quad \text{with } N_\kappa(l) = \begin{cases} K_\kappa - K_{\kappa-1} & \text{for } l < K_{\kappa-1}, \\ K_\kappa - l & \text{for } l \in [K_{\kappa-1}, K_\kappa], \\ 0 & \text{for } l > K_\kappa. \end{cases} \tag{12.1}$$

Note that $N_\kappa(l) = (K_\kappa - l)^+ - (K_{\kappa-1} - l)^+$, so $N_{t,\kappa}$ is equal to the sum of a long position in a put option on L_t with strike price K_κ and a short position in a put on L_t with strike price $K_{\kappa-1}$. Such positions are also known as put spreads.

We assume that in a stylized CDO the pay-off of tranche κ is equal to $N_{T,\kappa}$, the value of the tranche notional at the maturity date T. In Figure 12.2 we have graphed the pay-off for a stylized CDO with maturity $T = 5$ years and three tranches (equity, mezzanine, senior) on a homogeneous portfolio of $m = 1000$ firms, each with exposure one unit and loss given default $\delta_i = 0.5$. The attachment points are $K_1 = 20$, $K_2 = 40$, $K_3 = 60$, corresponding to 2%, 4% and 6% of the overall exposure; tranches with higher attachment points are ignored. We have plotted two distributions for L_T: first, a loss distribution corresponding to a five-year default probability of 5% and a five-year default correlation of 2%; second, a loss distribution with a five-year default probability of 5% but with independent defaults. In both cases the expected loss is given by $E(L_T) = 25$. Figure 12.2 illustrates how the value of different CDO tranches depends on the extent of the dependence between default events.

- For independent defaults, L_T is typically close to its mean due to diversification effects within the portfolio. It is therefore quite unlikely that a tranche κ with lower attachment point $K_{\kappa-1}$ substantially larger than $E(L_T)$ (such

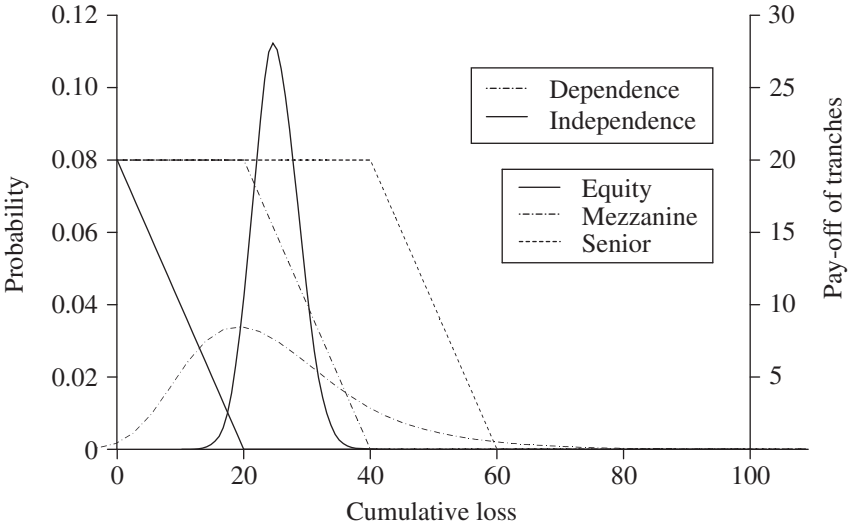

Figure 12.2. Pay-off of a stylized CDO contract and distribution of the five-year loss L_5 for a five-year default probability of 5% and different default correlations. Detailed explanations are given in the text.

as the senior tranche in Figure 12.2) suffers a loss, so the value of such a tranche is quite high. On the other hand, since the upper attachment point K_1 of the equity tranche is lower than $E(L_T) = 25$, it is quite unlikely that L_T is substantially smaller than K_1, and the value of the equity tranche is low.

- If defaults are dependent, diversification effects in the portfolio are less pronounced. Realizations of the loss L_T larger than the lower attachment point K_2 of the senior tranche are more likely, as are realizations of L_T smaller than the upper attachment point K_1 of the equity tranche. This reduces the value of tranches with high seniority and increases the value of the equity tranche compared with the case of independent defaults.
- The impact of changing default correlations on mezzanine tranches is unclear and cannot be predicted up front.

The relationship between default dependence and the value of CDO tranches carries over to the more complex structures that are actually traded, so dependence modelling is a key issue in any model for pricing CDO tranches (see also Section 12.3.2 below).

Pricing and the role of rating agencies. Before the financial crisis, rating agencies played a dominant role in the valuation of asset-backed CDOs. In fact, many CDO investors lacked the necessary sophistication and data to form an independent judgement of the riskiness of asset-backed CDOs, a problem that was compounded by the complex waterfall structure of most CDO issues. They therefore based their investment decision solely on the risk assessment of the rating agencies. This is particularly true for AAA-rated tranches, which appeared to be attractive investment

opportunities due to the relatively high offered yield (compared with the yield that could be earned on a standard AAA-rated bond, for example).

In relying on ratings, investors implicitly assumed that a high-quality rating such as AAA for a CDO tranche meant that the tranche had a similar risk profile to a AAA-rated bond. This perception is clearly wrong; since the loss distribution of an asset-backed CDO tranche is extremely sensitive with respect to the credit quality and the default correlation of the mortgages in the underlying asset pool, ratings for CDO tranches change rapidly with changes in these parameters. Moreover, while rating agencies have a lot of experience in rating corporate and sovereign debt, their experience with CDOs and other structured credit products was quite limited. As a result, ratings for CDO tranches turned out to be very unstable. In fact, at the onset of the crisis the rating of a large proportion of the traded ABS–CDOs (CDOs where the underlying asset pool consists of mortgage-based ABSs) was downgraded from investment grade to speculative grade, including default, within a very short period of time (Crouhy, Jarrow and Turnbull 2008). This massive rating change has sparked an intense debate about the appropriateness of rating methodologies, and about the role and the incentives of rating agencies more generally. We refer to Notes and Comments for further reading.

CDO-squared contracts. CDO-squared contracts are CDOs where the underlying asset pool itself consists of CDO tranches. These products are very complex and difficult, if not impossible, to value. For this reason they never became particularly popular on markets for synthetic CDOs. The situation is different in markets for asset-backed CDOs. Before the crisis there was intense trading activity in ABS–CDOs. The main reason for this was the fact that these products seemed to offer a way to create additional AAA-rated securities from the mezzanine tranches of the original ABSs, thus satisfying the high demand for AAA-rated securities. Investors in highly rated ABS–CDOs incurred severe losses during the credit crisis and, as shown by Hull and White (2010), the AAA rating carried by many of these structures was very hard to defend in retrospect. Many studies since the financial crisis have emphasized the need for simpler and more standardized financial products: see, for example, Crouhy, Jarrow and Turnbull (2008) and Hull (2009). ABS–CDOs are clearly a prime case in point.

12.1.2 Credit Indices and Index Derivatives

Credit index derivatives are standardized credit products whose pay-off is determined by the occurrence of credit events in a fixed pool of major firms that form the so-called credit index. A key requirement for the inclusion of a firm in a credit index is the existence of a liquid single-name CDS market in that firm. The availability of indices has helped to create a liquid market for certain credit index derivatives that has become a useful benchmark for model calibration and an important reference point for academic studies.

At present there are two major families of credit indices: the CDX family and the iTraxx family. CDX indices refer to American companies and iTraxx indices refer either to European firms or to Asian and Australian firms. Characteristics of

Table 12.1. Composition of main credit indices (taken from Couderc and Finger (2010)). The most important indices are the iTraxx Europe and CDX.NA.IG indices.

Name	Pool size	Region	Credit quality
CDX.NA.IG	125	North America	Investment grade
CDX.NA.IG.HVOL	30	North America	Low-quality investment grade
CDX.NA.HY	100	North America	Non-investment-grade
iTraxx Europe	125	Europe	Investment grade
iTraxx Europe	30	Europe	Low-quality investment grade

the main credit indices are given in Table 12.1. In order to reflect changes in the credit quality of the constituents, the composition of most credit indices changes every six months at the so-called roll dates (20 March and 20 September), and the pools corresponding to the roll dates are known as the different *series* of the index. Products on older series continue to trade but the market for products related to the current series is by far the most liquid.

Standardized index derivatives are credit index swaps and single-tranche CDOs with a standardized set of attachment points. The cash flow of these products bears some similarities to the cash flows of a single-name CDS as described in Sections 10.1.4 and 10.4.4. Each contract consists of a *premium payment leg* (payments made by the protection buyer) and a *default payment leg* (payments made by the protection seller). Premium payments are due at deterministic time points $0 < t_1 < \cdots < t_N = T$, where T is the maturity of the contract. Standardized index derivatives have quarterly premium payments, i.e. $t_n - t_{n-1} = 0.25$; the time to maturity at issuance is three, five, seven or ten years, with five-year products being the most liquid.

Next we describe the pay-offs of an index swap and a CDO tranche. We consider a fixed pool of m names ($m = 125$ for derivatives related to the iTraxx Europe and CDX.NA.IG indices) and we normalize the exposure of each firm to 1, so that the cumulative portfolio loss at time t equals $L_t = \sum_{i=1}^m \delta_i Y_{t,i}$.

Credit index swaps. At a default time $T_k \leqslant T$ there is a default payment of size δ_{ξ_k}, where $\xi_k \in \{1, \ldots, m\}$ is the identity of the name defaulting at T_k. The cumulative cash flows of the default payment leg up to time $t \leqslant T$ (received by the protection buyer) are therefore given by

$$\sum_{T_k \leqslant t} \delta_{\xi_k} = \sum_{\tau_i \leqslant t} \delta_i = \sum_{i=1}^m \delta_i Y_{t,i} = L_t.$$

Given an annualized swap spread x, the premium payment at time t_n (received by the protection seller) is given by

$$x(t_n - t_{n-1}) N_{t_n}^{\mathrm{Ind}},$$

where the *notional* N_t^{Ind} of the index swap is equal to the number of surviving firms at time t: that is, $N_t^{\mathrm{Ind}} = m - \sum_{i=1}^m Y_{t,i}$. This definition of the notional reflects the

Table 12.2. Standardized attachment points for single-tranche CDOs on the CDX.NA.IG and iTraxx Europe indices.

CDX.NA.IG	0–3%	3–7%	7–10%	10–15%	15–30%
iTraxx Europe	0–3%	3–6%	6–9%	9–12%	12–20%

fact that at a default time the defaulting entity is removed from the index. Moreover, at a default time $T_k \in (t_{n-1}, t_n]$, the protection buyer pays the protection seller the part of the premium that has accrued since the last regular premium payment date: that is, the quantity $x(T_k - t_{n-1})$. A credit event therefore has a double effect on the cash-flow structure of the index swap: it leads to a default payment and it reduces future premium payments.

Single-tranche CDOs. A single-tranche CDO on the reference portfolio is characterized by fixed lower and upper attachment points $0 \leqslant l < u \leqslant 1$, expressed as percentages of the overall notional m of the index pool. As in (12.1) we define the notional $N_t^{[l,u]}$ of the tranche by a put spread:

$$N_t^{[l,u]} = (um - L_t)^+ - (lm - L_t)^+. \tag{12.2}$$

In particular, for $L_0 = 0$ the initial notional of the tranche is equal to $m(u - l)$. The cumulative *tranche loss* up to time t is then given by

$$L_t^{[l,u]} := m(u - l) - N_t^{[l,u]} = (L_t - lm)^+ - (L_t - um)^+, \tag{12.3}$$

so the tranche loss can be viewed as a call spread on the cumulative portfolio loss. At a default time $T_k \leqslant T$ the protection seller makes a default payment of size

$$\Delta L_{T_k}^{[l,u]} := L_{T_k}^{[l,u]} - L_{T_k-}^{[l,u]}. \tag{12.4}$$

Again the premium payment leg consists of regular and accrued premium payments. Given an annualized tranche spread x, the regular premium payment at date t_n is given by $x(t_n - t_{n-1})N_{t_n}^{[l,u]}$. The accrued payment at a default time $T_k \in (t_{n-1}, t_n]$ equals $x(T_k - t_{n-1})\Delta L_{T_k}^{[l,u]}$. In order to simplify the exposition we will usually omit accrued premium payments below.

Single-tranche CDOs are sometimes called *synthetic*, as there is no physical transfer of credit-risky securities from the protection seller to the protection buyer, in contrast to asset-backed CDOs. There is a standardized set of attachment points for index tranches on the iTraxx Europe and CDX.NA.IG indices (see Table 12.2). In analogy with the terminology used for asset-backed CDOs, the tranche with lower attachment point $l = 0$ is known as the *equity tranche*; the equity tranche is clearly affected by the first losses in the underlying pool. The tranche with the highest attachment point is termed the *senior* tranche and the other tranches are known as *mezzanine* tranches of differing seniority. Tranches with non-standard maturity dates or attachment points and tranches on portfolios other than the constituents of a popular credit index are known as *bespoke* CDO tranches.

12.1.3 Basic Pricing Relationships for Index Swaps and CDOs

In this section we discuss some elementary model-independent pricing relationships for index swaps and CDOs that will be useful in this chapter. In our analysis we do not specify the underlying portfolio credit risk model in any detail; rather we take as given the joint distribution of the default times and of the LGDs calculated under some risk-neutral pricing measure Q. For simplicity we set the valuation date equal to $t = 0$. Moreover, we assume that the default-free interest rate $r(t)$ is deterministic, and we denote by

$$p_0(0,t) = \exp\left(-\int_0^t r(s)\,\mathrm{d}s\right), \quad t \geqslant 0,$$

the default-free discount factors or zero-coupon bond prices as seen from time $t = 0$. Deterministic interest rates are assumed in most of the literature on portfolio credit risk models, essentially because the additional complexity of stochastic interest rates is not warranted given the large amount of uncertainty surrounding the modelling of default dependence.

CDS index swaps. The market value V^{Def} of the default payments of a CDS index swap is given by the Q-expectation of the associated discounted cash-flow stream. The latter is given by $\sum_{T_k \leqslant T} p_0(0,T_k)\Delta L_{T_k}$, where $\Delta L_{T_k} = L_{T_k} - L_{T_k-} = \delta_{\xi_k}$. Since this sum may be written more succinctly as $\int_0^T p_0(0,t)\,\mathrm{d}L_t$ we get

$$V^{\mathrm{Def}} = E^Q\left(\int_0^T p_0(0,t)\,\mathrm{d}L_t\right) = \sum_{i=1}^m E^Q\left(\int_0^T p_0(0,t)\,\mathrm{d}L_{t,i}\right), \tag{12.5}$$

where $L_{t,i} = \delta_i Y_{t,i}$ is the cumulative loss process of firm i.

Given a generic spread x, the market value of the premium payments is given by $V^{\mathrm{Prem}}(x)$, where

$$\begin{aligned} V^{\mathrm{Prem}}(x) &= x\sum_{n=1}^N p_0(0,t_n)(t_n - t_{n-1})E^Q(N_t^{\mathrm{Ind}}) \\ &= x\sum_{i=1}^m\sum_{n=1}^N p_0(0,t_n)(t_n - t_{n-1})E^Q(1 - Y_{t_n,i}). \end{aligned} \tag{12.6}$$

Clearly, $V^{\mathrm{Prem}}(x) = xV^{\mathrm{Prem}}(1)$.

The market value at $t = 0$ of a protection buyer position in an index swap with given spread x is thus given by $V^{\mathrm{Def}} - V^{\mathrm{Prem}}(x)$. As in the case of single-name CDSs, the *fair index swap spread* x^{Ind} of the contract at a given point in time is set such that the market value of the contract at that date is 0. This leads to the formula

$$x^{\mathrm{Ind}} = \frac{V^{\mathrm{Def}}}{V^{\mathrm{Prem}}(1)}. \tag{12.7}$$

Next we consider the relationship between the index swap spread x^{Ind} and the fair CDS spread x^i for the single-name CDSs on the constituents of the index. It is tempting to conclude that the index swap spread is simply the arithmetic average of

the x^i, but this is wrong in general as the following analysis shows. For a single-name CDS on firm i with identical maturity and premium payment dates as the index swap, one has $V_i^{\mathrm{Def}} = V_i^{\mathrm{Prem}}(x^i)$ with

$$V_i^{\mathrm{Def}} = E^Q\bigg(\int_0^T p_0(0,t)\,\mathrm{d}L_{t,i}\bigg),$$
$$V_i^{\mathrm{Prem}}(x) = x\sum_{n=1}^N p_0(t,t_n)(t_n - t_{n-1})E^Q(1 - Y_{t_n,i}), \quad x > 0.$$

Comparing these relations with (12.5) and (12.6), we see that $V^{\mathrm{Def}} = \sum_{i=1}^m V_i^{\mathrm{Def}}$ and $V^{\mathrm{Prem}}(x) = \sum_{i=1}^m V_i^{\mathrm{Prem}}(x)$. For the fair index spread we therefore obtain

$$x^{\mathrm{Ind}} = \frac{V^{\mathrm{Def}}}{V^{\mathrm{Prem}}(1)} = \frac{\sum_{i=1}^m V_i^{\mathrm{Def}}}{\sum_{i=1}^m V_i^{\mathrm{Prem}}(1)} = \frac{\sum_{i=1}^m x^i V_i^{\mathrm{Prem}}(1)}{\sum_{i=1}^m V_i^{\mathrm{Prem}}(1)} =: \sum_{i=1}^m w_i x^i, \quad (12.8)$$

with weights given by $w_i := V_i^{\mathrm{Prem}}(1)/(\sum_{i=1}^m V_i^{\mathrm{Prem}}(1))$. The index spread is indeed therefore a weighted average of the single-name CDS spreads, but the weights are in general not equal to $1/m$. In fact, if firm i is of high credit quality and firm j of relatively low credit quality, one has

$$E^Q(1 - Y_{t,i}) = Q(\tau_i > t) > Q(\tau_j > t) = E^Q(1 - Y_{t,j}), \quad t > 0.$$

This implies that $V_i^{\mathrm{Prem}}(1) > V_j^{\mathrm{Prem}}(1)$ and hence that $w_i > w_j$, so that high-quality firms have a larger weight than low-quality firms. Of course, in the special case where all τ_i have the same distribution we get $w_1 = \cdots = w_m = 1/m$. An example is given by the simple model in which the default times are exponentially distributed with identical hazard rate γ^Q so that $Q(\tau_i > t) = \mathrm{e}^{-\gamma^Q t}$, and in which the LGD is deterministic and identical across firms. In that case we have $x^1 = \cdots = x^m = x^{\mathrm{ind}}$. Moreover, the parameter γ^Q can be calibrated from a market-observed index spread x^* using the same procedure as in the case of single-name CDS spreads (see Section 10.4.4). This setup is frequently employed by practitioners in the computation of implied correlations for single-tranche CDOs.

Single-tranche CDOs. Finally, we provide a more explicit description for the value of a single-tranche CDO. We begin with the premium payments. According to the definition of the tranche notional in (12.2), the market value of the regular premium payments for a generic CDO spread x is given by $V^{\mathrm{Prem}}(x)$, where

$$V^{\mathrm{Prem}}(x) = x\sum_{n=1}^N p_0(0,t_n)(t_n - t_{n-1})E^Q((um - L_{t_n})^+ - (lm - L_{t_n})^+). \quad (12.9)$$

Concerning the default payments we note from (12.4) that the discounted cash-flow stream of the default payment leg is given by

$$V^{\mathrm{Def}} = \sum_{T_k \leqslant T} p_0(0,T_k)\Delta L_{T_k}^{[l,u]} = \int_0^T p_0(0,t)\,\mathrm{d}L_t^{[l,u]}. \quad (12.10)$$

The integral on the right can be approximated by a Riemann sum, using the premium payment dates as gridpoints:

$$\int_0^T p_0(0,t)\,\mathrm{d}L_t^{[l,u]} \approx \sum_{n=1}^{N} p_0(0,t_n)(L_{t_n}^{[l,u]} - L_{t_{n-1}}^{[l,u]}). \tag{12.11}$$

In economic terms this approximation means that losses occurring during the period $[t_{n-1}, t_n]$ are paid only at time t_n (and are therefore discounted with a slightly higher factor). Since in practice premium payments typically occur quarterly, the error of this approximation is negligible. Recall that $L_t^{[l,u]}$ is a function of L_t, namely $L_t^{[l,u]} = v^{[l,u]}(L_t) := (L_t - lm)^+ - (L_t - um)^+$. Hence, with a slight abuse of notation we have

$$V^{\mathrm{Def}} = \sum_{n=1}^{N} p_0(0,t_n)(E^Q(v^{[l,u]}(L_{t_n})) - E^Q(v^{[l,u]}(L_{t_{n-1}}))). \tag{12.12}$$

Summarizing, we find that the evaluation of (12.12) and (12.9), and hence the determination of CDO spreads, reduces to computing call or put option prices on the cumulative loss process at the premium payment dates $t_1, \ldots, t_N$.

Notes and Comments

There are many contributions that discuss the pros and cons of securitization in the light of the subprime credit crisis of 2007 and 2008. Excellent descriptions of the events surrounding the crisis—including discussions of steps that should be implemented to prevent a repeat of it—are given by Crouhy, Jarrow and Turnbull (2008), Hull (2009) and, in an insurance context, Donnelly and Embrechts (2010) (see also Das, Embrechts and Fasen 2013). Hull and White (2010) test the ratings given to ABSs and ABS–CDOs before the crisis. They find that, whereas the AAA ratings assigned to ABSs were not unreasonable, the AAA ratings assigned to tranches of CDOs created from mezzanine tranches of ABSs cannot be justified by any proper quantitative analysis.

In his analysis of the credit crisis, Brunnermeier (2009) is particularly concerned with the various transmission mechanisms that caused losses in the relatively small American subprime mortgages market to be amplified in such a way that they created a global financial crisis. An interesting discussion of securitization in the light of the subprime crisis from a regulatory viewpoint is the well-known Turner Review (Lord Turner 2009). Incentive problems in the securitization of mortgages are discussed in Franke and Kahnen (2009). A more technical analysis of the value of securitization as a risk-management tool can be found in Frey and Seydel (2010). From the multitude of books we single out Dewatripont, Rochet and Tirole (2010) and Shin (2010). We also suggest that the interested reader look at the various documents published by the Bank for International Settlements and the Basel Committee on Banking Supervision (see www.bis.org/bcbs).

For alternative textbook treatments of CDOs and related index derivatives we refer to Bluhm and Overbeck (2007) and Brigo, Pallavicini and Torresetti (2010).

Both texts contain a wealth of institutional details on CDO markets. The book by Bluhm and Overbeck (2007) also discusses so-called *basket default swaps*, or, more technically, kth-to-default swaps. These products offer protection against the kth default in a portfolio with $m > k$ obligors (the basket). As in the case of an ordinary CDS, the premium payments on a kth-to-default swap take the form of a periodic payment stream, which stops at the kth default time T_k. The default payment is triggered if T_k is smaller than the maturity date of the swap. While first-to-default swaps are fairly common, higher-order default swaps are encountered only rarely.

12.2 Copula Models

Copula models are widely used in practice for the pricing of CDO tranches and basket credit derivatives. In this section we discuss this important class of models with a particular focus on models where the copula has a factor structure.

12.2.1 Definition and Properties

Definition 12.1 (copula model for default times). Let C be a copula and let $\gamma_i(t)$, $1 \leqslant i \leqslant m$, be nonnegative functions such that $\Gamma_i(t) := \int_0^t \gamma_i(s)\,\mathrm{d}s < \infty$ for all $t > 0$ and $\lim_{t\to\infty} \Gamma_i(t) = \infty$. The default times $\tau_1, \ldots, \tau_m$ then follow a copula model with *survival copula* C and *marginal hazard functions* $\gamma_i(t)$ if their joint survival function can be written in the form

$$\bar{F}(t_1, \ldots, t_m) = C(\mathrm{e}^{-\Gamma_1(t_1)}, \ldots, \mathrm{e}^{-\Gamma_m(t_m)}). \tag{12.13}$$

The marginal survival functions in a copula model for defaults are obviously given by $\bar{F}_i(t) = \mathrm{e}^{-\Gamma_i(t)}$. The marginal survival functions and marginal dfs are continuous, and it follows from Sklar's Theorem that both the copula and the survival copula C of the vector of default times $\boldsymbol{\tau} := (\tau_1, \ldots, \tau_m)'$ are unique. Of course, the joint distribution of the default times could also be described in terms of the copula of $\boldsymbol{\tau}$ and the marginal distribution functions $F_1, \ldots, F_m$. We focus on survival copulas as this ties in with a large part of the literature. If C in (12.13) is radially symmetric (see Definition 7.15), then C is also the copula of $\boldsymbol{\tau}$. This is true, in particular, if C is an elliptical copula, such as the Gauss copula or the t copula.

From a mathematical point of view it makes no difference whether we specify the survival copula and the marginal hazard functions separately or whether we specify the joint survival function $\bar{F}$ directly; every joint survival function $\bar{F}$ with absolutely continuous marginal distributions has a unique representation of the form (12.13). However, the representation (12.13) is convenient for the *calibration* of the model to prices of traded credit derivatives. The usual calibration process is carried out in two steps, as we now explain.

In the first step, the marginal hazard functions are calibrated to a given term structure of CDS spreads or spreads of defaultable bonds, as described in Section 10.4.4. In the second step, the parameters of the survival copula C are calibrated to the observed prices of traded portfolio credit derivatives, most notably CDO index tranches. The two-stage calibration is feasible because the second step has no effect

on the parameters calibrated in the first step. This is very advantageous from a numerical point of view, which is one of the reasons for the popularity of copula models among practitioners. We return to the calibration of copula models in Section 12.3.

In order to link our analysis to the discussion of copulas in threshold models in Section 11.1, we take a brief look at the one-period portfolio model that is implied by a copula model for the default times $\tau_1, \dots, \tau_m$. We fix a horizon T and set $Y_i := Y_{T,i}$. By definition, $Y_i = 1$ if and only if $\tau_i \leqslant T$, so the one-period model has the threshold representation $(\boldsymbol{\tau}, (T, \dots, T)')$. The dependence of the default events in the one-period model is governed by the dependence structure of the critical variables $\tau_1, \dots, \tau_m$. Note that in Section 11.1 this dependence was described in terms of the copula of the critical variables, whereas in this section we prefer to work with the survival copula of the default times.

Definition 12.1 immediately leads to an explicit construction of the random times $\tau_1, \dots, \tau_m$ via random thresholds; this in turn yields a generic simulation method for copula models. Consider a random vector $\boldsymbol{U} \sim C$ and define the default times by

$$\tau_i = \inf\{t \geqslant 0 \colon \mathrm{e}^{-\Gamma_i(t)} \leqslant U_i\}, \quad 1 \leqslant i \leqslant m, \tag{12.14}$$

so that firm i defaults at the first time point where the marginal survival function $\bar{F}_i(t)$ crosses the random threshold U_i. Equation (12.14) yields

$$\begin{aligned} P(\tau_1 > t_1, \dots, \tau_m > t_m) &= P(U_1 \leqslant \mathrm{e}^{-\Gamma_1(t_1)}, \dots, U_m \leqslant \mathrm{e}^{-\Gamma_m(t_m)}) \\ &= C(\mathrm{e}^{-\Gamma_1(t)}, \dots, \mathrm{e}^{-\Gamma_m(t)}), \end{aligned}$$

as required. To generate a realization of $\boldsymbol{\tau}$ we generate a realization of the random vector $\boldsymbol{U}$ and construct the τ_i according to (12.14). An alternative simulation algorithm for factor copula models is given at the end of this section.

Factor copula models. Most copula models used in practice have a factor structure. By this we mean that the threshold vector $\boldsymbol{U} \sim C$ used in (12.14) has a conditional independence structure in the sense of Definition 11.9, i.e. there is a p-dimensional random vector $\boldsymbol{V}$, $p < m$, such that the U_i are independent given $\boldsymbol{V}$.

The conditional independence of the U_i allows for an alternative representation of the joint survival function of the default times in terms of a mixture model. By conditioning on $\boldsymbol{V}$, it follows from (12.14) that we can write the joint survival function of $\boldsymbol{\tau}$ as

$$\begin{aligned} \bar{F}(t_1, \dots, t_m) &= E(P(U_1 \leqslant \bar{F}_1(t_1), \dots, U_m \leqslant \bar{F}_m(t_m) \mid \boldsymbol{V})) \\ &= E\biggl(\prod_{i=1}^{m} P(U_i \leqslant \bar{F}_i(t_i) \;\Big|\; \boldsymbol{V})\biggr). \end{aligned} \tag{12.15}$$

Moreover, by the definition of τ_i in (12.14) we have $P(\tau_i > t \mid \boldsymbol{V}) = P(U_i \leqslant \bar{F}_i(t) \mid \boldsymbol{V})$. The conditional survival function of τ_i given $\boldsymbol{V} = \boldsymbol{v}$ therefore satisfies

$$\bar{F}_{\tau_i|\boldsymbol{V}}(t \mid \boldsymbol{v}) = P(U_i \leqslant \bar{F}_i(t) \mid \boldsymbol{V} = \boldsymbol{v}),$$

and we may write $\bar{F}(t_1, \ldots, t_m) = E(\prod_{i=1}^m \bar{F}_{\tau_i|\boldsymbol{V}}(t_i \mid \boldsymbol{V}))$. Denoting the density or probability mass function of $\boldsymbol{V}$ by $g_{\boldsymbol{V}}$, we will usually write the joint survival function in the form

$$\bar{F}(t_1, \ldots, t_m) = \int_{\mathbb{R}^p} \prod_{i=1}^m \bar{F}_{\tau_i|\boldsymbol{V}}(t_i \mid \boldsymbol{v}) g_{\boldsymbol{V}}(\boldsymbol{v}) \, \mathrm{d}\boldsymbol{v}. \tag{12.16}$$

Note that the representation (12.16) is analogous to the representation of one-period threshold models with conditional independence structure as Bernoulli mixture models (see Section 11.2.4). In particular, (12.16) shows that for T fixed the default indicators follow a Bernoulli mixture model with factor vector $\boldsymbol{V}$ and conditional default probabilities $p_i(\boldsymbol{v}) = 1 - \bar{F}_{\tau_i|\boldsymbol{V}}(t \mid \boldsymbol{v})$.

The mixture-model representation of a factor copula model gives rise to the following generic algorithm for the simulation of default times.

Algorithm 12.2 (sampling from factor copula models).

(1) Generate a realization of $\boldsymbol{V}$.

(2) Generate independent rvs τ_i with df $1 - \bar{F}_{\tau_i|\boldsymbol{V}}(t \mid \boldsymbol{V})$, $1 \leqslant i \leqslant m$.

The importance-sampling techniques discussed in Section 11.4 in the context of one-period Bernoulli mixture models can be applied to improve the performance of Algorithm 12.2. These techniques are particularly useful if one is interested in rare events, such as in the pricing of CDO tranches with high attachment points.

12.2.2 Examples

In this section specific examples of factor copula models will be discussed. Any random vector with continuous marginal distributions and a p-dimensional conditional independence structure (see Definition 11.9) can be used to construct a factor copula model. Important examples in practice include factor copulas models based on the Gauss copula C_P^{Ga}, the LT-Archimedean copulas discussed in Section 7.4.2, and general one-factor copulas including the so-called double-t and double-NIG copulas.

Example 12.3 (one-factor Gauss copula). Factor copula models based on a Gauss copula C_P^{Ga} are frequently used in practice. We will compute the joint survival function for the default times in the one-factor case. Let

$$X_i = \sqrt{\rho_i} V + \sqrt{1 - \rho_i} \varepsilon_i, \tag{12.17}$$

where $\rho_i \in (0, 1)$ and where V and $(\varepsilon_i)_{1 \leqslant i \leqslant m}$ are iid standard normal rvs. The random vector $\boldsymbol{X} = (X_1, \ldots, X_m)'$ satisfies $\boldsymbol{X} \sim N_m(\boldsymbol{0}, P)$, where the (i, j)th element of P is given by $\rho_{ij} = \sqrt{\rho_i \rho_j}$. We set $U_i = \Phi(X_i)$ so that $\boldsymbol{U} = (U_1, \ldots, U_m)' \sim C_P^{\mathrm{Ga}}$. Both $\boldsymbol{X}$ and $\boldsymbol{U}$ have a one-factor conditional independence structure.

The conditional survival function of τ_i is easy to compute. Writing $d_i(t) := \Phi^{-1}(\bar{F}_i(t))$, we have that

$$\bar{F}_{\tau_i|V}(t \mid v) = P(U_i \leqslant \bar{F}_i(t) \mid V = v) = P\left(\varepsilon_i \leqslant \frac{d_i(t) - \sqrt{\rho_i} V}{\sqrt{1 - \rho_i}} \;\middle|\; V = v\right),$$

leading to $\bar{F}_{\tau_i|V}(t \mid v) = \Phi((d_i(t) - \sqrt{\rho_i}v)/(\sqrt{1-\rho_i}))$. Hence

$$\bar{F}(t_1, \ldots, t_m) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \prod_{i=1}^{m} \Phi\left(\frac{d_i(t_i) - \sqrt{\rho_i}v}{\sqrt{1-\rho_i}}\right) \mathrm{e}^{-v^2/2}\, \mathrm{d}v, \tag{12.18}$$

which is easily computed using one-dimensional numerical integration.

In applications of a one-factor Gauss copula model to the pricing of portfolio credit derivatives it is frequently assumed that $\rho_i = \rho$ for all i (the equicorrelation case). This model is known as the *exchangeable Gauss copula model*. In this case $\rho = \mathrm{corr}(X_i, X_j)$, so ρ is readily interpreted in terms of asset correlation. This makes the exchangeable Gauss copula model popular with practitioners. In fact, it is common practice on CDO markets to quote prices for tranches of synthetic CDOs in terms of *implied asset correlations* computed in an exchangeable Gauss copula model, as will be discussed in detail in Section 12.3.2.

Example 12.4 (general one-factor copula). This class of examples generalizes the construction of the one-factor Gauss copula in Example 12.3; some of the resulting models are useful in explaining prices of CDO tranches on credit indices. As in (12.17), one starts with random variables $X_i = \sqrt{\rho_i}V + \sqrt{1-\rho_i}\varepsilon_i$ for $\rho_i \in [0, 1]$ and *independent* rvs $V, \varepsilon_1, \ldots, \varepsilon_m$, but now V and the ε_i can have arbitrary continuous distributions (not necessary normal). The distribution F_{X_i} is then given by the convolution of the distributions of $\sqrt{\rho_i}V$ and of $\sqrt{1-\rho_i}\varepsilon_i$. The corresponding copula is the df of the random vector $\boldsymbol{U}$ with $U_i = F_{X_i}(X_i)$. A similar calculation to that in the case of the one-factor Gauss copula gives the following result for the conditional survival function:

$$\bar{F}_{\tau_i|V}(t \mid v) = F_\varepsilon\left(\frac{F_{X_i}^{-1}(\bar{F}_i(t)) - \sqrt{\rho_i}v}{\sqrt{1-\rho_i}}\right).$$

Popular examples include the so-called *double-t copula* (the case where V and ε_i follow a univariate t_ν distribution) and the *double-GH copula* (the case where V and ε_i both follow a univariate GH distribution as introduced in Section 6.2.3). We note that the double-t copula is not the same as the usual t copula since the rvs $V, \varepsilon_1, \ldots, \varepsilon_m$ do not have a multivariate t distribution (recall that uncorrelated bivariate t-distributed rvs are *not* independent, whereas V and ε_i are independent); the situation is similar for the double-GH copula.

The main computational challenge in working with the double-t and double-GH copulas is the determination of the convolution of V and ε_i and the corresponding quantile function $F_{X_i}^{-1}$. There exist a number of good solutions to this problem; references are given in Notes and Comments.

Example 12.5 (LT-Archimedean copulas). Consider a positive rv V with df G_V such that $G_V(0) = 0$ and denote by $\hat{G}_V$ the Laplace–Stieltjes transform of G_V. According to Definition 7.52 the associated LT-Archimedean copula is given by the formula

$$C(u_1, \ldots, u_m) = E\left(\exp\left(-V \sum_{i=1}^{m} \hat{G}_V^{-1}(u_i)\right)\right).$$

As usual, denote by $\bar{F}_i(t_i)$ the marginal survival function of τ_i. Using (12.13) the joint survival function of $\boldsymbol{\tau}$ can be calculated to be

$$\bar{F}(t_1, \ldots, t_m) = E\left(\prod_{i=1}^{m} \exp\{-V \hat{G}_V^{-1}(\bar{F}_i(t_i))\} \right), \tag{12.19}$$

which is obviously of the general form (12.16). Recall that in the special case of the Clayton copula with parameter θ we have $V \sim \mathrm{Ga}(1/\theta, 1)$; explicit formulas for $\hat{G}_V$ and $\hat{G}_V^{-1}$ in that case are given in Algorithm 7.53 for the simulation of LT-Archimedean copulas. Note that LT-Archimedean copulas are in general not radially symmetric, so the one-period version of the copula model for default times with survival function (12.19) is *not* the same as the threshold model with Archimedean copula presented in Example 11.4; it corresponds instead to a threshold model with Archimedean survival copula.

Notes and Comments

The first copula model for portfolio credit risk was given by Li (2000); his model is based on the Gauss copula. General copula models were introduced for the first time in Schönbucher and Schubert (2001). One of the first papers on factor copula models was Laurent and Gregory (2005). These models have received a lot of attention, mostly in connection with implied correlation skews (see also Section 12.3 below). The double-t copula was proposed by Hull and White (2004); numerical aspects of this copula model are discussed in Vrins (2009). Double-GH copulas are considered by Kalemanova, Schmid and Werner (2005), Guégan and Houdain (2005) and Eberlein, Frey and von Hammerstein (2008), among others.

12.3 Pricing of Index Derivatives in Factor Copula Models

In this section we are interested in the pricing of index derivatives (index swaps and CDOs) in factor copula models. These models are the market standard for dealing with CDO tranches. We begin with analytical and numerical methods for the valuation of index derivatives. In Section 12.3.2 we turn to qualitative properties of observed CDO spreads and discuss the well-known correlation skews. In Section 12.3.3 we consider an alternative factor model, known as an implied copula model, that has been developed in order to explain correlation skews.

12.3.1 Analytics

Throughout we consider a factor copula model with factor-dependent recovery risk. More precisely, we assume that under the risk-neutral pricing measure Q the survival function of the default times follows a factor copula model with mixture representation (12.16). Moreover, the loss given default of firm i is state dependent, so, given a realization of the factor $\boldsymbol{V} = \boldsymbol{v}$, the loss given default is of the form $\delta_i(\boldsymbol{v})$ for some function $\delta_i : \mathbb{R}^p \to (0, 1)$. A state-dependent loss given default provides extra flexibility for calibrating the model. Consider, for example, the case where V is a one-dimensional rv and where the conditional default probabilities are increasing in

V. By taking increasing functions $\delta_i(v)$ it is possible to model negative correlation between default probabilities and recovery rates. This extension is reasonable from an empirical viewpoint, as was discussed in Section 11.2.3.

Index swaps. We begin by computing the value of the default and the premium payment leg for index swaps. Recall that $V^{\mathrm{Def}}(x) = \sum_{i=1}^m V_i^{\mathrm{Def}}$ and $V^{\mathrm{Prem}}(x) = \sum_{i=1}^m V_i^{\mathrm{Prem}}(x)$, where V_i^{Def} and $V_i^{\mathrm{Prem}}(x)$ represent the value of the default and of the premium payment leg of the single-name CDS on obligor i. Moreover, V_i^{Def} and $V_i^{\mathrm{Prem}}(x)$ depend only on the loss process of firm i, which is given by $L_{t,i} = \delta_i Y_{t,i}$. The value of an index swap does not therefore actually depend on the copula of the default times; this is in contrast to the single-tranche CDO analysed below.

We now turn to the actual computations. If the recovery rates are constant, V_i^{Def} and $V_i^{\mathrm{Prem}}(x)$ can be computed using the results on the pricing of single-name CDSs in hazard rate models discussed in Section 10.4.4. In the general case one may proceed as follows. First, similar reasoning to that applied in (12.11) gives

$$V^{\mathrm{Def}} = E^Q\bigg(\int_0^T p_0(0,t)\,\mathrm{d}L_t\bigg) \approx \sum_{n=1}^N p_0(0,t_n)(E^Q(L_{t_n}) - E^Q(L_{t_{n-1}})). \tag{12.20}$$

Since $L_{t_n} = \sum_{i=1}^m \delta_i Y_{t_n,i}$, we find, by conditioning on $\boldsymbol{V}$, that

$$E^Q(L_{t_n}) = \int_{\mathbb{R}^p} \sum_{i=1}^m \delta_i(\boldsymbol{v}) E^Q(Y_{t_n,i} \mid \boldsymbol{V} = \boldsymbol{v}) g_{\boldsymbol{V}}(\boldsymbol{v})\,\mathrm{d}\boldsymbol{v}.$$

Now $Y_{t_n,i}$ is Bernoulli with $Q(Y_{t_n,i} = 1 \mid \boldsymbol{V} = \boldsymbol{v}) = 1 - \bar{F}_{\tau_i|\boldsymbol{V}}(t_n \mid \boldsymbol{v})$. Hence,

$$E^Q(L_{t_n}) = \int_{\mathbb{R}^p} \sum_{i=1}^m \delta_i(\boldsymbol{v})(1 - \bar{F}_{\tau_i|\boldsymbol{V}}(t_n \mid \boldsymbol{v})) g_{\boldsymbol{V}}(\boldsymbol{v})\,\mathrm{d}\boldsymbol{v},$$

and substitution of this expression into (12.20) gives the value of the default payment leg. Similar reasoning allows us to conclude that $V^{\mathrm{Prem}}(x)$, the value of the premium payments for a generic spread x, is equal to

$$V^{\mathrm{Prem}}(x) = x \sum_{n=1}^N p_0(0,t_n) \int_{\mathbb{R}^p} \sum_{i=1}^m \bar{F}_{\tau_i|\boldsymbol{V}}(t_n \mid \boldsymbol{v}) g_{\boldsymbol{V}}(\boldsymbol{v})\,\mathrm{d}\boldsymbol{v}. \tag{12.21}$$

In both cases the integration over $\boldsymbol{v}$ can be carried out with numerical quadrature methods. This is straightforward in the cases that are most relevant in practice, where V is a one-dimensional rv.

CDO tranches. Recall from Section 12.1.3 that the essential task in computing CDO spreads is to determine the price of call or put options on L_{t_n}. By a conditioning argument we obtain, for a generic function $f\colon \mathbb{R}^+ \to \mathbb{R}$ such as the pay-off of a call or put option, the formula

$$E^Q(f(L_{t_n})) = \int_{\mathbb{R}^p} E^Q\bigg(f\bigg(\sum_{i=1}^m \delta_i(\boldsymbol{v}) Y_{t_n,i} \;\bigg|\; \boldsymbol{V} = \boldsymbol{v}\bigg)\bigg) g_{\boldsymbol{V}}(\boldsymbol{v})\,\mathrm{d}\boldsymbol{v}.$$

Table 12.3. Fair tranche spreads computed in an exchangeable Gauss copula model using the LPA and the simple normal approximation. The exact spread value is computed by an extensive Monte Carlo approximation. The row labelled "asset correlation" reports the value of the correlation parameter that is used in the pricing of the tranche. These correlations are so-called implied tranche correlations (see Table 12.4). The normal approximation is a substantial improvement over the LPA for the tranches where asset correlation is low, such as the [3, 6] and [6, 9] tranches; for the other tranches both approximation methods perform reasonably well. The reason for this is the fact that for high asset correlation the loss distribution is essentially determined by the realization of V, whereas for low asset correlation the form of the conditional distribution of L_t matters. Numerical values are taken from Frey, Popp and Weber (2008).

Tranche	[0, 3]	[3, 6]	[6, 9]	[9, 12]	[12, 22]
Asset correlation	$\rho = 0.219$	$\rho = 0.042$	$\rho = 0.148$	$\rho = 0.223$	$\rho = 0.305$
LPA (%)	30.66	0.79	0.53	0.36	0.18
Normal approximation (%)	29.38	1.51	0.66	0.42	0.20
Exact spread (%)	28.38	1.55	0.68	0.42	0.20

To discuss this formula we fix t_n and introduce the simpler notation $q_i(\boldsymbol{v}) := Q(Y_{t_n,i} = 1 \mid \boldsymbol{V} = \boldsymbol{v})$. Since the rvs $Y_{t_n,i}$, $1 \leqslant i \leqslant m$, are conditionally independent given $\boldsymbol{V}$, computing the inner conditional expectation essentially amounts to finding the distribution of a sum of independent Bernoulli rvs. This is fairly straightforward for a homogeneous portfolio, where the LGD functions and the conditional default probabilities for all firms are equal, so that $q_i(\boldsymbol{v}) \equiv q(\boldsymbol{v})$ and $\delta_i(\boldsymbol{v}) \equiv \delta(\boldsymbol{v})$. In that case, $L_{t_n} = \delta(V)M_{t_n}$, where $M_t = \sum_{i=1}^m Y_{t,i}$ gives the number of firms that have defaulted up to time t. Given $\boldsymbol{V} = \boldsymbol{v}$, M_{t_n} is the sum of independent and (due to the homogeneity) identically distributed Bernoulli rvs, so M_{t_n} has a binomial distribution with parameters m and $q(\boldsymbol{v})$.

In a heterogeneous portfolio, on the other hand, the evaluation of the conditional loss distribution is not straightforward, essentially because one is dealing with sums of independent Bernoulli rvs that are *not* identically distributed. We discuss several approaches to this problem in the remainder of this section.

Large-portfolio approximation. In the so-called large-portfolio approximation (LPA), the conditional distribution of L_{t_n} given $\boldsymbol{V} = \boldsymbol{v}$ is replaced by a point mass at the conditional mean $\ell(\boldsymbol{v}) = \sum_{i=1}^m \delta_i(\boldsymbol{v})q_i(\boldsymbol{v})$. This leads to the approximation

$$E^Q(f(L_{t_n})) \approx \int_{\mathbb{R}^p} f(\ell(\boldsymbol{v}))g_V(\boldsymbol{v})\,\mathrm{d}\boldsymbol{v}. \tag{12.22}$$

The method is motivated by the asymptotic results of Section 11.3. In particular, in that section we have shown that, in the case of one-factor models, quantiles of the portfolio loss distribution can be well approximated by quantiles of the rv $\ell(V)$, provided that the portfolio is sufficiently large and not too inhomogeneous. One would therefore expect that for typical credit portfolios the approximation (12.22) performs reasonably well. Numerical results on the performance of the LPA are given in Table 12.3.

Normal and Poisson approximation. The LPA completely ignores the impact of the fluctuations of L_{t_n} around its conditional mean. One way of capturing these fluctuations is to replace the unknown conditional distribution of L_{t_n} with a known distribution with similar moments. In a simple normal approximation the conditional distribution of L_{t_n} is approximated by a normal distribution with mean $\ell(\boldsymbol{v})$ and variance

$$\sigma^2(\boldsymbol{v}) := \operatorname{var}(L_{t_n} \mid \boldsymbol{V} = \boldsymbol{v}) = \sum_{i=1}^{m} \delta_i^2(\boldsymbol{v}) q_i(\boldsymbol{v})(1 - q_i(\boldsymbol{v}))$$

(the choice of the normal distribution is of course motivated by the central limit theorem). Under the normal approximation one has

$$E^Q(f(L_{t_n})) \approx \int_{\mathbb{R}^p} \frac{1}{\sqrt{2\pi\sigma^2(\boldsymbol{v})}} \int_{\mathbb{R}} f(l) \exp\left(-\frac{(l - \ell(\boldsymbol{v}))^2}{2\sigma^2(\boldsymbol{v})}\right) \mathrm{d}l \, g_V(\boldsymbol{v}) \, \mathrm{d}\boldsymbol{v}.$$

The performances of the normal approximation and the LPA are illustrated in Table 12.3 for the case of an exchangeable Gauss copula model.

In a similar spirit, in a Poisson approximation the conditional loss distribution is approximated by a Poisson distribution with parameter $\lambda = \ell(\boldsymbol{v})$. These simple normal and Poisson approximation methods can be refined substantially (see El Karoui, Kurtz and Jiao (2008) for details). There are a number of other techniques for evaluating the conditional loss distribution. On the one hand, one may apply Fourier or Laplace inversion to the problem; on the other hand, there are a number of recursive techniques that can be used. References for both approaches can be found in Notes and Comments.

12.3.2 Correlation Skews

Investors typically express CDO spreads in terms of implied correlations computed in a simple homogeneous Gauss copula model known as the *benchmark model*. In this way tranche spreads can be compared across attachment points and maturities. This practice is similar to the use of implied volatilities as a common yardstick for comparing prices of equity and currency options. In this section we explain this quoting convention in more detail and discuss qualitative properties of market-observed implied correlations.

The benchmark model is an exchangeable Gauss copula model, as in Example 12.3. It is assumed that all default times are exponentially distributed with $Q(\tau_i > t) = \mathrm{e}^{-\gamma^Q t}$ for all i and that the LGDs are deterministic and equal to 60% for all firms. The hazard rate γ^Q is used to calibrate the model to the spread of the index swap with the same maturity as the tranche under consideration; see the discussion of calibration of a homogeneous model to index swaps in Section 12.1.3. The only free parameter in the benchmark model is the "asset correlation" ρ. This parameter can be used to calibrate the model to tranche spreads observed in the market, which leads to various notions of implied correlation.

Implied correlations. The market uses two notions of implied correlation in order to describe market-observed tranche spreads. *Implied tranche correlation*, also known as *compound correlation*, is the lowest number $\rho \in [0, 1]$ such that the spread computed in the benchmark model equals the spread observed in the market. For mezzanine tranches the notion of implied tranche correlation suffers from two problems. First, a correlation number ρ such that the model spread matches the market spread does not always exist; this problem was encountered frequently for the spreads that were quoted in the period 2008–9 (during the financial crisis). Second, for certain values of the market spread there is more than one matching value of $\rho \in [0, 1]$, and the convention to quote the lowest such number is arbitrary.

If market spreads are available for a complete set of tranches, as is the case for the standardized tranches on iTraxx and CDX, the so-called *base correlation* is frequently used as an alternative means of expressing tranche spreads. To compute base correlation we recursively compute implied correlations for a nested set of equity tranches so that they are consistent with the observed tranche spreads. The base correlation methodology is popular since it mitigates the calibration problems arising in the computation of implied tranche correlation. On the other hand, base correlations are more difficult to interpret and there may be theoretical inconsistencies, as we explain below. In the following algorithm the computation of base correlation is described in more detail.

Algorithm 12.6. Suppose that we observe market spreads x_κ^*, $\kappa = 1, \dots, K$, for a set of CDO tranches with attachment points (l_κ, u_κ), $\kappa = 1, \dots, k$, with $l_1 = 0$ and $l_{\kappa+1} = u_\kappa$. Denote by $V^{\mathrm{Prem}}(u, \rho, x)$ the value of the premium payment leg of an equity tranche with attachment point u and spread $x > 0$ in the benchmark model with correlation parameter $\rho \in [0, 1]$. Denote by $V^{\mathrm{Def}}(u, \rho)$ the value of the corresponding default payment leg. We carry out the following steps.

(1) Compute the base correlation ρ_1 as the implied tranche correlation of the $[0, u_1]$ tranche using the observed spread x_1. Set $\kappa = 1$.

(2) Given the base correlation ρ_κ and the spread $x_{\kappa+1}^*$ (the market spread of the $[l_{\kappa+1}, u_{\kappa+1}]$ tranche), compute the base correlation $\rho_{\kappa+1}$ as the solution of the following equation (an explanation is given after the algorithm):

$$V^{\mathrm{Def}}(u_{\kappa+1}, \rho_{\kappa+1}) - V^{\mathrm{Def}}(u_\kappa, \rho_\kappa) = V^{\mathrm{Prem}}(u_{\kappa+1}, \rho_{\kappa+1}, x_{\kappa+1}^*) - V^{\mathrm{Prem}}(u_\kappa, \rho_\kappa, x_{\kappa+1}^*). \quad (12.23)$$

(3) Replace κ with $\kappa + 1$ and return to step (2).

Equation (12.23) is derived from the following considerations. It is the basic premise of the base correlation approach that for all $1 \leqslant \kappa \leqslant K$ and for any spread x, the correct values of the default and premium legs of an equity tranche with upper attachment point u_κ are given by $V^{\mathrm{Def}}(u_\kappa, \rho_\kappa)$ and $V^{\mathrm{Prem}}(u_\kappa, \rho_\kappa, x)$, respectively. Under this assumption, the left-hand side of (12.23) gives the value of the default payment leg of the $[l_{\kappa+1}, u_{\kappa+1}]$ tranche and the right-hand side gives the value of the corresponding premium payment leg at the spread $x_{\kappa+1}^*$ observed in the market.

For the observed spread $x^*_{\kappa+1}$, the value of the default and of the premium payment leg of the $[l_{\kappa+1}, u_{\kappa+1}]$ tranche have to be equal, which leads to equation (12.23) for $\rho_{\kappa+1}$.

Note that, under the base-correlation methodology, two different models are used to derive the model value of the $[l_\kappa, u_\kappa]$ tranche: the benchmark model with correlation parameter ρ_κ is used for the upper attachment point and the benchmark model with correlation parameter $\rho_{\kappa-1}$ is used for the lower attachment point. As shown in Brigo, Pallavicini and Torresetti (2010), this may lead to inconsistencies if the base correlation approach is employed in the pricing of tranches with non-standard attachment points.

Empirical properties of implied correlations. In Table 12.4 we give market spreads and implied tranche correlations for iTraxx tranches for three different days in 2004, 2006 and 2008. These numbers are representative of three different periods in the credit index market: the early days of the market; before the credit crisis; and during the credit crisis. We see that implied correlations are quite unstable. In particular, a standard Gauss copula with a fixed correlation parameter ρ cannot explain all tranche spreads simultaneously (otherwise the implied correlation curves would be flat).

Before the financial crisis, implied correlations exhibited a typical form that became known as an *implied correlation skew*: implied tranche correlations showed a V-shaped relationship to attachment point; implied base correlations were strictly increasing; and implied correlations for senior tranches were comparatively high. The 2004 and 2006 rows of the table are typical of the behaviour of implied tranche correlations (see also Figure 12.3).

With the onset of the financial crisis, tranche spreads and implied tranche correlations became quite irregular. In particular, for mezzanine tranches it was often impossible to find a correlation number that would reproduce the spread that was observed on that day. For a more detailed description of the behaviour of implied correlation we refer the reader to the book by Brigo, Pallavicini and Torresetti (2010).

It can be argued that correlation skews reflect deficiencies of the benchmark model. To begin with, the Gauss copula is an ad hoc choice that is motivated mostly by analytical convenience and not by thorough data analysis. Moreover, it is highly unlikely that the dependence structure of the default times in a portfolio can be characterized by a single correlation number. Also, the assumption of constant recovery rates is at odds with reality.

Explaining observed CDO spreads in factor copula models. The fact that the benchmark model cannot reproduce observed CDO spreads for several tranches creates problems for the pricing of so-called *bespoke* tranches with non-standard maturities or attachment points and for the risk management of a book of tranches. In these applications one would like to take all available price information into account. There is a need for portfolio credit risk models that can be made consistent with observed spreads for several tranches simultaneously.

Table 12.4. Market quotes and implied tranche correlations for five-year tranches on the iTraxx Europe index on different dates. Note that the spread for the equity tranche corresponds to an upfront payment quoted as a percentage of the notional; the quarterly spread for the equity tranche is set to 5% by market convention.

Type of data	Year	Index	[0,3]	[3,6]	[6,9]	[9,12]	[12,22]
Market spread	2004	42 bp	27.6%	168 bp	70 bp	43 bp	20 bp
Tranche correlation	2004		22.4%	5.0%	15.3%	22.6%	30.6%
Market spread	2006	26 bp	14.5%	62 bp	18 bp	7 bp	3 bp
Tranche correlation	2006		18.6%	7.9%	14.1%	17.25%	23.54%
Market spread	2008	150 bp	46.5%	5.7%	3.7%	2.3%	1.45%
Tranche correlation	2008		51.1%	85.7%	95.3%	3.0%	17.6%

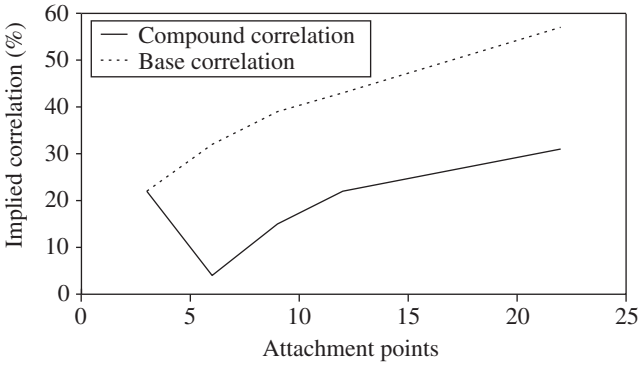

Figure 12.3. Compound correlation and base correlation corresponding to the CDO spreads in the 2004 row of Table 12.4. The spreads exhibit a typical correlation skew. (Data from Hull and White (2004).)

Quite naturally, many researchers have addressed this problem by considering more general factor copula models where some of the unrealistic assumptions made in the benchmark model are relaxed. To begin with, several authors have shown that by introducing state-dependent recovery rates that are negatively correlated with default probabilities, the correlation skew can be mitigated (but not eliminated completely). More importantly, alternative copula models have been employed, mostly models from the class of general one-factor copulas introduced in Example 12.4. It turns out that the double-t copula and, in particular, the double-GH copula model can be calibrated reasonably well to market-observed CDO tranche spreads (see, for example, Eberlein, Frey and von Hammerstein 2008). However, the empirical performance of these models worsened substantially with the onset of the credit crisis in 2007. For further details we refer to the references given in Notes and Comments.

12.3.3 The Implied Copula Approach

The class of implied copula models can be viewed as a generalization of the one-factor copula models considered in Section 12.2.2. In an implied copula model the factor variable V is modelled as a discrete rv whose probability mass function

is determined by calibration to market data such as observed index and tranche spreads. Since the support of V is usually taken as a fairly large set, this creates some additional flexibility that helps to explain observed market spreads for CDO tranches.

Definition and key properties. The structure of the joint survival function of $\tau_1, \dots, \tau_m$ in an implied copula model is similar to the mixture representation (12.16) in a factor copula model. It is assumed that the τ_i are conditionally independent given a mixing variable V that takes the discrete values $v_1, \dots, v_K$ (the *states* of the system); the conditional survival probabilities are of the form

$$Q(\tau_i > t \mid V = v_k) = \exp(-\lambda_i(v_k)t), \quad 1 \leqslant i \leqslant m,\ t \geqslant 0, \tag{12.24}$$

for functions $\lambda_i : \{v_1, \dots, v_K\} \to (0, \infty)$. The probability mass function of V is denoted by $\boldsymbol{\pi} = (\pi_1, \dots, \pi_k)$ with $\pi_k = Q(V = v_k), k = 1, \dots, K$. It follows that the joint survival function of $\tau_1, \dots, \tau_m$ is given by

$$\bar{F}(t_1, \dots, t_m) = \sum_{k=1}^{K} \pi_k \left\{ \prod_{i=1}^{m} \exp(-\lambda_i(v_k)t_i) \right\}. \tag{12.25}$$

It is possible to make the functions $\lambda_i(\cdot)$ in (12.24) time dependent, and this is in fact necessary if the model is to be calibrated simultaneously to tranche spreads of different maturities. We will describe the time-independent case for ease of exposition.

Note that the name "implied copula model" is somewhat misleading, since both the dependence structure and the marginal distributions of the τ_i change if the distribution of V is changed. For instance, for $\boldsymbol{\pi} = (1, 0, \dots, 0)$ we have $Q(\tau_i > t) = \exp(-\lambda_i(v_1)t)$, whereas for $\boldsymbol{\pi} = (0, \dots, 0, 1)$ we have $Q(\tau_i > t) = \exp(-\lambda_i(v_K)t)$. A better name for (12.25) would therefore be "a model with an implied factor distribution", but the label "implied copula model" has become standard, essentially because of the influential paper by Hull and White (2006).

In order to compute spreads of CDSs, index swaps or single-tranche CDOs in an implied copula model, we proceed in a similar fashion to the factor copula models discussed in the previous section. First we compute the conditional values of the default payment leg and the premium payment leg given $V = v_k$, denoted by $V^{\mathrm{Def}}(v_k)$ and $V^{\mathrm{Prem}}(v_k, x)$, where x represents a generic spread level; the methods described in Section 12.3.1 can be used for this purpose. The unconditional values of the default payment and the premium payment legs are then given by averaging over the different states: that is,

$$V^{\mathrm{Def}} = \sum_{k=1}^{K} \pi_k V^{\mathrm{Def}}(v_k) \quad \text{and} \quad V^{\mathrm{Prem}}(x) = \sum_{k=1}^{K} \pi_k V^{\mathrm{Prem}}(v_k, x).$$

The fair spread of the structure is equal to $x^* = V^{\mathrm{Def}} / V^{\mathrm{Prem}}(1)$.

Implied copula models are mostly used for the pricing of CDO tranches with non-standard attachment points or maturities. It is also possible to value tranches where

the underlying pool is different from a standard credit index, but the valuation of such products is somewhat subjective; we refer to the references given in Notes and Comments. In Section 17.4.3 we explain how to embed an implied copula model into a dynamic portfolio credit risk model. In that extension of the model it is possible to price options on standard index products such as CDS index swaps.

Example 12.7 (implied copula models for a homogeneous portfolio). In a homogeneous portfolio the default intensities of all firms are identical, $\lambda_i(\cdot) = \lambda(\cdot)$, and we may parametrize the model directly in terms of the values $\lambda_1, \dots, \lambda_K$ of the default intensity in the different states. Moreover, the states can be ordered so that $\lambda_1 < \lambda_2 < \cdots < \lambda_K$. In this way, state 1 can be viewed as the best state of the economy (the one with the lowest default intensity) and state K corresponds to the worst state. In our calibration example below we use a model with $K = 9$ states and we assume that the default intensity takes values in the set $\{0.01\%, 0.3\%, 0.6\%, 1.2\%, 2.5\%, 4.0\%, 8.0\%, 20\%, 70\%\}$.

For heterogeneous portfolios we require more sophisticated parametrizations. An example can be found in Rosen and Saunders (2009); further references are given in Notes and Comments.

In practical implementations of the model the set of states $\{v_1, \dots, v_K\}$ is typically chosen at the outset of the analysis and is kept fixed during model calibration. Moreover, in order to obtain pricing results for bespoke index products that are robust with respect to the precise values of the v_k it is advisable to work on a fine grid, and hence with a fairly large number of states K. Of course, this means that the determination of the probability mass function $\boldsymbol{\pi}$ becomes a high-dimensional problem, but the calibration of $\boldsymbol{\pi}$ is comparatively easy, as we now explain.

Calibration of $\boldsymbol{\pi}$. In the implied copula framework we need to determine the probability mass function $\boldsymbol{\pi}$ from observed market data. As is usual in model calibration, $\boldsymbol{\pi}$ is found by minimizing some distance between market prices and model prices. This is substantially facilitated by the observation that model prices are linear in $\boldsymbol{\pi}$. The set of all probability mass functions consistent with the price information at a given point in time t can therefore be described in terms of a set of linear inequalities. We now explain this for the case of a corporate bond and a single-name CDS; similar arguments apply for index swaps and CDO tranches.

- Consider a zero-coupon bond issued by firm i with maturity T, and denote by $p_i(v_k) = \exp(-\lambda_i(v_k)(T-t))$ the value of the bond in state v_k (we assume a recovery rate equal to zero for simplicity). Suppose that we observe bid and ask quotes $\underline{p} \leqslant \overline{p}$ for the bond. In order to be consistent with this information, a probability mass function $\boldsymbol{\pi}$ needs to satisfy the following linear inequalities:

$$\underline{p} \leqslant \sum_{k=1}^{K} \pi_k p_i(v_k) \leqslant \overline{p}.$$

- Consider next a CDS contract on firm i; this is a simple example of a contract where two cash-flow streams are exchanged. Suppose that at time t we

observe bid and ask spreads $\underline{x} \leqslant \overline{x}$ for the contract. Denote by $V_i^{\mathrm{Def}}(v_k)$ and $V_i^{\mathrm{Def}}(v_k, x)$ the values of the premium payment and default payment legs of the contract in state k for a generic CDS spread x. Then $\boldsymbol{\pi}$ must satisfy the following two inequalities:

$$\sum_{k=1}^{K} \pi_k (V_i^{\mathrm{prem}}(v_k, \underline{x}) - V_i^{\mathrm{def}}(v_k)) \leqslant 0,$$
$$\sum_{k=1}^{K} \pi_k (V_i^{\mathrm{prem}}(v_k, \overline{x}) - V_i^{\mathrm{def}}(k)) \geqslant 0.$$

Moreover, $\boldsymbol{\pi}$ needs to satisfy the obvious linear constraints $\pi_k \geqslant 0$ for all k and $\sum_{k=1}^{K} \pi_k = 1$.

It follows from the above discussion that the constraints on $\boldsymbol{\pi}$ at time t can be written in the generic form

$$A\boldsymbol{\pi} \leqslant b$$

for some matrix $A \in \mathbb{R}^{N \times K}$, some vector $b \in \mathbb{R}^N$ and some $N \in \mathbb{N}$. In order to find a vector $\boldsymbol{\pi}$ that satisfies this system of linear inequalities, fix a vector $\boldsymbol{c} = (c_1, \ldots, c_K)$ of weights and consider the linear programming problem

$$\min_{\boldsymbol{\pi} \in \mathbb{R}^K} \sum_{i=1}^{K} c_k \pi_k \quad \text{subject to } A\boldsymbol{\pi} \leqslant b. \tag{12.26}$$

Clearly, every solution of (12.26) is a probability mass function that is consistent with the given price information in the sense that it solves the system $A\boldsymbol{\pi} \leqslant b$. Note that problem (12.26) can be solved with standard linear programming software.

If the number of states K is large compared with the number of constraints, there will typically be more than one probability mass function $\boldsymbol{\pi}$ that solves the system $A\boldsymbol{\pi} \leqslant b$. An easy way to check this is to vary the weight vector $\boldsymbol{c}$ in (12.26), since different weight vectors usually correspond to different solutions of the system $A\boldsymbol{\pi} \leqslant b$. In that case a unique solution $\boldsymbol{\pi}^*$ of the calibration problem can be determined by a suitable *regularization procedure*. For instance, one could choose $\boldsymbol{\pi}^*$ by minimizing the *relative entropy* to the uniform distribution on the set $\{v_1, \ldots, v_K\}$. This leads to the convex optimization problem

$$\min_{\boldsymbol{\pi} \in \mathbb{R}^K} \sum_{i=1}^{K} \pi_k \ln \pi_k \quad \text{subject to } A\boldsymbol{\pi} \leqslant b, \tag{12.27}$$

which can be addressed with standard convex programming algorithms (see, for example, Bertsimas and Tsitsiklis 1997). Model calibration via entropy minimization has a number of attractive features, as is explained, for example, in Avellaneda (1998). In particular, the prices of bespoke tranches (the model output) depend smoothly on the spreads observed on the market (the model input).

We close this section on implied copula models by presenting results from a calibration exercise of Frey and Schmidt (2011). In that paper the homogeneous

Table 12.5. Results of the calibration of a homogeneous implied copula model to iTraxx spread data (index and tranches) for different data sets from several years; the components of $\boldsymbol{\pi}^*$ are expressed as percentages. The numerical results are from Frey and Schmidt (2011).

	λ (in %)								
	0.01	0.3	0.6	1.2	2.5	4.0	8.0	20	70
$\boldsymbol{\pi}^*$, data from 2004	12.6	22.9	42.0	17.6	2.5	1.45	0.54	0.13	0.03
$\boldsymbol{\pi}^*$, data from 2006	22.2	29.9	39.0	7.6	1.2	0.16	0.03	0.03	0.05
$\boldsymbol{\pi}^*$, data from 2008	1.1	7.9	57.6	10.8	11.7	4.9	1.26	1.79	2.60
$\boldsymbol{\pi}^*$, data from 2009	0.0	13.6	6.35	42.2	22.3	12.5	0.0	0.00	3.06

model of Example 12.7 was calibrated to iTraxx tranche and index spread data for the years 2004, 2006, 2008 and 2009; all contracts had a maturity of five years. The data from 2004 and 2006 are typical for tranche and index spreads before the credit crisis; the data from 2008 and 2009 represent the state of the market during the crisis. Entropy minimization was used in order to determine a solution $\boldsymbol{\pi}^*$ of the calibration problem. The resulting values of $\boldsymbol{\pi}^*$ are given in Table 12.5. We clearly see that with the emergence of the credit crisis the calibration procedure puts more mass on states where the default intensity is high; in particular, the extreme state where $\lambda = 70\%$ gets a probability of around 3%. This reflects the increased awareness of future defaults and the increasing risk aversion in the market after the beginning of the crisis.

The implied copula model can usually be calibrated very well to tranche and index spreads with a single maturity; calibrating tranche spreads for several maturities simultaneously is more involved (see, for example, Brigo, Pallavicini and Torresetti (2010) for details).

Notes and Comments

Semianalytic approaches for the pricing of synthetic CDOs in factor copula models have been developed by Laurent and Gregory (2005), Hull and White (2004), Gibson (2004) and Andersen and Sidenius (2004), among others. Laurent and Gregory exploit the conditional independence structure of factor copula models and develop methods based on Fourier analysis; Andersen and Sidenius, Gibson and Hull and White propose recursive methods. The LPA is originally due to Vasicek (1997). Frey, Popp and Weber (2008) propose the normal approximation as a simple yet efficient alternative to the LPA. A very deep study of normal and Poisson approximations for the sum of independent random variables with applications to CDO pricing is El Karoui, Kurtz and Jiao (2008).

There is a rich literature on general one-factor copula models in relation to implied correlation skews: the double-t copula has been studied in Hull and White (2004) and Vrins (2009); double-GH copulas have been analysed by Eberlein, Frey and von Hammerstein (2008), Guégan and Houdain (2005) and Kalemanova, Schmid and Werner (2005), among others. The idea of using state-dependent recovery rates to improve the fit of CDO pricing models is explored, for instance, in Hull and

White (2006). The base-correlation approach is used in a number of (fairly dubious) extrapolation procedures for the pricing of non-standard tranches. An in-depth discussion of implied correlation skews in credit risk can be found in Brigo, Pallavicini and Torresetti (2010).

The implied copula model is due to Hull and White (2006); similar ideas can be found in Rosen and Saunders (2009) and in Frey and Schmidt (2012). Hull and White (2006) and Rosen and Saunders (2009) discuss the pricing of bespoke CDO tranches in the context of the model. The calibration of implied copula models to inhomogeneous portfolios is discussed in detail in Rosen and Saunders (2009) and in Frey and Schmidt (2012). An empirical assessment of the calibration properties of implied copula models is also given in Brigo, Pallavicini and Torresetti (2010). Calibration methods based on entropy minimization are discussed by Avellaneda (1998).

13
Operational Risk and Insurance Analytics

We have so far concentrated on the modelling of market and credit risk, which reflects the historical development of quantitative risk management in the banking context. Some of the techniques we have discussed are also relevant in operational risk modelling, particularly the statistical models of extreme value theory (EVT) in Chapter 5 and the aggregation methodology of Chapter 8. But we also need other techniques tailored specifically to operational risk, and we believe that actuarial models used in non-life insurance are particularly relevant.

In the first half of this chapter (Section 13.1) we examine the Basel requirements for the quantitative modelling of operational risk in banks, discussing various potential approaches. On the basis of industry data we highlight the challenges involved in implementing a so-called advanced measurement (AM) approach based on modelling loss distributions, also known as the loss distribution approach (LDA).

In operational risk there is no consensus concerning the best modelling approach. In contrast to market and credit risk, the data sources are more limited and the overall statistical properties of available data show a high degree of non-homogeneity and non-stationarity, defying straightforward applications of statistical tools. Rather than offering specific models, the current chapter aims to provide a set of tools that can be used to learn more about this important but difficult-to-model risk category.

In Section 13.2 we summarize the techniques from actuarial modelling that are relevant to operational risk, under the heading of *insurance analytics*. Our discussion in that section, though motivated by quantitative modelling of operational risk, has much wider applicability in quantitative risk management. For example, some techniques have implicitly been used in the credit risk chapters. The Notes and Comments section at the end of the chapter gives an overview of further techniques from insurance mathematics that have potential applications in the broader field of quantitative risk management.

13.1 Operational Risk in Perspective

13.1.1 An Important Risk Class

In our overview of the development of the Basel regulatory framework in Section 1.2.2 we explained how *operational risk* was introduced under Basel II as a new risk class for which financial institutions were bound to set aside regulatory capital. It has also been incorporated into the Solvency II framework for insurers, although we will concentrate on the banking treatment in this chapter.

We first recall the Basel definition as it appears in the final comprehensive version of the Basel II document (Basel Committee on Banking Supervision 2006).

> Operational risk is defined as the risk of loss resulting from inadequate or failed internal processes, people and systems or from external events. This definition includes legal risk, but excludes strategic and reputational risk.

Examples of losses due to operational risk. Examples of losses falling within this category are, for instance, fraud (internal as well as external), losses due to IT failures, errors in settlements of transactions, litigation and losses due to external events like flooding, fire, earthquake or terrorism. Losses due to unfortunate management decisions, such as many of the mergers and acquisitions in the two decades leading up to the 2007–9 financial crisis, are definitely not included. However, the fact that legal risk is part of the definition has had a considerable impact on the financial industry in the aftermath of the crisis.

An early case that touched upon almost all aspects of the above definition was that of Barings (see also Section 1.2.2). From insufficient internal checks and balances (processes), to fraud (human risk), to external events (the Kobe earthquake), many operational risk factors contributed to the downfall of this once-renowned merchant bank. Further examples include the $691 million *rogue trading* loss at Allfirst Financial, the $484 million settlement due to misleading sales practices at Household Finance, and the estimated $140 million loss for the Bank of New York stemming from the September 11 attacks.

More recent examples in which legal risk has been involved include LIBOR rigging, for which the European Commission fined eight large financial institutions a total of $2.3 billion, and the possible rigging of rates in the foreign-exchange market, a market with an estimated $5.35 trillion daily turnover. Following the 2007–9 financial crisis several financial institutions faced fines for misselling of financial products—particularly securitized credit products—on the basis of inaccurate or misleading information about their risks. In the latter category, the Bank of America was fined the record amount of $16.65 billion on 21 August 2014.

Increasing use of algorithmic and high-frequency trading has resulted in operational losses. Examples include the 2010 Flash Crash and an estimated $440 million loss from a computer-trading glitch at Knight Capital Group. A number of prominent cases involving large losses attributable to rogue traders have also been widely reported in the press, including Fabrice Tourre at Goldman Sachs, Kweku Adoboli at UBS, Jérôme Kerviel at Société Générale and Bruno Iksil (also known as the London Whale) at JPMorgan Chase.

All the examples cited above, and the seriousness with which they have been taken by regulators worldwide, offer clear proof of the fundamental importance of operational risk as a risk class. Many banks have been forced to increase the capital they hold for operational risk (in some cases by up to 50%). An example of a regulatory document that probes more deeply into a trading loss at UBS is FINMA (2012).

Distinctiveness of operational risk. An essential difference between operational risk, on the one hand, and market and credit risk, on the other, is that operational risk has no upside for a bank. It comes about through the malfunctioning of parts of daily business and hence is as much a question of quality control as anything else. Although banks try as hard as possible to reduce operational risk, operational losses continue to occur.

Despite their continuing occurrence, a lack of publicly available, high-quality operational loss data has been a major issue in the development of operational risk models. This is similar to the problem faced by underwriters of catastrophe insurance. The insurance industry's answer to the problem has involved data pooling across industry participants, and similar developments are now taking place in the banking industry. Existing sources of pooled data include the Quantitative Impact Studies (QISs) of the Basel Committee, the database compiled by the Federal Reserve Bank of Boston, and subscription-based services for members like that of the ORX (Operational Riskdata eXchange Association). Increasingly, private companies are also providing data. However, the lack of more widely accessible data at the individual bank level remains a major problem. As data availability improves, many of the methods discussed in this book (such as the extreme value models of Chapter 5 and the insurance analytics of Section 13.2) will become increasingly useful.

Elementary versus advanced measurement approaches. In Section 13.1.3 we discuss the *advanced measurement* (AM) approach, which is typically adopted by larger banks that have access to high-quality operational loss data. This approach is often referred to as the *loss distribution approach* (LDA), since a series of distributional models or stochastic processes are typically fitted to operational loss data that have been categorized into different types of loss.

First, however, we discuss the so-called *elementary approaches* to operational risk modelling. In these approaches, aimed at smaller banks, the detailed modelling of loss distributions for different loss types is not required; a fairly simple volume-based capital charge is proposed.

We note that, as in the case of credit risk, the approaches proposed in the Basel framework for the calculation of regulatory capital represent a gradation in complexity. Recall that, for credit risk, banks must implement either the standardized approach or the internal-ratings-based (IRB) approach, as discussed in Section 1.3.1. The field of regulation is in a constant state of flux and the detail of the methods we describe below may change over time but the underlying principles are likely to continue to hold.

13.1.2 The Elementary Approaches

There are two elementary approaches to operational risk measurement. Under the *basic-indicator* (BI) approach, banks must hold capital for operational risk equal to the average over the previous three years of a fixed percentage (denoted by α) of positive annual gross income (GI). Figures for any year in which annual gross income is negative or zero should be excluded from both the numerator and denominator

when calculating the average. The risk capital under the BI approach for operational risk in year t is therefore given by

$$\mathrm{RC}^t_{\mathrm{BI}}(\mathrm{OR}) = \frac{1}{Z_t} \sum_{i=1}^{3} \alpha \max(\mathrm{GI}^{t-i}, 0), \tag{13.1}$$

where $Z_t = \sum_{i=1}^{3} I_{\{\mathrm{GI}^{t-i}>0\}}$ and GI^{t-i} stands for gross income in year $t-i$. Note that an operational risk capital charge is calculated on a yearly basis. The BI approach gives a fairly straightforward, volume-based, one-size-fits-all capital charge. Based on the various QISs, the Basel Committee suggests a value of α of 15%.

Under the *standardized* approach, banks' activities are divided into eight *business lines*: corporate finance; trading & sales; retail banking; commercial banking; payment & settlement; agency services; asset management; and retail brokerage. Precise definitions of these business lines are to be found in the final Basel II document (Basel Committee on Banking Supervision 2006). Within each business line, gross income is a broad indicator that serves as a proxy for the scale of business operations and thus the likely scale of operational risk exposure. The capital charge for each business line is calculated by multiplying gross income by a factor (denoted by β) assigned to that business line. As in (13.1), the total capital charge is calculated as a three-year average over positive GIs, resulting in the following capital charge formula:

$$\mathrm{RC}^t_{\mathrm{S}}(\mathrm{OR}) = \frac{1}{3} \sum_{i=1}^{3} \max\left[\sum_{j=1}^{8} \beta_j \mathrm{GI}^{t-i}_j, 0 \right]. \tag{13.2}$$

It may be noted that in formula (13.2), in any given year $t-i$, negative capital charges (resulting from negative gross income) in some business line j can offset positive capital charges in other business lines (albeit at the discretion of the national supervisor). This kind of "netting" should induce banks to go from the basic-indicator approach to the standardized approach; the word "netting" is of course to be used with care here. Based on the QISs, the Basel Committee has set the beta coefficients as in Table 13.1. Moscadelli (2004) gives a critical analysis of these beta factors, based on the full database of more than 47 000 operational losses of the second QIS of the summer of 2002 (see also Section 13.1.4). Concerning the use of GI in (13.1) and (13.2), see Notes and Comments.

13.1.3 Advanced Measurement Approaches

Under an AM approach, the regulatory capital is determined by a bank's own internal risk-measurement system according to a number of quantitative and qualitative criteria set forth in the regulatory documentation (Basel Committee on Banking Supervision 2006). We will not detail every relevant step in the procedure that leads to the acceptance of an AM approach for an internationally active bank and its subsidiaries; the Basel Committee's documents give a clear and readable account of this. We focus instead on the methodological aspects of a full quantitative approach to operational risk measurement. It should be stated, however, that, as in the case of

Table 13.1. Beta factors for the standardized approach.

Business line (j)	Beta factors (β_j)
$j = 1$, corporate finance	18%
$j = 2$, trading & sales	18%
$j = 3$, retail banking	12%
$j = 4$, commercial banking	15%
$j = 5$, payment & settlement	18%
$j = 6$, agency services	15%
$j = 7$, asset management	12%
$j = 8$, retail brokerage	12%

market and credit risk, the adoption of an AM approach to operational risk is subject to approval and continuing quality checking by the national supervisor.

While the BI and standardized approaches prescribe the explicit formulas (13.1) and (13.2), the AM approach lays down general guidelines. In the words of the Basel Committee (see Basel Committee on Banking Supervision 2006, paragraph 667):

> Given the continuing evolution of analytical approaches for operational risk, the Committee is not specifying the approach or distributional assumptions used to generate the operational risk measure for regulatory capital purposes. However, a bank must be able to demonstrate that its approach captures potentially severe "tail" loss events. Whatever approach is used, a bank must demonstrate that its operational risk measure meets a soundness standard comparable to that of the internal ratings-based approach for credit risk (comparable to a one year holding period and the 99.9 percent confidence interval).

In the usual LDA interpretation of an AM approach, operational losses are typically categorized according to the eight business lines mentioned in Section 13.1.2 as well as the following seven *loss-event types*: internal fraud; external fraud; employment practices & workplace safety; clients, products & business practices; damage to physical assets; business disruption & system failures; and execution, delivery & process management. While the categorization of losses in terms of eight business lines and seven loss-event types is standard, banks may deviate from this format if appropriate.

Banks are expected to gather internal data on repetitive, high-frequency losses (three to five years of data), as well as relevant external data on non-repetitive low-frequency losses. Moreover, they must add stress scenarios both at the level of loss severity (parameter shocks to model parameters) and correlation between loss types. In the absence of detailed joint models for different loss types, risk measures for the aggregate loss should be calculated by summing across the different loss categories. In general, both so-called *expected* and *unexpected* losses should be taken into account (i.e. risk-measure estimates cannot be reduced by subtraction of an expected loss amount).

We now describe a skeletal version of a typical AM solution for the calculation of an operational risk charge for year t. We assume that historical loss data from previous years have been collected in a data warehouse with the structure

$$\{X_k^{t-i,b,\ell} : i = 1, \dots, T;\ b = 1, \dots, 8;\ \ell = 1, \dots, 7;\ k = 1, \dots, N^{t-i,b,\ell}\}, \tag{13.3}$$

where $X_k^{t-i,b,\ell}$ stands for the kth loss of type ℓ for business line b in year $t-i$; $N^{t-i,b,\ell}$ is the number of such losses and $T \geqslant 5$ years, say. Note that thresholds may be imposed for each (i, b, ℓ) category, and small losses less than the threshold may be neglected; a threshold is typically of the order of €10 000. The total historical loss amount for business line b in year $t-i$ is obviously

$$L^{t-i,b} = \sum_{\ell=1}^{7} \sum_{k=1}^{N^{t-i,b,\ell}} X_k^{t-i,b,\ell}, \tag{13.4}$$

and the total loss amount for year $t-i$ is

$$L^{t-i} = \sum_{b=1}^{8} L^{t-i,b}. \tag{13.5}$$

The problem in the AM approach is to use the loss data to estimate the distribution of L_t for year t and to calculate risk measures such as VaR or expected shortfall (see Section 2.3) for the estimated distribution. Writing ϱ_α for the risk measure at a confidence level α, the regulatory capital is determined by

$$\mathrm{RC}_{\mathrm{AM}}^t(\mathrm{OR}) = \varrho_\alpha(L^t), \tag{13.6}$$

where α would typically take a value in the range 0.99–0.999 imposed by the local regulator. Because the joint distributional structure of the losses in (13.4) and (13.5) for any given year is generally unknown, we would typically resort to simple aggregation of risk measures across loss categories to obtain a formula of the form

$$\mathrm{RC}_{\mathrm{AM}}^t(\mathrm{OR}) = \sum_{b=1}^{8} \varrho_\alpha(L^{t,b}). \tag{13.7}$$

In view of our discussions in Chapter 8, the choice of an additive rule in (13.7) can be understood. Indeed, for any coherent risk measure ϱ_α, the right-hand side of (13.7) yields an upper bound for the total risk $\varrho_\alpha(L^t)$. In the important case of VaR, the right-hand side of (13.7) corresponds to the comonotonic scenario (see Proposition 7.20). The optimization results of Section 8.4.4 can be used to calculate bounds for $\varrho_\alpha(L^t)$ under different dependence scenarios for the business lines (see, in particular, Example 8.40).

Reduced to its most stylized form in the case when $\varrho_\alpha = \mathrm{VaR}_\alpha$ and $\alpha = 0.999$, a capital charge under the AM approach requires the calculation of a quantity of the type

$$\mathrm{VaR}_{0.999}\left(\sum_{k=1}^{N} X_k\right), \tag{13.8}$$

where (X_k) is some sequence of loss *severities* and N is an rv describing the *frequency* with which operational losses occur. Random variables of the type (13.8) are one of the prime examples of the actuarial models that we treat in Section 13.2.2. Before we move on to those models in the next section, we highlight some "stylized facts" concerning operational loss data.

13.1.4 Operational Loss Data

In order to reliably estimate (13.6), (13.7) or, in a stylized version, a quantity like (13.8), we need extensive data. The data situation for operational risk is much worse than that for credit risk, and is clearly an order of magnitude worse than for market risk, where vast quantities of data are publicly available. As discussed in Section 13.1.1, banks have not been gathering data for long and pooling initiatives are still in their infancy. As far as we know, no reliable *publicly* available data source on operational risk exists.

Our discussion below is based on industry data that we have been able to analyse as well as on the findings in Moscadelli (2004) for the QIS database and the results of the 2004 loss-data collection exercise by the Federal Reserve Bank of Boston (see Federal Reserve Bank of Boston 2005). An excellent overview of some of the data characteristics is to be found in the Basel Committee's report (Basel Committee on Banking Supervision 2003). As the latter report states:

> Despite this progress, inferences based on the data should still be made with caution.... In addition, the most recent data collection exercise provides data for only one year and, even under the best of circumstances, a one-year collection window will provide an incomplete picture of the full range of potential operational risk events, especially of rare but significant "tail events".

Further information is to be found in the increasing number of papers on the topic, and particularly in the various papers coming out of the ORX Consortium (see Notes and Comments).

In Figure 13.1 we have plotted operational loss data obtained from several sources; parts (a)–(c) show losses for three business lines for the period 1992–2001. It is less important for the reader to know the exact loss type—it is sufficient to accept that the data are typical for (b, ℓ) categories in (13.3). In part (d) the data from the three previous figures have been pooled.

Exploratory data analysis reveals the following stylized facts (confirmed in several other studies):

- loss severities have a heavy-tailed distribution;
- losses occur randomly in time; and
- loss frequency may vary substantially over time.

The third observation is partly explained by the fact that banks have gathered an increasing amount of operational loss data since the Basel II rules were first announced. There is therefore a considerable amount of *reporting bias*, resulting

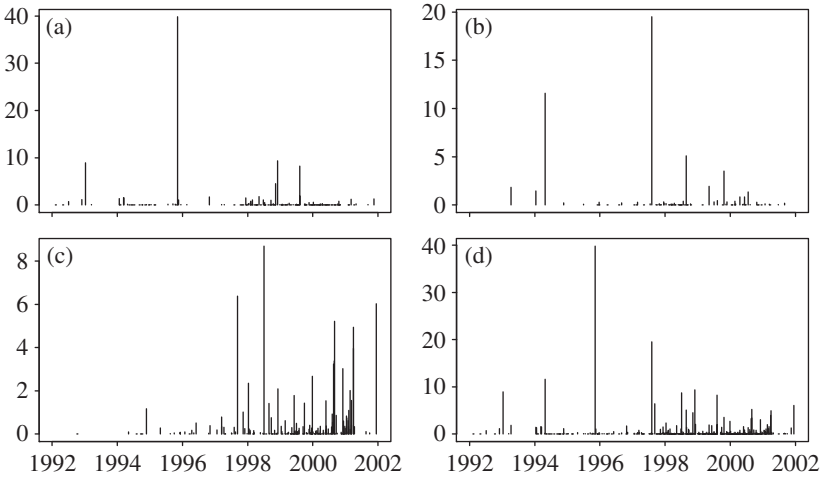

Figure 13.1. Operational risk losses: (a) type 1, $n = 162$; (b) type 2, $n = 80$; (c) type 3, $n = 175$; and (d) pooled losses $n = 417$.

in fewer losses in the first half of the 1990s and more losses afterwards. Moreover, several classes of loss may have a considerable cyclical component and/or may depend on changing economic covariables. For instance, back-office errors may depend on volume traded and fraud may be linked to the overall level of the economy (depressions versus boom cycles). Moreover, there may be a rise in legal losses in the aftermath of a severe crisis, as has been observed for the 2007–9 credit crisis. This clear inhomogeneity in the loss frequency makes an immediate application of statistical methodology difficult. However, it may be reasonable to at least assume that the (inflation-adjusted) loss sizes have a common severity distribution, which would allow, for instance, the application of methods from Chapter 5.

In Figure 13.2 we have plotted the sample mean excess functions (5.16) for the data in Figure 13.1. This figure clearly indicates the first stylized fact of heavy-tailed loss severities. The mean excess plots in (a) and (b) are clearly increasing in an approximately linear fashion, pointing to Pareto-type behaviour. This contrasts with (c), where the plot appears to level off from a threshold of 1. This hints at a loss distribution with finite upper limit, but this can only be substantiated by more detailed knowledge of the type of loss concerned. Pooling the data in (d) masks the different kinds of behaviour, and perhaps illustrates the dangers of naive statistical analyses that do not consider the data-generating mechanism.

Moscadelli (2004) performed a detailed EVT analysis (including a first attempt to solve the frequency problem) of the full QIS data set of more than 47 000 operational losses and concluded that the loss dfs are well fitted by generalized Pareto distributions (GPDs) in the upper-tail area (see Section 5.2.2 for the necessary statistical background). The estimated tail parameters (ξ in (5.14)) for the different business lines range from 0.85 for asset management to 1.39 for commercial banking. Six of the business lines have an estimate of ξ greater than 1, corresponding to an *infinite-mean model*! Based on these QIS data, the estimated risk capital/GI ratios (the β in

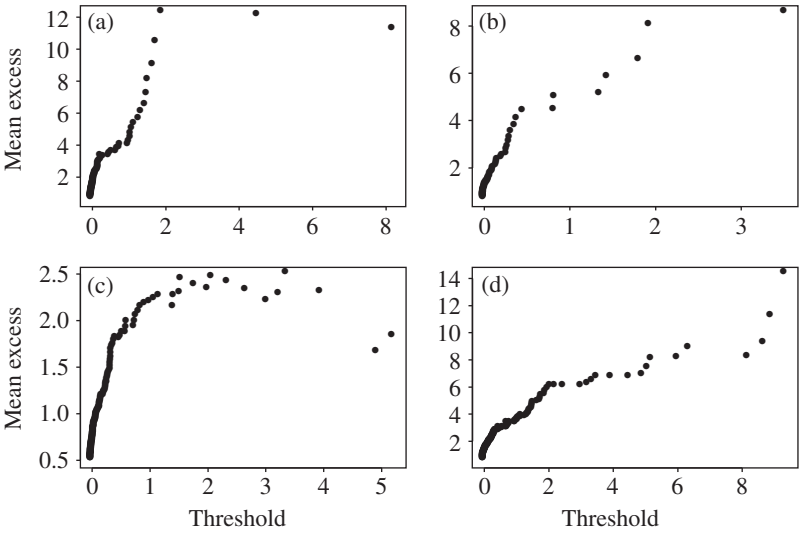

Figure 13.2. Corresponding sample mean excess plots for the data in Figure 13.1: (a) type 1; (b) type 2; (c) type 3; and (d) pooled.

Table 13.1) range from 8.3% for retail banking to 33.3% for payment & settlement, with an overall alpha value (see (13.1)) of 13.3%, slightly below the Basel II value of 15% used in the BI approach. Note the much wider range of values of β that emerge from the analysis of the QIS data compared with the prescribed range of 12–18% for the standardized approach in Table 13.1.

Notes and Comments

Several textbooks on operational risk have been published: see, for example, Cruz (2002, 2004), King (2001), the Risk Books publication edited by Jenkins and Roberts (2004), and chapters in Ong (2003) and Crouhy, Galai and Mark (2001). In particular, Chapter 4 of Cruz (2004), written by Carolyn Currie, gives an excellent overview of the regulatory issues surrounding operational risk. Further textbooks include Shevchenko (2011), Cruz, Peters and Shevchenko (2015) and Panjer (2006). The *Journal of Operational Risk* publishes relevant research from academia and practice, including informative papers coming out of the ORX Consortium: see, for example, Cope and Antonini (2008) and Cope and Labbi (2008).

Papers on implementing operational risk models in practice include Ebnöther et al. (2003); Frachot, Georges and Roncalli (2001), which discusses the loss distribution approach to operational risk; Doebeli, Leippold and Vanini (2003), which shows that a good operational risk framework may lead to an overall improvement in the quality of business operations; and Aue and Kalkbrener (2006), which provides a comprehensive description of an approach based on loss distributions that was developed at Deutsche Bank. Excellent data-analytic papers using published operational risk losses are de Fontnouvelle et al. (2003) and Moscadelli (2004). Rosenberg and Schuermann (2006) address the aggregation of market, credit and operational risk measures. A comprehensive study, combining internal and external

data, together with expert opinion, in a scenario-generation context is de Jongh et al. (2014); this paper also contains an excellent list of references.

Figure 13.1 is taken from Embrechts, Kaufmann and Samorodnitsky (2004). This paper also stresses the important difference between so-called repetitive and non-repetitive losses. For the former (to some extent less important) losses, statistical modelling can be very useful. For non-repetitive, low-probability, high-severity losses, much more care has to be taken before a statistical analysis can be performed (see Pézier 2003a,b).

EVT methods for operational risk quantification have been used by numerous authors (see, for example, Coleman 2002, 2003; Medova 2000; Medova and Kyriacou 2000). Because of the non-stationarity of operational loss data over several years, more refined EVT models are called for: see, for example, Chavez-Demoulin and Embrechts (2004) and Chavez-Demoulin, Embrechts and Hofert (2014) for some examples of such models. For a critical article on the use of EVT for the calculation of an operational risk capital charge, see Embrechts, Furrer and Kaufmann (2003), which contains a simulation study of the number of data needed to come up with a reasonable estimate of a high quantile. The use of statistical methods other than EVT are discussed in the textbooks referred to above. These methods include linear predictive models, Bayesian belief networks, discriminant analysis, and tools and techniques from reliability theory.

We have noted that severity models for operational risk are typically (extremely) heavy tailed. Several publications report infinite-mean models. For an early discussion of this issue, see Nešlehová, Embrechts and Chavez-Demoulin (2006). The reader interested in a more philosophical discussion of the economic sense of such models should carry out an internet search for "the dismal theorem" and read some of the material that has been written on this topic. The term "dismal theorem" was coined by Martin Weitzman, who introduced it in relation to the economics of catastrophic climate change; an interesting paper on the topic is Weitzman (2011).

It should be clear from what we learned in earlier chapters about risk measures, their aggregation and their statistical estimation that calculating a 99.9% yearly VaR for operational risk is, to put it mildly, a daunting task. As a consequence, Ames, Schuermann and Scott (2014) suggest several regulatory policy changes leading to simpler, more standardized, more stable and more robust methodologies, at least until our understanding of operational risk has increased. In a recent publication, the Basel Committee on Banking Supervision (2014) proposes a change from the gross income (GI) indicator in (13.1) and (13.2) to a new business indicator (BI).

13.2 Elements of Insurance Analytics

13.2.1 The Case for Actuarial Methodology

Actuarial tools and techniques for the modelling, pricing and reserving of insurance products in the traditional fields of life insurance, non-life insurance and reinsurance have a long history going back more than a century. More recently, the border between financial and insurance products has become blurred, examples of this

process being equity-linked life products and the transfer of insurance risks to the capital markets via securitization (see Chapter 1, particularly Section 1.5.3 and the Notes and Comments).

Whereas some of the combined bank-assurance products have not met with the success that was originally hoped for, it remains true that there exists an increasing need for financial and actuarial professionals who can close the methodological gaps between the two fields. In the sections that follow we discuss insurance-analytical tools that we believe the more traditional finance-oriented risk manager ought to be aware of; the story behind the name insurance analytics can be found in Embrechts (2002).

It is not only the occasional instance of joint product development between the banking and insurance worlds that prompts us to make a case for actuarial methodology in QRM, but also the observation that many of the concepts and techniques of QRM described in the preceding chapters are in fact borrowed from the actuarial literature.

- Risk measures like expected shortfall (Definition 2.12) have been studied in a systematic way in the insurance literature. Expected shortfall is also the standard risk measure to be used under the Solvency II guidelines.
- Many of the dependence modelling tools presented in Chapter 7 saw their first applications in the realm of insurance. Moreover, notions like comonotonicity of risk factors have their origins in actuarial questions.
- In Section 2.3.5 we discussed the axiomatization of financial risk measures and mentioned the parallel development of insurance premium principles (often with very similar goals and results).
- The statistical modelling of extremal events has been a bread-and-butter subject for actuaries since the start of insurance. Many of the tools presented in Chapter 5 are therefore well known to actuaries.
- Within the world of credit risk management, the industry model CreditRisk$^+$ (Section 11.2.5) is known as an actuarial model.
- The actuarial approach to the modelling of operational risk is apparent in the AM approach of Section 13.1.3.

In the sections that follow we give a brief discussion of relevant actuarial techniques. The material presented should enable the reader to transfer actuarial concepts to QRM in finance more easily. We do not strive for a full treatment of the relevant tools as that would fill a separate (voluminous) textbook (see, for example, Denuit and Charpentier (2004), Mikosch (2004) and Partrat and Besson (2004) for excellent accounts of many of the relevant techniques).

13.2.2 The Total Loss Amount

Reconsider formula (13.8), where a random number N of random losses or severities X_k occurring in a given time period are summed. To apply a risk measure like

VaR we need to make assumptions about the (X_k) and N, which leads us to one of the fundamental concepts of (non-life) insurance mathematics.

Definition 13.1 (total loss amount and distribution). Denote by $N(t)$ the (random) number of losses over a fixed time period $[0, t]$ and write $X_1, X_2, \ldots$ for the individual losses. The *total loss amount* (or *aggregate loss*) is defined as

$$S_{N(t)} = \sum_{k=1}^{N(t)} X_k, \tag{13.9}$$

with df $F_{S_{N(t)}}(x) = P(S_{N(t)} \leqslant x)$, the *total* (or *aggregate*) *loss df*. Whenever t is fixed, $t = 1$ say, we may drop the time index from the notation and simply write S_N and F_{S_N}.

Remark 13.2. The definition of (13.9) as an rv is to be understood as $S_{N(t)}(\omega) = \sum_{k=1}^{N(t)(\omega)} X_k(\omega)$, $\omega \in \Omega$, and is referred to as a random (or randomly indexed) sum.

A prime goal of this section will be the analytical and numerical calculation of F_{S_N}, which requires further assumptions about the (X_k) and N.

Assumption 13.3 (independence, compound sums). *We assume that the rvs (X_k) are iid with common df G, $G(0) = 0$. We further assume that the rvs N and (X_k) are independent; in that case we refer to (13.9) as a compound sum. The probability mass function of N is denoted by $p_N(k) = P(N = k)$, $k = 0, 1, 2, \ldots$. The rv N is referred to as the compounding rv.*

Proposition 13.4 (compound distribution). *Let S_N be a compound sum and suppose that Assumption 13.3 holds. Then, for all $x \geqslant 0$,*

$$F_{S_N}(x) = P(S_N \leqslant x) = \sum_{k=0}^{\infty} p_N(k) G^{(k)}(x), \tag{13.10}$$

where $G^{(k)}(x) = P(S_k \leqslant x)$, the kth convolution of G. Note that $G^{(0)}(x) = 1$ for $x \geqslant 0$, and $G^{(0)}(x) = 0$ for $x < 0$.

Proof. Suppose that $x \geqslant 0$. Then

$$F_{S_N}(x) = \sum_{k=0}^{\infty} P(S_N \leqslant x \mid N = k) P(N = k) = \sum_{k=0}^{\infty} p_N(k) G^{(k)}(x).$$

□

Although formula (13.10) is explicit, its actual calculation in specific cases is difficult because the convolution powers $G^{(k)}$ of a df G are not generally available in closed form. One therefore resorts to (numerical) approximation methods. A first class of these uses the fact that the Laplace–Stieltjes transform of a convolution is the product of the Laplace–Stieltjes transforms. Using the usual notation

$\hat{F}(s) = \int_0^\infty \mathrm{e}^{-sx}\,\mathrm{d}F(x)$, where $s \geqslant 0$ for Laplace–Stieltjes transforms, we have that $\widehat{G^{(k)}}(s) = (\hat{G}(s))^k$. It follows from Proposition 13.4 that

$$\hat{F}_{S_N}(s) = \sum_{k=0}^{\infty} p_N(k)\hat{G}^k(s) = \Pi_N(\hat{G}(s)), \quad s \geqslant 0, \tag{13.11}$$

where Π_N denotes the *probability-generating function* of N, defined by $\Pi_N(s) = \sum_{k=1}^{\infty} p_N(k)s^k$.

Example 13.5 (the compound Poisson df). Suppose that N has a Poisson df with intensity parameter $\lambda > 0$, denoted by $N \sim \mathrm{Poi}(\lambda)$. In that case, $p_N(k) = \mathrm{e}^{-\lambda}\lambda^k/k!$, $k \geqslant 0$, and, for $s > 0$,

$$\Pi_N(s) = \sum_{k=0}^{\infty} \mathrm{e}^{-\lambda}\frac{\lambda^k}{k!}s^k = \mathrm{e}^{-\lambda(1-s)}.$$

From (13.11) it therefore follows that, for $s \geqslant 0$,

$$\hat{F}_{S_N}(s) = \exp(-\lambda(1 - \hat{G}(s))).$$

In this case, the df of S_N is referred to as the *compound Poisson df* and we write $S_N \sim \mathrm{CPoi}(\lambda, G)$.

Example 13.6 (the compound negative binomial df). Suppose that N has a negative binomial df with parameters $\alpha > 0$ and $0 < p < 1$, denoted by $N \sim \mathrm{NB}(\alpha, p)$. The probability mass function is given by (A.18) and we get, for $0 < s < (1 - p)^{-1}$,

$$\Pi_N(s) = \sum_{k=0}^{\infty} \binom{\alpha + k - 1}{k} p^\alpha (1 - p)^k s^k = \left(\frac{p}{1 - s(1 - p)}\right)^\alpha.$$

From (13.11) it therefore follows that, for $s \geqslant 0$,

$$\hat{F}_{S_N}(s) = \left(\frac{p}{1 - \hat{G}(s)(1 - p)}\right)^\alpha.$$

In this case, the df of S_N is referred to as the *compound negative binomial df* and we write $S_N \sim \mathrm{CNB}(\alpha, p, G)$.

Formula (13.11) facilitates the calculation of moments of S_N and lends itself to numerical evaluation through Fourier inversion, using a technique known as the fast Fourier transform (FFT) (see Notes and Comments for references on the latter). For the calculation of moments, note that, under the assumption of the existence of sufficiently high moments and hence differentiability of $\hat{G}$ and Π_N, we obtain

$$\frac{\mathrm{d}^k}{\mathrm{d}s^k}\Pi_N(s)\bigg|_{s=1} = E(N(N-1)\cdots(N-k+1))$$

and

$$(-1)^k\frac{\mathrm{d}^k}{\mathrm{d}s^k}\hat{G}(s)\bigg|_{s=0} = E(X_1^k) = \mu_k.$$

Example 13.7 (continuation of Example 13.5). In the case of the compound Poisson df one obtains

$$E(S_N) = (-1)\frac{\mathrm{d}}{\mathrm{d}s}\hat{F}_{S_N}(s)\bigg|_{s=0} = \exp(-\lambda(1-\hat{G}(0)))\lambda(-\hat{G}'(0))$$
$$= \lambda\mu_1 = E(N)E(X_1).$$

Similar calculations yield $\operatorname{var}(S_N) = E(S_N^2) - (E(S_N))^2 = \lambda\mu_2$.

For the general compound case one obtains the following useful result.

Proposition 13.8 (moments of compound dfs). *Under Assumption 13.3 and assuming that $E(N) < \infty$, $\mu_2 < \infty$, we have that*

$$E(S_N) = E(N)E(X_1) \quad \textit{and} \quad \operatorname{var}(S_N) = \operatorname{var}(N)(E(X_1))^2 + E(N)\operatorname{var}(X_1). \tag{13.12}$$

Proof. This follows readily from (13.11), differentiating with respect to s. The following direct proof avoids the use of transforms. Conditioning on N and using Assumption 13.3 one obtains

$$E(S_N) = E(E(S_N \mid N)) = E\bigg(E\bigg(\sum_{k=1}^{N} X_k \;\bigg|\; N\bigg)\bigg)$$
$$= E\bigg(\sum_{k=1}^{N} E(X_k)\bigg) = E(N)E(X_1)$$

and, similarly,

$$E(S_N^2) = E\bigg(E\bigg(\bigg(\sum_{k=1}^{N} X_k\bigg)^2 \;\bigg|\; N\bigg)\bigg) = E\bigg(E\bigg(\sum_{k=1}^{N}\sum_{\ell=1}^{N} X_k X_\ell \;\bigg|\; N\bigg)\bigg)$$
$$= E(N\mu_2 + N(N-1)\mu_1^2) = E(N)\mu_2 + (E(N^2) - E(N))\mu_1^2$$
$$= E(N)\operatorname{var}(X_1) + E(N^2)(E(X_1))^2,$$

so $\operatorname{var}(S_N) = E(S_N^2) - (E(S_N))^2 = E(N)\operatorname{var}(X_1) + \operatorname{var}(N)(E(X_1))^2$. □

Remark 13.9. Formula (13.12) elegantly combines the randomness of the frequency ($\operatorname{var}(N)$) with that of the severity ($\operatorname{var}(X_1)$). In the compound Poisson case it reduces to the formula $\operatorname{var}(S_N) = \lambda E(X_1^2) = \lambda\mu_2$, as in Example 13.7. In the deterministic-sum case, when $P(N = n) = 1$, say, we find the well-known results $E(S_N) = n\mu_1$ and $\operatorname{var}(S_N) = n\operatorname{var}(X_1)$; indeed, in this degenerate case, $\operatorname{var}(N) = 0$.

The compound Poisson model is a basic model for aggregate financial or insurance risk losses. The ubiquitousness of the Poisson distribution in insurance can be understood as follows. Consider a time interval [0, 1] and let N denote the total number of losses in that interval. Suppose further that we have a number of potential loss generators (transactions, credit positions, insurance policies, etc.) that can produce, with probability p_n, one loss or, with probability $1 - p_n$, no loss in each

small subinterval $((k-1)/n, k/n]$ for $k = 1, \ldots, n$. Moreover, suppose that the occurrence or non-occurrence of a loss in any particular subinterval is not influenced by the occurrence of losses in other intervals. The number N_n of losses then has a binomial df with parameters n and p_n, so

$$P(N_n = k) = \binom{n}{k} p_n^k (1-p_n)^{n-k}, \quad k = 0, \ldots, n.$$

Combined with a loss-severity distribution this frequency distribution gives rise, in (13.10), to the so-called *binomial loss model*. Next suppose that $n \to \infty$ in such a way that $\lim_{n\to\infty} np_n = \lambda > 0$. It follows from *Poisson's theorem of rare events* (see also Section 5.3.1) that

$$\lim_{n\to\infty} P(N_n = k) = \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \ldots,$$

i.e. $N_\infty \sim \mathrm{Poi}(\lambda)$, explaining why the Poisson model assumption is very natural as a frequency distribution and why the compound Poisson model is a common aggregate loss model. The compound Poisson model has several nice properties, one of which concerns aggregation and is useful in the operational risk context in situations such as (13.5).

Proposition 13.10 (sums of compound Poisson rvs). *Suppose that the compound sums $S_{N_i} \sim \mathrm{CPoi}(\lambda_i, G_i)$, $i = 1, \ldots, d$, and that these rvs are independent. Then $S_N = \sum_{i=1}^d S_{N_i} \sim \mathrm{CPoi}(\lambda, G)$, where $\lambda = \sum_{i=1}^d \lambda_i$ and $G = \sum_{i=1}^d (\lambda_i/\lambda) G_i$.*

Proof. (For $d = 2$, the general case being similar.) Because of independence and Example 13.5 we have, for the Laplace–Stieltjes transform of S_N,

$$\begin{aligned}
\hat{F}_{S_N}(s) &= \hat{F}_{S_{N_1}}(s) \hat{F}_{S_{N_2}}(s) \\
&= \exp\left(-(\lambda_1+\lambda_2)\left(1 - \frac{1}{\lambda_1+\lambda_2}(\lambda_1 \hat{G}_1(s) + \lambda_2 \hat{G}_2(s))\right)\right) \\
&= \exp(-\lambda(1-\hat{G}(s))),
\end{aligned}$$

where $\lambda = \lambda_1 + \lambda_2$ and

$$G = \frac{\lambda_1}{\lambda_1+\lambda_2} G_1 + \frac{\lambda_2}{\lambda_1+\lambda_2} G_2.$$

The result follows since the Laplace–Stieltjes transform uniquely determines the underlying df. □

The new intensity λ is just the sum of the old ones, whereas the new severity df G is a *discrete mixture* of the loss dfs G_i with weights λ_i/λ, $i = 1, \ldots, d$. We can easily simulate losses from such a model through a two-stage procedure: first draw i ($i = 1, \ldots, d$) with probability λ_i/λ, and then draw a loss with df G_i.

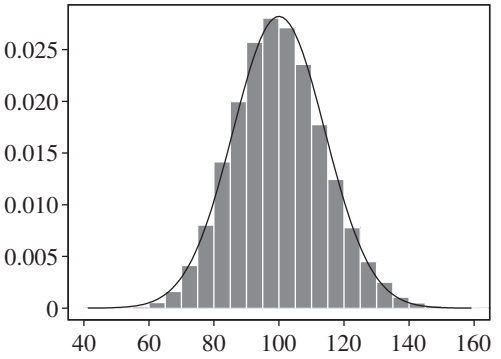

Figure 13.3. Histogram of simulated compound loss data ($n = 100\,000$) for $S_N \sim \text{CPoi}(100, \text{Exp}(1))$ together with normal approximation (13.14).

Beyond the Poisson model. The Poisson model can serve as a stylized representation of the loss-generating mechanism from which more realistic models can be derived. For instance, we may wish to introduce a time parameter in N to capture different occurrence patterns over time (see Section 13.2.6). Also, the intensity parameter λ may be assumed to be random (see Example 13.21). Indeed, a further step is to turn λ into a stochastic process, which gives rise to such models as doubly stochastic (or Cox) processes (see Section 10.5.1) or self-exciting processes, as encountered in Section 16.2.1. Furthermore, various forms of dependence among the X_k rvs or between N and (X_k) could be modelled. Finally, multiline portfolios require multivariate models for vectors of the type $(S_{N_1}, \ldots, S_{N_d})$. An ultimate goal of the AM approach to operational risk would be to model such random vectors where, for instance, d might stand for seven risk types, eight business lines, or a total of 56 loss category cells.

13.2.3 Approximations and Panjer Recursion

As mentioned in Section 13.2.2, the analytic calculation of F_{S_N} is not possible for the majority of reasonable models, which has led actuaries to come up with several numerical approximations. Below we review some of these approximations and illustrate their use for several choices of the severity df G. The basic example we look at is the compound Poisson case, $S_N \sim \text{CPoi}(\lambda, G)$, though most of the approximations discussed can be adjusted to deal with other distributions for N. Given λ and G we can easily simulate F_{S_N} and, by repeating this many times, we can get an empirical estimate that is close to the true df. Figure 13.3 contains a simulation of $n = 100\,000$ realizations of $S_N \sim \text{CPoi}(100, \text{Exp}(1))$. Although the histogram exhibits mild skewness (which can easily be shown theoretically (see (13.15))), a clear central limit effect takes place. This is used in the first approximation below.

Normal approximation. As the loss rvs X_i are iid (with finite second moment, say) and S_N is a (random) sum of the X_i variables, one can apply Theorem 2.5.16 from Embrechts, Klüppelberg and Mikosch (1997) and Proposition 13.8 to obtain

the following approximation, for general N:

$$F_{S_N}(x) \approx \Phi\left(\frac{x - E(N)E(X_1)}{\sqrt{\mathrm{var}(N)(E(X_1))^2 + E(N)\,\mathrm{var}(X_1)}}\right). \tag{13.13}$$

Here, and in the approximations below, "$\approx$" has no specific mathematical interpretation beyond "there exists a limit result justifying the right-hand side to be used as approximation of the left-hand side". In particular, for the compound Poisson case above, (13.13) reduces to

$$F_{S_N}(x) \approx \Phi\left(\frac{x - 100}{\sqrt{200}}\right), \tag{13.14}$$

where Φ is the standard normal df, as usual. It is this normal approximation that is superimposed on the histogram in Figure 13.3. Clearly, there are conditions that must be satisfied in order to obtain the approximation (13.13): for example, claims should not be too heavy tailed (see Theorem 13.22).

For CPoi(λ, G) it is not difficult to show that the skewness parameter satisfies

$$\frac{E((S_N - E(S_N))^3)}{(\mathrm{var}(S_N))^{3/2}} = \frac{E(X_1^3)}{\sqrt{\lambda(E(X_1^2))^3}} > 0 \tag{13.15}$$

(note that $X_1 \geqslant 0$ almost surely), so an approximation by a df with positive skewness may improve the approximation (13.14), especially in the tail area. This is indeed the case and leads to the next approximation.

Translated-gamma approximation. We approximate S_N by $k + Y$, where k is a translation parameter and $Y \sim \mathrm{Ga}(\alpha, \beta)$ has a gamma distribution (see Section A.2.4). The parameters (k, α, β) are found by matching the mean, the variance and the skewness of $k + Y$ and S_N. It is not difficult to check that the following equations result:

$$k + \frac{\alpha}{\beta} = \lambda E(X_1), \qquad \frac{\alpha}{\beta^2} = \lambda E(X_1^2), \qquad \frac{2}{\sqrt{\alpha}} = \frac{E(X_1^3)}{\sqrt{\lambda(E(X_1^2))^3}}.$$

In our case, where $\lambda = 100$ and X_1 has a standard exponential distribution, these yield the equations $k + \alpha/\beta = 100$, $\alpha/\beta^2 = 200$ and $2/\sqrt{\alpha} = 0.2121$ with solution $\alpha = 88.89$, $\beta = 0.67$, $k = -32.72$.

Commentary on these approximations. Both approximations work reasonably well for the bulk of the data. However, for risk-management purposes we are mainly interested in upper tail risk; in Figure 13.4 we have therefore plotted both approximations for $x \geqslant 120$ on a log–log scale. This corresponds to the tail area beyond the 90% quantile of F_{S_N}. Similar plots were routinely used in Chapter 5 on EVT (see, for example, Figure 5.6). It becomes clear that, as can be expected, the gamma approximation works better in this upper tail area where the normal approximation underestimates the loss potential.

Of course, for loss data with heavier tails than exponential (lognormal or Pareto, say), even the translated-gamma approximation will be insufficient, and other

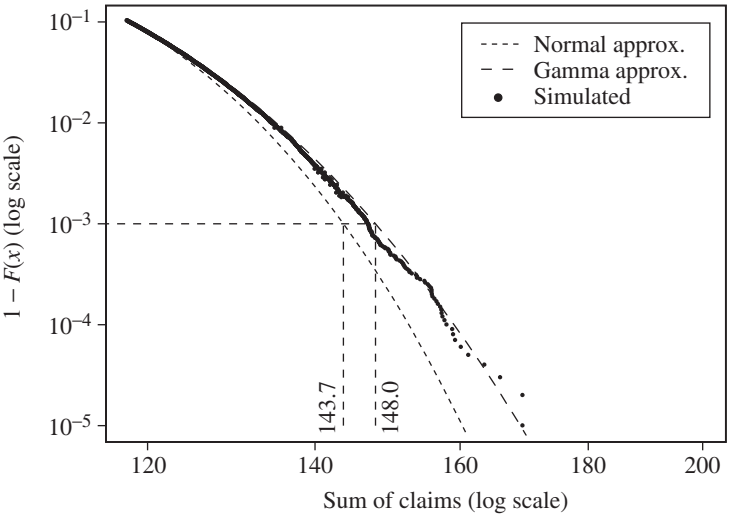

Figure 13.4. Simulated CPoi(100, Exp(1)) data together with normal and translated-gamma approximations (log–log scale). The 99.9% quantile estimates are also given.

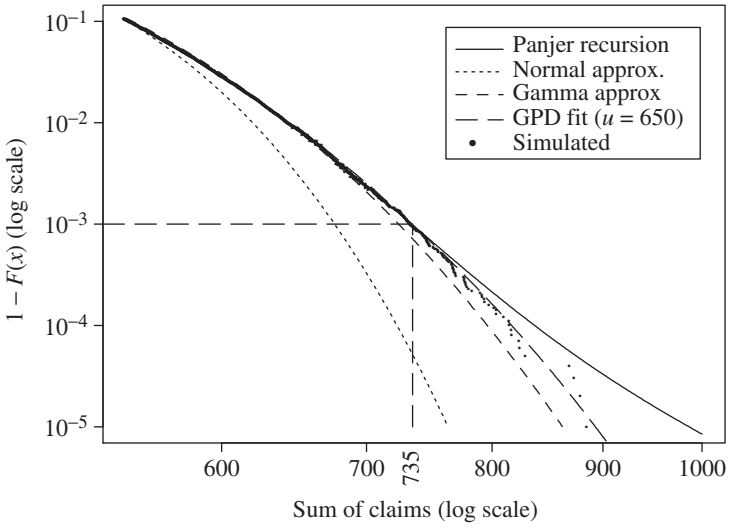

Figure 13.5. Simulated CPoi(100, LN(1, 1)) data ($n = 100\,000$) with normal, translated-gamma, GPD and Panjer recursion (see Example 13.18) approximations (log–log scale).

approximations can be devised based on heavier-tailed distributions, such as translated F, inverse gamma or generalized Pareto.

Another approach could be based on Monte Carlo simulation of aggregate losses S_N to which an appropriate heavy-tailed loss distribution could then be fitted. One possible approach would be to model the tail of these simulated compound losses with the GPD using the methodology of Section 5.2.2. This is what has been done in Figures 13.5 and 13.6, where we have plotted various approximations

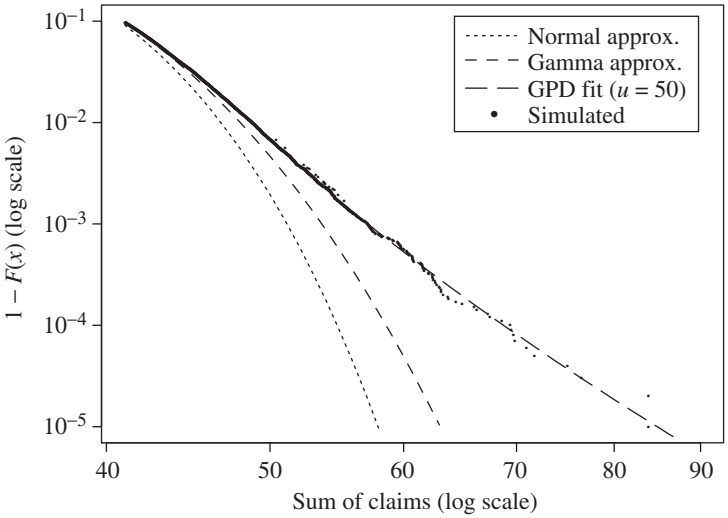

Figure 13.6. Simulated CPoi(100, Pa(4, 1)) data ($n = 100\,000$) with normal, translated-gamma, and GPD approximations (log–log scale).

for CPoi(100, LN(1, 1)) and CPoi(100, Pa(4, 1)). The former corresponds to a standard industry model for operational risk (see Frachot 2004). The latter corresponds to a class of operational risk models used in Moscadelli (2004). The message of these figures is clear: if the data satisfy the compound Poisson assumption, then the GPD yields a superior fit for high quantiles.

We now turn to an important class of approximations based on recursive methods. In the case where the loss sizes (X_i) are discrete and the distribution function of N satisfies a specific condition (see Definition 13.11 below) a reliable recursive method can be worked out.

Suppose that X_1 has a discrete distribution so that $P(X_1 \in \mathbb{N}_0) = 1$ with $g_k = P(X_1 = k)$, $p_k = P(N = k)$ (for notational convenience we write p_k for $p_N(k)$) and $s_k = P(S_N = k)$. For simplicity assume that $g_0 = 0$ and let

$$g_k^{(n)} = P(X_1 + \cdots + X_n = k),$$

the discrete convolution of the probability mass function g_k. Note that, by definition, $g_k^{(n+1)} = \sum_{i=1}^{k-1} g_i^{(n)} g_{k-i}$. We immediately obtain the following identities:

$$\left.\begin{aligned} s_0 &= P(S_N = 0) = P(N = 0) = p_0, \\ s_n &= P(S_N = n) = \sum_{k=1}^{\infty} p_k g_n^{(k)}, \quad n \geqslant 1, \end{aligned}\right\} \tag{13.16}$$

where the latter formula corresponds to Proposition 13.4 but now in the discrete case. As in Proposition 13.4 we note that (13.16) is difficult to calculate, mainly due to the convolutions $g_n^{(k)}$. However, for an important class of counting variables N, (13.16) can be reduced to a simple recursion. For this we introduce the so-called *Panjer classes*.

Definition 13.11 (Panjer class). The probability mass function (p_k) of N belongs to the Panjer(a, b) class for some $a, b \in \mathbb{R}$ if the following relationship holds for $r \geqslant 1$: $p_r = (a + (b/r))p_{r-1}$.

Example 13.12 (binomial). If $N \sim \mathrm{B}(n, p)$, then its probability mass function is $p_r = \binom{n}{r} p^r (1-p)^{n-r}$ for $0 \leqslant r \leqslant n$ and it can be easily checked that

$$\frac{p_r}{p_{r-1}} = -\frac{p}{1-p} + \frac{(n+1)p}{r(1-p)},$$

showing that N belongs to the Panjer(a, b) class with $a = -p/(1-p)$ and $b = (n+1)p/(1-p)$.

Example 13.13 (Poisson). If $N \sim \mathrm{Poi}(\lambda)$, then its probability mass function $p_r = \mathrm{e}^{-\lambda}\lambda^r/r!$ satisfies $p_r/p_{r-1} = \lambda/r$, so N belongs to the Panjer(a, b) class with $a = 0$ and $b = \lambda$.

Example 13.14 (negative binomial). If N has a negative binomial distribution, denoted by $N \sim \mathrm{NB}(\alpha, p)$, then its probability mass function is

$$p_r = \binom{\alpha + r - 1}{r} p^\alpha (1-p)^r, \quad r \geqslant 0,\ \alpha > 0,\ 0 < p < 1$$

(see Section A.2.7 for further details). We can easily check that

$$\frac{p_r}{p_{r-1}} = 1 - p + \frac{(\alpha - 1)(1-p)}{r}.$$

Hence N belongs to the Panjer(a, b) class with $a = 1 - p$ and $b = (\alpha - 1)(1 - p)$. In Proposition 13.21 we will show that the negative binomial model follows very naturally from the Poisson model when one randomizes the intensity parameter of the latter using a gamma distribution.

Remark 13.15. One can show that, neglecting degenerate models for (p_k), the above three examples are the *only* counting distributions satisfying Definition 13.11. This result goes back to Johnson and Kotz (1969) and was formulated explicitly in the actuarial literature in Sundt and Jewell (1982).

Theorem 13.16 (Panjer recursion). *Suppose that N satisfies the* Panjer(a, b) *class condition and that $g_0 = P(X_1 = 0) = 0$, then $s_0 = p_0$ and, for $r \geqslant 1$, $s_r = \sum_{i=1}^r (a + (bi/r)) g_i s_{r-i}$.*

Proof. We already know that $s_0 = p_0$ from (13.16), so suppose that $r \geqslant 1$. Noting that $X_1, \ldots, X_n$ are iid, we require the following well-known identity for exchangeable rvs:

$$\begin{aligned} E\left(X_1 \,\middle|\, \sum_{i=1}^n X_i = r\right) &= \frac{1}{n}\sum_{j=1}^n E\left(X_j \,\middle|\, \sum_{i=1}^n X_i = r\right) \\ &= \frac{1}{n} E\left(\sum_{j=1}^n X_j \,\middle|\, \sum_{i=1}^n X_i = r\right) = \frac{r}{n}. \end{aligned} \tag{13.17}$$

Moreover, using the fact that $g_0^{(n-1)} = 0$ for $n \geqslant 2$, we make the preliminary calculation that

$$\begin{aligned}
p_{n-1} \sum_{i=1}^{r-1} \left(a + \frac{bi}{r}\right) g_i g_{r-i}^{(n-1)} &= p_{n-1} \sum_{i=1}^{r} \left(a + \frac{bi}{r}\right) g_i g_{r-i}^{(n-1)} \\
&= p_{n-1} \sum_{i=1}^{r} \left(a + \frac{bi}{r}\right) P\left(X_1 = i,\ \sum_{j=2}^{n} X_j = r - i\right) \\
&= p_{n-1} \sum_{i=1}^{r} \left(a + \frac{bi}{r}\right) P\left(X_1 = i,\ \sum_{j=1}^{n} X_j = r\right) \\
&= p_{n-1} \sum_{i=1}^{r} \left(a + \frac{bi}{r}\right) P\left(X_1 = i \,\middle|\, \sum_{j=1}^{n} X_j = r\right) g_r^{(n)} \\
&= p_{n-1} E\left(a + \frac{bX_1}{r} \,\middle|\, \sum_{j=1}^{n} X_j = r\right) g_r^{(n)} \\
&= p_{n-1} \left(a + \frac{b}{n}\right) g_r^{(n)} = p_n g_r^{(n)},
\end{aligned}$$

where (13.17) is used in the final step. Therefore, the identity (13.16) yields

$$\begin{aligned}
s_r &= \sum_{n=1}^{\infty} p_n g_r^{(n)} = p_1 g_r + \sum_{n=2}^{\infty} p_n g_r^{(n)} \\
&= (a+b) p_0 g_r + \sum_{n=2}^{\infty} \sum_{i=1}^{r-1} \left(a + \frac{bi}{r}\right) g_i p_{n-1} g_{r-i}^{(n-1)} \\
&= (a+b) s_0 g_r + \sum_{i=1}^{r-1} \left(a + \frac{bi}{r}\right) g_i \sum_{n=2}^{\infty} p_{n-1} g_{r-i}^{(n-1)} \\
&= (a+b) g_r s_0 + \sum_{i=1}^{r-1} \left(a + \frac{bi}{r}\right) g_i s_{r-i} \\
&= \sum_{i=1}^{r} \left(a + \frac{bi}{r}\right) g_i s_{r-i}.
\end{aligned}$$

□

Remark 13.17. In the case of both the FFT method and the Panjer recursion, an initial discretization of the loss df G generally has to be made, which introduces an approximation error. An in-depth discussion of discretization errors for the computation of compound distributions is to be found in Grübel and Hermesmeier (1999, 2000) (see also references therein for a comparison of these approaches). A slight correction to Theorem 13.16 has to be made if $g_0 = P(X_1 = 0) > 0$. One obtains $s_0 = \sum_{k=0}^{\infty} p_k g_0^k$ and, for $r \geqslant 1$, $s_r = (1 - a g_0)^{-1} \sum_{i=1}^{r} (a + bi/r) g_i s_{r-i}$ (see Mikosch 2004, Theorem 3.3.10). We give further references in Notes and Comments.

Example 13.18 (Panjer recursion for the CPoi(100, LN(1, 1)) case). In Figure 13.5 we have included the Panjer approximation for the CPoi(100, LN(1, 1)) case. In order to apply Theorem 13.16 we first have to discretize the lognormal df. An equispaced discretization of about 0.5 yields the Panjer approximation in Figure 13.5, which is excellent for quantile values around 0.999, relevant for applications. The 99.9% quantile estimate based on the Panjer recursion is 735, a value very close to the GPD estimate. Far out in the tail, beyond 0.999, say, rounding errors become important (the tail drifts off) and one has to be more careful; we give some references in Notes and Comments on how to improve recursive methods far out in the tail.

13.2.4 Poisson Mixtures

Poisson mixture models have been used in both credit and operational risk modelling; for an example in the latter case see Cruz (2002, Section 5.2.2) as well as that book's jacket, which features a negative binomial distribution (a particular Poisson mixture model). Poisson mixtures have been used by actuaries for a long time; the negative binomial made its first appearance in the actuarial literature as the distribution of the number of repeated accidents suffered by an individual in a given time span (see Seal 1969).

In Example 13.5 we introduced the compound Poisson model CPoi(λ, G), where $N \sim \text{Poi}(\lambda)$ counts the number of losses and G is the loss severity df. One disadvantage of the Poisson frequency distribution is that $\text{var}(N) = \lambda = E(N)$, whereas count data often exhibit so-called *overdispersion*, meaning that they indicate a model where $\text{var}(N) > E(N)$. A standard way to achieve this is by mixing the intensity λ over some df $F_\Lambda(\lambda)$, i.e. assume that $\lambda > 0$ is a realization of a positive rv Λ with this df so that, by definition,

$$\begin{aligned} p_N(k) = P(N = k) &= \int_0^\infty P(N = k \mid \Lambda = \lambda)\, \mathrm{d}F_\Lambda(\lambda) \\ &= \int_0^\infty \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!}\, \mathrm{d}F_\Lambda(\lambda). \end{aligned} \tag{13.18}$$

Definition 13.19 (the mixed Poisson distribution). The rv N with df (13.18) is called a *mixed Poisson* rv with *structure* (or *mixing*) distribution F_Λ.

A consequence of the next result is that mixing leads to overdispersion.

Proposition 13.20. *Suppose that N is mixed Poisson with structure df F_Λ. Then $E(N) = E(\Lambda)$ and $\text{var}(N) = E(\Lambda) + \text{var}(\Lambda)$, i.e. for Λ non-degenerate, N is overdispersed.*

Proof. One immediately obtains

$$E(N) = \sum_{k=0}^\infty k p_N(k) = \int_0^\infty \sum_{k=0}^\infty k \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!}\, \mathrm{d}F_\Lambda(\lambda) = \int_0^\infty \lambda\, \mathrm{d}F_\Lambda(\lambda) = E(\Lambda).$$

And, similarly,

$$E(N^2) = \sum_{k=0}^{\infty} k^2 p_N(k) = E(\Lambda) + E(\Lambda^2),$$

so the result follows. □

We now give a concrete example of a mixed Poisson distribution, which is particularly important in both operational risk and credit risk modelling. Indeed we have already used the following result when describing the industry credit risk model CreditRisk$^+$ in Section 11.2.5.

Proposition 13.21 (negative binomial as Poisson mixture). *Suppose that the rv N has a mixed Poisson distribution with a gamma-distributed mixing variable $\Lambda \sim \mathrm{Ga}(\alpha, \beta)$. Then N has a negative binomial distribution $N \sim \mathrm{NB}(\alpha, \beta/(\beta+1))$.*

Proof. Using the definition of a gamma distribution in Section A.2.4 we have

$$P(N=k) = \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\lambda^k}{k!} \mathrm{e}^{-\lambda} \lambda^{\alpha-1} \mathrm{e}^{-\beta\lambda}\, \mathrm{d}\lambda = \frac{\beta^\alpha}{k!\Gamma(\alpha)} \int_0^\infty \lambda^{\alpha+k-1} \mathrm{e}^{-(\beta+1)\lambda}\, \mathrm{d}\lambda.$$

Substituting $u = (\beta+1)\lambda$, the integral can be evaluated to be

$$\int_0^\infty (\beta+1)^{-(\alpha+k)} u^{\alpha+k-1} \mathrm{e}^{-u}\, \mathrm{d}u = \frac{\Gamma(\alpha+k)}{(\beta+1)^{\alpha+k}}.$$

This yields

$$P(N=k) = \left(\frac{\beta}{\beta+1}\right)^\alpha \left(\frac{1}{\beta+1}\right)^k \frac{\Gamma(\alpha+k)}{k!\Gamma(\alpha)}.$$

Using the relation $\Gamma(\alpha+k) = (\alpha+k-1)\cdots\alpha\Gamma(\alpha)$, we see that this is equal to the probability mass function of a negative binomial rv with $p := \beta/(\beta+1)$ (see Section A.2.7). □

Recall the definition of compound sums from Section 13.2.2 (Assumption 13.3 and Proposition 13.4). In the special case of mixed Poisson rvs, compounding leads to so-called *compound mixed Poisson distributions*, such as the compound negative binomial distribution of Example 13.6. There is much literature on dfs of this type (see Notes and Comments).

13.2.5 Tails of Aggregate Loss Distributions

In Section 5.1.2 we defined the class of rvs with regularly varying or power tails. If the (claim size) df G is regularly varying with index $\alpha > 0$, then there exists a slowly varying function L (Definition 5.7) such that $\bar{G}(x) = 1 - G(x) = x^{-\alpha} L(x)$. The next result shows that, for a wide class of counting dfs $(p_N(k))$, the df of the compound sum S_N, F_{S_N}, inherits the power-like behaviour of G.

Theorem 13.22 (power-like behaviour of compound-sum distribution). *Suppose that S_N is a compound sum with $E(N) = \lambda$ and suppose that there exists an $\varepsilon > 0$*

such that $\sum_{k=0}^{\infty}(1+\varepsilon)^k p_N(k) < \infty$. *If* $\bar{G}(x) = x^{-\alpha}L(x)$ *with* $\alpha \geqslant 0$ *and* L *slowly varying, then*

$$\lim_{x\to\infty} \frac{\bar{F}_{S_N}(x)}{\bar{G}(x)} = \lambda,$$

so $\bar{F}_{S_N}$ *inherits the power-like behaviour of* $\bar{G}$.

Proof. This result holds more generally for *subexponential* dfs; a proof together with further discussion can be found in Embrechts, Klüppelberg and Mikosch (1997, Section 1.3.3). □

Example 13.23 (negative binomial). It is not difficult to show that the negative binomial case satisfies the condition on N in Theorem 13.22. The kind of argument that is required is to be found in Embrechts, Klüppelberg and Mikosch (1997, Example 1.3.11). Hence, if $\bar{G}(x) = x^{-\alpha}L(x)$, the tail of the compound-sum df behaves like the tail of G, i.e.

$$\bar{F}_{S_N}(x) \sim \frac{\alpha}{\beta}\bar{G}(x), \quad \text{as } x \to \infty.$$

(For details, see Embrechts, Klüppelberg and Mikosch (1997, Section 1.3.3).)

Under the conditions of Theorem 13.22 the asymptotic behaviour of $\bar{F}_{S_N}(x)$ in the case of a Pareto loss df is again Pareto with the same index. This is clearly seen in Figure 13.6 in the linear behaviour of the simulated losses as well as the fitted GPD. In the case of Figure 13.5, one can show that $\bar{F}_{S_N}(x)$ decays like a lognormal tail; see the reference given in the proof of Theorem 13.22 for details. Note that the GPD is able to pick up the features of the tail in both cases.

13.2.6 The Homogeneous Poisson Process

In the previous sections we looked at counting rvs N over a fixed time interval $[0, 1]$, say. Without any additional difficulty we could have looked at $N(t)$, counting the number of events in $[0, t]$ for $t \geqslant 0$. In the Poisson case this would correspond to $N(t) \sim \text{Poi}(\lambda t)$; hence, for fixed t and on replacing λ by λt, all of the previous results concerning $\text{Poi}(\lambda)$ rvs can be suitably adapted.

In this section we want to integrate the rvs $N(t)$, $t \geqslant 0$, into a stochastic process framework. The less mathematically trained reader should realize that there is a big difference between a family of rvs indexed by time, for which we only specify the one-dimensional dfs (which is what we have done so far), and a stochastic process with a specific structure in which these rvs are embedded. This difference is akin to the difference between marginal and joint distributions, a topic we have highlighted as very important in Chapter 7 through the notion of copulas; of course, in the stochastic process case, there also has to be some probabilistic consistency across time. In a certain sense, the finite-dimensional problem of Chapter 7 becomes an infinite-dimensional problem.

After these words of warning on the difference between rvs and stochastic processes, we now take some methodological shortcuts to arrive at our goal. The interested reader wanting to learn more will have to delve deeper into the mathematical

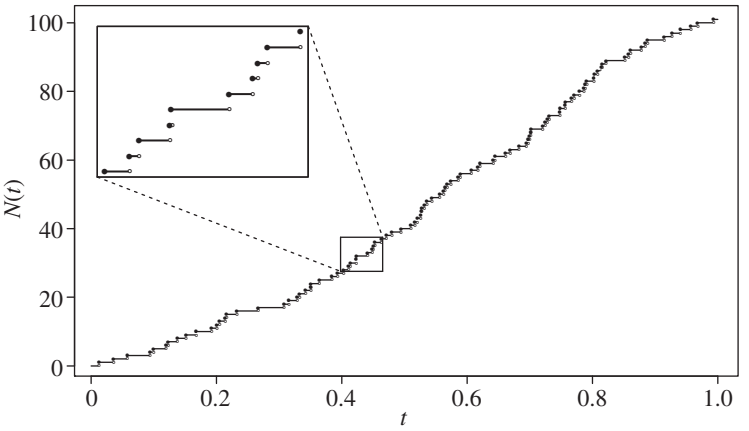

Figure 13.7. A sample path of a counting process.

background of stochastic processes in general and counting processes in particular. The Notes and Comments section contains some references.

Definition 13.24 (counting processes). A stochastic process $N = (N(t))_{t \geqslant 0}$ is a *counting process* if its sample paths are right continuous with left limits existing and if there exists a sequence of rvs $T_0 = 0, T_1, T_2, \ldots$ tending almost surely to ∞ such that $N(t) = \sum_{k=1}^{\infty} I_{\{T_k \leqslant t\}}$.

A typical realization of such a process is given in Figure 13.7. We now define the homogeneous Poisson process as a special counting process.

Definition 13.25 (homogeneous Poisson process). A stochastic process $N = (N(t))_{t \geqslant 0}$ is a *homogeneous Poisson process* with intensity (rate) $\lambda > 0$ if the following properties hold:

(i) N is a counting process;

(ii) $N(0) = 0$, almost surely;

(iii) N has stationary and independent increments; and

(iv) for each $t > 0$, $N(t) \sim \text{Poi}(\lambda t)$.

Remark 13.26. Note that conditions (iii) and (iv) imply that, for $0 < u < v < t$, the rvs $N(v) - N(u)$ and $N(t) - N(v)$ are independent and that, for $k \geqslant 0$,

$$\begin{aligned} P(N(v) - N(u) = k) &= P(N(v-u) = k) \\ &= \mathrm{e}^{-\lambda(v-u)} \frac{(\lambda(v-u))^k}{k!}. \end{aligned}$$

The rv $N(v) - N(u)$ counts the number of events (claims, losses) in the interval $(u, v]$; by stationarity, it has the same df as $N(v-u)$. In Figure 13.8 we have generated ten realizations of a homogeneous Poisson process on $[0, 1]$ with $\lambda = 100$. Note the rather narrow band within which the various sample paths fall.

For practical purposes, the following result contains the main properties of the homogeneous Poisson process.

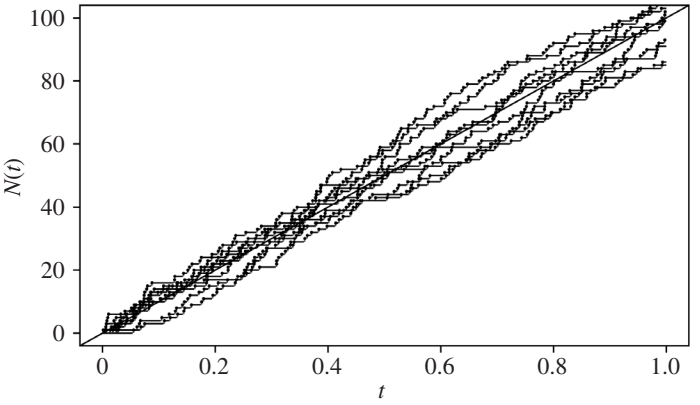

Figure 13.8. Ten realizations of a homogeneous Poisson process with $\lambda = 100$.

Theorem 13.27 (characterizations of the homogeneous Poisson process). *Suppose that N is a counting process. The following statements are then equivalent:*

(1) *N is a homogeneous Poisson process with rate $\lambda > 0$;*

(2) *N has stationary and independent increments and*

$$P(N(t) = 1) = \lambda t + o(t), \quad \text{as } t \downarrow 0,$$
$$P(N(t) \geqslant 2) = o(t), \quad \text{as } t \downarrow 0;$$

(3) *the inter-event-times $(\Delta_k = T_k - T_{k-1})_{k \geqslant 1}$ are iid with distribution* $\mathrm{Exp}(\lambda)$*; and*

(4) *for all $t > 0$, $N(t) \sim \mathrm{Poi}(\lambda t)$ and, given that $N(t) = k$, the occurrence times $T_1, T_2, \ldots, T_k$ have the same distribution as the ordered sample from k independent rvs, uniformly distributed on $[0, t]$; as a consequence, we can write the conditional joint density as*

$$f_{T_1,\ldots,T_k|N(t)=k}(t_1, \ldots, t_k) = \frac{k!}{t^k} I_{\{0<t_1<\cdots<t_k<t\}}.$$

Proof. Many standard textbooks on stochastic processes contain proofs of this important theorem (see, for example, Mikosch 2004; Resnick 1992). □

Discussion. Statement (2) in Theorem 13.27 implies that λ can indeed be interpreted as a rate or intensity: $\lambda = \lim_{t \downarrow 0}(1/t)P(N(t) = 1)$. Moreover, the same statement implies that a homogeneous Poisson process does not allow for clustering of events: $\lim_{t \downarrow 0} P(N(t) \geqslant 2) = 0$. Statement (3) gives an event-time definition of a homogeneous Poisson process. It follows immediately that the first event-time has an $\mathrm{Exp}(\lambda)$ df: $P(T_1 > t) = P(N(t) = 0) = \mathrm{e}^{-\lambda t}$, $t \geqslant 0$. Statement (3), however, goes well beyond this by stating that the inter-event-times Δ_k are iid with $\Delta_k \sim \mathrm{Exp}(\lambda)$. This leads to a straightforward way of simulating a stream of loss events from a homogeneous Poisson process with rate λ. Moreover, this equivalent definition immediately yields a generalization by assuming that the Δ_k are still iid

but that $\Delta_k \sim F_\Delta$, a general df. The resulting process is a so-called *renewal process* (note that the only Markovian renewal process is the homogeneous Poisson process).

Finally, statement (4) yields an easy algorithm to generate the occurrences of homogeneous Poisson times over the interval $[0, t]$ given that we have a total of k events up to t—we simply generate k uniform rvs on $[0, t]$ and order them.

Multivariate Poisson processes. In many applications we want to model the frequencies of different loss types with a number of Poisson processes while considering possible dependence between loss frequencies for different loss types. More generally, we might want to construct a number of compound Poisson processes where loss severities for the different business lines were also dependent. A natural approach to modelling this dependence is to assume that all losses can be related to a series of underlying and independent Poisson *shock* processes. In insurance these shocks might be natural catastrophes; in credit risk modelling they might be a variety of economic events, such as local or global recessions; in operational risk modelling they might be the failure of various IT systems. When a shock occurs this may cause losses of several different types; the common shock causes the numbers of losses of each type to be dependent. See Lindskog and McNeil (2003), Pfeifer and Nešlehová (2004) and Chavez-Demoulin and Embrechts (2004) for models of this kind.

13.2.7 Processes Related to the Poisson Process

Using the fundamental building block of the homogeneous Poisson process, one can construct more general counting processes that are useful for loss-event modelling in finance and insurance. Such generalizations include the following.

Renewal processes (mentioned above). The exponential waiting time distribution is replaced by a general df F_Δ.

Inhomogeneous Poisson processes. The constant intensity λ is replaced by a deterministic function $\lambda(\cdot)$.

Mixed Poisson processes. The deterministic constant intensity λ is replaced by an rv Λ.

Doubly stochastic or Cox processes. λ is replaced by a stochastic process $\{\lambda_t : t \geqslant 0\}$ in accordance with notation used in Chapter 10 (see, for example, Definition 10.15).

Self-exciting or Hawkes processes. λ is replaced by a stochastic process depending only on previous event-times. See Section 16.2.1 for a concrete example.

Below, we highlight some features of some of these processes.

Inhomogeneous Poisson processes.

Definition 13.28 (inhomogeneous Poisson). A counting process N is an *inhomogeneous Poisson process* if, for some deterministic function $\lambda(s) \geqslant 0$, the following conditions hold:

(i) $N(0) = 0$, almost surely;

(ii) N has independent increments; and

(iii) for all $t \geqslant 0$,

$$P(N(t+h) - N(t) = 1) = \lambda(t)h + o(h), \quad h \downarrow 0,$$
$$P(N(t+h) - N(t) \geqslant 2) = o(h), \quad h \downarrow 0.$$

The function $\lambda(\cdot)$ is referred to as the *intensity* or *rate function*. The integral $\Lambda(t) = \int_0^t \lambda(s)\,\mathrm{d}s$ is referred to as the *intensity measure* (or *cumulative intensity function*).

Remark 13.29. A characterization theorem, similar to Theorem 13.27, can be derived. In particular, we find that, for $0 < s < t$, $N(t) - N(s) \sim \mathrm{Poi}(\Lambda(t) - \Lambda(s))$.

The inhomogeneous Poisson process is a useful tool in loss modelling whenever a deterministic trend or seasonality component is to be modelled in the loss frequency. The next example also shows that this process naturally emerges as a counting process for record losses.

Example 13.30 (records). The world of finance and insurance abounds with statements on *record events*: the largest single-day drop in the dollar/yen, the most expensive hurricane, the three best fund managers during the last year, the second largest loss due to internal fraud, the biggest one-day change in the credit spread of a particular company, etc. Likewise, the world of records is intimately related to the (general) theory of Poisson processes. In Notes and Comments we shall give several references for this. Below we indicate how an easy example related to a question on records leads to an inhomogeneous Poisson process as a model.

Suppose that the loss rvs $X_i \geqslant 0$ are iid with density function $f(x) > 0$, $x \geqslant 0$. Define the counting process N:

$$N(t) = \sum_{i=1}^{\infty} I_{\{X_i \leqslant t \text{ and } X_i > X_{i-j},\ j=1,\ldots,i-1\}}.$$

$N(t)$ counts the number of records in the sequence $(X_i)_{i\geqslant 1}$ of size less than t, and $(N(t))$ is referred to as the *record process*. It follows that, for $h, t > 0$,

$$\begin{aligned} P(N(t+h) - N(t) \geqslant 1) &= \sum_{i=1}^{\infty} P(X_i \in (t, t+h] \text{ and } X_{i-1} \leqslant t, \ldots, X_1 \leqslant t) \\ &= \sum_{i=1}^{\infty} (F(t+h) - F(t))(F(t))^{i-1} \\ &= \frac{F(t+h) - F(t)}{1 - F(t)} \\ &= \frac{f(t)}{1 - F(t)} h + o(h), \quad \text{as } h \downarrow 0. \end{aligned}$$

Moreover, for $h, t > 0$,

$$\begin{aligned} P(N(t+h) - N(t) \geqslant 2) &\leqslant \sum_{i<j} P(X_1 \leqslant t, \ldots, X_{i-1} \leqslant t,\ X_i \in (t, t+h], \\ &\qquad X_{i+1} \leqslant t+h, \ldots, X_{j-1} \leqslant t+h,\ X_j \in (t, t+h]) \\ &= \left(\int_t^{t+h} f(s)\,\mathrm{d}s \right)^2 \sum_{i<j} (F(t))^{i-1} (F(t))^{j-i-1} \\ &= o(h^2), \quad \text{as } h \downarrow 0. \end{aligned}$$

From these calculations one deduces that the record process N is inhomogeneous Poisson with rate function $\lambda(t) = f(t)/(1 - F(t))$, the so-called *hazard rate* of F, a notion that we encountered in Section 10.4.1.

Suppose now that, as in most practical cases, $\Lambda(t)$ is strictly increasing, so $\Lambda(\Lambda^{-1}(t)) = \Lambda^{-1}(\Lambda(t)) = t$. We can then always transform an inhomogeneous Poisson process N with intensity measure Λ into a homogeneous Poisson process with intensity 1 by a *change of time*.

Proposition 13.31 (time change, operational time). *Suppose that N is an inhomogeneous Poisson process with Λ strictly increasing, and define, for $t \geqslant 0$, $\tilde{N}(t) = N(\Lambda^{-1}(t))$. We then have that $\tilde{N}$ is homogeneous Poisson with intensity* 1.

Proof. For $t > 0$ fixed and $k \geqslant 0$,

$$P(\tilde{N}(t) = k) = P(N(\Lambda^{-1}(t)) = k) = \mathrm{e}^{-\Lambda(\Lambda^{-1}(t))} \frac{(\Lambda(\Lambda^{-1}(t)))^k}{k!} = \mathrm{e}^{-t} \frac{t^k}{k!},$$

so $\tilde{N}(t) \sim \mathrm{Poi}(t)$. By definition, the increments of $\tilde{N}$ are independent; moreover, for $0 < u < v$ we have that

$$\begin{aligned} P(\tilde{N}(v) - \tilde{N}(u) = k) &= P(N(\Lambda^{-1}(v)) - N(\Lambda^{-1}(u)) = k) \\ &= \mathrm{e}^{-(\Lambda(\Lambda^{-1}(v)) - \Lambda(\Lambda^{-1}(u)))} \frac{(\Lambda(\Lambda^{-1}(v)) - \Lambda(\Lambda^{-1}(u)))^k}{k!} \\ &= \mathrm{e}^{-(v-u)} \frac{(v-u)^k}{k!}, \end{aligned}$$

from which stationarity follows. □

This is one of the many examples in insurance and finance where a more complicated process N can be reduced to a standard (easier) model $\tilde{N}$ through the careful choice of a new time clock (a so-called *time change construction*) (see also Section 10.5.1 on credit risk). Proposition 13.31 can be formulated more generally for Λ not strictly increasing, and the converse also holds. Proposition 13.31 justifies the common simplifying assumption that a loss frequency model is homogeneous (unit rate) Poisson, albeit in many cases only in operational time. The original time-scale of N is slowed down or speeded up in such a way that, on average, $\tilde{N}$ has one claim per time unit, whereas N has, on average, $\Lambda(1)$ claims.

Remark 13.32. A standard way in which an inhomogeneous Poisson process can be obtained from a homogeneous Poisson process is by random sampling. Suppose an intensity function λ satisfies $\lambda(s) \leqslant c < \infty$ for $s \geqslant 0$. Start from a homogeneous Poisson process with rate $c > 0$ and denote its arrival times by $T_0 = 0, T_1, T_2, \ldots$. Construct a new process $\tilde{N}$ from $(T_i)_{i \geqslant 0}$ through deletion of each T_i independently of the other T_j with probability $1 - (\lambda(T_i)/c)$. The so-called *thinned* counting process $\tilde{N}$ consists of the remaining (undeleted) points. It can be shown that this process is inhomogeneous Poisson with intensity function $\lambda(\cdot)$.

Mixed Poisson processes. The mixed Poisson rvs of Section 13.2.4 can be embedded into a so-called *mixed Poisson process*. A single realization of such a process cannot be distinguished through statistical means from a realization of a homogeneous Poisson process; indeed, to simulate a sample path, one first draws a realization of the random intensity $\lambda = \Lambda(\omega)$ and then draws the sample path of the homogeneous Poisson process with rate λ. (Here, Λ denotes an rv and not the intensity measure in the inhomogeneous Poisson case above.) Only by repeating this simulation more frequently does one see the different probabilistic nature of the mixed Poisson process: compare Figure 13.9 with Figure 13.8. In the former we have simulated ten sample paths from a mixed Poisson process with mixing variable $\Lambda \sim \mathrm{Ga}(100, 1)$ so that $E(\Lambda) = 100$. Note the much greater variability in the paths.

Example 13.33. When counting processes are used in credit risk modelling the times T_k typically correspond to credit events, for instance default or downgradings. More precisely, a credit event can be constructed as the first jump of a counting process N. The df of the time to the credit event can be easily derived by observing that $P(T_1 > t) = P(N(t) = 0)$. This probability can be calculated in a straightforward way for a homogeneous Poisson process with intensity λ; we obtain $P(N(t) = 0) = \mathrm{e}^{-\lambda t}$. When N is a mixed Poisson process with mixing df F_Λ we obtain

$$P(T_1 > t) = P(N(t) = 0) = \int_0^\infty \mathrm{e}^{-t\lambda} \, \mathrm{d}F_\Lambda(\lambda) = \hat{F}_\Lambda(t),$$

the Laplace–Stieltjes transform of F_Λ in t. In the special case when $\Lambda \sim \mathrm{Ga}(\alpha, \beta)$, the negative binomial case treated in Proposition 13.21, one finds that

$$\begin{aligned} P(T_1 > t) &= \int_0^\infty \mathrm{e}^{-t\lambda} \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} \mathrm{e}^{-\beta\lambda} \, \mathrm{d}\lambda \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} (t+\beta)^{-\alpha} \int_0^\infty \mathrm{e}^{-s} s^{\alpha-1} \, \mathrm{d}s \\ &= \beta^\alpha (t+\beta)^{-\alpha}, \quad t \geqslant 0, \end{aligned}$$

so that T_1 has a Pareto distribution $T_1 \sim \mathrm{Pa}(\alpha, \beta)$ (see Section A.2.8).

Processes with stochastic intensity. A further important class of models is obtained when λ in the homogeneous Poisson case is replaced by a general stochastic process (λ_t), yielding a two-tier stochastic model or so-called *doubly stochastic* process.

For example, one could take λ_t to be a diffusion or, alternatively, a finite-state Markov chain. The latter case gives rise to a *regime-switching* model: in each state of

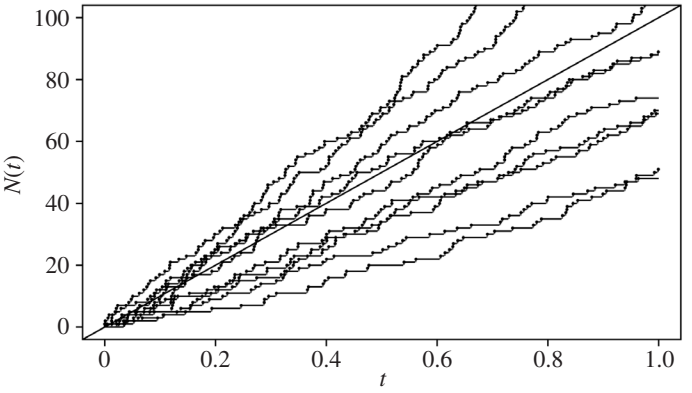

Figure 13.9. Ten realizations of a mixed Poisson process with $\Lambda \sim \text{Ga}(100, 1)$.

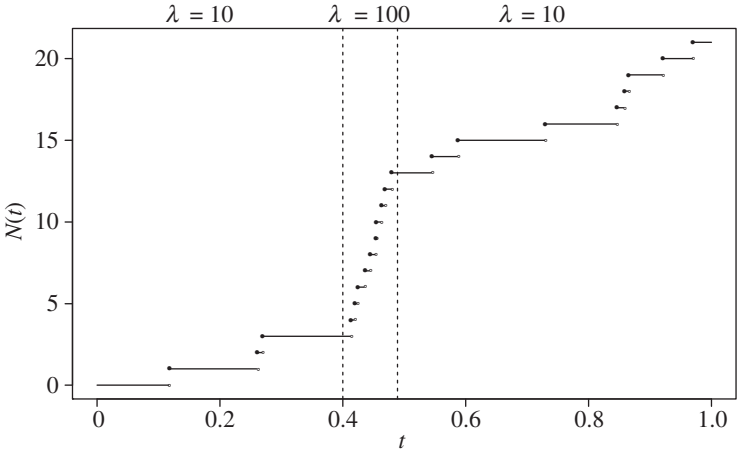

Figure 13.10. Realization of a counting process with a regime switch from $\lambda = 10$ to $\lambda = 100$.

the Markov chain the intensity has a different constant level and the process remains in that state for an exponential length of time, before jumping to another state. In Figure 13.10 we have simulated the sample path of such a process randomly switching between $\lambda = 10$ and $\lambda = 100$. In Section 10.5.1 we looked at doubly stochastic random times, which correspond to the first jump of a doubly stochastic Poisson process.

Notes and Comments

The story behind the name *insurance analytics* is told in Embrechts (2002). A good place to start a search for actuarial literature is the website of the International Actuarial Association: www.actuaries.org. Several interesting books can be found on the website of the Society of Actuaries, www.soa.org (whose offices happen to be on North *Martingale* Road, Schaumburg, Illinois). A standard Society of Actuaries textbook on actuarial mathematics is Bowers et al. (1986); financial economics for

actuaries is to be found in Panjer and Boyle (1998). For our purposes, excellent texts are Mikosch (2004) and Partrat and Besson (2004). A most readable and informative (online) set of lecture notes, with many historical links, is Wüthrich (2013). Rolski et al. (1999) gives a broad, more technical overview of the relevant stochastic process models. In Chapters 7 and 8 we have given several references to actuarial tools relevant to the study of risk measures, including distortion risk measures (Section 8.2.1), comonotonicity (Section 7.2.1) and Fréchet bounds (Sections 7.2.1 and 8.4.4). Finally, an overview of the state of the art of actuarial modelling is to be found in Teugels and Sundt (2004).

Actuarial textbooks dealing in particular with the modelling of loss distributions in insurance are Hogg and Klugman (1984), Klugman, Panjer and Willmot (2012) and Klugman, Panjer and Willmot (2013). Besides the general references above, an early textbook discussion of the use of numerical methods for the calculation of the df of total loss amount rvs is Feilmeier and Bertram (1987); Bühlmann (1984) contains a first comparison between the FFT method and Panjer recursion. More extensive comparisons, taking rounding and discretization errors into account, are found in Grübel and Hermesmeier (1999, 2000). A discussion of the use of the FFT in insurance is given in Embrechts, Gruebel and Pitts (1993). Algorithms for the FFT are freely available on the web, as a search will quickly reveal. The original paper by Panjer (1981) also contains a density version of Theorem 13.16. For an application of Panjer recursion to credit risk measurement within the CreditRisk$^+$ framework, see Credit Suisse Financial Products (1997). Based on Giese (2003), Haaf, Reiss and Schoenmakers (2004) propose an alternative recursive method. For more recent work on Panjer recursion, especially in the multivariate case, see, for example, Hesselager (1996) and Sundt (1999, 2000). For a textbook treatment of recursive methods in insurance, see Sundt and Vernic (2009).

Asymptotic approximation methods going beyond the normal approximation (13.13) are known in statistics under the names Berry–Esséen, Edgeworth and *saddle-point*. The former two are discussed, for example, in Embrechts, Klüppelberg and Mikosch (1997) and are of more theoretical importance. The saddle-point technique is very useful: see Jensen (1995) for an excellent summary, and Embrechts et al. (1985) for an application to compound distributions. Gordy (2002) discusses the importance of saddle-point methods for credit risk modelling, again within the context of CreditRisk$^+$. Wider applications within risk management can be found in Studer (2001) and Glasserman (2004); Glasserman and Li (2003)

Poisson mixture models with insurance applications in mind are summarized in Grandell (1997) (see also Bening and Korolev 2002). In order to delve more deeply into the world of counting processes, one has to study the theory of point processes. Very comprehensive and readable accounts are Daley and Vere-Jones (2003) and Karr (1991). A study of this theory is both mathematically demanding and practically rewarding. Such models are being used increasingly in credit risk. The notion of time change is fundamental to many applications in insurance and finance; for an example of how it can be used to model operational risk, see Embrechts, Kaufmann and Samorodnitsky (2004). For its introduction into finance, see Ané and Geman

(2000) and Dacorogna et al. (2001). An excellent survey is to be found in Peeters (2004).

What have we not included in our brief account of the elements of insurance analytics? We have not treated ruin theory and the general stochastic process theory of insurance risk, credibility theory, dynamic financial analysis, also referred to as dynamic solvency testing, or reinsurance, to name but a few omissions.

The stochastic process theory of insurance risk has a long tradition. The first fundamental summary came through the pioneering work of Cramér (1994a,b). Bühlmann (1970) made the field popular to several generations of actuaries. This early work has now been generalized in every way possible. A standard textbook on *ruin theory* is Asmussen and Albrecher (2010). The modelling of large claims and its consequences for ruin estimates can be found in Embrechts, Klüppelberg and Mikosch (1997).

Credibility theory concerns premium calculation for non-homogeneous portfolios and has a very rich history rooted in non-life insurance mathematics. Its basic concepts were first developed by American actuaries in the 1920s; pioneering papers in this early period were Mowbray (1914) and Whitney (1918). Further important work is found in the papers of Bailey (1945), Robbins (1955, 1964) and Bühlmann (1967, 1969, 1971). An excellent review article tracing the historical development of the basic ideas is Norberg (1979); see also Jewell (1990) for a more recent review. Various textbook versions exist: Bühlmann and Gisler (2005) give an authoritative account of its actuarial usage and hint at applications to financial risk management.

Dynamic financial analysis, also referred to as *dynamic solvency testing*, is a systematic approach, based on large-scale computer simulations, for the integrated financial modelling of non-life insurance and reinsurance companies aimed at assessing the risks and benefits associated with strategic decisions (see Blum 2005; Blum and Dacorogna 2004). An easy introduction can be found in Kaufmann, Gadmer and Klett (2001). The interested reader can consult the website of the Casualty Actuarial Society (www.casact.org/research/drm).

Part IV

Special Topics

14
Multivariate Time Series

In this chapter we consider multivariate time-series models for multiple series of financial risk-factor change data, such as differenced logarithmic price series. The presentation closely parallels the presentation of the corresponding ideas for univariate time series in Chapter 4.

An introduction to concepts in the analysis of multivariate time series and a discussion of multivariate ARMA models is found in Section 14.1, while Section 14.2 presents some of the more important examples of multivariate GARCH models.

14.1 Fundamentals of Multivariate Time Series

Among the fundamentals of multivariate time series discussed in this section are the concepts of cross-correlation, stationarity of multivariate time series, multivariate white noise processes and multivariate ARMA models.

14.1.1 Basic Definitions

A multivariate time-series model for multiple risk factors is a stochastic process $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$, i.e. a family of random vectors, indexed by the integers and defined on some probability space $(\Omega, \mathcal{F}, P)$.

Moments of a multivariate time series. Assuming they exist, we define the *mean function* $\boldsymbol{\mu}(t)$ and the *covariance matrix function* $\Gamma(t,s)$ of $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ by

$$\begin{aligned} \boldsymbol{\mu}(t) &= E(\boldsymbol{X}_t), && t \in \mathbb{Z}, \\ \Gamma(t,s) &= E((\boldsymbol{X}_t - \boldsymbol{\mu}(t))(\boldsymbol{X}_s - \boldsymbol{\mu}(s))'), && t,s \in \mathbb{Z}. \end{aligned}$$

Analogously to the univariate case, we have $\Gamma(t,t) = \operatorname{cov}(\boldsymbol{X}_t)$. By observing that the elements $\gamma_{ij}(t,s)$ of $\Gamma(t,s)$ satisfy

$$\gamma_{ij}(t,s) = \operatorname{cov}(X_{t,i}, X_{s,j}) = \operatorname{cov}(X_{s,j}, X_{t,i}) = \gamma_{ji}(s,t), \tag{14.1}$$

it is clear that $\Gamma(t,s) = \Gamma(s,t)'$ for all t, s. However, the matrix Γ need not be symmetric, so in general $\Gamma(t,s) \neq \Gamma(s,t)$, which is in contrast to the univariate case. Lagged values of one of the component series can be more strongly correlated with future values of another component series than vice versa. This property, when observed in empirical data, is known as a *lead–lag* effect and is discussed in more detail in Example 14.7.

Stationarity. The multivariate models we consider will be stationary in one or both of the following senses.

Definition 14.1 (strict stationarity). The multivariate time series $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ is *strictly* stationary if

$$(\boldsymbol{X}'_{t_1}, \ldots, \boldsymbol{X}'_{t_n}) \stackrel{\mathrm{d}}{=} (\boldsymbol{X}'_{t_1+k}, \ldots, \boldsymbol{X}'_{t_n+k})$$

for all $t_1, \ldots, t_n, k \in \mathbb{Z}$ and for all $n \in \mathbb{N}$.

Definition 14.2 (covariance stationarity). The multivariate time series $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ is *covariance* stationary (or *weakly* or *second-order* stationary) if the first two moments exist and satisfy

$$\begin{aligned} \boldsymbol{\mu}(t) &= \boldsymbol{\mu}, & t &\in \mathbb{Z}, \\ \Gamma(t,s) &= \Gamma(t+k, s+k), & t, s, k &\in \mathbb{Z}. \end{aligned}$$

A strictly stationary multivariate time series with finite covariance matrix is covariance stationary, but we note that, as in the univariate case, it is possible to define infinite-variance processes (including certain multivariate ARCH and GARCH processes) that are strictly stationary but not covariance stationary.

Serial correlation and cross-correlation in stationary multivariate time series. The definition of covariance stationarity implies that for all s, t we have $\Gamma(t-s, 0) = \Gamma(t, s)$, so that the covariance between X_t and X_s only depends on their temporal separation $t - s$, which is known as the *lag*. In contrast to the univariate case, the sign of the lag is important.

For a covariance-stationary multivariate process we write the covariance matrix function as a function of one variable: $\Gamma(h) := \Gamma(h, 0), \forall h \in \mathbb{Z}$. Noting that $\Gamma(0) = \mathrm{cov}(\boldsymbol{X}_t)$, $\forall t$, we can now define the correlation matrix function of a covariance-stationary process. For this definition we recall the operator $\Delta(\cdot)$, defined in (6.4), which when applied to $\Sigma = (\sigma_{ij}) \in \mathbb{R}^{d\times d}$ returns a diagonal matrix with the values $\sqrt{\sigma_{11}}, \ldots, \sqrt{\sigma_{dd}}$ on the diagonal.

Definition 14.3 (correlation matrix function). Writing $\Delta := \Delta(\Gamma(0))$, where $\Delta(\cdot)$ is the operator defined in (6.4), the correlation matrix function $P(h)$ of a covariance-stationary multivariate time series $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ is

$$P(h) := \Delta^{-1}\Gamma(h)\Delta^{-1}, \quad \forall h \in \mathbb{Z}. \tag{14.2}$$

The diagonal entries $\rho_{ii}(h)$ of this matrix-valued function give the autocorrelation function of the ith component series $(X_{t,i})_{t\in\mathbb{Z}}$. The off-diagonal entries give so-called cross-correlations between different component series at different times. It follows from (14.1) that $P(h) = P(-h)'$, but $P(h)$ need not be symmetric, and in general $P(h) \neq P(-h)$.

White noise processes. As in the univariate case, *multivariate white noise* processes are building blocks for practically useful classes of time-series model.

Definition 14.4 (multivariate white noise). $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ is multivariate white noise if it is covariance stationary with correlation matrix function given by

$$P(h) = \begin{cases} P, & h = 0, \\ 0, & h \neq 0, \end{cases}$$

for some positive-definite correlation matrix P.

A multivariate white noise process with mean 0 and covariance matrix $\Sigma = \text{cov}(\boldsymbol{X}_t)$ will be denoted by WN($\mathbf{0}$, Σ). Such a process has no cross-correlation between component series, except for contemporaneous cross-correlation at lag zero. A simple example is a series of iid random vectors with finite covariance matrix, and this is known as a *multivariate strict white noise*.

Definition 14.5 (multivariate strict white noise). $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ is multivariate strict white noise if it is a series of iid random vectors with finite covariance matrix.

A strict white noise process with mean 0 and covariance matrix Σ will be denoted by SWN($\mathbf{0}$, Σ).

The martingale-difference noise concept may also be extended to higher dimensions. As before we assume that the time series $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ is adapted to some *filtration* $(\mathcal{F}_t)$, typically the natural filtration $(\sigma(\{\boldsymbol{X}_s : s \leqslant t\}))$, which represents the information available at time t.

Definition 14.6 (multivariate martingale difference). $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ has the multivariate martingale-difference property with respect to the filtration $(\mathcal{F}_t)$ if $E|\boldsymbol{X}_t| < \infty$ and

$$E(\boldsymbol{X}_t \mid \mathcal{F}_{t-1}) = \mathbf{0}, \quad \forall t \in \mathbb{Z}.$$

The unconditional mean of such a process is obviously also zero and, if $\text{cov}(\boldsymbol{X}_t) < \infty$ for all t, the covariance matrix function satisfies $\Gamma(t, s) = 0$ for $t \neq s$. If the covariance matrix is also constant for all t, then a process with the multivariate martingale-difference property is a multivariate white noise process.

14.1.2 Analysis in the Time Domain

We now assume that we have a random sample $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ from a covariance-stationary multivariate time-series model $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$. In the time domain we construct empirical estimators of the covariance matrix function and the correlation matrix function from this random sample.

The *sample covariance matrix function* is calculated according to

$$\hat{\Gamma}(h) = \frac{1}{n}\sum_{t=1}^{n-h}(\boldsymbol{X}_{t+h} - \bar{\boldsymbol{X}})(\boldsymbol{X}_t - \bar{\boldsymbol{X}})', \quad 0 \leqslant h < n,$$

where $\bar{\boldsymbol{X}} = \sum_{t=1}^{n} \boldsymbol{X}_t/n$ is the sample mean, which estimates $\boldsymbol{\mu}$, the mean of the time series. Writing $\hat{\Delta} := \Delta(\hat{\Gamma}(0))$, where $\Delta(\cdot)$ is the operator defined in (6.4), the *sample correlation matrix function* is

$$\hat{P}(h) = \hat{\Delta}^{-1}\hat{\Gamma}(h)\hat{\Delta}^{-1}, \quad 0 \leqslant h < n.$$

The information contained in the elements $\hat{\rho}_{ij}(h)$ of the sample correlation matrix function is generally displayed in the *cross-correlogram*, which is a $d \times d$ matrix of plots (see Figure 14.1 for an example). The ith diagonal plot in this graphical display is simply the correlogram of the ith component series, given by $\{(h, \hat{\rho}_{ii}(h)) : h = 0, 1, 2, \dots\}$. For the off-diagonal plots containing the estimates of *cross-correlation* there are various possible presentations and we will consider the following convention: for $i < j$ we plot $\{(h, \hat{\rho}_{ij}(h)) : h = 0, 1, 2, \dots\}$; for $i > j$ we plot $\{(-h, \hat{\rho}_{ij}(h)) : h = 0, 1, 2, \dots\}$. An interpretation of the meaning of the off-diagonal pictures is given in Example 14.7.

It can be shown that for causal processes driven by multivariate strict white noise innovations (see Section 14.1.3) the estimates that comprise the components of the sample correlation matrix function $\hat{P}(h)$ are consistent estimates of the underlying theoretical quantities. For example, if the data themselves are from an SWN process, then the cross-correlation estimators $\hat{\rho}_{ij}(h)$ for $h \neq 0$ converge to zero as the sample size is increased. However, results concerning the asymptotic distribution of cross-correlation estimates are, in general, more complicated than the univariate result for autocorrelation estimates given in Theorem 4.13. Some relevant theory is found in Chapter 11 of Brockwell and Davis (1991) and Chapter 7 of Brockwell and Davis (2002). It is standard to plot the off-diagonal pictures with Gaussian confidence bands at $(-1.96/\sqrt{n}, 1.96/\sqrt{n})$, but these bands should be used as rough guidance for the eye rather than being relied upon too heavily to draw conclusions.

Example 14.7 (cross-correlogram of trivariate index returns). In Figure 14.1 the cross-correlogram of daily log-returns is shown for the Dow Jones, Nikkei and Swiss Market indices for 26 July 1996–25 July 2001. Although every vector observation in this trivariate time series relates to the same trading day, the returns are of course not properly synchronized due to time zones. This picture therefore shows interpretable lead–lag effects that help us to understand the off-diagonal pictures in the cross-correlogram.

Part (b) of the figure shows estimated correlations between the Dow Jones index return on day $t + h$ and the Nikkei index return on day t, for $h \geqslant 0$; these estimates are clearly small and lie mainly within the confidence band, with the obvious exception of the correlation estimate for returns on the same trading day $\hat{P}_{12}(0) \approx 0.14$. Part (d) shows estimated correlations between the Dow Jones index return on day $t + h$ and the Nikkei index return on day t, for $h \leqslant 0$; the estimate corresponding to $h = -1$ is approximately 0.28 and can be interpreted as showing how the American market leads the Japanese market. Comparing parts (c) and (g) we see, unsurprisingly, that the American market also leads the Swiss market, so that returns on day $t - 1$ in the former are quite strongly correlated with returns on day t in the latter.

14.1.3 Multivariate ARMA Processes

We provide a brief excursion into multivariate ARMA models to indicate how the ideas of Section 4.1.2 generalize to higher dimensions. For daily data, capturing multivariate ARMA effects is much less important than capturing multivariate volatility effects (and dynamic correlation effects) through multivariate GARCH

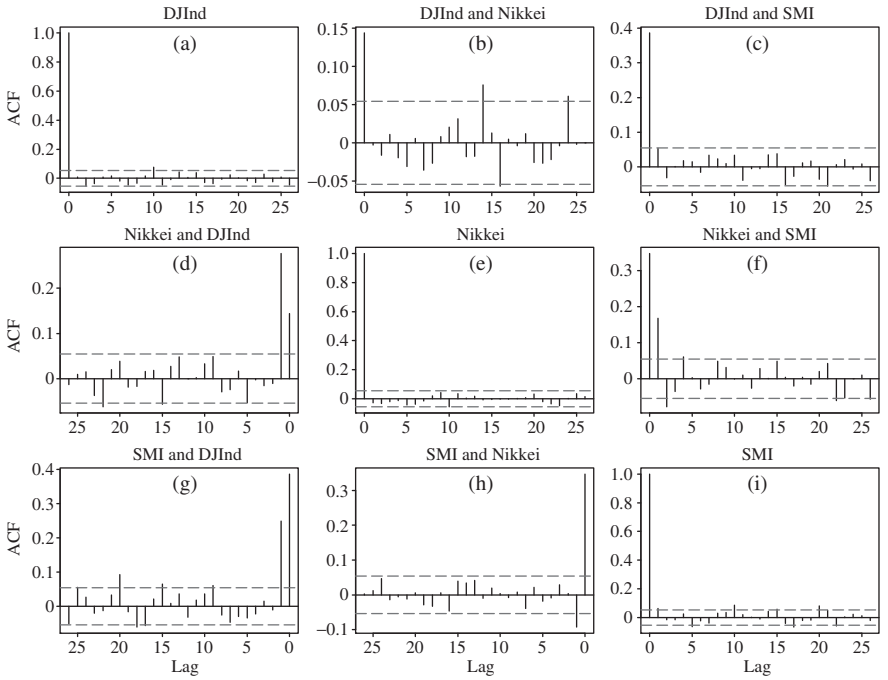

Figure 14.1. Cross-correlogram of daily log-returns of the Dow Jones, Nikkei and Swiss Market indices for 26 July 1996–25 July 2001 (see Example 14.7 for commentary).

modelling, but, for longer-period returns, the more traditional ARMA processes become increasingly useful. In the econometrics literature they are more commonly known as vector ARMA (or VARMA) processes.

Definition 14.8 (VARMA process). Let $(\boldsymbol{\varepsilon}_t)_{t\in\mathbb{Z}}$ be WN$(\mathbf{0}, \Sigma_\varepsilon)$. The process $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ is a zero-mean VARMA(p, q) process if it is a covariance-stationary process satisfying difference equations of the form

$$\boldsymbol{X}_t - \Phi_1 \boldsymbol{X}_{t-1} - \cdots - \Phi_p \boldsymbol{X}_{t-p} = \boldsymbol{\varepsilon}_t + \Theta_1 \boldsymbol{\varepsilon}_{t-1} + \cdots + \Theta_q \boldsymbol{\varepsilon}_{t-q}, \quad \forall t \in \mathbb{Z},$$

for parameter matrices Φ_i and Θ_j in $\mathbb{R}^{d\times d}$. $(\boldsymbol{X}_t)$ is a VARMA process with mean $\boldsymbol{\mu}$ if the centred series $(\boldsymbol{X}_t - \boldsymbol{\mu})_{t\in\mathbb{Z}}$ is a zero-mean VARMA(p, q) process.

Consider a zero-mean VARMA(p, q) process. For practical applications we again consider only causal processes (see Section 4.1.2), which are processes where the solution of the defining equations has a representation of the form

$$\boldsymbol{X}_t = \sum_{i=0}^{\infty} \Psi_i \boldsymbol{\varepsilon}_{t-i}, \tag{14.3}$$

where $(\Psi_i)_{i\in\mathbb{N}_0}$ is a sequence of matrices in $\mathbb{R}^{d\times d}$ with absolutely summable components, meaning that, for any j and k,

$$\sum_{i=0}^{\infty} |\psi_{i,jk}| < \infty. \tag{14.4}$$

As in the univariate case (see Proposition 4.9) it can be verified by direct calculation that such linear processes are covariance stationary. For $h \geqslant 0$ the covariance matrix function is given by

$$\Gamma(t+h,t) = \operatorname{cov}(\boldsymbol{X}_t + h, \boldsymbol{X}_t) = E\bigg(\sum_{i=0}^{\infty} \Psi_i \boldsymbol{\varepsilon}_{t+h-i} \sum_{j=0}^{\infty} \boldsymbol{\varepsilon}'_{t-j} \Psi'_j\bigg).$$

Arguing much as in the univariate case it is easily shown that this depends only on h and not on t and that it is given by

$$\Gamma(h) = \sum_{i=0}^{\infty} \Psi_{i+h} \Sigma_\varepsilon \Psi'_i, \quad h = 0, 1, 2, \ldots. \tag{14.5}$$

The correlation matrix function is easily derived from (14.5) and (14.2).

The requirement that a VARMA process be causal imposes conditions on the values that the parameter matrices Φ_i (in particular) and Θ_j may take. The theory is quite similar to univariate ARMA theory. We will give a single useful example from the VARMA class; this is the first-order vector autoregressive (or VAR(1)) model.

Example 14.9 (VAR(1) process). The first-order VAR process satisfies the set of vector difference equations

$$\boldsymbol{X}_t = \Phi \boldsymbol{X}_{t-1} + \boldsymbol{\varepsilon}_t, \quad \forall t. \tag{14.6}$$

It is possible to find a causal process satisfying (14.3) and (14.4) that is a solution of (14.6) if and only if all eigenvalues of the matrix Φ are less than 1 in absolute value. The causal process

$$\boldsymbol{X}_t = \sum_{i=0}^{\infty} \Phi^i \boldsymbol{\varepsilon}_{t-i} \tag{14.7}$$

is then the unique solution. This solution can be thought of as an infinite-order vector moving-average process, a so-called VMA(∞) process. The covariance matrix function of this process follows from (14.3) and (14.5) and is given by

$$\Gamma(h) = \sum_{i=0}^{\infty} \Phi^{i+h} \Sigma_\varepsilon \Phi^{i\prime} = \Phi^h \Gamma(0), \quad h = 0, 1, 2, \ldots.$$

In practice, full VARMA models are less common than models from the VAR and VMA subfamilies, one reason being that identifiability problems arise when estimating parameters. For example, we can have situations where the first-order VARMA(1, 1) model $\boldsymbol{X}_t - \Phi \boldsymbol{X}_{t-1} = \boldsymbol{\varepsilon}_t + \Theta \boldsymbol{\varepsilon}_{t-1}$ can be rewritten as $\boldsymbol{X}_t - \Phi^* \boldsymbol{X}_{t-1} = \boldsymbol{\varepsilon}_t + \Theta^* \boldsymbol{\varepsilon}_{t-1}$ for completely different parameter matrices Φ^* and Θ^* (see Tsay (2002, p. 323) for an example). Of the two subfamilies, VAR models are easier to estimate. Fitting options for VAR models range from multivariate least-squares estimation without strong assumptions concerning the distribution of the driving white noise, to full ML estimation; models combining VAR and multivariate GARCH features can be estimated using a conditional ML approach in a very similar manner to that described for univariate models in Section 4.2.4.

Notes and Comments

Many standard texts on time series also handle the multivariate theory: see, for example, Brockwell and Davis (1991, 2002) or Hamilton (1994). A key reference aimed at an econometrics audience is Lütkepohl (1993). For examples in the area of finance see Tsay (2002) and Zivot and Wang (2003).

14.2 Multivariate GARCH Processes

In this section we first establish general notation for multivariate GARCH (or MGARCH) models before going on to consider models that are defined via their conditional correlation matrix in Section 14.2.2 and models that are defined via their conditional covariance matrix in Section 14.2.4. We also provide brief notes on model estimation, strategies for dimension reduction, and the use of multivariate GARCH models in quantitative risk management.

14.2.1 General Structure of Models

For the following definition we recall that the Cholesky factor $\Sigma^{1/2}$ of a positive-definite matrix Σ is the lower-triangular matrix satisfying $AA' = \Sigma$ (see the discussion at end of Section 6.1).

Definition 14.10. Let $(\boldsymbol{Z}_t)_{t\in\mathbb{Z}}$ be SWN$(\mathbf{0}, I_d)$. The process $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ is said to be a multivariate GARCH process if it is strictly stationary and satisfies equations of the form

$$\boldsymbol{X}_t = \Sigma_t^{1/2}\boldsymbol{Z}_t, \quad t \in \mathbb{Z}, \tag{14.8}$$

where $\Sigma_t^{1/2} \in \mathbb{R}^{d\times d}$ is the Cholesky factor of a positive-definite matrix Σ_t that is measurable with respect to $\mathcal{F}_{t-1} = \sigma(\{\boldsymbol{X}_s : s \leqslant t-1\})$, the history of the process up to time $t-1$.

Conditional moments. It is easily calculated that a covariance-stationary process of this type has the multivariate martingale-difference property

$$E(\boldsymbol{X}_t \mid \mathcal{F}_{t-1}) = E(\Sigma_t^{1/2}\boldsymbol{Z}_t \mid \mathcal{F}_{t-1}) = \Sigma_t^{1/2}E(\boldsymbol{Z}_t) = \mathbf{0},$$

and it must therefore be a multivariate white noise process, as argued in Section 14.1. Moreover, Σ_t is the *conditional covariance matrix* since

$$\begin{aligned}\operatorname{cov}(\boldsymbol{X}_t \mid \mathcal{F}_{t-1}) &= E(\boldsymbol{X}_t\boldsymbol{X}_t' \mid \mathcal{F}_{t-1})\\ &= \Sigma_t^{1/2}E(\boldsymbol{Z}_t\boldsymbol{Z}_t')(\Sigma_t^{1/2})'\\ &= \Sigma_t^{1/2}(\Sigma_t^{1/2})'\\ &= \Sigma_t. \end{aligned}\tag{14.9}$$

The conditional covariance matrix Σ_t in a multivariate GARCH model corresponds to the squared volatility σ_t^2 in a univariate GARCH model. The use of the Cholesky factor of Σ_t to describe the relationship to the driving noise in (14.8) is not important, and in fact any type of "square root" of Σ_t could be used (such as the root derived from a symmetric decomposition). (The only practical consequence is that

different choices would lead to different series of residuals when the model is fitted in practice.) We denote the elements of Σ_t by $\sigma_{t,ij}$ and also use the notation $\sigma_{t,i} = \sqrt{\sigma_{t,ii}}$ to denote the conditional standard deviation (or volatility) of the ith component series $(X_{t,i})_{t\in\mathbb{Z}}$.

We recall that we can write $\Sigma_t = \Delta_t P_t \Delta_t$, where

$$\Delta_t = \Delta(\Sigma_t) = \operatorname{diag}(\sigma_{t,1}, \ldots, \sigma_{t,d}), \quad P_t = \wp(\Sigma_t), \tag{14.10}$$

using the operator notation defined in (6.4) and (6.5). The diagonal matrix Δ_t is known as the *volatility matrix* and P_t is known as the *conditional correlation matrix*. The art of building multivariate GARCH models is to specify the dependence of Σ_t (or of Δ_t and P_t) on the past in such a way that Σ_t always remains symmetric and positive definite. A covariance matrix must of course be symmetric and positive semidefinite, and in practice we restrict our attention to the positive-definite case. This facilitates fitting, since the conditional distribution of $\boldsymbol{X}_t \mid \mathcal{F}_{t-1}$ never has a singular covariance matrix.

Unconditional moments. The *unconditional covariance matrix* Σ of a process of this type is given by

$$\Sigma = \operatorname{cov}(\boldsymbol{X}_t) = E(\operatorname{cov}(\boldsymbol{X}_t \mid \mathcal{F}_{t-1})) + \operatorname{cov}(E(\boldsymbol{X}_t \mid \mathcal{F}_{t-1})) = E(\Sigma_t),$$

from which it can be calculated that the *unconditional correlation matrix* P has elements

$$\rho_{ij} = \frac{E(\sigma_{t,ij})}{\sqrt{E(\sigma_{t,ii})E(\sigma_{t,jj})}} = \frac{E(\rho_{t,ij}\sigma_{t,i}\sigma_{t,j})}{\sqrt{E(\sigma_{t,i}^2)E(\sigma_{t,j}^2)}}, \tag{14.11}$$

which is in general difficult to evaluate and is usually not simply the expectation of the conditional correlation matrix.

Innovations. In practical work the innovations are generally taken to be from either a multivariate Gaussian distribution ($\boldsymbol{Z}_t \sim N_d(\boldsymbol{0}, I_d)$) or, more realistically for daily returns, an appropriately scaled spherical multivariate t distribution ($\boldsymbol{Z}_t \sim t_d(\nu, \boldsymbol{0}, (\nu-2)I_d/\nu)$). Any distribution with mean 0 and covariance matrix I_d is permissible, and appropriate members of the normal mixture family of Section 6.2 or the spherical family of Section 6.3.1 may be considered.

Presentation of models. In the following sections we present some of the more important multivariate GARCH specifications. In doing this we concentrate on the following aspects of the models.

- The form of the dynamic equations, with economic arguments and criticisms where appropriate.
- The conditions required to guarantee that the conditional covariance matrix Σ_t remains positive definite. Other mathematical properties of these models, such as conditions for covariance stationarity, are difficult to derive with full mathematical rigour; references in Notes and Comments contain further information.

- The parsimoniousness of the parametrization. A major problem with most multivariate GARCH specifications is that the number of parameters tends to explode with the dimension of the model, making them unsuitable for analyses of many risk factors.
- Simple intuitive fitting methods where available. All models can be fitted by a general global-maximization approach described in Section 14.2.4 but certain models lend themselves to estimation in stages, particularly the models of Section 14.2.2.

14.2.2 Models for Conditional Correlation

In this section we present models that focus on specifying the conditional correlation matrix P_t while allowing volatilities to be described by univariate GARCH models; we begin with a popular and relatively parsimonious model where P_t is assumed to be constant for all t.

Constant conditional correlation (CCC).

Definition 14.11. The process $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ is a CCC–GARCH process if it is a process with the general structure given in Definition 14.10 such that the conditional covariance matrix is of the form $\Sigma_t = \Delta_t P_c \Delta_t$, where

(i) P_c is a constant, positive-definite correlation matrix; and

(ii) Δ_t is a diagonal volatility matrix with elements $\sigma_{t,k}$ satisfying

$$\sigma_{t,k}^2 = \alpha_{k0} + \sum_{i=1}^{p_k} \alpha_{ki} X_{t-i,k}^2 + \sum_{j=1}^{q_k} \beta_{kj} \sigma_{t-j,k}^2, \quad k = 1, \ldots, d, \tag{14.12}$$

where $\alpha_{k0} > 0$, $\alpha_{ki} \geqslant 0$, $i = 1, \ldots, p_k$, $\beta_{kj} \geqslant 0$, $j = 1, \ldots, q_k$.

The CCC–GARCH specification represents a simple way of combining univariate GARCH processes. This can be seen by observing that in a CCC–GARCH model observations and innovations are connected by equations $\boldsymbol{X}_t = \Delta_t P_c^{1/2} \boldsymbol{Z}_t$, which may be rewritten as $\boldsymbol{X}_t = \Delta_t \tilde{\boldsymbol{Z}}_t$ for an SWN($\boldsymbol{0}$, P_c) process $(\tilde{\boldsymbol{Z}}_t)_{t\in\mathbb{Z}}$. Clearly, the component processes are univariate GARCH.

Proposition 14.12. *The CCC–GARCH model is well defined in the sense that Σ_t is almost surely positive definite for all t. Moreover, it is covariance stationary if and only if $\sum_{i=1}^{p_k} \alpha_{ki} + \sum_{j=1}^{q_k} \beta_{kj} < 1$ for $k = 1, \ldots, d$.*

Proof. For a vector $\boldsymbol{v} \neq \boldsymbol{0}$ in $\mathbb{R}^d$ we have

$$\boldsymbol{v}' \Sigma_t \boldsymbol{v} = (\Delta_t \boldsymbol{v})' P_c (\Delta_t \boldsymbol{v}) > 0,$$

since P_c is positive definite and the strict positivity of the individual volatility processes ensures that $\Delta_t \boldsymbol{v} \neq \boldsymbol{0}$ for all t.

If $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ is covariance stationary, then each component series $(X_{t,k})_{t\in\mathbb{Z}}$ is a covariance-stationary GARCH process for which a necessary and sufficient condition is $\sum_{i=1}^{p_k} \alpha_{ki} + \sum_{j=1}^{q_k} \beta_{kj} < 1$ by Proposition 4.21. Conversely, if the component

series are covariance stationary, then for all i and j the Cauchy–Schwarz inequality implies

$$\sigma_{ij} = E(\sigma_{t,ij}) = \rho_{ij} E(\sigma_{t,i}\sigma_{t,j}) \leqslant \rho_{ij}\sqrt{E(\sigma_{t,i}^2)}\sqrt{E(\sigma_{t,j}^2)} < \infty.$$

Since $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ is a multivariate martingale difference with finite, non-time-dependent second moments σ_{ij}, it is a covariance-stationary white noise. □

The CCC model is often a useful starting point from which to proceed to more complex models. In some empirical settings it gives an adequate performance, but it is generally accepted that the constancy of conditional correlation in this model is an unrealistic feature and that the impact of news on financial markets requires models that allow a dynamic evolution of conditional correlation as well as a dynamic evolution of volatilities. A further criticism of the model (which in fact applies to the majority of MGARCH specifications) is the fact that the individual volatility dynamics (14.12) do not allow for the possibility that large returns in one component series at a particular point in time can contribute to the increased volatility of another component time series at future points in time.

To describe a simple method of fitting the CCC model, we introduce the notion of a *devolatized* process. For any multivariate time-series process $\boldsymbol{X}_t$, the devolatized process is the process $\boldsymbol{Y}_t = \Delta_t^{-1}\boldsymbol{X}_t$, where Δ_t is, as usual, the diagonal matrix of volatilities. In the case of a CCC model it is easily seen that the devolatized process $(\boldsymbol{Y}_t)_{t\in\mathbb{Z}}$ is an SWN$(\mathbf{0}, P_c)$ process.

This structure suggests a simple two-stage fitting method in which we first estimate the individual volatility processes for the component series by fitting univariate GARCH processes to data $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$; note that, although we have specified in Definition 14.11 that the individual volatilities should follow standard GARCH models, we could of course extend the model to allow any of the univariate models in Section 4.2.3 to be used, such as GARCH with leverage or threshold GARCH.

In a second stage we construct an estimate of the devolatized process by taking $\hat{\boldsymbol{Y}}_t = \hat{\Delta}_t^{-1}\boldsymbol{X}_t$ for $t = 1, \ldots, n$, where $\hat{\Delta}_t^{-1}$ is the estimate of Δ_t; in other words, we collect the standardized residuals from the univariate GARCH models. Note that we have also described this construction in slightly more general terms in the context of a multivariate approach to dynamic historical simulation in Section 9.2.4. If the CCC–GARCH model is adequate, then the $\hat{\boldsymbol{Y}}_t$ data should behave like a realization from an SWN$(\mathbf{0}, P_c)$ process, and this can be investigated with the correlogram and cross-correlogram applied to raw and absolute values. Assuming that the model is adequate, the conditional correlation matrix P_c can then be estimated from the standardized residuals using methods from Chapter 6.

A special case of CCC–GARCH that we call a pure diagonal model occurs when $P_c = I_d$. A covariance-stationary model of this kind is a multivariate white noise where the contemporaneous components $X_{t,i}$ and $X_{t,j}$ are also uncorrelated for $i \neq j$. This subsumes the case of independent GARCH models for each component series. Indeed, if we assume that the driving SWN$(\mathbf{0}, I_d)$ process is multivariate Gaussian, then the component series are independent. However, if, for example, we

assume that the innovations satisfy $\boldsymbol{Z}_t \sim t_d(\nu, \mathbf{0}, ((\nu-2)/\nu)I_d)$, then the component processes are dependent.

Dynamic conditional correlation (DCC). This model generalizes the CCC model to allow conditional correlations to evolve dynamically according to a relatively parsimonious scheme, but it is constructed in a way that still allows estimation in stages using univariate GARCH models. Its formal analysis as a stochastic process is difficult due to the use of the correlation matrix extraction operator $\wp$ in its definition.

Definition 14.13. The process $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ is a DCC–GARCH process if it is a process with the general structure given in Definition 14.10, where the volatilities comprising Δ_t follow univariate GARCH specifications as in (14.12) and the conditional correlation matrices P_t satisfy, for $t \in \mathbb{Z}$, the equations

$$P_t = \wp\left(\left(1 - \sum_{i=1}^{p}\alpha_i - \sum_{j=1}^{q}\beta_j\right)P_{\mathrm{c}} + \sum_{i=1}^{p}\alpha_i \boldsymbol{Y}_{t-i}\boldsymbol{Y}_{t-i}' + \sum_{j=1}^{q}\beta_j P_{t-j}\right), \quad (14.13)$$

where P_{c} is a positive-definite correlation matrix, $\wp$ is the operator in (6.5), $\boldsymbol{Y}_t = \Delta_t^{-1}\boldsymbol{X}_t$ denotes the devolatized process, and the coefficients satisfy $\alpha_i \geqslant 0$, $\beta_j \geqslant 0$ and $\sum_{i=1}^{p}\alpha_i - \sum_{j=1}^{q}\beta_j < 1$.

Observe first that if all the α_i and β_j coefficients in (14.13) are zero, then the model reduces to the CCC model. If one makes an analogy with a covariance-stationary univariate GARCH model with unconditional variance σ^2, for which the volatility equation can be written

$$\sigma_t^2 = \left(1 - \sum_{i=1}^{p}\alpha_i - \sum_{j=1}^{q}\beta_j\right)\sigma^2 + \sum_{i=1}^{p}\alpha_i X_{t-i} + \sum_{j=1}^{q}\beta_j\sigma_{t-j}^2,$$

then the correlation matrix P_{c} in (14.13) can be thought of as representing the long-run correlation structure. Although this matrix could be estimated by fitting the DCC model to data by ML estimation in one step, it is quite common to estimate it using an empirical correlation matrix calculated from the devolatized data, as in the estimation of the CCC model.

Observe also that the dynamic equation (14.13) preserves the positive definiteness of P_t. If we define

$$Q_t := \left(1 - \sum_{i=1}^{p}\alpha_i - \sum_{j=1}^{q}\beta_j\right)P_{\mathrm{c}} + \sum_{i=1}^{p}\alpha_i \boldsymbol{Y}_{t-i}\boldsymbol{Y}_{t-i}' + \sum_{j=1}^{q}\beta_j P_{t-j}$$

and assume that $P_{t-q}, \ldots, P_{t-1}$ are positive definite, then it follows that, for a vector $\boldsymbol{v} \neq \mathbf{0}$ in $\mathbb{R}^d$, we have

$$\boldsymbol{v}'Q_t\boldsymbol{v} = \left(1 - \sum_{i=1}^{p}\alpha_i - \sum_{j=1}^{q}\beta_j\right)\boldsymbol{v}'P_{\mathrm{c}}\boldsymbol{v} + \sum_{i=1}^{p}\alpha_i \boldsymbol{v}'\boldsymbol{Y}_{t-i}\boldsymbol{Y}_{t-i}'\boldsymbol{v} + \sum_{j=1}^{q}\beta_j \boldsymbol{v}'P_{t-j}\boldsymbol{v} > 0,$$

since the first term is strictly positive and the second and third terms are non-negative. If Q_t is positive definite, then so is P_t.

The usual estimation method for the DCC model is as follows.

(1) Fit univariate GARCH-type models to the component series to estimate the volatility matrix Δ_t. Form an estimated realization of the devolatized process by taking $\hat{\boldsymbol{Y}}_t = \hat{\Delta}_t \boldsymbol{X}_t$.

(2) Estimate P_c by taking the sample correlation matrix of the devolatized data (or, better still, some robust estimator of correlation).

(3) Estimate the remaining parameters α_i and β_j in equation (14.13) by fitting a model with structure $\boldsymbol{Y}_t = P_t^{1/2} \boldsymbol{Z}_t$ to the devolatized data. We leave this step vague for the time being and note that this will be a simple application of the methodology for fitting general multivariate GARCH models in Section 14.2.4; in a first-order model ($p = q = 1$), there will be only two remaining parameters to estimate.

14.2.3 Models for Conditional Covariance

The models of this section specify explicitly a dynamic structure for the conditional covariance matrix Σ_t. These models are not designed for multiple-stage estimation based on univariate GARCH estimation procedures.

Vector GARCH models (VEC and DVEC). The most general vector GARCH model—the VEC model—has too many parameters for practical purposes and our task will be to simplify the model by imposing various restrictions on parameter matrices.

Definition 14.14 (VEC model). The process $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ is a VEC process if it has the general structure given in Definition 14.10, and the dynamics of the conditional covariance matrix Σ_t are given by the equations

$$\operatorname{vech}(\Sigma_t) = \boldsymbol{a}_0 + \sum_{i=1}^{p} \bar{A}_i \operatorname{vech}(\boldsymbol{X}_{t-i}\boldsymbol{X}'_{t-i}) + \sum_{j=1}^{q} \bar{B}_j \operatorname{vech}(\Sigma_{t-j}) \tag{14.14}$$

for a vector $\boldsymbol{a}_0 \in \mathbb{R}^{d(d+1)/2}$ and matrices $\bar{A}_i$ and $\bar{B}_j$ in $R^{(d(d+1)/2)\times(d(d+1)/2)}$.

In this definition "vech" denotes the *vector half* operator, which stacks the columns of the lower triangle of a symmetric matrix in a single column vector of length $d(d+1)/2$. Thus (14.14) should be understood as specifying the dynamics for the lower-triangular portion of the conditional covariance matrix, and the remaining elements of the matrix are determined by symmetry.

In this very general form the model has $(1 + (p+q)d(d+1)/2)d(d+1)/2$ parameters; this number grows rapidly with dimension so that even a trivariate model has 78 parameters. The most common simplification has been to restrict attention to cases when $\bar{A}_i$ and $\bar{B}_j$ are diagonal matrices, which gives us the diagonal VEC or DVEC model. This special case can be written very elegantly in terms of a different kind of matrix product, namely the *Hadamard product*, denoted by "$\circ$", which signifies element-by-element multiplication of two matrices of the same size.

We obtain the representation

$$\Sigma_t = A_0 + \sum_{i=1}^{p} A_i \circ (\boldsymbol{X}_{t-i}\boldsymbol{X}_{t-i}') + \sum_{j=1}^{q} B_j \circ \Sigma_{t-j}, \tag{14.15}$$

where A_0 and the A_i and B_j must all be symmetric matrices in $\mathbb{R}^{d\times d}$ such that A_0 has positive diagonal elements and all other matrices have non-negative diagonal elements (standard univariate GARCH assumptions). This representation emphasizes structural similarities with the univariate GARCH model of Definition 4.20.

To better understand the dynamic implications of (14.15), consider a bivariate model of order (1, 1) and write $a_{0,ij}$, $a_{1,ij}$ and b_{ij} for the elements of A_0, A_1 and B_1, respectively. The model amounts to the following three simple equations:

$$\left.\begin{aligned} \sigma_{t,1}^2 &= a_{0,11} + a_{1,11}X_{t-1,1}^2 + b_{11}\sigma_{t-1,1}^2, \\ \sigma_{t,12} &= a_{0,12} + a_{1,12}X_{t-1,1}X_{t-1,2} + b_{12}\sigma_{t-1,12}, \\ \sigma_{t,2}^2 &= a_{0,22} + a_{1,22}X_{t-1,2}^2 + b_{22}\sigma_{t-1,2}^2. \end{aligned}\right\} \tag{14.16}$$

The volatilities of the two component series ($\sigma_{t,1}$ and $\sigma_{t,2}$) follow univariate GARCH updating patterns, and the conditional covariance $\sigma_{t,12}$ has a similar structure driven by the products of the lagged values $X_{t-1,1}X_{t-1,2}$. As for the CCC and DCC models, the volatility of a single component series is only driven by large lagged values of that series and cannot be directly affected by large lagged values in another series; the more general but overparametrized VEC model would allow this feature.

The requirement that Σ_t in (14.15) should be a proper positive-definite covariance matrix does impose conditions on the A_0, A_i and B_j matrices that we have not yet discussed. In practice, in some software implementations of this model, formal conditions are not imposed, other than that the matrices should be symmetric with non-negative diagonal elements; the positive definiteness of the resulting estimates of the conditional covariance matrices can be checked after model fitting.

However, a sufficient condition for Σ_t to be almost surely positive definite is that A_0 should be positive definite and the matrices $A_1, \ldots, A_p, B_1, \ldots, B_q$ should all be positive semidefinite (see Notes and Comments), and this condition is easy to impose. We can constrain all parameter matrices to have a form based on a Cholesky decomposition; that is, we can parametrize the model in terms of lower-triangular Cholesky factor matrices $A_0^{1/2}$, $A_i^{1/2}$ and $B_j^{1/2}$ such that

$$A_0 = A_0^{1/2}(A_0^{1/2})', \qquad A_i = A_i^{1/2}(A_i^{1/2})', \qquad B_j = B_j^{1/2}(B_j^{1/2})'. \tag{14.17}$$

Because the sufficient condition only prescribes that $A_1, \ldots, A_p$ and $B_1, \ldots, B_q$ should be positive semidefinite, we can in fact also consider much simpler parametrizations, such as

$$A_0 = A_0^{1/2}(A_0^{1/2})', \qquad A_i = \boldsymbol{a}_i\boldsymbol{a}_i', \qquad B_j = \boldsymbol{b}_j\boldsymbol{b}_j', \tag{14.18}$$

where $\boldsymbol{a}_i$ and $\boldsymbol{b}_j$ are vectors in $\mathbb{R}^d$. An even cruder model, satisfying the requirement of positive definiteness of Σ_t, would be

$$A_0 = A_0^{1/2}(A_0^{1/2})', \qquad A_i = a_i I_d, \qquad B_j = b_j I_d, \tag{14.19}$$

where a_i and b_j are simply positive constants. In fact, the specifications of the multivariate ARCH and GARCH effects in (14.17)–(14.19) can be mixed and matched in obvious ways.

The BEKK model of Baba, Engle, Kroner and Kraft. The next family of models have the great advantage that their construction ensures the positive definiteness of Σ_t without the need for further conditions.

Definition 14.15. The process $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ is a BEKK process if it has the general structure given in Definition 14.10 and if the conditional covariance matrix Σ_t satisfies, for all $t \in \mathbb{Z}$,

$$\Sigma_t = A_0 + \sum_{i=1}^{p} A_i' \boldsymbol{X}_{t-i} \boldsymbol{X}_{t-i}' A_i + \sum_{j=1}^{q} B_j' \Sigma_{t-j} B_j, \tag{14.20}$$

where all coefficient matrices are in $\mathbb{R}^{d\times d}$ and A_0 is symmetric and positive definite.

Proposition 14.16. *In the BEKK model (14.20), the conditional covariance matrix Σ_t is almost surely positive definite for all t.*

Proof. Consider a first-order model for simplicity. For a vector $\boldsymbol{v} \neq \boldsymbol{0}$ in $\mathbb{R}^d$ we have

$$\boldsymbol{v}'\Sigma_t\boldsymbol{v} = \boldsymbol{v}'A_0\boldsymbol{v} + (\boldsymbol{v}'A_1'\boldsymbol{X}_{t-1})^2 + (B_1\boldsymbol{v})'\Sigma_{t-1}(B_1\boldsymbol{v}) > 0,$$

since the first term is strictly positive and the second and third terms are non-negative. □

To gain an understanding of the BEKK model it is again useful to consider the bivariate special case of order (1, 1) and to consider the dynamics that are implied while comparing these with equations (14.16):

$$\begin{aligned}\sigma_{t,1}^2 &= a_{0,11} + a_{1,11}^2 X_{t-1,1}^2 + 2a_{1,11}a_{1,12}X_{t-1,1}X_{t-1,2} + a_{1,12}^2 X_{t-1,2}^2 \\ &\quad + b_{11}^2\sigma_{t-1,1}^2 + 2b_{11}b_{12}\sigma_{t-1,12} + b_{12}^2\sigma_{t-1,2}^2,\end{aligned} \tag{14.21}$$

$$\begin{aligned}\sigma_{t,12} &= a_{0,12} + (a_{1,11}a_{1,22} + a_{1,12}a_{1,21})X_{t-1,1}X_{t-1,2} \\ &\quad + a_{1,11}a_{1,21}X_{t-1,1}^2 + a_{1,22}a_{1,12}X_{t-1,2}^2 \\ &\quad + (b_{11}b_{22} + b_{12}b_{21})\sigma_{t-1,12} + b_{11}b_{21}\sigma_{t-1,1}^2 + b_{22}b_{12}\sigma_{t-1,2}^2,\end{aligned} \tag{14.22}$$

$$\begin{aligned}\sigma_{t,2}^2 &= a_{0,22} + a_{1,22}^2 X_{t-1,2}^2 + 2a_{1,22}a_{1,21}X_{t-1,1}X_{t-1,2} + a_{1,21}^2 X_{t-1,1}^2 \\ &\quad + b_{22}^2\sigma_{t-1,2}^2 + 2b_{22}b_{21}\sigma_{t-1,21} + b_{21}^2\sigma_{t-1,1}^2.\end{aligned} \tag{14.23}$$

From (14.21) it follows that we now have a model where a large lagged value of the second component $X_{t-1,2}$ can influence the volatility of the first series $\sigma_{t,1}$. The BEKK model has more parameters than the DVEC model and appears to have much richer dynamics. Note, however, that the DVEC model cannot be obtained as a special case of the BEKK model as we have defined it. To eliminate all crossover effects in the conditional variance equations of the BEKK model in (14.21) and (14.23), we would have to set the diagonal terms $a_{1,12}$, $a_{1,21}$, b_{12} and b_{21} to be zero and the parameters governing the individual volatilities would also govern the conditional covariance $\sigma_{t,12}$ in (14.22).

Table 14.1. Summary of numbers of parameters in various multivariate GARCH models: in CCC it is assumed that the numbers of ARCH and GARCH terms for all volatility equations are, respectively, p and q; in DCC it is assumed that the conditional correlation equation has $p + q$ parameters. The second column gives the general formula; the final columns give the numbers for models of dimensions 2, 5 and 10 when $p = q = 1$. Additional parameters in the innovation distribution are not considered.

Model	Parameter count	2	5	10
VEC	$d(d+1)(1+(p+q)d(d+1)/2)/2$	21	465	6105
BEKK	$d(d+1)/2 + d^2(p+q)$	11	65	255
DVEC as in (14.17)	$d(d+1)(1+p+q)/2$	9	45	165
DCC	$d(d+1)/2 + (d+1)(p+q)$	9	27	77
CCC	$d(d+1)/2 + d(p+q)$	7	25	75
DVEC as in (14.18)	$d(d+1)/2 + d(p+q)$	7	25	75
DVEC as in (14.19)	$d(d+1)/2 + (p+q)$	5	17	57

Remark 14.17. A broader definition of the BEKK class, which does subsume all DVEC models, was originally given by Engle and Kroner (1995). In this definition we have

$$\Sigma_t = A_0 A_0' + \sum_{k=1}^{K} \sum_{i=1}^{p} A_{k,i}' \boldsymbol{X}_{t-i} \boldsymbol{X}_{t-i}' A_{k,i} + \sum_{k=1}^{K} \sum_{j=1}^{q} B_{k,j}' \Sigma_{t-j} B_{k,j},$$

where $\frac{1}{2}d(d+1) > K \geqslant 1$ and the choice of K determines the richness of the model. This model class is of largely theoretical interest and tends to be too complex for practical applications; even the case $K = 1$ is difficult to fit in higher dimensions.

In Table 14.1 we have summarized the numbers of parameters in these models. Broad conclusions concerning the practical implications are as follows: the general VEC model is of purely theoretical interest; the BEKK and general DVEC models are for very low-dimensional use; and the remaining models are the most practically useful.

14.2.4 Fitting Multivariate GARCH Models

Model fitting. We have already given notes on fitting some models in stages and it should be stressed that in the high-dimensional applications of risk management this may in fact be the only feasible strategy. Where interest centres on a multivariate risk-factor return series of more modest dimension (fewer than ten dimensions, say), we can attempt to fit multivariate GARCH models by maximizing an appropriate likelihood with respect to all parameters in a single step. The procedure follows from the method for univariate time series described in Section 4.2.4.

The method of building a likelihood for a generic multivariate GARCH model $\boldsymbol{X}_t = \Sigma_t^{1/2} \boldsymbol{Z}_t$ is completely analogous to the univariate case; consider again a first-order model ($p = q = 1$) for simplicity and assume that our data are labelled $\boldsymbol{X}_0, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$. A conditional likelihood is based on the conditional joint density of $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ given $\boldsymbol{X}_0$ and an initial value Σ_0 for the conditional covariance

matrix. This conditional joint density is

$$f_{\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n|\boldsymbol{X}_0,\Sigma_0}(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n \mid \boldsymbol{x}_0, \Sigma_0) = \prod_{t=1}^{n} f_{\boldsymbol{X}_t|\boldsymbol{X}_{t-1},\ldots,\boldsymbol{X}_0,\Sigma_0}(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1},\ldots,\boldsymbol{x}_0, \Sigma_0).$$

If we denote the multivariate innovation density of $\boldsymbol{Z}_t$ by $g(z)$, then we have

$$f_{\boldsymbol{X}_t|\boldsymbol{X}_{t-1},\ldots,\boldsymbol{X}_0,\Sigma_0}(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1},\ldots,\boldsymbol{x}_0, \Sigma_0) = |\Sigma_t|^{-1/2} g(\Sigma_t^{-1/2}\boldsymbol{x}_t),$$

where Σ_t is a matrix-valued function of $\boldsymbol{x}_{t-1},\ldots,\boldsymbol{x}_0$ and Σ_0. Most common choices of $g(z)$ are in the spherical family, so by (6.39) we have $g(z) = h(z'z)$ for some function h of a scalar variable (known as a density generator), yielding a conditional likelihood of the form

$$L(\boldsymbol{\theta}; \boldsymbol{X}_1,\ldots,\boldsymbol{X}_n) = \prod_{t=1}^{n} |\Sigma_t|^{-1/2} h(\boldsymbol{X}_t'\Sigma_t^{-1}\boldsymbol{X}_t),$$

where all parameters appearing in the volatility equation and the innovation distribution are collected in $\boldsymbol{\theta}$. It would of course be possible to add a constant mean term or a conditional mean term with, say, vector autoregressive structure to the model and to adapt the likelihood accordingly.

Evaluation of the likelihood requires us to input a value for Σ_0. Maximization can again be performed in practice using a modified Newton–Raphson procedure, such as that of Berndt et al. (1974). References concerning properties of estimators are given in Notes and Comments, although the literature for multivariate GARCH is small.

Model checking and comparison. Residuals are calculated according to $\hat{Z}_t = \hat{\Sigma}_t^{-1/2}\boldsymbol{X}_t$ and should behave like a realization of an SWN$(\boldsymbol{0}, I_d)$ process. The usual univariate procedures (correlograms, correlograms of absolute values, and portmanteau tests such as Ljung–Box) can be applied to the component series of the residuals. Also, there should not be any evidence of cross-correlations at any lags for either the raw or the absolute residuals in the cross-correlogram.

Model selection is usually performed by a standard comparison of AIC numbers, although there is limited literature on theoretical aspects of the use of the AIC in a multivariate GARCH context.

14.2.5 Dimension Reduction in MGARCH

In view of the large number of parameters in MGARCH models and the practical difficulties involved in their estimation, as well as the fact that many time series of risk-factor changes are likely to have high levels of cross-correlation at lag zero, it makes sense to consider how dimension-reduction strategies from the area of factor modelling (see Section 6.4.1) can be applied in the context of multivariate GARCH.

As discussed in Section 6.4.2 there are a number of different statistical approaches to building a factor model. We give brief notes on strategies for constructing (1) a *macroeconomic* factor model and (2) a *statistical* factor model based on *principal component analysis*.

Macroeconomic factor model. In this approach we attempt to identify a small number of common factors $\boldsymbol{F}_t$ that can explain the variation in many risk factors $\boldsymbol{X}_t$; we might, for example, use stock index returns to explain the variation in individual equity returns. These common factors could be modelled using relatively detailed multivariate models incorporating ARMA and GARCH features.

The dependence of the individual returns on the factor returns can then be modelled by calibrating a factor model of the type

$$\boldsymbol{X}_t = \boldsymbol{a} + B\boldsymbol{F}_t + \boldsymbol{\varepsilon}_t, \quad t = 1, \ldots, n.$$

In Section 6.4.3 we showed how this may be done in a static way using regression techniques. We now assume that, conditional on the factors $\boldsymbol{F}_t$, the errors $\boldsymbol{\varepsilon}_t$ form a multivariate white noise process with GARCH volatility structure.

In a perfect factor model (corresponding to Definition 6.31) these errors would have a diagonal covariance matrix and would be attributable to idiosyncratic effects alone. In GARCH terms we could assume they follow a pure diagonal model, i.e. a CCC model where the constant conditional correlation matrix is the identity matrix. A pure diagonal model can be fitted in two ways, which correspond to the two ways of estimating a static regression model in Section 6.4.3.

(1) Fit univariate models to the component series $X_{1,k}, \ldots, X_{n,k}$, $k = 1, \ldots, d$. For each k assume that

$$X_{t,k} = \mu_{t,k} + \varepsilon_{t,k}, \quad \mu_{t,k} = a_k + \boldsymbol{b}_k' \boldsymbol{F}_t, \quad t = 1, \ldots, n,$$

where the errors $\varepsilon_{t,k}$ follow some univariate GARCH specification. Most software for GARCH estimation allows covariates to be incorporated into the model for the conditional mean term $\mu_{t,k}$.

(2) Fit in one step the multivariate model

$$\boldsymbol{X}_t = \boldsymbol{\mu}_t + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\mu}_t = \boldsymbol{a} + B\boldsymbol{F}_t, \quad t = 1, \ldots, n,$$

where the errors $\boldsymbol{\varepsilon}_t$ follow a pure diagonal CCC model and the SWN($\mathbf{0}$, I_d) process driving the GARCH model is some non-Gaussian spherical distribution, such as an appropriate scaled t distribution. (If the SWN is Gaussian, approaches (1) and (2) give the same results.)

In practice, it is never possible to find the "right" common factors such that the idiosyncratic errors have a diagonal covariance structure. The pure diagonal assumption can be examined by looking at the errors from the GARCH modelling, estimating their correlation matrix and assessing its closeness to the identity matrix. In the case where correlation structure remains, the formal concept of the factor model can be loosened by allowing errors with a CCC–GARCH structure, which can be calibrated by two-stage estimation as described in Section 14.2.2.

Principal components GARCH (PC-GARCH). In Section 6.4.5 we explained the application of principal component analysis to time-series data under the assumption that the data came from a stationary multivariate time-series process. The idea of principal components GARCH is that we first compute time series of sample principal components and then model the most important series—the ones with highest sample variance—using GARCH models. Usually we attempt to fit a series of independent univariate GARCH models to the sample principal components.

In terms of an underlying model we are implicitly assuming that the data are realizations from a so-called PC-GARCH or orthogonal GARCH model defined as follows.

Definition 14.18. The process $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ follows a PC-GARCH (or orthogonal GARCH) model if there exists some orthogonal matrix $\Gamma \in \mathbb{R}^{d\times d}$ satisfying $\Gamma\Gamma' = \Gamma'\Gamma = I_d$ such that $(\Gamma'\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ follows a pure diagonal GARCH model. The process $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ follows a PC-GARCH model with mean $\boldsymbol{\mu}$ if $(\boldsymbol{X}_t - \boldsymbol{\mu})_{t\in\mathbb{Z}}$ follows a PC-GARCH model.

If $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ follows a PC-GARCH process for some matrix Γ, then we can introduce the process $(\boldsymbol{Y}_t)_{t\in\mathbb{Z}}$ defined by $\boldsymbol{Y}_t = \Gamma'\boldsymbol{X}_t$, which satisfies $\boldsymbol{Y}_t = \Delta_t\boldsymbol{Z}_t$, where $(\boldsymbol{Z}_t)_{t\in\mathbb{Z}}$ is SWN$(\mathbf{0}, I_d)$ and Δ_t is a (diagonal) volatility matrix with elements that are updated according to univariate GARCH schemes and past values of the components of $\boldsymbol{Y}_t$. Since $\boldsymbol{X}_t = \Gamma\Delta_t\boldsymbol{Z}_t$, the conditional and unconditional covariance matrices have the structure

$$\Sigma_t = \Gamma\Delta_t^2\Gamma', \qquad \Sigma = \Gamma E(\Delta_t^2)\Gamma', \tag{14.24}$$

and they are obviously symmetric and positive definite.

Comparing (14.24) with (6.59) we see that the PC-GARCH model implies a spectral decomposition of the conditional and unconditional covariance matrices. The *eigenvalues* of the conditional covariance matrix, which are the elements of the diagonal matrix Δ_t^2, are given a GARCH updating structure. The *eigenvectors* form the columns of Γ and are used to construct the time series $(\boldsymbol{Y}_t)_{t\in\mathbb{Z}}$, the principal components transform of $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$. It should be noted that despite the simple structure of (14.24), the conditional correlation matrix of $\boldsymbol{X}_t$ is not constant in this model.

As noted above, this model is estimated in two stages. Let us suppose we have risk-factor change data $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ assumed to come from a PC-GARCH model with mean $\boldsymbol{\mu}$. In the first step we calculate the spectral decomposition of the sample covariance matrix S of the data as in Section 6.4.5; this gives us an estimator G of Γ. We then rotate the data to obtain sample principal components $\{\boldsymbol{Y}_t = G'(\boldsymbol{X}_t - \bar{\boldsymbol{X}}) : t = 1, \ldots, n\}$, where $\bar{\boldsymbol{X}} = n^{-1}\sum_{t=1}^n \boldsymbol{X}_t$. These should be consistent with a pure diagonal model if the PC-GARCH model is appropriate for the original data; there should be no cross-correlation between the series at any lag. In a second stage we fit univariate GARCH models to the time series of principal components; the residuals from these GARCH models should behave like SWN$(\mathbf{0}, I_d)$.

To turn the model output into a factor model we use the idea embodied in equation (6.65)—that the first k loading vectors in the matrix G specify the most important

principal components—and we write these columns in the submatrix $G_1 \in \mathbb{R}^{d\times k}$. We define the factors to be $\boldsymbol{F}_t := (Y_{t,1}, \ldots, Y_{t,k})'$, $t = 1, \ldots, n$, the first k principal component time series. We have now calibrated an approximate factor model of the form $\boldsymbol{X}_t = \bar{\boldsymbol{X}} + G_1\boldsymbol{F}_t + \boldsymbol{\varepsilon}_t$, where the components of $\boldsymbol{F}_t$ are assumed to follow a pure diagonal GARCH model of dimension k. We can use this model to forecast the behaviour of the risk-factor changes $\boldsymbol{X}_t$ by first forecasting the behaviour of the factors $\boldsymbol{F}_t$; the error term is usually ignored in practice.

14.2.6 MGARCH and Conditional Risk Measurement

Suppose we calibrate an MGARCH model (possibly with VARMA conditional mean structure) having the general structure $\boldsymbol{X}_t = \boldsymbol{\mu}_t + \Sigma_t^{1/2}\boldsymbol{Z}_t$ to historical risk-factor return data $\boldsymbol{X}_{t-n+1}, \ldots, \boldsymbol{X}_t$. We are interested in the loss distribution of $L_{t+1} = l_{[t]}(\boldsymbol{X}_{t+1})$ conditional on $\mathcal{F}_t = \sigma(\{\boldsymbol{X}_s : s \leqslant t\})$, as described in Sections 9.1 and 9.2.1. (We may also be interested in longer-period losses as in Section 9.2.7.)

A general method that could be applied is the Monte Carlo method of Section 9.2.5: we could simulate many times the next value $\boldsymbol{X}_{t+1}$ (and subsequent values for losses over longer periods) of the stochastic process $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ using estimates of $\boldsymbol{\mu}_{t+1}$ and Σ_{t+1}.

Alternatively, a variance–covariance calculation could be made as in Section 9.2.2. Considering a linearized loss operator with the general form $l^{\Delta}_{[t]}(\boldsymbol{x}) = -(c_t + \boldsymbol{b}_t'\boldsymbol{x})$, the moments of the conditional loss distribution would be

$$E(L^{\Delta}_{t+1} \mid \mathcal{F}_t) = -c_t - \boldsymbol{b}_t'\boldsymbol{\mu}_{t+1}, \qquad \operatorname{cov}(L^{\Delta}_{t+1} \mid \mathcal{F}_t) = \boldsymbol{b}_t'\Sigma_{t+1}\boldsymbol{b}_t.$$

Under an assumption of multivariate Gaussian innovations, the conditional distribution of L^{Δ}_{t+1} given $\mathcal{F}_t$ would be univariate Gaussian as in equation (2.14). Under an assumption of (scaled) multivariate t innovations, it would be univariate t. To calculate VaR and ES we can follow the calculations in Examples 2.11, 2.14 and 2.15 but we would first need to compute estimates of Σ_{t+1} and $\boldsymbol{\mu}_{t+1}$ using our multivariate time-series model, as indicated in Example 14.19.

Example 14.19. Consider the simple stock portfolio in Example 2.1, where $c_t = 0$ and $\boldsymbol{b}_t = V_t\boldsymbol{w}_t$, and suppose our time-series model is a first-order DVEC model with a constant mean term. The model takes the form

$$\boldsymbol{X}_t - \boldsymbol{\mu} = \Sigma_t^{1/2}\boldsymbol{Z}_t, \qquad \Sigma_t = A_0 + A_1 \circ ((\boldsymbol{X}_{t-1} - \boldsymbol{\mu})(\boldsymbol{X}_{t-1}' - \boldsymbol{\mu}')) + B \circ \Sigma_{t-1}. \tag{14.25}$$

Suppose we assume that the innovations are multivariate Student t. The standard risk measures applied to the linearized loss distribution would take the form

$$\mathrm{VaR}^t_\alpha = -V_t\boldsymbol{w}_t'\boldsymbol{\mu} + V_t\sqrt{\frac{\boldsymbol{w}_t'\Sigma_{t+1}\boldsymbol{w}_t(\nu-2)}{\nu}}\,t_\nu^{-1}(\alpha),$$

$$\mathrm{ES}^t_\alpha = -V_t\boldsymbol{w}_t'\boldsymbol{\mu} + V_t\sqrt{\frac{\boldsymbol{w}_t'\Sigma_{t+1}\boldsymbol{w}_t(\nu-2)}{\nu}}\,\frac{g_\nu(t_\nu^{-1}(\alpha))}{1-\alpha}\left(\frac{\nu + (t_\nu^{-1}(\alpha))^2}{\nu-1}\right),$$

where the notation is as in Example 2.15. Estimates of the risk measures are obtained by replacing $\boldsymbol{\mu}$, ν and Σ_{t+1} by estimates. The latter can be calculated iteratively from (14.25) using estimates of A_0, A_1 and B and a starting value for Σ_0.

As an alternative to formal estimation of MGARCH models we can also use the technique of multivariate exponentially weighted moving average (EWMA) forecasting to compute an estimate of Σ_{t+1}. The recursive procedure has been described in Section 9.2.2.

Notes and Comments

The CCC–GARCH model was suggested by Bollerslev (1990), who used it to model European exchange-rate data before and after the introduction of the European Monetary System (EMS) and came to the expected conclusion that conditional correlations after the introduction of the EMS were higher. The idea of the DCC model is explored by Engle (2002), Engle and Sheppard (2001) and Tse and Tsui (2002). Fitting in stages is promoted in the formulation of Engle and Sheppard (2001), and asymptotic statistical theory for this procedure is given. Hafner and Franses (2009) suggest that the dynamics of CCC are too simple for collections of many asset returns and give a generalization. See also the textbook by Engle (2009).

The DVEC model was proposed by Bollerslev, Engle and Wooldridge (1988). The more general (but overparametrized) VEC model is discussed in Engle and Kroner (1995) alongside the BEKK model, named after these two authors as well as Baba and Kraft, who co-authored an earlier unpublished manuscript. The condition for the positive definiteness of Σ_t in (14.15), which suggests the parametrizations (14.17)–(14.19), is described in Attanasio (1991).

There is limited work on statistical properties of quasi-maximum likelihood estimates in multivariate models: Jeantheau (1998) shows consistency for a general formulation and Comte and Lieberman (2003) show asymptotic normality for the BEKK formulation.

The PC-GARCH model was first described by Ding (1994) in a PhD thesis; under the name of orthogonal GARCH it has been extensively investigated by Alexander (2001). The latter shows how PC-GARCH can be used as a dimension-reduction tool for expressing the conditional covariances of a number of asset return series in terms of a much smaller number of principal component return series.

Survey articles by Bollerslev, Engle and Nelson (1994) and Bauwens, Laurent and Rombouts (2006) are useful sources of additional information and references for all of these multivariate models.

15
Advanced Topics in Multivariate Modelling

This chapter contains selected advanced topics that extend the presentation of multivariate models and copulas in Chapters 6 and 7. In Section 15.1 we treat topics that relate to multivariate normal mixture distributions and elliptical distributions. In Section 15.2 we extend the theory of Archimedean copulas and consider models that are more general or more flexible than those presented in Section 7.4.

15.1 Normal Mixture and Elliptical Distributions

This section addresses two statistical issues that were raised in Chapter 6. In Section 15.1.1 we give the algorithm for fitting multivariate generalized hyperbolic distributions to data, which was used to create the examples in Section 6.2.4. In Section 15.1.2 we discuss empirical tests for elliptical symmetry.

15.1.1 Estimation of Generalized Hyperbolic Distributions

While univariate GH models have been fitted to return data in many empirical studies, there has been relatively little applied work with the multivariate distributions. However, normal mixture distributions of the kind we have described may be fitted with algorithms of the EM (expectation–maximization) type. In this section we present an algorithm for that purpose and sketch the ideas behind it. Similar methods have been developed independently by other authors and references may be found in Notes and Comments.

Assume we have iid data $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ and wish to fit the multivariate GH distribution, or one of its special cases. Summarizing the parameters by $\boldsymbol{\theta} = (\lambda, \chi, \psi, \boldsymbol{\mu}, \Sigma, \boldsymbol{\gamma})'$, the problem is to maximize

$$\ln L(\boldsymbol{\theta}; \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) = \sum_{i=1}^{n} \ln f_{\boldsymbol{X}}(\boldsymbol{X}_i; \boldsymbol{\theta}), \tag{15.1}$$

where $f_{\boldsymbol{X}}(\boldsymbol{x}; \boldsymbol{\theta})$ denotes the GH density in (6.29).

This problem is not particularly easy at first sight due to the number of parameters and the necessity of maximizing over covariance matrices Σ. However, if we were able to "observe" the latent mixing variables $W_1, \ldots, W_n$ coming from the mixture representation in (6.24), it would be much easier. Since the joint density of any pair $\boldsymbol{X}_i$ and W_i is given by

$$f_{\boldsymbol{X},W}(\boldsymbol{x}, w; \boldsymbol{\theta}) = f_{\boldsymbol{X}|W}(\boldsymbol{x} \mid w; \boldsymbol{\mu}, \Sigma, \boldsymbol{\gamma}) h_W(w; \lambda, \chi, \psi), \tag{15.2}$$

we could construct the likelihood

$$\ln \tilde{L}(\boldsymbol{\theta}; \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n, W_1, \ldots, W_n) = \sum_{i=1}^{n} \ln f_{\boldsymbol{X}|W}(\boldsymbol{X}_i \mid W_i; \boldsymbol{\mu}, \Sigma, \boldsymbol{\gamma}) + \sum_{i=1}^{n} \ln h_W(W_i; \lambda, \chi, \psi), \quad (15.3)$$

where the two terms could be maximized separately with respect to the parameters they involve. The apparently more problematic parameters of Σ and $\boldsymbol{\gamma}$ are in the first term of the likelihood, and estimates are relatively easy to derive due to the Gaussian form of this term.

To overcome the latency of the W_i data the EM algorithm is used. This is an iterative procedure consisting of an E-step, or expectation step (where essentially W_i is replaced with an estimate given the observed data and current parameter estimates), and an M-step, or maximization step (where the parameter estimates are updated). Suppose at the beginning of step k we have the vector of parameter estimates $\boldsymbol{\theta}^{[k]}$. We proceed as follows.

E-step. We calculate the conditional expectation of the so-called augmented likelihood (15.3) given the data $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ using the parameter values $\boldsymbol{\theta}^{[k]}$. This results in the objective function

$$Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{[k]}) = E(\ln \tilde{L}(\boldsymbol{\theta}; \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n, W_1, \ldots, W_n) \mid \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; \boldsymbol{\theta}^{[k]}).$$

M-step. We maximize the objective function with respect to $\boldsymbol{\theta}$ to obtain the next set of estimates $\boldsymbol{\theta}^{[k+1]}$.

Alternating between these steps, the EM algorithm produces improved parameter estimates at each step (in the sense that the value of the original likelihood (15.1) is continually increased) and they converge to the maximum likelihood (ML) estimates.

In practice, performing the E-step amounts to replacing any functions $g(W_i)$ of the latent mixing variables that arise in (15.3) by the quantities $E(g(W_i) \mid \boldsymbol{X}_i; \boldsymbol{\theta}^{[k]})$. To calculate these quantities we can observe that the conditional density of W_i given $\boldsymbol{X}_i$ satisfies $f_{W|\boldsymbol{X}}(w \mid \boldsymbol{x}; \boldsymbol{\theta}) \propto f_{W,\boldsymbol{X}}(w, \boldsymbol{x}; \boldsymbol{\theta})$, up to some constant of proportionality. It may therefore be deduced from (15.2) that

$$W_i \mid \boldsymbol{X}_i \sim N^-(\lambda - \tfrac{1}{2}d, (\boldsymbol{X}_i - \boldsymbol{\mu})' \tilde{\Sigma}^{-1} (\boldsymbol{X}_i - \boldsymbol{\mu}) + \chi, \psi + \boldsymbol{\gamma}' \Sigma^{-1} \boldsymbol{\gamma}). \quad (15.4)$$

If we write out the likelihood (15.3) using (6.25) for the first term and the generalized inverse Gaussian (GIG) density (A.14) for the second term, we find that the functions $g(W_i)$ arising in (15.3) are $g_1(w) = w$, $g_2(w) = 1/w$ and $g_3(w) = \ln(w)$. The conditional expectation of these functions in model (15.4) may be evaluated using information about the GIG distribution in Section A.2.5; note that $E(\ln(W_i) \mid \boldsymbol{X}_i; \boldsymbol{\theta}^{[k]})$ involves derivatives of a Bessel function with respect to order and must be approximated numerically. We will introduce the notation

$$\delta_i^{[\cdot]} = E(W_i^{-1} \mid \boldsymbol{X}_i; \boldsymbol{\theta}^{[\cdot]}), \quad \eta_i^{[\cdot]} = E(W_i \mid \boldsymbol{X}_i; \boldsymbol{\theta}^{[\cdot]}), \quad \xi_i^{[\cdot]} = E(\ln(W_i) \mid \boldsymbol{X}_i; \boldsymbol{\theta}^{[\cdot]}), \quad (15.5)$$

which allows us to describe the basic EM scheme as well as a variant below.

In the M-step there are two terms to maximize, coming from the two terms in (15.3); we write these as $Q_1(\boldsymbol{\mu}, \Sigma, \boldsymbol{\gamma}; \boldsymbol{\theta}^{[k]})$ and $Q_2(\lambda, \chi, \psi; \boldsymbol{\theta}^{[k]})$. To address the identifiability issue mentioned in Section 6.2.3 we constrain the determinant of Σ to be some fixed value (in practice, we take the determinant of the sample covariance matrix S) in the maximization of Q_1. The maximizing values of $\boldsymbol{\mu}$, Σ and $\boldsymbol{\gamma}$ may then be derived analytically by calculating partial derivatives and setting these equal to zero; the resulting formulas are embedded in Algorithm 15.1 below (see steps (3) and (4)). The maximization of $Q_2(\lambda, \chi, \psi; \boldsymbol{\theta}^{[k]})$ with respect to the parameters of the mixing distribution is performed numerically; the function $Q_2(\lambda, \chi, \psi; \boldsymbol{\theta}^{[\cdot]})$ is

$$(\lambda - 1)\sum_{i=1}^{n} \xi_i^{[\cdot]} - \tfrac{1}{2}\chi \sum_{i=1}^{n} \delta_i^{[\cdot]} - \tfrac{1}{2}\psi \sum_{i=1}^{n} \eta_i^{[\cdot]} \\ - \tfrac{1}{2}n\lambda \ln(\chi) + \tfrac{1}{2}n\lambda \ln(\psi) - n \ln(2K_\lambda(\sqrt{\chi\psi})). \quad (15.6)$$

This would complete one iteration of a standard EM algorithm. However, there are a couple of variants on the basic scheme; both involve modification of the final step described above, namely the maximization of Q_2.

Assuming the parameters $\boldsymbol{\mu}$, Σ and $\boldsymbol{\gamma}$ have been updated first in iteration k, we define

$$\boldsymbol{\theta}^{[k,2]} = (\lambda^{[k]}, \chi^{[k]}, \psi^{[k]}, \boldsymbol{\mu}^{[k+1]}, \Sigma^{[k+1]}, \boldsymbol{\gamma}^{[k+1]})',$$

recalculate the weights $\delta_i^{[k,2]}$, $\eta_i^{[k,2]}$ and $\xi_i^{[k,2]}$ in (15.5), and then maximize the function $Q_2(\lambda, \eta, \xi; \boldsymbol{\theta}^{[k,2]})$ in (15.6). This results in a so-called MCECM algorithm (multi-cycle, expectation, conditional maximization), which is the one we present below.

Alternatively, instead of maximizing Q_2 we may maximize the original likelihood (15.1) with respect to λ, χ and ψ with the other parameters held fixed at the values $\boldsymbol{\mu}^{[k]}$, $\Sigma^{[k]}$ and $\boldsymbol{\gamma}^{[k]}$; this results in an ECME algorithm.

Algorithm 15.1 (EM estimation of GH distribution).

(1) Set iteration count $k = 1$ and select starting values for $\boldsymbol{\theta}^{[1]}$. In particular, reasonable starting values for $\boldsymbol{\mu}$, $\boldsymbol{\gamma}$ and Σ, respectively, are the sample mean, the zero vector and the sample covariance matrix S.

(2) Calculate weights $\delta_i^{[k]}$ and $\eta_i^{[k]}$ using (15.5), (15.4) and (A.15). Average the weights to get

$$\bar{\delta}^{[k]} = n^{-1}\sum_{i=1}^{n} \delta_i^{[k]} \quad \text{and} \quad \bar{\eta}^{[k]} = n^{-1}\sum_{i=1}^{n} \eta_i^{[k]}.$$

(3) For a symmetric model set $\boldsymbol{\gamma}^{[k+1]} = \mathbf{0}$. Otherwise set

$$\boldsymbol{\gamma}^{[k+1]} = \frac{n^{-1}\sum_{i=1}^{n} \delta_i^{[k]}(\bar{\boldsymbol{X}} - \boldsymbol{X}_i)}{\bar{\delta}^{[k]}\bar{\eta}^{[k]} - 1}.$$

(4) Update estimates of the location vector and dispersion matrix by

$$\boldsymbol{\mu}^{[k+1]} = \frac{n^{-1}\sum_{i=1}^{n}\delta_i^{[k]}\boldsymbol{X}_i - \boldsymbol{\gamma}^{[k+1]}}{\bar{\delta}^{[k]}},$$

$$\Psi = \frac{1}{n}\sum_{i=1}^{n}\delta_i^{[k]}(\boldsymbol{X}_i - \boldsymbol{\mu}^{[k+1]})(\boldsymbol{X}_i - \boldsymbol{\mu}^{[k+1]})' - \bar{\eta}^{[k]}\boldsymbol{\gamma}^{[k+1]}\boldsymbol{\gamma}^{[k+1]\prime},$$

$$\Sigma^{[k+1]} = \frac{|S|^{1/d}\Psi}{|\Psi|^{1/d}}.$$

(5) Set

$$\boldsymbol{\theta}^{[k,2]} = (\lambda^{[k]}, \chi^{[k]}, \psi^{[k]}, \boldsymbol{\mu}^{[k+1]}, \Sigma^{[k+1]}, \boldsymbol{\gamma}^{[k+1]})'.$$

Calculate weights $\delta_i^{[k,2]}$, $\eta_i^{[k,2]}$ and $\xi_i^{[k,2]}$ using (15.5), (15.4) and information in Section A.2.5.

(6) Maximize $Q_2(\lambda, \chi, \psi; \boldsymbol{\theta}^{[k,2]})$ in (15.6) with respect to λ, χ and ψ to complete the calculation of $\boldsymbol{\theta}^{[k,2]}$. Increment iteration count $k \to k+1$ and go to step (2).

This algorithm may be easily adapted to fit special cases of the GH distribution. This involves holding certain parameters fixed throughout and maximizing with respect to the remaining parameters: for the hyperbolic distribution we set $\lambda = 1$; for the NIG distribution $\lambda = -\frac{1}{2}$; for the t distribution $\psi = 0$; for the VG distribution $\chi = 0$. In the case of the t and VG distributions, in step (6) we have to work with the function Q_2 that results from assuming an inverse gamma or gamma density for h_W.

15.1.2 Testing for Elliptical Symmetry

The general problem of this section is to test whether a sample of identically distributed data vectors $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ has an elliptical distribution $E_d(\boldsymbol{\mu}, \Sigma, \psi)$ for some $\boldsymbol{\mu}$, Σ and generator ψ. In all of the methods we require estimates of $\boldsymbol{\mu}$ and Σ, and these can be obtained using approaches discussed in Section 6.3.4, such as fitting t distributions, calculating M-estimates or perhaps using (6.49) in the bivariate case. We denote the estimates simply by $\hat{\boldsymbol{\mu}}$ and $\hat{\Sigma}$.

In finance we cannot generally assume that the observations are of iid random vectors, but we assume that they at least have an identical distribution. Note that, even if the data were independent, the fact that we generally estimate $\boldsymbol{\mu}$ and Σ from the whole data set would introduce dependence in the procedures that we describe below.

Stable correlation estimates: an exploratory method. An easy exploratory graphical method can be based on Proposition 6.28. We could attempt to estimate

$$\rho(\boldsymbol{X} \mid h(\boldsymbol{X}) \geqslant c), \qquad h(\boldsymbol{x}) = (\boldsymbol{x} - \hat{\boldsymbol{\mu}})'\hat{\Sigma}^{-1}(\boldsymbol{x} - \hat{\boldsymbol{\mu}})$$

for various values of $c \geqslant 0$. We expect that for elliptically distributed data the estimates will remain roughly stable over a range of different c values. Of course the estimates of this correlation should again be calculated using some method that is more efficient than the standard correlation estimator for heavy-tailed data. The

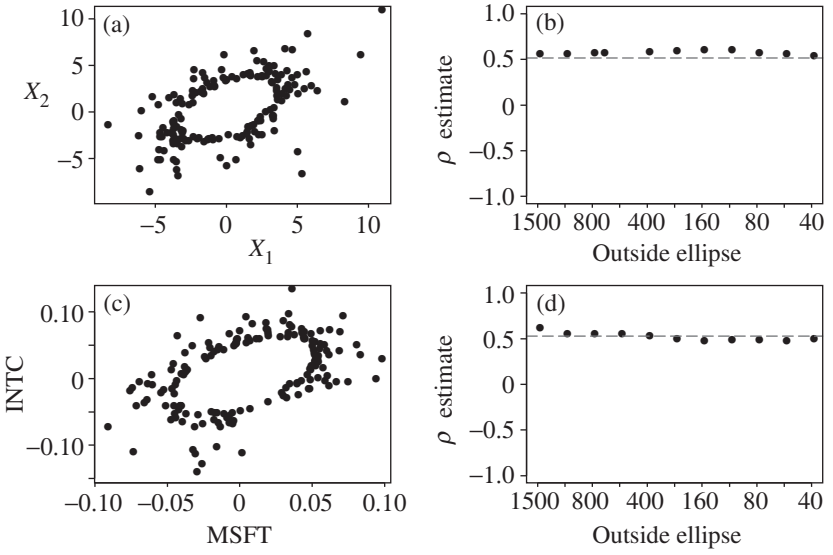

Figure 15.1. Correlations are estimated using the Kendall's tau method for points lying outside ellipses of progressively larger size (as shown in (a) and (c)). (a), (b) Two thousand t-distributed data with four degrees of freedom and $\rho = 0.5$. (c), (d) Two thousand daily log-returns on Microsoft and Intel. Points show estimates for an ellipse that is allowed to grow until there are only forty points outside; dashed lines show estimates of correlation for all data.

method is most natural as a bivariate method, and in this case the correlation of $\boldsymbol{X} \mid h(\boldsymbol{X}) \geqslant c$ can be estimated by applying the Kendall's tau transform method to those data points $\boldsymbol{X}_i$ that lie outside the ellipse defined by $h(\boldsymbol{x}) = c$. In Figure 15.1 we give an example with both simulated and real data, neither of which show any marked departure from the assumption of stable correlations. The method is of course exploratory and does not allow us to come to any formal conclusion.

Q–Q plots. The remaining methods that we describe rely on the link (6.41) between non-singular elliptical and spherical distributions. If $\boldsymbol{\mu}$ and Σ were known, then we would test for elliptical symmetry by testing the data $\{\Sigma^{-1/2}(\boldsymbol{X}_i - \boldsymbol{\mu}) : i = 1, \ldots, n\}$ for spherical symmetry. Replacing these parameters by estimates as above, we consider whether the data

$$\{\boldsymbol{Y}_i = \hat{\Sigma}^{-1/2}(\boldsymbol{X}_i - \hat{\boldsymbol{\mu}}) : i = 1, \ldots, n\} \tag{15.7}$$

are consistent with a spherical distribution, while ignoring the effect of estimation error.

Some graphical methods based on Q–Q plots have been suggested by Li, Fang and Zhu (1997) and these are particularly useful for large d. These essentially rely on the following result.

Lemma 15.2. *Suppose that $T(\boldsymbol{Y})$ is a statistic such that, almost surely,*

$$T(a\boldsymbol{Y}) = T(\boldsymbol{Y}) \quad \text{for every } a > 0. \tag{15.8}$$

Then $T(\boldsymbol{Y})$ has the same distribution for every spherical vector $\boldsymbol{Y} \sim S_d^+(\psi)$.

Proof. From Theorem 6.21 we have $T(\boldsymbol{Y}) \stackrel{\mathrm{d}}{=} T(R\boldsymbol{S})$, and $T(R\boldsymbol{S}) \stackrel{\text{a.s.}}{=} T(\boldsymbol{S})$ follows from (15.8). Since the distribution of $T(\boldsymbol{Y})$ only depends on $\boldsymbol{S}$ and not R, it must be the same for all $\boldsymbol{Y} \sim S_d^+(\psi)$. □

We exploit this result by looking for statistics $T(\boldsymbol{Y})$ with the property (15.8), whose distribution we know when $\boldsymbol{Y} \sim N_d(\boldsymbol{0}, I_d)$. Two examples are

$$\left.\begin{aligned} T_1(\boldsymbol{Y}) &= \frac{d^{1/2}\bar{Y}}{\sqrt{(1/(d-1))\sum_{i=1}^d (Y_i - \bar{Y})^2}}, \quad \bar{Y} = \frac{1}{d}\sum_{i=1}^d Y_i, \\ T_2(\boldsymbol{Y}) &= \frac{\sum_{i=1}^k Y_i^2}{\sum_{i=1}^d Y_i^2}. \end{aligned}\right\} \tag{15.9}$$

For $\boldsymbol{Y} \sim N_d(\boldsymbol{0}, I_d)$, and hence for $\boldsymbol{Y} \sim S_d^+(\psi)$, we have $T_1(\boldsymbol{Y}) \sim t_{d-1}$ and $T_2(\boldsymbol{Y}) \sim \text{Beta}(\frac{1}{2}k, \frac{1}{2}(d-k))$.

Our experience suggests that the beta plot is the more revealing of the resulting Q–Q plots. Li, Fang and Zhu (1997) suggest choosing k such that it is roughly equal to $d - k$. In Figure 15.2 we show examples of the Q–Q plots obtained for 2000 simulated data from a ten-dimensional t distribution with four degrees of freedom and for the daily, weekly and monthly return data on ten Dow Jones 30 stocks that were analysed in Example 6.3 and Section 6.2.4. The curvature in the plots for daily and weekly returns seems to be evidence against the elliptical hypothesis.

Numerical tests. We restrict ourselves to simple ideas for bivariate tests; references to more general test ideas are found in Notes and Comments. If we neglect the error involved in estimating location and dispersion, testing for elliptical symmetry amounts to testing the $\boldsymbol{Y}_i$ data in (15.7) for spherical symmetry. For $i = 1, \ldots, n$, if we set $R_i = \|\boldsymbol{Y}_i\|$ and $\boldsymbol{S}_i = \boldsymbol{Y}_i / \|\boldsymbol{Y}_i\|$, then under the null hypothesis the $\boldsymbol{S}_i$ data should be uniformly distributed on the unit sphere $\mathcal{S}^{d-1}$, and the paired data $(R_i, \boldsymbol{S}_i)$ should form realizations of independent pairs.

In the bivariate case, testing for uniformity on the unit circle $\mathcal{S}^1$ amounts to a univariate test of uniformity on $[0, 2\pi]$ for the angles Θ_i described by the points $\boldsymbol{S}_i = (\cos\Theta_i, \sin\Theta_i)'$ on the perimeter of the circle; equivalently, we may test the data $\{U_i := \Theta_i/(2\pi) \colon i = 1, \ldots, n\}$ for uniformity on $[0, 1]$. Neglecting issues of serial dependence in the data, this may be done, for instance, by a standard chi-squared goodness-of-fit test (see Rice 1995, p. 241) or a Kolmogorov–Smirnov test (see Conover 1999). Again neglecting issues of serial dependence, the independence of the components of the pairs $\{(R_i, U_i) \colon i = 1, \ldots, n\}$ could be examined by performing a test of association with Spearman's rank correlation coefficient (see, for example, Conover 1999, pp. 312–328).

We have performed these tests for the two data sets used in Figure 15.1, these being 2000 simulated bivariate t data with four degrees of freedom and 2000 daily log-returns for Intel and Microsoft. In Figure 15.3 the transformed data on the unit circle $\boldsymbol{S}_i$ and the implied angles U_i on the $[0, 1]$ scale are shown; the dispersion matrices have been estimated using the construction (6.49) based on Kendall's tau.

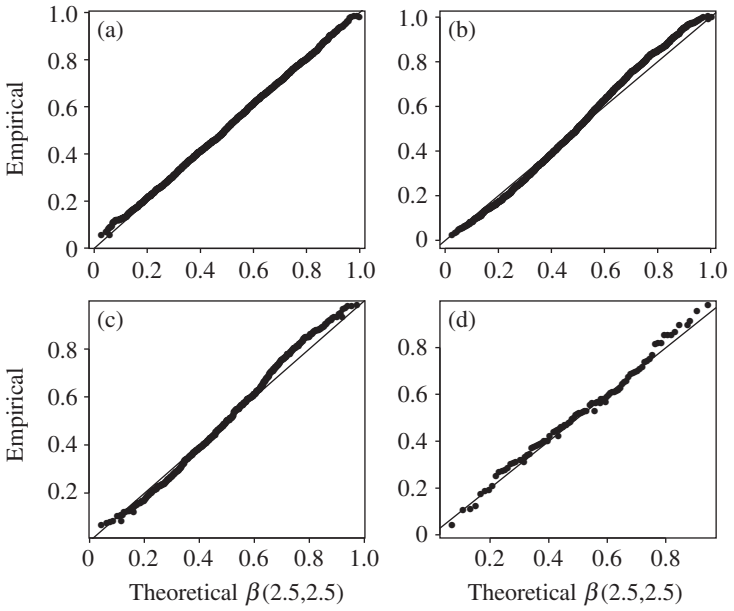

Figure 15.2. Q–Q plots of the beta statistic (15.9) for four data sets with dimension $d = 10$; we have set $k = 5$. (a) Two thousand simulated observations from a t distribution with four degrees of freedom. (b) Daily, (c) weekly and (d) monthly returns on Dow Jones stocks as analysed in Example 6.3 and Section 6.2.4. Daily and weekly returns show evidence against elliptical symmetry.

Neither of these data sets shows significant evidence against the elliptical hypothesis. For the bivariate t data the p-values for the chi-squared and Kolmogorov–Smirnov tests of uniformity and the Spearman's rank test of association are, respectively, 0.99, 0.90 and 0.10. For the stock-return data they are 0.08, 0.12 and 0.19. Note that simulated data from lightly skewed members of the GH family often fail these tests.

Notes and Comments

EM algorithms for the multivariate GH distribution have been independently proposed by Protassov (2004) and Neil Shephard (personal communication). Our approach is based on EM-type algorithms for fitting the multivariate t distribution with unknown degrees of freedom. A good starter reference on this subject is Liu and Rubin (1995), where the use of the MCECM algorithm of Meng and Rubin (1993) and the ECME algorithm proposed in Liu and Rubin (1994) is discussed. Further refinements of these algorithms are discussed in Liu (1997) and Meng and van Dyk (1997).

The Q–Q plots for testing spherical symmetry were suggested by Li, Fang and Zhu (1997). There is a large literature on tests of spherical symmetry, including Smith (1977), Kariya and Eaton (1977), Beran (1979) and Baringhaus (1991). This work is also related to tests of uniformity for directional data: see Mardia (1972), Giné (1975) and Prentice (1978).

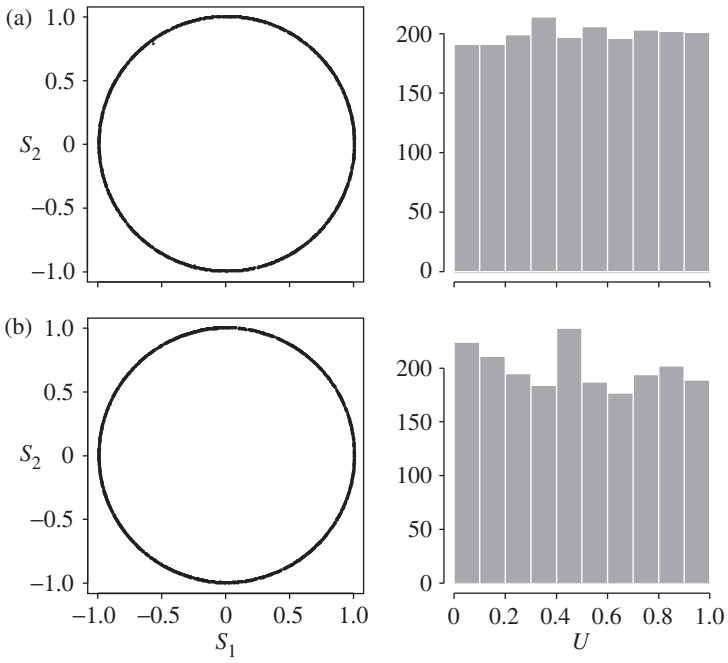

Figure 15.3. Illustration of the transformation of bivariate data to points on the unit circle $\mathcal{S}^1$ using the transformation $\boldsymbol{S}_i = \boldsymbol{Y}_i/\|\boldsymbol{Y}_i\|$, where the $\boldsymbol{Y}_i$ data are defined in (15.7); the angles of these points are then transformed on to the [0, 1] scale, where they can be tested for uniformity. (a) Two thousand simulated t data with four degrees of freedom. (b) Two thousand Intel and Microsoft log-returns. Neither show strong evidence against elliptical symmetry.

15.2 Advanced Archimedean Copula Models

In Section 7.4.2 we gave a characterization result for Archimedean copula generators that can be used to generate copulas in any dimension d (Theorem 7.50). In Section 15.2.1 we provide a more general result that characterizes the larger class of generators that may be used in a given dimension d.

The Archimedean copulas discussed so far have all been models for exchangeable random vectors. We discuss non-exchangeable, asymmetric extensions of the Archimedean family in Section 15.2.2.

15.2.1 Characterization of Archimedean Copulas

Recall that multivariate Archimedean copulas take the form

$$C(u_1, \ldots, u_d) = \psi(\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_d)),$$

where ψ is an Archimedean generator function, i.e. a decreasing, continuous, convex function $\psi : [0, \infty) \to [0, 1]$ satisfying $\psi(0) = 1$ and $\lim_{t\to\infty} \psi(t) = 0$. The necessary and sufficient condition for ψ to generate a copula in every dimension $d \geqslant 2$ is that it should be a completely monotonic function satisfying

$$(-1)^k \frac{\mathrm{d}^k}{\mathrm{d}t^k}\psi(t) \geqslant 0, \quad k \in \mathbb{N},\ t \in (0, \infty).$$

We also recall that these generators may be characterized as the Laplace–Stieltjes transforms of dfs G on $[0, \infty)$ such that $G(0) = 0$.

While Archimedean copulas with completely monotonic generators can be used in any dimension, if we are interested in Archimedean copulas in a given dimension d, we can relax the requirement of complete monotonicity and substitute the weaker requirement of *d-monotonicity*. A generator ψ is d-monotonic if it is differentiable up to order $(d-2)$ on $(0, \infty)$ with derivatives satisfying

$$(-1)^k \frac{\mathrm{d}^k}{\mathrm{d}t^k}\psi(t) \geqslant 0, \quad k = 0, 1, \ldots, d-2,$$

and if $(-1)^{d-2}\psi^{(d-2)}$ is decreasing and convex.

Theorem 15.3. *If $\psi : [0, \infty) \to [0, 1]$ is an Archimedean copula generator, then the construction (7.46) gives a copula in dimension d if and only if ψ is d-monotonic.*

Proof. See McNeil and Nešlehová (2009). □

If ψ is a d-monotonic generator, we write $\psi \in \Psi_d$. Examples of generators in Ψ_d that are not in Ψ_{d+1} (or Ψ_∞) are $\psi(t) = \max((1-t)^{d-1}, 0)$ and the Clayton generator $\psi(t) = \max((1+\theta t)^{-1/\theta}, 0)$ for $-1/(d-1) \leqslant \theta < -1/d$ (see McNeil and Nešlehová (2009) for details). An elegant way of characterizing the d-monotonic copula generators is in terms of a lesser-known integral transform.

Let G be a df on $[0, \infty)$ satisfying $G(0) = 0$ and let X be an rv with df G. Then, for $t \geqslant 0$ and $d \geqslant 2$, the Williamson d-transform of G (or X) is the function

$$\begin{aligned}\mathfrak{W}_d G(t) &= \int_0^\infty \max\left(\left(1 - \frac{t}{x}\right)^{d-1}, 0\right) \mathrm{d}G(x) \\ &= E\left(\max\left(\left(1 - \frac{t}{X}\right)^{d-1}, 0\right)\right).\end{aligned}$$

Every Williamson d-transform of a df G satisfying $G(0) = 0$ is a d-monotonic copula generator, and every d-monotonic copula generator is the Williamson d-transform of a unique df G. In the same way that a rich variety of copulas for arbitrary dimensions can be created by taking Laplace–Stieltjes transforms of particular distributions, a rich variety of copulas for dimension d can be created by taking Williamson d-transforms. For example, in McNeil and Nešlehová (2010) it is shown that by taking X to have a gamma, inverse-gamma or Pareto distribution we can obtain interesting families of copulas that give rise to a wide range of Spearman's or Kendall's rank correlation values.

There is an elegant connection between d-monotonic copulas generators and *simplex* distributions, which can form the basis of a sampling algorithm for the corresponding copulas. A d-dimensional random vector $\boldsymbol{X}$ is said to have a simplex distribution if $\boldsymbol{X} \overset{\mathrm{d}}{=} R\boldsymbol{S}_d$, where $\boldsymbol{S}_d$ is an rv that is distributed uniformly on the unit simplex $\mathcal{S}_d = \{\boldsymbol{x} \in \mathbb{R}_+^d : x_1 + \cdots + x_d = 1\}$ and R is an independent non-negative scalar rv with df F_R known as the radial distribution of the simplex distribution. A key theorem shows that the class of d-dimensional Archimedean copulas is equivalent

to the class of survival copulas of d-dimensional simplex distributions (excluding those with point mass at the origin).

Theorem 15.4. *If $\boldsymbol{X}$ has a simplex distribution with radial distribution F_R satisfying $F_R(0) = 0$, then the survival copula of $\boldsymbol{X}$ is Archimedean with generator $\psi = \mathfrak{W}_d F_R$.*

Proof. Let $\boldsymbol{x} \in \mathbb{R}^d_+$ and write $\boldsymbol{S}_d = (S_1, \ldots, S_d)'$. We use the fact that the survival function of $\boldsymbol{S}_d$ is given by

$$P(S_1 > x_1, \ldots, S_d > x_d) = \max\left(\left(1 - \sum_{i=1}^{d} x_i\right)^{d-1}, 0\right)$$

(Fang, Kotz and Ng 1990, p. 115) to establish that

$$\begin{aligned} P(X_1 > x_1, \ldots, X_d > x_d) &= E\left(P\left(S_1 > \frac{x_1}{R}, \ldots, S_d > \frac{x_d}{R}\right)\right) \\ &= E\left(\max\left(\left(1 - \frac{x_1 + \cdots + x_d}{R}\right)^{d-1}, 0\right)\right) \\ &= \psi(x_1 + \cdots + x_d). \end{aligned}$$

The marginal survival functions $P(X_i > x) = \psi(x)$ are continuous and strictly decreasing on $\{x\colon \psi(x) > 0\}$ and it may be verified that, for any $\boldsymbol{x} \in \mathbb{R}^d_+$, we have $\psi(x_1 + \cdots + x_d) = \psi(\psi^{-1}(\psi(x_1)) + \cdots + \psi^{-1}(\psi(x_d)))$, so we can write

$$P(X_1 > x_1, \ldots, X_d > x_d) = \psi(\psi^{-1}(\psi(x_1)) + \cdots + \psi^{-1}(\psi(x_d))),$$

which proves the assertion; for more technical details see McNeil and Nešlehová (2009). □

15.2.2 Non-exchangeable Archimedean Copulas

A copula obtained from construction (7.46) is obviously an *exchangeable* copula conforming to (7.20). While exchangeable bivariate Archimedean copulas are widely used in modelling applications, their exchangeable multivariate extensions represent a very specialized form of dependence structure and have more limited applications. An exception to this is in the area of credit risk, although even here more general models with group structures are also needed. It is certainly natural to enquire whether there are extensions to the Archimedean class that are not rigidly exchangeable, and we devote this section to a short discussion of some possible extensions.

Asymmetric Archimedean copulas. Let C_θ be any exchangeable d-dimensional copula. A parametric family of asymmetric copulas $C_{\theta,\alpha_1,\ldots,\alpha_d}$ is then obtained by setting

$$C_{\theta,\alpha_1,\ldots,\alpha_d}(u_1, \ldots, u_d) = C_\theta(u_1^{\alpha_1}, \ldots, u_d^{\alpha_d}) \prod_{i=1}^{d} u_i^{1-\alpha_i}, \quad (u_1, \ldots, u_d) \in \mathbb{R}^d, \tag{15.10}$$

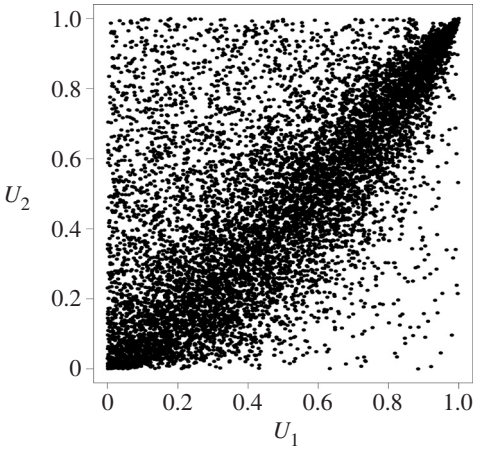

Figure 15.4. Pairwise scatterplots of 10 000 simulated points from an asymmetric Gumbel copula $C^{\mathrm{Gu}}_{4,0.95,0.7}$ based on construction (15.10). This is simulated using Algorithm 15.5.

where $0 \leqslant \alpha_i \leqslant 1$ for all i. Only in the special case $\alpha_1 = \cdots = \alpha_d$ is the copula (15.10) exchangeable. Note also that when the α_i parameters are 0, $C_{\theta,0,\ldots,0}$ is the independence copula, and when the α_i parameters are 1, $C_{\theta,1,\ldots,1}$ is simply C_θ. When C_θ is an Archimedean copula, we refer to copulas constructed by (15.10) as asymmetric Archimedean copulas.

We check that (15.10) defines a copula by constructing a random vector with this df and observing that its margins are standard uniform. The construction is presented as a simulation algorithm.

Algorithm 15.5 (asymmetric Archimedean copula).

(1) Generate a random vector $(V_1, \ldots, V_d)$ with df C_θ.

(2) Generate, independently of $(V_1, \ldots, V_d)$, independent standard uniform variates $\tilde{U}_i$ for $i = 1, \ldots, d$.

(3) Return $U_i = \max(V_i^{1/\alpha_i}, \tilde{U}_i^{1/(1-\alpha_i)})$ for $i = 1, \ldots, d$.

It may be easily verified that $(U_1, \ldots, U_d)$ have the df (15.10). See Figure 15.4 for an example of simulated data from an asymmetric bivariate copula based on Gumbel's copula. Note that an alternative copula may be constructed by taking $(\tilde{U}_1, \ldots, \tilde{U}_d)$ in Algorithm 15.5 to be distributed according to some copula other than the independence copula.

Nested Archimedean copulas. Non-exchangeable, higher-dimensional Archimedean copulas with exchangeable bivariate margins can be constructed by recursive application (or nesting) of Archimedean generators and their inverses, and we will give examples in this section. The biggest problem with these constructions lies in checking that they lead to valid multivariate distributions satisfying (7.1). The necessary theory is complicated and we will simply indicate the nature of the conditions that are necessary without providing justification; a comprehensive reference is Joe (1997). It turns out that with some care we can construct situations of *partial*

exchangeability. We give three- and four-dimensional examples that indicate the pattern of construction.

Example 15.6 (three-dimensional non-exchangeable Archimedean copulas). Suppose that ϕ_1 and ϕ_2 are two Archimedean copula generators and consider

$$C(u_1, u_2, u_3) = \psi_2(\psi_2^{-1} \circ \psi_1(\psi_1^{-1}(u_1) + \psi_1^{-1}(u_2)) + \psi_2^{-1}(u_3)). \tag{15.11}$$

Conditions that ensure that this is a copula are that the generators ψ_1 and ψ_2 are completely monotonic functions, as in (7.47), and that the composition $\psi_2^{-1} \circ \psi_1 \colon [0, \infty) \to [0, \infty)$ is a function whose derivative is a completely monotonic function.

Observe that when $\psi_2 = \psi_1 = \psi$ we are back in the situation of full exchangeability, as in (7.46). Otherwise, if $\psi_1 \neq \psi_2$ and (U_1, U_2, U_3) is a random vector with df given by (15.11), then only U_1 and U_2 are exchangeable, i.e. $(U_1, U_2, U_3) \stackrel{\mathrm{d}}{=} (U_2, U_1, U_3)$, but no other swapping of subscripts is possible. All bivariate margins of (15.11) are themselves Archimedean copulas. The margins C_{13} and C_{23} have generator ψ_2 and C_{12} has generator ψ_1.

Example 15.7 (four-dimensional non-exchangeable Archimedean copulas). A possible four-dimensional construction is

$$\begin{aligned} C(u_1, u_2, u_3, u_4) = \psi_3(\psi_3^{-1} \circ \psi_1(\psi_1^{-1}(u_1) + \psi_1^{-1}(u_2)) \\ + \psi_3^{-1} \circ \psi_2(\psi_2^{-1}(u_3) + \psi_2^{-1}(u_4))), \end{aligned} \tag{15.12}$$

where ψ_1, ψ_2 and ψ_3 are three distinct, completely monotonic Archimedean copula generators and we assume that the composite functions $\psi_3^{-1} \circ \psi_1$ and $\psi_3^{-1} \circ \psi_2$ have derivatives that are completely monotonic to obtain a proper distribution function. This is not the only possible four-dimensional construction (Joe 1997), but it is a useful construction because it gives two exchangeable groups. If (U_1, U_2, U_3, U_4) has the df (15.12), then U_1 and U_2 are exchangeable, as are U_3 and U_4.

The same kinds of construction can be extended to higher dimensions, subject again to complete monotonicity conditions on the compositions of generator inverses and generators (see Notes and Comments).

LT-Archimedean copulas with p-factor structure. Recall from Definition 7.52 the family of LT-Archimedean copulas. The arguments of Proposition 7.51 establish that these have the form

$$\begin{aligned} C(u_1, \ldots, u_d) &= \hat{G}(\hat{G}^{-1}(u_1) + \cdots + \hat{G}^{-1}(u_d)) \\ &= E\left(\exp\left(-V \sum_{i=1}^{d} \hat{G}^{-1}(u_i)\right)\right) \end{aligned} \tag{15.13}$$

for a strictly positive rv V with Laplace–Stieltjes transform $\hat{G}$.

It is possible to generalize the construction (15.13) to obtain a larger family of non-exchangeable copulas. An LT-Archimedean copula with p-factor structure is

constructed from a p-dimensional random vector $\boldsymbol{V} = (V_1, \ldots, V_p)'$ with independent, strictly positive components and a matrix $A \in \mathbb{R}^{d\times p}$ with elements $a_{ij} > 0$ as follows:

$$C(u_1, \ldots, u_d) = E\left(\exp\left(-\sum_{i=1}^{d} \boldsymbol{a}_i' \boldsymbol{V} \hat{G}_i^{-1}(u_i)\right)\right), \tag{15.14}$$

where $\boldsymbol{a}_i$ is the ith row of A and $\hat{G}_i^{-1}$ is the Laplace–Stieltjes transform of the strictly positive rv $\boldsymbol{a}_i'\boldsymbol{V}$.

We can write (15.14) in a different way, which facilitates the computation of $C(u_1, \ldots, u_d)$. Note that

$$\sum_{i=1}^{d} \boldsymbol{a}_i' \boldsymbol{V} \hat{G}_i^{-1}(u_i) = \sum_{j=1}^{p} V_j \sum_{i=1}^{d} a_{ij} \hat{G}_i^{-1}(u_i).$$

It follows from the independence of the V_j that

$$\begin{aligned} C(u_1, \ldots, u_d) &= \prod_{j=1}^{p} E\left(\exp\left(-V_j \sum_{i=1}^{d} a_{ij} \hat{G}_i^{-1}(u_i)\right)\right) \\ &= \prod_{j=1}^{p} \hat{G}_{V_j}\left(\sum_{i=1}^{d} a_{ij} \hat{G}_i^{-1}(u_i)\right). \end{aligned} \tag{15.15}$$

Note that (15.15) is fairly easy to evaluate when $\hat{G}_{V_j}$, the Laplace–Stieltjes transform of the V_j, is available in closed form, because $\hat{G}_i(t) = \prod_{j=1}^{p} \hat{G}_{V_j}(a_{ij}t)$ by the independence of the V_j.

Notes and Comments

The relationship between d-monotonic functions and Archimedean copulas in dimension d, as well as the link to simplex distributions, is developed in McNeil and Nešlehová (2009); see also McNeil and Nešlehová (2010), which provides many examples of such copulas and generalizes the theory to so-called Liouville copulas, the survival copulas of Liouville distributions.

For more details on the asymmetric copulas obtained from construction (15.10) and ideas for more general asymmetric copulas see Genest, Ghoudi and Rivest (1998). These copula classes were introduced in the PhD thesis of Khoudraji (1995). For additional theory concerning nested higher-dimensional Archimedean copulas and sampling algorithms see Joe (1997), McNeil (2008), Hofert (2008, 2011, 2012) and Hofert and Mächler (2011). LT-Archimedean copulas with p-factor structure were proposed by Rogge and Schönbucher (2003) with applications in credit risk in mind. Krupskii and Joe (2013) extend the idea to suggest even more general ways of constructing factor copula models.

16
Advanced Topics in Extreme Value Theory

This chapter extends the treatment of EVT in Chapter 5. In Section 16.1 we provide more information about the tails of some of the distributions and models that are prominent in this book, including the tails of normal variance mixture models (Chapter 6) and strictly stationary GARCH models (Chapter 4).

In Section 16.2 we build on the point process framework for describing the occurrence and magnitude of threshold exceedances (the POT model) that was described in Section 5.3. We show how self-exciting processes for extremes may be developed based on the idea of Hawkes processes.

Sections 16.3 and 16.4 provide a concise summary of the more important ideas in multivariate EVT; they deal, respectively, with multivariate maxima and multivariate threshold exceedances. The novelty of these sections is that the ideas are presented as far as possible using the copula methodology of Chapter 7. The style is similar to Sections 5.1 and 5.2, with the main results being mostly stated without proof and an emphasis being given to examples relevant for applications.

16.1 Tails of Specific Models

In this section we survey the tails of some of the more important distributions and models that we have encountered in this book.

16.1.1 Domain of Attraction of the Fréchet Distribution

As stated in Section 5.1.2, the domain of attraction of the Fréchet distribution consists of distributions with regularly varying tails of the form $\bar{F}(x) = x^{-\alpha}L(x)$ for $\alpha > 0$, where α is known as the tail index. These are heavy-tailed models where higher-order moments cease to exist. Normalized maxima of random samples from such distributions converge to a Fréchet distribution with shape parameter $\xi = 1/\alpha$, and excesses over sufficiently high thresholds converge to a generalized Pareto distribution with shape parameter $\xi = 1/\alpha$.

We now show that the Student t distribution and the inverse gamma distribution are in this class; we analyse the former because of its general importance in financial modelling and the latter because it appears as the mixing distribution that yields the Student t in the class of normal variance mixture models (see Example 6.7). In Section 16.1.3 we will see that the mixing distribution in a normal variance mixture model essentially determines the tail of that model.

Both the t and inverse gamma distributions are presented in terms of their density, and the analysis of their tails proves to be a simple application of a useful result known as Karamata's Theorem, which is given in Section A.1.4.

Example 16.1 (Student t distribution). It is easily verified that the standard univariate t distribution with $\nu \geqslant 1$ has a density of the form $f_\nu(x) = x^{-(\nu+1)}L(x)$ where L is a slowly varying function. Karamata's Theorem (see Theorem A.7) therefore allows us to calculate the form of the tail $\bar{F}_\nu(x) = \int_x^\infty f_\nu(y)\,\mathrm{d}y$ by essentially treating the slowly varying function as a constant and taking it out of the integral. We get

$$\bar{F}_\nu(x) = \int_x^\infty y^{-(\nu+1)}L(y)\,\mathrm{d}y \sim \nu^{-1}x^{-\nu}L(x), \quad x \to \infty,$$

from which we conclude that the df F_ν of a t distribution has tail index ν and $F_\nu \in \mathrm{MDA}(H_{1/\nu})$ by Theorem 5.8.

Example 16.2 (inverse gamma distribution). The density of the inverse gamma distribution is given in (A.17). It is of the form $f_{\alpha,\beta}(x) = x^{-(\alpha+1)}L(x)$, since $\mathrm{e}^{-\beta/x} \to 1$ as $x \to \infty$. Using the same technique as in Example 16.1, we deduce that this distribution has tail index α, so $F_{\alpha,\beta} \in \mathrm{MDA}(H_{1/\alpha})$.

16.1.2 Domain of Attraction of the Gumbel Distribution

The Gumbel class consists of the so-called *von Mises* distribution functions and any other distributions that are *tail equivalent* to von Mises distributions (see Embrechts, Klüppelberg and Mikosch 1997, pp. 138–150). We give the definitions of both of these concepts below. Note that distributions in this class can have both infinite and finite right endpoints; again we write $x_F = \sup\{x \in \mathbb{R}\colon F(x) < 1\} \leqslant \infty$ for the right endpoint of F.

Definition 16.3 (von Mises distribution function). Suppose there exists some $z < x_F$ such that F has the representation

$$\bar{F}(x) = c\exp\left\{-\int_z^x \frac{1}{a(t)}\,\mathrm{d}t\right\}, \quad z < x < x_F,$$

where c is some positive constant, $a(t)$ is a positive and absolutely continuous function with derivative a', and $\lim_{x\uparrow x_F} a'(x) = 0$. Then F is called a von Mises distribution function.

Definition 16.4 (tail equivalence). Two dfs F and G are called *tail equivalent* if they have the same right endpoints (i.e. $x_F = x_G$) and $\lim_{x\uparrow x_F} \bar{F}(x)/\bar{G}(x) = c$ for some constant $0 < c < \infty$.

To decide whether a particular df F is a von Mises df, the following condition is extremely useful. Assume there exists some $z < x_F$ such that F is twice differentiable on (z, x_F) with density $f = F'$ and $F'' < 0$ in (z, x_F). Then F is a von Mises df if and only if

$$\lim_{x\uparrow x_F} \frac{\bar{F}(x)F''(x)}{f^2(x)} = -1. \tag{16.1}$$

We now use this condition to show that the gamma df is a von Mises df.

Example 16.5 (gamma distribution). The density $f = f_{\alpha,\beta}$ of the gamma distribution is given in (A.13), and a straightforward calculation yields $F''(x) = f'(x) = -f(x)(\beta + (1-\alpha)/x) < 0$, provided $x > \max((\alpha - 1)/\beta, 0)$. Clearly, $\lim_{x\to\infty} F''(x)/f(x) = -\beta$. Moreover, using L'Hôpital's rule we get $\lim_{x\to\infty} \bar{F}(x)/f(x) = \lim_{x\to\infty} -f(x)/f'(x) = \beta^{-1}$. Combining these two limits establishes (16.1).

Example 16.6 (GIG distribution). The density of an rv $X \sim N^-(\lambda, \chi, \psi)$ with the GIG distribution is given in (A.14). Let $F_{\lambda,\chi,\psi}(x)$ denote the df and consider the case where $\psi > 0$. If $\psi = 0$, then the GIG is an inverse gamma distribution, which was shown in Example 16.2 to be in the Fréchet class. If $\psi > 0$, then $\lambda \geqslant 0$, and a similar technique to Example 16.5 could be used to establish that the GIG is a von Mises df. In the case where $\lambda > 0$ it is easier to demonstrate tail equivalence with a gamma distribution, which is the special case when $\chi = 0$. We observe that

$$\lim_{x\to\infty} \frac{\bar{F}_{\lambda,\chi,\psi}(x)}{\bar{F}_{\lambda,0,\psi}(x)} = \lim_{x\to\infty} \frac{f_{\lambda,\chi,\psi}(x)}{f_{\lambda,0,\psi}(x)} = c_{\lambda,\chi,\psi}$$

for some constant $0 < c_{\lambda,\chi,\psi} < \infty$. It follows that $F_{\lambda,\chi,\psi} \in \mathrm{MDA}(H_0)$.

16.1.3 Mixture Models

In this book we have considered a number of models for financial risk-factor changes that arise as mixtures (or products) of rvs. In Chapter 6 we introduced multivariate normal variance mixture models including the Student t and (symmetric) GH distributions, which have the general structure given in (6.18). A one-dimensional normal variance mixture (or the marginal distribution of a d-dimensional normal variance mixture) is of the same type (see Section A.1.1) as an rv X satisfying

$$X \overset{\mathrm{d}}{=} \sqrt{W}Z, \tag{16.2}$$

where $Z \sim N(0, 1)$ and W is an independent, positive-valued scalar rv. We would like to know more about the tails of distributions satisfying (16.2).

More generally, to understand the tails of the marginal distributions of elliptical distributions it suffices to consider spherical distributions, which have the stochastic representation

$$\boldsymbol{X} \overset{\mathrm{d}}{=} R\boldsymbol{S} \tag{16.3}$$

for a random vector $\boldsymbol{S}$ that is uniformly distributed on the unit sphere $\mathcal{S}^{d-1} = \{\boldsymbol{s} \in \mathbb{R}^d : \boldsymbol{s}'\boldsymbol{s} = 1\}$, and for an independent radial variate R (see Section 6.3.1, and Theorem 6.21 in particular). Again we would like to know more about the tails of the marginal distributions of the vector $\boldsymbol{X}$ in (16.3).

In Section 4.2 we considered strictly stationary stochastic processes (X_t), such as GARCH processes satisfying equations of the form

$$X_t = \sigma_t Z_t, \tag{16.4}$$

where (Z_t) are strict white noise innovations, typically with a Gaussian distribution or (more realistically) a scaled Student t distribution, and where σ_t is an

$\mathcal{F}_{t-1}$-measurable rv representing volatility. These models can also be seen as mixture models and we would like to know something about the tail of the stationary distribution of (X_t).

A useful result for analysing the tails of mixtures is the following theorem due to Breiman (1965), which we immediately apply to spherical distributions.

Theorem 16.7 (tails of mixture distributions). *Let X be given by $X = YZ$ for independent, non-negative rvs Y and Z such that*

(1) *Y has a regularly varying tail with tail index α;*

(2) *$E(Z^{\alpha+\varepsilon}) < \infty$ for some $\varepsilon > 0$.*

Then X has a regularly varying tail with tail index α and

$$P(X > x) \sim E(Z^{\alpha})P(Y > x), \quad x \to \infty.$$

Proposition 16.8 (tails of spherical distributions). *Let $\boldsymbol{X} \stackrel{\mathrm{d}}{=} R\boldsymbol{S} \sim S_d(\psi)$ have a spherical distribution. If R has a regularly varying tail with tail index α, then so does $|X_i|$ for $i = 1, \ldots, d$. If $E(R^k) < \infty$ for all $k > 0$, then $|X_i|$ does not have a regularly varying tail.*

Proof. Suppose that R has a regularly varying tail with tail index α and consider RS_i. Since $|S_i|$ is a non-negative rv with finite support $[0, 1]$ and finite moments, it follows from Theorem 16.7 that $R|S_i|$, and hence $|X_i|$, are regularly varying with tail index α. If $E(R^k) < \infty$ for all $k > 0$, then $E|X_i|^k < \infty$ for all $k > 0$, so that $|X_i|$ cannot have a regularly varying tail. □

Example 16.9 (tails of normal variance mixtures). Suppose that $\boldsymbol{X} \stackrel{\mathrm{d}}{=} \sqrt{W}\boldsymbol{Z}$ with $\boldsymbol{Z} \sim N_d(\mathbf{0}, I_d)$ and W an independent scalar rv, so that both $\boldsymbol{Z}$ and $\boldsymbol{X}$ have spherical distributions and $\boldsymbol{X}$ has a normal variance mixture distribution. The vector $\boldsymbol{Z}$ has the spherical representation $\boldsymbol{Z} \stackrel{\mathrm{d}}{=} \tilde{R}\boldsymbol{S}$, where $\tilde{R}^2 \sim \chi^2_d$ (see Example 6.23). The vector $\boldsymbol{X}$ has the spherical representation $\boldsymbol{X} \stackrel{\mathrm{d}}{=} R\boldsymbol{S}$, where $R \stackrel{\mathrm{d}}{=} \sqrt{W}\tilde{R}$.

Now, the chi-squared distribution (being a gamma distribution) is in the domain of attraction of the Gumbel distribution, so $E(\tilde{R}^k) = E((\tilde{R}^2)^{k/2}) < \infty$ for all $k > 0$. We first consider the case when W has a regularly varying tail with tail index α so that $\bar{F}_W(w) = L(w)w^{-\alpha}$. It follows that $P(\sqrt{W} > x) = P(W > x^2) = L_2(x)x^{-2\alpha}$, where $L_2(x) := L(x^2)$ is also slowly varying, so that $\sqrt{W}$ has a regularly varying tail with tail index 2α. By Theorem 16.7, $R \stackrel{\mathrm{d}}{=} \sqrt{W}\tilde{R}$ also has a regularly varying tail with tail index 2α, and by Proposition 16.8, so do the components of $|\boldsymbol{X}|$.

To consider a particular case, suppose that $W \sim \mathrm{Ig}(\frac{1}{2}\nu, \frac{1}{2}\nu)$, so that, by Example 16.2, W is regularly varying with tail index $\frac{1}{2}\nu$. Then $\sqrt{W}$ has a regularly varying tail with tail index ν and so does $|X_i|$; this is hardly surprising because $\boldsymbol{X} \sim t_d(\nu, \mathbf{0}, I_d)$, implying that X_i has a univariate Student t distribution with ν degrees of freedom, and we already know from Example 16.1 that this has tail index ν.

On the other hand, if $F_W \in \mathrm{MDA}(H_0)$, then $E(R^k) < \infty$ for all $k > 0$ and $|X_i|$ cannot have a regularly varying tail by Proposition 16.8. This means, for example,

Table 16.1. Approximate theoretical values of the tail index κ solving (16.5) for various GARCH(1, 1) processes with Gaussian and Student t innovation distributions.

Parameters	Gauss	t distribution $\nu = 8$	t distribution $\nu = 4$
$\alpha_1 = 0.2,\ \beta = 0.75$	4.4	3.5	2.7
$\alpha_1 = 0.1,\ \beta = 0.85$	9.1	5.8	3.4
$\alpha_1 = 0.05,\ \beta = 0.95$	21.1	7.9	3.9

that univariate GH distributions do not have power tails (except for the special boundary case corresponding to Student t) because the GIG is in the maximum domain of attraction of the Gumbel distribution, as was shown in Example 16.6.

Analysis of the tails of the stationary distribution of GARCH-type models is more challenging. In view of Theorem 16.7 and the foregoing examples, it is clear that when the innovations (Z_t) are Gaussian, then the law of the process (X_t) in (16.4) will have a regularly varying tail if the volatility σ_t has a regularly varying tail. Mikosch and Stărică (2000) analyse the GARCH(1, 1) model (see Definition 4.20), where the squared volatility satisfies $\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta\sigma_{t-1}^2$. They show that under relatively weak conditions on the innovation distribution of (Z_t), the volatility σ_t has a regularly varying tail with tail index κ given by the solution of the equation

$$E((\alpha_1 Z_t^2 + \beta)^{\kappa/2}) = 1. \tag{16.5}$$

In Table 16.1 we have calculated approximate values of κ for various innovation distributions and parameter values using numerical integration and root-finding procedures. By Theorem 16.7 these are the values of the tail index for the stationary distribution of the GARCH(1, 1) model itself.

Two main findings are obvious: for any fixed set of parameter values, the tail index gets smaller and the tail of the GARCH model gets heavier as we move to heavier-tailed innovation distributions; for any fixed innovation distribution, the tail of the GARCH model gets lighter as we decrease the ARCH effect (α_1) and increase the GARCH effect (β).

Tail dependence in elliptical distributions. We close this section by giving a result that reveals an interesting connection between tail dependence in elliptical distributions and regular variation of the radial rv R in the representation $\boldsymbol{X} \stackrel{\mathrm{d}}{=} \boldsymbol{\mu} + R\boldsymbol{A}\boldsymbol{S}$ of an elliptically symmetric distribution given in Proposition 6.27.

Theorem 16.10. *Let $\boldsymbol{X} \stackrel{\mathrm{d}}{=} \boldsymbol{\mu} + R\boldsymbol{A}\boldsymbol{S} \sim E_d(\boldsymbol{\mu}, \Sigma, \psi)$, where $\boldsymbol{\mu}$, R, A and $\boldsymbol{S}$ are as in Proposition 6.27 and we assume that $\sigma_{ii} > 0$ for all $i = 1, \ldots, d$. If R has a regularly varying tail with tail index $\alpha > 0$, then the coefficient of upper and lower tail dependence between X_i and X_j is given by*

$$\lambda(X_i, X_j) = \frac{\int_{\pi/2 - \arcsin \rho_{ij}}^{\pi/2} \cos^\alpha(t)\, \mathrm{d}t}{\int_0^{\pi/2} \cos^\alpha(t)\, \mathrm{d}t}, \tag{16.6}$$

where ρ_{ij} is the (i, j)th element of $P = \wp(\Sigma)$ and $\wp$ is the correlation operator defined in (6.5).

An example where R has a regularly varying tail occurs in the case of the multivariate t distribution $\boldsymbol{X} \sim t_d(\nu, \boldsymbol{\mu}, \Sigma)$. It is obvious from the arguments used in Example 16.9 that the tail of the df of R is regularly varying with tail index $\alpha = \nu$. Thus (16.6) with α replaced by ν gives an alternative expression to (7.38) for calculating tail-dependence coefficients for the t copula $C^t_{\nu,P}$.

Arguably, the original expression (7.38) is easier to work with, since the df of a univariate t distribution is available in statistical software packages. Moreover, the equivalence of the two formulas allows us to conclude that we can use (7.38) to evaluate tail-dependence coefficients for any bivariate elliptical distribution with correlation parameter ρ when the distribution of the radial rv R has a regularly varying tail with tail index ν.

Notes and Comments

Section 16.1 is a highly selective account tailored to the study of a number of very specific models, and all of the theoretical subjects touched upon—regular variation, von Mises distributions, tails of products, tails of stochastic recurrence equations—can be studied in much greater detail.

For more about regular variation, slow variation and Karamata's Theorem see Bingham, Goldie and Teugels (1987) and Seneta (1976). A summary of the more important ideas with regard to the study of extremes is found in Resnick (2008). Section 16.1.2, with the exception of the examples, is taken from Embrechts, Klüppelberg and Mikosch (1997), and detailed references to results on von Mises distributions and the maximum domain of attraction of the Gumbel distribution are found therein.

Theorem 16.7 follows from results of Breiman (1965). Related results on distributions of products are found in Embrechts and Goldie (1980). The discussion of tails of GARCH models is based on Mikosch and Stărică (2000); the theory involves the study of stochastic recurrence relations and is essentially due to Kesten (1973). See also Mikosch (2003) for an excellent introduction to these ideas.

The formula for tail-dependence coefficients in elliptical distributions when the radial rv has a regularly varying tail is taken from Hult and Lindskog (2002). Similar results were derived independently by Schmidt (2002); see also Frahm, Junker and Szimayer (2003) for a discussion of the applicability of such results to financial returns.

16.2 Self-exciting Models for Extremes

The models described in this section build on the point process formulation of the POT model for threshold exceedances in Section 5.3. However, instead of assuming Poisson behaviour as we did before, we now attempt to explain the clustering of extreme events in financial return data in terms of self-exciting behaviour.

16.2.1 Self-exciting Processes

We first consider modelling the occurrence of threshold exceedances in time using a simple form of self-exciting point process known as a Hawkes process. The basic idea is that instead of modelling the instantaneous risk of a threshold exceedance as being constant over time, we assume that threshold exceedances in the recent past cause the instantaneous risk of a threshold exceedance to be higher. The main area of application of these models has traditionally been in the modelling of earthquakes and their aftershocks, although there is a growing number of applications in financial modelling (see Notes and Comments).

Given data $X_1, \ldots, X_n$ and a threshold u, we will assume as usual that there are N_u exceedances, comprising the data $\{(i, X_i) : 1 \leqslant i \leqslant n,\ X_i > u\}$. Note that from now on we will express the time of an exceedance on the natural timescale of the time series, so if, for example, the data are daily observations, then our times are expressed in days. It will also be useful to have the alternative notation $\{(T_j, \tilde{X}_j) : j = 1, \ldots, N_u\}$, which enumerates exceedances consecutively.

First we consider a model for exceedance times only. In point process notation we let $Y_i = i I_{\{X_i > u\}}$, so Y_i returns an exceedance time, in the event that one takes place at time i, and returns 0 otherwise. The point process of exceedances is the process $N(\cdot)$ with state space $\mathcal{X} = (0, n]$ given by $N(A) = \sum_{i=1}^{n} I_{\{Y_i \in A\}}$ for $A \subset \mathcal{X}$.

We assume that the point process $N(\cdot)$ is a self-exciting process with *conditional intensity*

$$\lambda^*(t) = \tau + \psi \sum_{j\,:\,0<T_j<t} h(t - T_j, \tilde{X}_j - u), \tag{16.7}$$

where $\tau > 0$, $\psi \geqslant 0$ and h is some positive-valued function. Each previous exceedance $(T_j, \tilde{X}_j)$ contributes to the conditional intensity and the amount that it contributes can depend on both the elapsed time $(t - T_j)$ since that exceedance and the amount of the excess loss $(\tilde{X}_j - u)$ over the threshold. Informally, we understand the conditional intensity as expressing the instantaneous chance of a new exceedance of the threshold at time t, like the rate or intensity of an ordinary Poisson process. However, in the self-exciting model, the conditional intensity is itself a stochastic process that depends on ω, the state of nature, through the history of threshold exceedances up to (but not including) time t.

Possible parametric specifications of the h function are

- $h(s, x) = \mathrm{e}^{\delta x - \gamma s}$, where $\delta, \gamma > 0$; or
- $h(s, x) = \mathrm{e}^{\delta x}(s + \gamma)^{-(\rho+1)}$, where $\delta, \gamma, \rho > 0$.

Collecting all parameters in $\boldsymbol{\theta}$, the likelihood takes the form

$$L(\boldsymbol{\theta}; \text{data}) = \exp\left(-\int_0^n \lambda^*(s)\,\mathrm{d}s\right) \prod_{i=1}^{N_u} \lambda^*(T_i),$$

and may be maximized numerically to obtain parameter estimates.

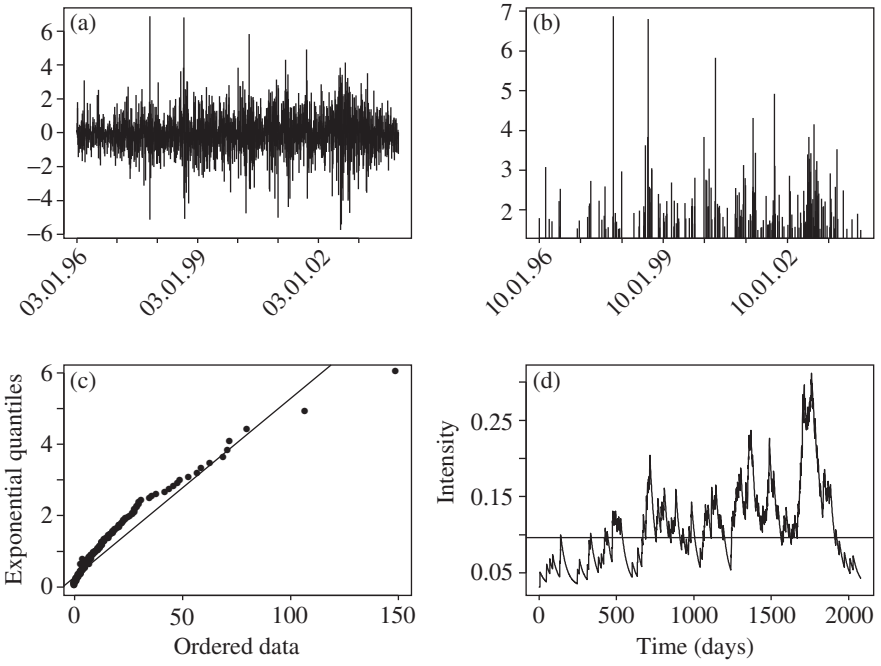

Figure 16.1. (a) S&P daily percentage loss data. (b) Two hundred largest losses. (c) A Q–Q plot of inter-exceedance times against an exponential reference distribution. (d) The estimated intensity of exceeding the threshold in a self-exciting model. See Example 16.11 for details.

Example 16.11 (S&P daily percentage losses 1996–2003). We apply the self-exciting process methodology to all daily percentage losses incurred by the Standard & Poor's index in the eight-year period 1996–2003 (2078 values). In Figure 16.1 the loss data are shown as well as the point process of the 200 largest daily losses exceeding a threshold of 1.5%. Clearly, there is clustering in the pattern of exceedance data, and the Q–Q plot shows that the inter-exceedance times are not exponential.

We fit the simpler self-exciting model with $h(s, x) = \mathrm{e}^{\delta x - \gamma s}$. The parameter estimates (and standard errors) are $\hat{\tau} = 0.032(0.011)$, $\hat{\psi} = 0.016(0.0069)$, $\hat{\gamma} = 0.026(0.011)$, $\hat{\delta} = 0.13(0.27)$, suggesting that all parameters except δ are significant. The log-likelihood for the fitted model is -648.2, whereas the log-likelihood for a homogeneous Poisson model is -668.2; the Poisson special case can therefore clearly be rejected in a likelihood ratio test with a p-value less than 0.001. Figure 16.1 (d) shows the estimated intensity $\lambda^*(t)$ of crossing the threshold throughout the data observation period, which seems to reflect the pattern of exceedances observed.

Note that a simple refinement of this model (and those of the following section) would be to consider a self-exciting structure where both extreme negative and extreme positive returns contributed to the conditional intensity; this would involve setting upper and lower thresholds and considering exceedances of both.

16.2.2 A Self-exciting POT Model

We now consider how the POT model of Section 5.3.2 might be generalized to incorporate a self-exciting component. We first develop a marked self-exciting model where marks have a generalized Pareto distribution but are *unpredictable*, meaning that the excess losses are iid GPD. In the second model we consider the case of *predictable* marks. In this model the excess losses are conditionally generalized Pareto, given the exceedance history up to the time of the mark, with a scaling parameter that depends on that history. In this way we get a model where, in a period of excitement, both the temporal intensity of occurrence and the magnitude of the exceedances increase.

In point process language, our models are processes $N(\cdot)$ on a state space of the form $\mathcal{X} = (0, n] \times (u, \infty)$ such that $N(A) = \sum_{i=1}^{n} I_{\{(i,X_i)\in A\}}$ for sets $A \subset \mathcal{X}$. To build these models we start with the intensity of the reparametrized version of the standard POT model given in (5.31). We recall that this model simply says that exceedances of the threshold u occur as a homogeneous Poisson process with rate τ and that excesses have a generalized Pareto distribution with df $G_{\xi,\beta}$.

Model with unpredictable marks. We first introduce the notation

$$v^*(t) = \sum_{j:\, 0<T_j<t} h(t - T_j, \tilde{X}_j - u)$$

for the *self-excitement function*, where the function h is as in Section 16.2.1. We generalize (5.31) and consider a self-exciting model with conditional intensity

$$\lambda^*(t, x) = \frac{\tau + \psi v^*(t)}{\beta}\left(1 + \xi\frac{x-u}{\beta}\right)^{-1/\xi-1} \tag{16.8}$$

on a state space $\mathcal{X} = (0, n] \times (u, \infty)$, where $\tau > 0$ and $\psi \geqslant 0$. Effectively, we have combined the one-dimensional intensity in (16.7) with a GPD density. When $\psi = 0$ we have an ordinary POT model with no self-exciting structure.

It is easy to calculate that the conditional rate of crossing the threshold $x \geqslant u$ at time t, given information up to that time, is

$$\tau^*(t, x) = \int_x^\infty \lambda^*(t, y)\,\mathrm{d}y = (\tau + \psi v^*(t))\left(1 + \xi\frac{x-u}{\beta}\right)^{-1/\xi}, \tag{16.9}$$

which, for fixed x, is simply a one-dimensional self-exciting process of the form (16.7). The implied distribution of the excess losses when an exceedance takes place is generalized Pareto, because

$$\frac{\tau^*(t, u + x)}{\tau^*(t, u)} = \left(1 + \frac{\xi x}{\beta}\right)^{-1/\xi} = \bar{G}_{\xi,\beta}(x), \tag{16.10}$$

independently of t. Statistical fitting of this model is performed by maximizing a likelihood of the form

$$L(\boldsymbol{\theta}; \text{data}) = \exp\left(-n\tau - \psi\int_0^n v^*(s)\,\mathrm{d}s\right)\prod_{j=1}^{N_u}\lambda^*(T_j, \tilde{X}_j). \tag{16.11}$$

A model with predictable marks. A model with predictable marks can be obtained by generalizing (16.8) to get

$$\lambda^*(t,x) = \frac{\tau + \psi v^*(t)}{\beta + \alpha v^*(t)} \left(1 + \xi \frac{x-u}{\beta + \alpha v^*(t)}\right)^{-1/\xi - 1}, \tag{16.12}$$

where $\beta > 0$ and $\alpha \geqslant 0$. For simplicity we have assumed that the GPD scaling is also linear in the self-excitement function $v^*(t)$. The properties of this model follow immediately from the model with unpredictable marks. The conditional crossing rate of the threshold $x \geqslant u$ at time t is as in (16.9) with the parameter β replaced by the time-dependent self-exciting function $\beta + \alpha v^*(t)$. By repeating the calculation in (16.10) we find that the distribution of the excess loss over the threshold, given that an exceedance takes place at time t and given the history of exceedances up to time t, is generalized Pareto with df $G_{\xi,\beta+\alpha v^*(t)}$. The likelihood for fitting the model is again (16.11), where the function $\lambda^*(t,x)$ is now given by (16.12). Note that by comparing a model with $\alpha = 0$ and a model with $\alpha > 0$ we can formally test the hypothesis that the marks are unpredictable using a likelihood ratio test.

Example 16.12 (self-exciting POT model for S&P daily loss data). We continue the analysis of the data of Example 16.11 by fitting self-exciting POT models with both unpredictable and predictable marks to the 200 exceedances of the threshold $u = 1.5\%$. The former is equivalent to fitting a self-exciting model to the exceedance times as in Example 16.11 and then fitting a GPD to the excess losses over the threshold; the estimated intensity of crossing the threshold is therefore identical to the one shown in Figure 16.1. The log-likelihood for this model is -783.4, whereas a model with predictable marks gives a value of -779.3 for one extra parameter α; in a likelihood ratio test the p-value is 0.004, showing a significant improvement.

In Figure 16.2 we show the exceedance data as well as the estimated intensity $\tau^*(t,u)$ of exceeding the threshold in the model with predictable marks. We also show the estimated mean of the GPD for the conditional distribution of the excess loss above the threshold, given that an exceedance takes place at time t. The GPD mean $(\beta + \alpha v^*(t))/(1-\xi)$ and the intensity $\tau^*(t,u)$ are both affine functions of the self-excitement function $v^*(t)$ and obviously follow its path.

Calculating conditional risk measures. Finally, we note that self-exciting POT models can be used to estimate a conditional VaR and also a conditional expected shortfall. If we have analysed n daily data ending on day t and want to calculate, say, a 99% VaR, then we treat the problem as a (conditional) return-level problem; we look for the level at which the conditional exceedance intensity at a time point just after t (denoted by $t+$) is 0.01. In general, to calculate a conditional estimate of VaR_α^t (for α sufficiently large) we would attempt to solve the equation $\tau^*(t+,x) = (1-\alpha)$ for some value of x satisfying $x \geqslant u$. In the model with predictable marks this is

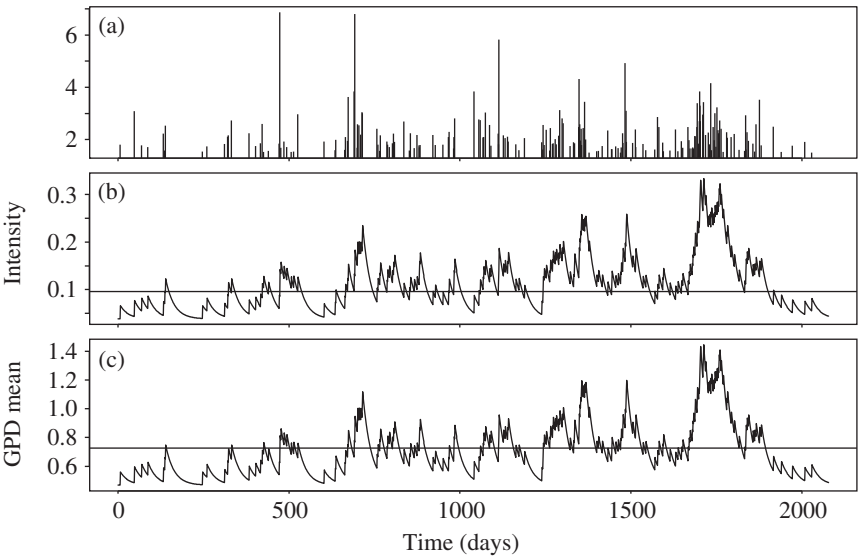

Figure 16.2. (a) Exceedance pattern for 200 largest daily losses in S&P data. (b) Estimated intensity of exceeding the threshold in a self-exciting POT model with predictable marks. (c) Mean of the conditional generalized Pareto distribution of the excess loss above the threshold. See Example 16.12 for details.

possible if $\tau + \psi v^*(t+) > 1 - \alpha$ and gives the formula

$$\mathrm{VaR}_\alpha^t = u + \frac{\beta + \alpha v^*(t+)}{\xi}\left(\left(\frac{1-\alpha}{\tau + \psi v^*(t+)}\right)^{-\xi} - 1\right).$$

The associated conditional expected shortfall could then be calculated by observing that the conditional distribution of excess losses above VaR_α^t given information up to time t is GPD with shape parameter ξ and scaling parameter given by $\beta + \alpha v^*(t+) + \xi(\mathrm{VaR}_\alpha^t - u)$.

Notes and Comments

The original reference to the Hawkes self-exciting process is Hawkes (1971). There is a large literature on the application of such processes to earthquake modelling; a starter reference is Ogata (1988). To our knowledge, the earliest contribution on Hawkes processes in financial econometrics was Bowsher (2002); this paper finally appeared as Bowsher (2007). The application to extremes in financial time series was suggested in Chavez-Demoulin, Davison and McNeil (2005). Embrechts, Liniger and Lin (2011) give an application of multivariate Hawkes processes to multiple financial time series. Self-exciting processes have also been applied in credit risk (Errais, Giesecke and Goldberg 2010) and in the modelling of high-frequency financial data (Chavez-Demoulin and McGill 2012).

The idea detailed in Section 16.2.2 (of a POT model with self-exciting structure) was first proposed in the first edition of this textbook. Grothe, Korniichuk and Manner (2014) extend the idea and consider multivariate models for extremes where component processes may excite themselves or other component processes.

16.3 Multivariate Maxima

In this section we give a brief overview of the theory of multivariate maxima, stating the main results in terms of copulas. A class of copulas known as extreme value copulas emerges as the class of natural limiting dependence structures for multivariate maxima. These provide useful dependence structures for modelling the joint tail behaviour of risk factors that appear to show tail dependence. A useful reference is Galambos (1987), which is one of the few texts to treat the theory of multivariate maxima as a copula theory (although Galambos does not use the word, referring to copulas simply as dependence functions).

16.3.1 Multivariate Extreme Value Copulas

Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be iid random vectors in $\mathbb{R}^d$ with joint df F and marginal dfs $F_1, \ldots, F_d$. We label the components of these vectors $\boldsymbol{X}_i = (X_{i,1}, \ldots, X_{i,d})'$ and interpret them as losses of d different types. We define the maximum of the jth component to be $M_{n,j} = \max(X_{1,j}, \ldots, X_{n,j})$, $j = 1, \ldots, d$. In classical multivariate EVT, the object of interest is the vector of *componentwise block maxima*: $\boldsymbol{M}_n = (M_{n,1}, \ldots, M_{n,d})'$. In particular, we are interested in the possible multivariate limiting distributions for $\boldsymbol{M}_n$ under appropriate normalizations, much as in the univariate case. It should, however, be observed that the vector $\boldsymbol{M}_n$ will in general not correspond to any of the vector observations $\boldsymbol{X}_i$.

We seek limit laws for

$$\frac{\boldsymbol{M}_n - \boldsymbol{d}_n}{\boldsymbol{c}_n} = \left(\frac{M_{n,1} - d_{n,1}}{c_{n,1}}, \ldots, \frac{M_{n,d} - d_{n,d}}{c_{n,d}} \right)'$$

as $n \to \infty$, where $\boldsymbol{c}_n = (c_{n,1}, \ldots, c_{n,d})'$ and $\boldsymbol{d}_n = (d_{n,1}, \ldots, d_{n,d})'$ are vectors of normalizing constants, the former satisfying $\boldsymbol{c}_n > \boldsymbol{0}$. Note that in this and other statements in this section, arithmetic operations on vectors of equal length are understood as componentwise operations. Supposing that $(\boldsymbol{M}_n - \boldsymbol{d}_n)/\boldsymbol{c}_n$ converges in distribution to a random vector with joint df H, we have

$$\lim_{n\to\infty} P\left(\frac{\boldsymbol{M}_n - \boldsymbol{d}_n}{\boldsymbol{c}_n} \leqslant \boldsymbol{x} \right) = \lim_{n\to\infty} F^n(\boldsymbol{c}_n \boldsymbol{x} + \boldsymbol{d}_n) = H(\boldsymbol{x}). \tag{16.13}$$

Definition 16.13 (multivariate extreme value distribution and domain of attraction). If (16.13) holds for some F and some H, we say that F is in the maximum domain of attraction of H, written $F \in \mathrm{MDA}(H)$, and we refer to H as a multivariate extreme value (MEV) distribution.

The convergence issue for multivariate maxima is already partly solved by the univariate theory. If H has non-degenerate margins, then these must be univariate extreme value distributions of Fréchet, Gumbel or Weibull type. Since these are continuous, Sklar's Theorem tells us that H must have a unique copula. The following theorem asserts that this copula C must have a particular kind of scaling behaviour.

Theorem 16.14 (extreme value copula). *If (16.13) holds for some F and some H with GEV margins, then the unique copula C of H satisfies*

$$C(\boldsymbol{u}^t) = C^t(\boldsymbol{u}), \quad \forall t > 0. \tag{16.14}$$

Any copula with the property (16.14) is known as an *extreme value (EV) copula* and can be the copula of an MEV distribution. The independence and comonotonicity copulas are EV copulas and the Gumbel copula provides an example of a parametric EV copula family. The bivariate version in (7.12) obviously has property (16.14), as does the exchangeable higher-dimensional Gumbel copula based on (7.46) as well as the non-exchangeable versions based on (15.10)–(15.12).

There are a number of mathematical results characterizing MEV distributions and EV copulas. One such result is the following.

Theorem 16.15 (Pickands representation). *The copula C is an EV copula if and only if it has the representation*

$$C(\boldsymbol{u}) = \exp\left\{B\left(\frac{\ln u_1}{\sum_{k=1}^d \ln u_k}, \ldots, \frac{\ln u_d}{\sum_{k=1}^d \ln u_k}\right)\sum_{i=1}^d \ln u_i\right\}, \tag{16.15}$$

where $B(\boldsymbol{w}) = \int_{\mathcal{S}_d} \max(x_1 w_1, \ldots, x_d w_d)\,\mathrm{d}S(\boldsymbol{x})$ and S is a finite measure on the d-dimensional simplex, i.e. the set $\mathcal{S}_d = \{\boldsymbol{x}\colon x_i \geqslant 0,\ i = 1, \ldots, d,\ \sum_{i=1}^d x_i = 1\}$.

The function $B(\boldsymbol{w})$ is sometimes referred to as the dependence function of the EV copula. In the general case, such functions are difficult to visualize and work with, but in the bivariate case they have a simple form that we discuss in more detail.

In the bivariate case we redefine $B(\boldsymbol{w})$ as a function of a scalar argument by setting $A(w) := B((w, 1-w)')$ with $w \in [0, 1]$. It follows from Theorem 16.15 that a bivariate copula is an EV copula if and only if it takes the form

$$C(u_1, u_2) = \exp\left\{(\ln u_1 + \ln u_2)A\left(\frac{\ln u_1}{\ln u_1 + \ln u_2}\right)\right\}, \tag{16.16}$$

where $A(w) = \int_0^1 \max((1-x)w, x(1-w))\,\mathrm{d}S(x)$ for a measure S on $[0, 1]$. It can be inferred that such bivariate dependence functions must satisfy

$$\max(w, 1-w) \leqslant A(w) \leqslant 1, \quad 0 \leqslant w \leqslant 1, \tag{16.17}$$

and that they must, moreover, be convex. Conversely, a differentiable convex function $A(w)$ satisfying (16.17) can be used to construct an EV copula using (16.16).

The upper and lower bounds in (16.17) have intuitive interpretations. If $A(w) = 1$ for all w, then the copula (16.16) is clearly the independence copula, and if $A(w) = \max(w, 1-w)$, then it is the comonotonicity copula. It is also useful to note, and easy to show, that we can extract the dependence function from the EV copula in (16.16) by setting

$$A(w) = -\ln C(\mathrm{e}^{-w}, \mathrm{e}^{-(1-w)}), \quad w \in [0, 1]. \tag{16.18}$$

Example 16.16 (Gumbel copula). We consider the asymmetric version of the bivariate Gumbel copula defined by (7.12) and construction (15.10), i.e. the copula

$$C^{\mathrm{Gu}}_{\theta,\alpha,\beta}(u_1, u_2) = u_1^{1-\alpha} u_2^{1-\beta} \exp\{-((-\alpha \ln u_1)^\theta + (-\beta \ln u_2)^\theta)^{1/\theta}\}.$$

As already remarked, this copula has the scaling property (16.14) and is an EV copula. Using (16.18) we calculate that the dependence function is given by

$$A(w) = (1-\alpha)w + (1-\beta)(1-w) + ((\alpha w)^\theta + (\beta(1-w))^\theta)^{1/\theta}.$$

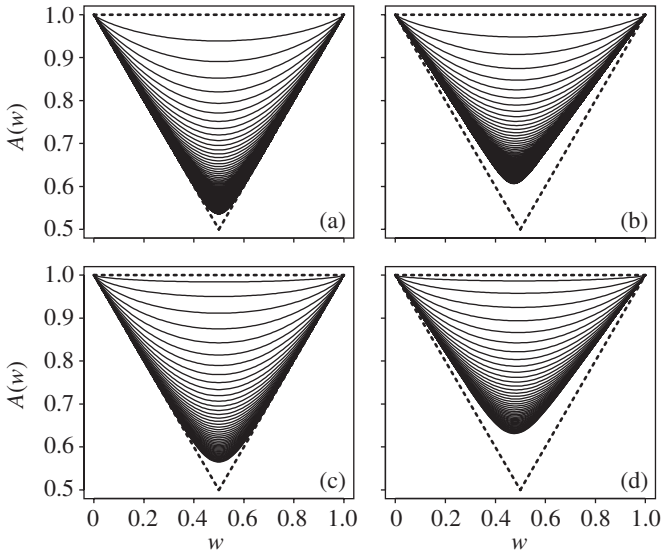

Figure 16.3. Plot of dependence functions for the (a) symmetric Gumbel, (b) asymmetric Gumbel, (c) symmetric Galambos and (d) asymmetric Galambos copulas (asymmetric cases have $\alpha = 0.9$ and $\beta = 0.8$) as described in Examples 16.16 and 16.17. Dashed lines show boundaries of the triangle in which the dependence function must reside; solid lines show dependence functions for a range of parameter values.

We have plotted this function in Figure 16.3 for a range of θ values running from 1.1 to 10 in steps of size 0.1. Part (a) shows the standard symmetric Gumbel copula with $\alpha = \beta = 1$; the dependence function essentially spans the whole range from independence, represented by the upper edge of the dashed triangle, to comonotonicity, represented by the two lower edges of the dashed triangle, which comprise the function $A(w) = \max(w, 1 - w)$. Part (b) shows an asymmetric example with $\alpha = 0.9$ and $\beta = 0.8$; in this case we still have independence when $\theta = 1$, but the limit as $\theta \to \infty$ is no longer the comonotonicity model. The Gumbel copula model is also sometimes known as the logistic model.

Example 16.17 (Galambos copula). This time we begin with the dependence function given by

$$A(w) = 1 - ((\alpha w)^{-\theta} + (\beta(1 - w))^{-\theta})^{-1/\theta}, \tag{16.19}$$

where $0 \leqslant \alpha, \beta \leqslant 1$ and $0 < \theta < \infty$. It can be verified that this is a convex function satisfying $\max(w, 1 - w) \leqslant A(w) \leqslant 1$ for $0 \leqslant w \leqslant 1$, so it can be used to create an EV copula in (16.16). We obtain the copula

$$C^{\mathrm{Gal}}_{\theta,\alpha,\beta}(u_1, u_2) = u_1 u_2 \exp\{((-\alpha \ln u_1)^{-\theta} + (-\beta \ln u_2)^{-\theta})^{-1/\theta}\},$$

which has also been called the negative logistic model. We have plotted this function in Figure 16.3 for a range of θ values running from 0.2 to 5 in steps of size 0.1. Part (c) shows the standard symmetric case with $\alpha = \beta = 1$ spanning the whole range from independence to comonotonicity. Part (d) shows an asymmetric example

with $\alpha = 0.9$ and $\beta = 0.8$; in this case we still approach independence as $\theta \to 0$, but the limit as $\theta \to \infty$ is no longer the comonotonicity model.

A number of other bivariate EV copulas have been described in the literature (see Notes and Comments).

16.3.2 Copulas for Multivariate Minima

The structure of limiting copulas for multivariate minima can be easily inferred from the structure of limiting copulas for multivariate maxima; moving from maxima to minima essentially involves the same considerations that we made at the end of Section 5.1.1 and uses identity (5.2) in particular.

Normalized componentwise minima of iid random vectors $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ with df F will converge in distribution to a non-degenerate limit if the df $\tilde{F}$ of the random vectors $-\boldsymbol{X}_1, \ldots, -\boldsymbol{X}_n$ is in the maximum domain of attraction of an MEV distribution (see Definition 16.13), written $\tilde{F} \in \mathrm{MDA}(H)$. Of course, for a radially symmetric distribution, $\tilde{F}$ coincides with F.

Let $\boldsymbol{M}_n^*$ be the vector of componentwise maxima of $-\boldsymbol{X}_1, \ldots, -\boldsymbol{X}_n$ such that $M_{n,j}^* = \max(-X_{1,j}, \ldots, -X_{n,j})$. If $\tilde{F} \in \mathrm{MDA}(H)$ for some non-degenerate H, we have

$$\lim_{n\to\infty} P\left(\frac{\boldsymbol{M}_n^* - \boldsymbol{d}_n}{\boldsymbol{c}_n} \leqslant \boldsymbol{x}\right) = \lim_{n\to\infty} \tilde{F}^n(\boldsymbol{c}_n\boldsymbol{x} + \boldsymbol{d}_n) = H(\boldsymbol{x}) \tag{16.20}$$

for appropriate sequences of normalizing vectors $\boldsymbol{c}_n$ and $\boldsymbol{d}_n$, and an MEV distribution H of the form $H(\boldsymbol{x}) = C(H_{\xi_1}(x_1), \ldots, H_{\xi_d}(x_d))$, where H_{ξ_j} denotes a GEV distribution with shape parameter ξ_j and C is an EV copula satisfying (16.14).

Defining the vector of componentwise minima by $\boldsymbol{m}_n$ and using (5.2), it follows from (16.20) that

$$\lim_{n\to\infty} P\left(\frac{\boldsymbol{m}_n + \boldsymbol{d}_n}{\boldsymbol{c}_n} \geqslant \boldsymbol{x}\right) = H(-\boldsymbol{x}),$$

so normalized minima converge in distribution to a limit with survival function $H(-\boldsymbol{x}) = C(H_{\xi_1}(-x_1), \ldots, H_{\xi_d}(-x_d))$. It follows that the copula of the limiting distribution of the minima is the survival copula of C (see Section 7.1.5 for discussion of survival copulas). In general, the limiting copulas for minima are *survival copulas of EV copulas* and concrete examples of such copulas are the Gumbel and Galambos survival copulas.

In the special case of a radially symmetric underlying distribution, the limiting copula of the minima is precisely the survival copula of the limiting EV copula of the maxima.

16.3.3 Copula Domains of Attraction

As in the case of univariate maxima we would like to know which underlying multivariate dfs F are attracted to which MEV distributions H. We now give a useful result in terms of copulas that is essentially due to Galambos (see Notes and Comments).

Theorem 16.18. *Let $F(\boldsymbol{x}) = C(F_1(x_1), \ldots, F_d(x_d))$ for continuous marginal dfs $F_1, \ldots, F_d$ and some copula C. Let $H(\boldsymbol{x}) = C_0(H_1(x_1), \ldots, H_d(x_d))$ be an MEV distribution with EV copula C_0. Then $F \in \mathrm{MDA}(H)$ if and only if $F_i \in \mathrm{MDA}(H_i)$ for $1 \leqslant i \leqslant d$ and*

$$\lim_{t\to\infty} C^t(u_1^{1/t}, \ldots, u_d^{1/t}) = C_0(u_1, \ldots, u_d), \quad \boldsymbol{u} \in [0,1]^d. \tag{16.21}$$

This result shows that the copula C_0 of the limiting MEV distribution is determined solely by the copula C of the underlying distribution according to (16.21); the marginal distributions of F determine the margins of the MEV limit but are irrelevant to the determination of its dependence structure. This motivates us to introduce the concept of a copula domain of attraction.

Definition 16.19. If (16.21) holds for some C and some EV copula C_0, we say that C is in the copula domain of attraction of C_0, written $C \in \mathrm{CDA}(C_0)$.

There are a number of equivalent ways of writing (16.21). First, by taking logarithms and using the asymptotic identity $\ln(x) \sim x - 1$ as $x \to 1$, we get, for $\boldsymbol{u} \in (0,1]^d$,

$$\left.\begin{aligned} \lim_{t\to\infty} t(1 - C(u_1^{1/t}, \ldots, u_d^{1/t})) &= -\ln C_0(u_1, \ldots, u_d), \\ \lim_{s\to 0^+} \frac{1 - C(u_1^s, \ldots, u_d^s)}{s} &= -\ln C_0(u_1, \ldots, u_d). \end{aligned}\right\} \tag{16.22}$$

By inserting $u_i = \mathrm{e}^{-x_i}$ in the latter identity and using $\mathrm{e}^{-sx} \sim 1 - sx$ as $s \to 0$, we get, for $\boldsymbol{x} \in [0,\infty)^d$,

$$\lim_{s\to 0^+} \frac{1 - C(1 - sx_1, \ldots, 1 - sx_d)}{s} = -\ln C_0(\mathrm{e}^{-x_1}, \ldots, \mathrm{e}^{-x_d}). \tag{16.23}$$

Example 16.20 (limiting copula for bivariate Pareto distribution). In Example 7.14 we saw that the bivariate Pareto distribution has univariate Pareto margins $F_i(x) = 1 - (\kappa_i/(\kappa_i + x))^\alpha$ and a Clayton survival copula. It follows from Example 5.6 that $F_i \in \mathrm{MDA}(H_{1/\alpha})$, $i = 1, 2$. Using (7.16), the Clayton survival copula is calculated to be $C(u_1, u_2) = u_1 + u_2 - 1 + ((1-u_1)^{-1/\alpha} + (1-u_2)^{-1/\alpha} - 1)^{-\alpha}$. Using (16.23), it is easily calculated that $C_0(u_1, u_2) = u_1 u_2 \exp(((-\ln u_1)^{-1/\alpha} + (-\ln u_2)^{-1/\alpha})^{-\alpha})$, which is the standard exchangeable Galambos copula of Example 16.17. The limiting distribution of maxima therefore consists of two Fréchet dfs connected by the Galambos copula.

The coefficients of upper tail dependence play an interesting role in the copula domain of attraction theory. In particular, they can help us to recognize copulas that lie in the copula domain of attraction of the independence copula.

Proposition 16.21. *Let C be a bivariate copula with upper tail-dependence coefficient λ_u, and assume that C satisfies $C \in \mathrm{MDA}(C_0)$ for some EV copula C_0. Then λ_u is also the upper tail-dependence coefficient of C_0 and is related to its dependence function by $\lambda_\mathrm{u} = 2(1 - A(\frac{1}{2}))$.*

Proof. We use (7.35) and (7.16) to see that

$$\lambda_u = \lim_{q\to 1^-} \frac{\hat{C}(1-q,1-q)}{1-q} = 2 - \lim_{q\to 1^-} \frac{1-C(q,q)}{1-q}.$$

By using the asymptotic identity $\ln x \sim x-1$ as $x \to 1$ and the CDA condition (16.22) we can calculate

$$\begin{aligned}
\lim_{q\to 1^-} \frac{1-C_0(q,q)}{1-q} &= \lim_{q\to 1^-} \frac{\ln C_0(q,q)}{\ln q} \\
&= \lim_{q\to 1^-} \lim_{s\to 0^+} \frac{1-C(q^s,q^s)}{-s\ln q} \\
&= \lim_{q\to 1^-} \lim_{s\to 0^+} \frac{1-C(q^s,q^s)}{-\ln(q^s)} \\
&= \lim_{v\to 1^-} \frac{1-C(v,v)}{1-v},
\end{aligned}$$

which shows that C and C_0 share the same coefficient of upper tail dependence. Using the formula $\lambda_u = 2 - \lim_{q\to 1^-} \ln C_0(q,q)/\ln q$ and the representation (16.16) we easily obtain that $\lambda_u = 2(1-A(\frac{1}{2}))$. □

In the case when $\lambda_u = 0$ we must have $A(\frac{1}{2}) = 1$, and the convexity of dependence functions dictates that $A(w)$ is identically 1, so C_0 must be the independence copula. In the higher-dimensional case this is also true: if C is a d-dimensional copula with all upper tail-dependence coefficients equal to 0, then the bivariate margins of the limiting copula C_0 must all be independence copulas, and, in fact, it can be shown that C_0 must therefore be the d-dimensional independence copula (see Notes and Comments).

As an example consider the limiting distribution of multivariate maxima of Gaussian random vectors. Since the pairwise coefficients of tail dependence of Gaussian vectors are 0 (see Example 7.38), the limiting distribution is a product of marginal Gumbel distributions. The convergence is extremely slow, but ultimately normalized componentwise maxima are independent in the limit.

Now consider the multivariate t distribution, which has been an important model throughout this book. If $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ are iid random vectors with a $t_d(\nu, \boldsymbol{\mu}, \Sigma)$ distribution, we know from Example 16.1 that univariate maxima of the individual components are attracted to univariate Fréchet distributions with parameter $1/\nu$. Moreover, we know from Example 7.39 that tail dependence coefficients for the t copula are strictly positive; the limiting EV copula cannot be the independence copula.

In fact, the limiting EV copula for t-distributed random vectors can be calculated using (16.23), although the calculations are tedious. In the bivariate case it is found that the limiting copula, which we call the t–EV copula, has dependence function

$$A(w) = wt_{\nu+1}\left(\frac{(w/(1-w))^{1/\nu}-\rho}{\sqrt{(1-\rho^2)/(\nu+1)}}\right) + (1-w)t_{\nu+1}\left(\frac{((1-w)/w)^{1/\nu}-\rho}{\sqrt{(1-\rho^2)/(\nu+1)}}\right), \tag{16.24}$$

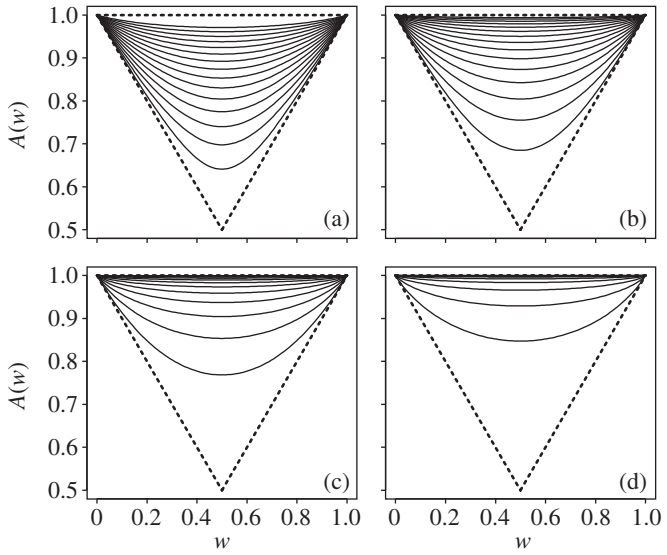

Figure 16.4. Plots of the dependence function for the t–EV copula for (a) $\nu = 2$, (b) $\nu = 4$, (c) $\nu = 10$ and (d) $\nu = 20$, and with various values of ρ.

where ρ is the off-diagonal component of $P = \wp(\Sigma)$. This dependence function is shown in Figure 16.4 for four different values of ν and for ρ values ranging from -0.5 to 0.9 with increments of 0.1. As $\rho \to 1$ the t–EV copula converges to the comonotonicity copula; as $\rho \to -1$ or as $\nu \to \infty$ it converges to the independence copula.

16.3.4 Modelling Multivariate Block Maxima

A multivariate block maxima method analogous to the univariate method of Section 5.1.4 could be developed, although similar criticisms apply, namely that the block maxima method is not the most efficient way of making use of extreme data. Also, the kind of inference that this method allows may not be exactly what is desired in the multivariate case, as will be seen.

Suppose we divide our underlying data into blocks as before and we denote the realizations of the block maxima vectors by $\boldsymbol{M}_{n,1}, \ldots, \boldsymbol{M}_{n,m}$, where m is the total number of blocks. The distributional model suggested by the univariate and multivariate maxima theory consists of GEV margins connected by an extreme value copula.

In the multivariate theory there is, in a sense, a "correct" EV copula to use, which is the copula C_0 to which the copula C of the underlying distribution of the raw data is attracted. However, the underlying copula C is unknown and so the approach is generally to work with any tractable EV copula that appears appropriate for the task in hand. In a bivariate application, if we restrict to exchangeable copulas, then we have at our disposal the Gumbel, Galambos and t–EV copulas, and a number of other possibilities for which references in Notes and Comments should be consulted. As will be apparent from Figures 16.3 and 16.4, the essential functional form of all

these families is really very similar; it is mostly sufficient to work with either the Gumbel copula or the Galambos copula as these have simple forms that permit a relatively easy calculation of the copula density (which is needed for likelihood inference). Even if the "true" underlying copula were t, it would seldom make sense to use the more complicated t–EV copula, since the dependence function in (16.24) can be accurately approximated by the dependence function of a Gumbel copula for many values of ν and ρ.

The Gumbel copula also allows us to explore the possibility of asymmetry by using the general non-exchangeable family described in Example 16.16. For applications in dimensions higher than two, the higher-dimensional extensions of Gumbel discussed in Sections 7.4.2 and 15.2.2 may be useful, although we should stress again that multivariate extreme value models are best suited to low-dimensional applications.

Putting these considerations together, data on multivariate maxima could be modelled using the df $H_{\boldsymbol{\xi},\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\theta}}(\boldsymbol{x}) = C_{\boldsymbol{\theta}}(H_{\xi_1,\mu_1,\sigma_1}(x_1), \ldots, H_{\xi_d,\mu_d,\sigma_d}(x_d))$ for some tractable parametric EV copula $C_{\boldsymbol{\theta}}$. The usual method involves maximum likelihood inference and the maximization can either be performed in one step for all parameters of the margins and copula or broken into two steps, whereby marginal models are estimated first and then a parametric copula is fitted using the ideas in Sections 7.5.2 and 7.5.3. The following bivariate example gives an idea of the kind of inference that can be made with such a model.

Example 16.22. Let $M_{65,1}$ represent the quarterly maximum of daily percentage falls of the US dollar against the euro and let $M_{65,2}$ represent the quarterly maximum of daily percentage falls of the US dollar against the yen. We define a stress event for each of these daily return series: for the dollar against the euro we might be concerned about a 4% fall in any one day; for the dollar against the yen we might be concerned about a 5% fall in any one day. We want to estimate the unconditional probability that one or both of these stress events occurs over any quarter. The probability p of interest is given by $p = 1 - P(M_{65,1} \leqslant 4\%, M_{65,2} \leqslant 5\%)$ and is approximated by $1 - H_{\boldsymbol{\xi},\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\theta}}(0.04, 0.05)$, where the parameters are estimated from the block maxima data. Of course, a more worrying scenario might be that both of these stress events should occur on the *same* day. To calculate the probability of simultaneous extreme events we require a different methodology, which is developed in Section 16.4.

Notes and Comments

Early works on distributions for bivariate extremes include Geffroy (1958), Tiago de Oliveira (1958) and Sibuya (1960). A selection of further important papers in the development of the subject include Galambos (1975), de Haan and Resnick (1977), Balkema and Resnick (1977), Deheuvels (1980) and Pickands (1981). The texts by Galambos (1987) and Resnick (2008) have both been influential; our presentation more closely resembles the former.

Theorem 16.14 is proved in Galambos (1987): see Theorem 5.2.1 and Lemma 5.4.1 therein (see also Joe 1997, p. 173). Theorem 16.15 is essentially a result of Pickands (1981). A complete version of the proof is given in Theorem 5.4.5 of

Galambos (1987), although it is given in the form of a characterization of MEV distributions with Gumbel margins. This is easily reformulated as a characterization of the EV copulas. In the bivariate case, necessary and sufficient conditions for $A(w)$ in (16.16) to define a bivariate EV copula are given in Joe (1997, Theorem 6.4).

The copula of Example 16.17 appears in Galambos (1975). A good summary of other bivariate and multivariate extreme value copulas is found in Kotz and Nadarajah (2000); they are presented as MEV distributions with unit Fréchet margins but the EV copulas are easily inferred from this presentation. See also Joe (1997, Chapters 5 and 6), in which EV copulas and their higher-dimensional extensions are discussed. Many parametric models for extremes have been suggested by Tawn (1988, 1990).

Theorem 16.18 is found in Galambos (1987), where the necessary and sufficient copula convergence criterion is given as $\lim_{n\to\infty} C^n(\boldsymbol{u}^{1/n}) = C_0(\boldsymbol{u})$ for positive integers n; by noting that for any $t > 0$ we have the inequalities

$$C^{[t]+1}(\boldsymbol{u}^{1/[t]}) \leqslant C^t(\boldsymbol{u}^{1/t}) \leqslant C^{[t]}(\boldsymbol{u}^{1/([t]+1)}),$$

it can be inferred that this is equivalent to $\lim_{t\to\infty} C^t(\boldsymbol{u}^{1/t}) = C_0(\boldsymbol{u})$. Further equivalent CDA conditions are found in Takahashi (1994). The idea of a domain of attraction of an EV copula also appears in Abdous, Ghoudi and Khoudraji (1999). Not every copula is in a copula domain of attraction; a counterexample may be found in Schlather and Tawn (2002).

We have shown that pairwise asymptotic independence for the components of random vectors implies pairwise independence of the corresponding components in the limiting MEV distribution of the maxima. Pairwise independence for an MEV distribution in fact implies mutual independence, as recognized and described by a number of authors: see Galambos (1987, Corollary 5.3.1), Resnick (2008, Theorem 5.27), and the earlier work of Geffroy (1958) and Sibuya (1960).

16.4 Multivariate Threshold Exceedances

In this section we describe practically useful models for multivariate extremes (again in low-dimensional applications) that build on the basic idea of modelling excesses over high thresholds with the generalized Pareto distribution (GPD) as in Section 5.2. The idea is to use GPD-based tail models of the kind discussed in Section 5.2.3 together with appropriate copulas to obtain models for multivariate threshold exceedances.

16.4.1 Threshold Models Using EV Copulas

Assume that the vectors $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ have unknown joint distribution $F(\boldsymbol{x}) = C(F_1(x_1), \ldots, F_d(x_d))$ for some unknown copula C and margins $F_1, \ldots, F_d$, and that F is in the domain of attraction of an MEV distribution. Much as in the univariate case we would like to approximate the upper tail of $F(\boldsymbol{x})$ above some vector of high thresholds $\boldsymbol{u} = (u_1, \ldots, u_d)'$. The univariate theory of Sections 5.2.2 and 5.2.3 tells us that, for $x_j \geqslant u_j$ and u_j high enough, the tail of the marginal distribution

F_j may be approximated by a GPD-based functional form

$$\tilde{F}_j(x_j) = 1 - \lambda_j \left(1 + \xi_j \frac{x_j - u_j}{\beta_j}\right)^{-1/\xi_j}, \tag{16.25}$$

where $\lambda_j = \bar{F}_j(u_j)$. This suggests that for $\boldsymbol{x} \geqslant \boldsymbol{u}$ we use the approximation $F(\boldsymbol{x}) \approx C(\tilde{F}_1(x_1), \ldots, \tilde{F}_d(x_d))$. But C is also unknown and must itself be approximated in the tail. The following heuristic argument suggests that we should be able to replace C by its limiting copula C_0.

The CDA condition (16.21) suggests that for any value $\boldsymbol{v} \in (0, 1)^d$ and t sufficiently large we may make the approximation $C(\boldsymbol{v}^{1/t}) \approx C_0^{1/t}(\boldsymbol{v})$. If we now write $\boldsymbol{w} = \boldsymbol{v}^{1/t}$, we have

$$C(\boldsymbol{w}) \approx C_0^{1/t}(\boldsymbol{w}^t) = C_0(\boldsymbol{w}) \tag{16.26}$$

by the scaling property of EV copulas. The approximation (16.26) will be best for large values of $\boldsymbol{w}$, since $\boldsymbol{v}^{1/t} \to \mathbf{1}$ as $t \to \infty$.

We assume then that we can substitute the copula C with its EV limit C_0 in the tail, and this gives us the overall model

$$\tilde{F}(\boldsymbol{x}) = C_0(\tilde{F}_1(x_1), \ldots, \tilde{F}_d(x_d)), \quad \boldsymbol{x} \geqslant \boldsymbol{u}. \tag{16.27}$$

We complete the model specification by choosing a flexible and tractable parametric EV copula for C_0. As before, the Gumbel copula family is particularly convenient.

16.4.2 Fitting a Multivariate Tail Model

Assume we have observations $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ from a df F with a tail that permits the approximation (16.27). Of these observations, only a minority are likely to be in the joint tail ($\boldsymbol{x} \geqslant \boldsymbol{u}$); other observations may exceed some of the individual thresholds but lie below others. The usual way of making inferences about all the parameters of such a model (the marginal parameters $\xi_j, \beta_j, \lambda_j$ for $j = 1, \ldots, d$ and the copula parameter (or parameter vector) θ is to maximize a likelihood for *censored data*.

Let us suppose that m_i components of the data vector $\boldsymbol{X}_i$ exceed their respective thresholds in the vector $\boldsymbol{u}$. The only relevant information that the remaining components convey is that they lie below their thresholds; such a component $X_{i,j}$ is said to be censored at the value u_j. The contribution to the likelihood of $\boldsymbol{X}_i$ is given by

$$L_i = L_i(\boldsymbol{\xi}, \boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\theta}; \boldsymbol{X}_i) = \left. \frac{\partial^{m_i} \tilde{F}(x_1, \ldots, x_d)}{\partial x_{j_1} \cdots \partial x_{j_{m_i}}} \right|_{\max(\boldsymbol{X}_i, \boldsymbol{u})},$$

where the indices $j_1, \ldots, j_{m_i}$ are those of the components of $\boldsymbol{X}_i$ exceeding their thresholds.

For example, in a bivariate model with Gumbel copula (7.12), the likelihood contribution would be

$$L_i = \begin{cases} C_\theta^{\mathrm{Gu}}(1 - \lambda_1, 1 - \lambda_2), & X_{i,1} \leqslant u_1,\ X_{i,2} \leqslant u_2, \\ C_{\theta,1}^{\mathrm{Gu}}(\tilde{F}_1(X_{i,1}), 1 - \lambda_2)\tilde{f}_1(X_{i,1}), & X_{i,1} > u_1,\ X_{i,2} \leqslant u_2, \\ C_{\theta,2}^{\mathrm{Gu}}(1 - \lambda_1, \tilde{F}_2(X_{i,2}))\tilde{f}_2(X_{i,2}), & X_{i,1} \leqslant u_1,\ X_{i,2} > u_2, \\ c_\theta^{\mathrm{Gu}}(\tilde{F}_1(X_{i,1}), \tilde{F}_2(X_{i,2}))\tilde{f}_1(X_{i,1})\tilde{f}_2(X_{i,2}), & X_{i,1} > u_1,\ X_{i,2} > u_2, \end{cases} \tag{16.28}$$

Table 16.2. Parameter estimates and standard errors (in brackets) for a bivariate tail model fitted to exchange-rate return data; see Example 16.23 for details.

	\$/€	\$/¥
u	0.75	1.00
N_u	189	126
λ	0.094 (0.0065)	0.063 (0.0054)
ξ	−0.049 (0.066)	0.095 (0.11)
β	0.33 (0.032)	0.38 (0.053)
θ	1.10 (0.030)	

where $\tilde{f}_j$ denotes the density of the univariate tail model $\tilde{F}_j$ in (16.25), $c_\theta^{\mathrm{Gu}}(u_1, u_2)$ denotes the Gumbel copula density, and $C_{\theta,j}^{\mathrm{Gu}}(u_1, u_2) := (\partial/\partial u_j)C_\theta^{\mathrm{Gu}}(u_1, u_2)$ denotes a conditional distribution of the copula, as in (7.17). The overall likelihood is a product of such contributions and is maximized with respect to all parameters of the marginal models and copula.

In a simpler approach, parameters of the marginal GPD models could be estimated as in Section 5.2.3 and only the parameters of the copula obtained from the above likelihood. In fact, this is also a sensible way of getting starting values before going on to the global maximization over all parameters.

The model described by the likelihood (16.28) has been studied in some detail by Ledford and Tawn (1996) and a number of related models have been studied in the statistical literature on multivariate EVT (see Notes and Comments for more details).

Example 16.23 (bivariate tail model for exchange-rate return data). We analyse daily percentage falls in the value of the US dollar against the euro and the Japanese yen, taking data for the eight-year period 1996–2003. We have 2008 daily returns and choose to set thresholds at 0.75% and 1.00%, giving 189 and 126 exceedances, respectively. In a full maximization of the likelihood over all parameters we obtained the estimates and standard errors shown in Table 16.2. The value of the maximized log-likelihood is −1064.7, compared with −1076.4 in a model where independence in the tail is assumed (i.e. a Gumbel copula with $\theta = 1$), showing strong evidence against an independence assumption.

We can now use the fitted model (16.27) to make various calculations about stress events. For example, an estimate of the probability that on any given day the dollar falls by more than 2% against both currencies is given by

$$p_{12} := 1 - \tilde{F}_1(2.00) - \tilde{F}_2(2.00) + C_\theta^{\mathrm{Gu}}(\tilde{F}_1(2.00), \tilde{F}_2(2.00)) = 0.000\,315,$$

with $\tilde{F}_j$ as in (16.25), making this approximately a 13-year event (assuming 250 trading days per year). The marginal probabilities of falls in value of this magnitude are $p_1 := 1 - \tilde{F}_1(2.00) = 0.0014$ and $p_2 := 1 - \tilde{F}_2(2.00) = 0.0061$. We can use this information to calculate so-called spillover probabilities for the conditional occurrence of stress events; for example, the probability that the dollar falls 2% against the yen given that it falls 2% against the euro is estimated to be $p_{12}/p_1 = 0.23$.

16.4.3 Threshold Copulas and Their Limits

An alternative approach to multivariate extremes looks explicitly at the kind of copulas we get when we condition observations to lie above or below extreme thresholds. Just as the GPD is a natural limiting model for univariate threshold exceedances, so we can find classes of copula that are natural limiting models for the dependence structure of multivariate exceedances.

The theory has been studied in most detail in the case of exchangeable bivariate copulas, and we concentrate on this case. Moreover, it proves slightly easier to switch our focus at this stage and first consider the lower-left tail of a probability distribution, before showing how the theory is adapted to the upper-right tail.

Lower threshold copulas and their limits. Consider a random vector (X_1, X_2) with continuous margins F_1 and F_2 and an exchangeable copula C. We consider the distribution of (X_1, X_2) conditional on both being below their v-quantiles, an event we denote by $A_v = \{X_1 \leqslant F_1^{\leftarrow}(v),\ X_2 \leqslant F_2^{\leftarrow}(v)\}$, $0 < v \leqslant 1$. Assuming $C(v, v) \neq 0$, the probability that X_1 lies below its x_1-quantile and X_2 lies below its x_2-quantile conditional on this event is

$$P(X_1 \leqslant F_1^{\leftarrow}(x_1), X_2 \leqslant F_2^{\leftarrow}(x_2) \mid A_v) = \frac{C(x_1, x_2)}{C(v, v)}, \quad x_1, x_2 \in [0, v].$$

Considered as a function of x_1 and x_2 this defines a bivariate df on $[0, v]^2$, and by Sklar's Theorem we can write

$$\frac{C(x_1, x_2)}{C(v, v)} = C_v^0(F_{(v)}(x_1), F_{(v)}(x_2)), \quad x_1, x_2 \in [0, v], \tag{16.29}$$

for a unique copula C_v^0 and continuous marginal distribution functions

$$F_{(v)}(x) = P(X_1 \leqslant F_1^{\leftarrow}(x) \mid A_v) = \frac{C(x, v)}{C(v, v)}, \quad 0 \leqslant x \leqslant v. \tag{16.30}$$

This unique copula may be written as

$$C_v^0(u_1, u_2) = \frac{C(F_{(v)}^{\leftarrow}(u_1), F_{(v)}^{\leftarrow}(u_2))}{C(v, v)}, \tag{16.31}$$

and it will be referred to as the *lower threshold copula* of C at level v. Juri and Wüthrich (2002), who developed the approach we describe in this section, refer to it as a lower tail dependence copula. It is of interest to attempt to evaluate limits for this copula as $v \to 0$; such a limit will be known as a *limiting lower threshold copula*.

Much like the GPD in Example 5.19, limiting lower threshold copulas must possess a stability property under the operation of calculating lower threshold copulas in (16.31). A copula C is a limiting lower threshold copula if, for any threshold $0 < v \leqslant 1$, it satisfies

$$C_v^0(u_1, u_2) = C(u_1, u_2). \tag{16.32}$$

Example 16.24 (Clayton copula as limiting lower threshold copula). For the standard bivariate Clayton copula in (7.13) we can easily calculate that $F_{(v)}$ in (16.30) is

$$F_{(v)}(x) = \frac{(x^{-\theta} + v^{-\theta} - 1)^{-1/\theta}}{(2v^{-\theta} - 1)^{-1/\theta}}, \quad 0 \leqslant x \leqslant v,$$

and its inverse is

$$F_{(v)}^{\leftarrow}(u) = u(2v^{-\theta} - 1 + u^{\theta}(1 - v^{-\theta}))^{-1/\theta}, \quad 0 \leqslant u \leqslant 1.$$

The lower threshold copula for the Clayton copula can therefore be calculated from (16.31) and it may be verified that this is again the Clayton copula. In other words, the Clayton copula is a limiting lower threshold copula because (16.32) holds.

Upper threshold copulas. To define upper threshold copulas we consider again a random vector (X_1, X_2) with copula C and margins F_1 and F_2. We now condition on the event $\bar{A}_v = \{X_1 > F_1^{\leftarrow}(v),\ X_2 > F_2^{\leftarrow}(v)\}$ for $0 \leqslant v < 1$. We have the identity

$$P(X_1 > F_1^{\leftarrow}(x_1), X_2 > F_2^{\leftarrow}(x_2) \mid \bar{A}_v) = \frac{\bar{C}(x_1, x_2)}{\bar{C}(v, v)}, \quad x_1, x_2 \in [v, 1].$$

Since $\bar{C}(x_1, x_2)/\bar{C}(v, v)$ defines a bivariate survival function on $[v, 1]^2$, by (7.14) we can write

$$\frac{\bar{C}(x_1, x_2)}{\bar{C}(v, v)} = \hat{C}_v^1(\bar{G}_{(v)}(x_1), \bar{G}_{(v)}(x_2)), \quad x_1, x_2 \in [v, 1], \tag{16.33}$$

for some survival copula $\hat{C}_v^1$ of a copula C_v^1 and marginal survival functions

$$\bar{G}_{(v)}(x) = P(X_1 > F_1^{\leftarrow}(x) \mid \bar{A}_v) = \frac{\bar{C}(x, v)}{\bar{C}(v, v)}, \quad v \leqslant x \leqslant 1. \tag{16.34}$$

The copula C_v^1 is known as the *upper threshold copula* at level v and it is now of interest to find limits as $v \to 1$, which are known as limiting upper threshold copulas. In fact, as the following lemma shows, it suffices to study either lower or upper threshold copulas because results for one follow easily from results for the other.

Lemma 16.25. *The survival copula of the upper threshold copula of C at level v is the lower threshold copula of $\hat{C}$ at level $1 - v$.*

Proof. We use the identity $\bar{C}(u_1, u_2) = \hat{C}(1 - u_1, 1 - u_2)$ and (16.34) to rewrite (16.33) as

$$\frac{\hat{C}(1 - x_1, 1 - x_2)}{\hat{C}(1 - v, 1 - v)} = \hat{C}_v^1\left(\frac{\hat{C}(1 - x_1, 1 - v)}{\hat{C}(1 - v, 1 - v)}, \frac{\hat{C}(1 - v, 1 - x_2)}{\hat{C}(1 - v, 1 - v)}\right).$$

Writing $y_1 = 1 - x_1$, $y_2 = 1 - x_2$ and $w = 1 - v$ we have

$$\frac{\hat{C}(y_1, y_2)}{\hat{C}(w, w)} = \hat{C}_{1-w}^1\left(\frac{\hat{C}(y_1, w)}{\hat{C}(w, w)}, \frac{\hat{C}(w, y_2)}{\hat{C}(w, w)}\right), \quad y_1, y_2 \in [0, w],$$

and comparison with (16.29) and (16.30) shows that $\hat{C}_{1-w}^1$ must be the lower threshold copula of $\hat{C}$ at the level $w = 1 - v$. □

It follows that the survival copulas of limiting lower threshold copulas are limiting upper threshold copulas. The Clayton survival copula is a limiting upper threshold copula.

Relationship between limiting threshold copulas and EV copulas. We give one result that shows how limiting upper threshold copulas may be calculated for underlying exchangeable copulas C that are in the domain of attraction of EV copulas with tail dependence, thus linking the study of threshold copulas to the theory of Section 16.3.3.

Theorem 16.26. *If C is an exchangeable copula with upper tail-dependence coefficient $\lambda_u > 0$ satisfying $C \in \mathrm{CDA}(C_0)$, then C has a limiting upper threshold copula that is the survival copula of the df*

$$G(x_1, x_2) = \frac{(x_1 + x_2)(1 - A(x_1/(x_1 + x_2)))}{\lambda_u}, \tag{16.35}$$

where A is the dependence function of C_0. Also, $\hat{C}$ has a limiting lower threshold copula that is the copula of G.

Example 16.27 (upper threshold copula of Galambos copula). We use this result to calculate the limiting upper threshold copula for the Galambos copula. We recall that this is an EV copula with dependence function given in (16.19) and consider the standard exchangeable case with $\alpha = \beta = 1$. Using the methods of Section 7.2.4 it may easily be calculated that the coefficient of upper tail dependence of this copula is $\lambda_u = 2^{-1/\theta}$. The bivariate distribution $G(x_1, x_2)$ in (16.35) is therefore given by

$$G(x_1, x_2) = (\tfrac{1}{2}(x_1^{-\theta} + x_2^{-\theta}))^{-1/\theta}, \quad (x_1, x_2) \in (0, 1]^2,$$

the copula of which is the Clayton copula. The limiting upper threshold copula in this case is therefore the Clayton survival copula. Moreover, the limiting lower threshold copula of the Galambos survival copula is the Clayton copula.

The Clayton copula turns out to be an important attractor for a large class of underlying exchangeable copulas. Juri and Wüthrich (2003) have shown that all Archimedean copulas whose generators are regularly varying at 0 with negative parameter (meaning that $\phi(t)$ satisfies $\lim_{t\to 0} \phi(xt)/\phi(t) = x^{-\alpha}$ for all x and some $\alpha > 0$) share the Clayton copula C_α^{Cl} as their limiting lower threshold copula.

It is of interest to calculate limiting lower and upper threshold copulas for the t copula, and this can be done using Theorem 16.26 and the expression for the dependence function in (16.24). However, the resulting limit is not convenient for practical purposes because of the complexity of this dependence function. We have already remarked in Section 16.3.4 that the dependence function of the t–EV copula can be well approximated by the dependence functions of other exchangeable EV copulas, such as Gumbel and Galambos, for most practical purposes. Theorem 16.26 therefore suggests that instead of working with the true limiting upper threshold copula of the t copula we could instead work with the limiting upper threshold copula of, say, the Galambos copula, i.e. the Clayton survival copula. Similarly, we could work with the Clayton copula as an approximation for the true limiting lower threshold copula of the t copula.

Limiting threshold copulas in practice Limiting threshold copulas in dimensions higher than two have not yet been extensively studied, nor have limits for non-exchangeable bivariate copulas or limits when we define two thresholds v_1 and v_2 and let these tend to zero (or one) at different rates. The practical use of these ideas is therefore largely confined to bivariate applications when thresholds are set at approximately similar quantiles and a symmetric dependence structure is assumed.

Let us consider a situation where we have a bivariate distribution that appears to exhibit tail dependence in both the upper-right and lower-left corners. While true lower and upper limiting threshold copulas may exist for this unknown distribution, we could in practice simply adopt a tractable and flexible parametric limiting threshold copula family. It is particularly easy to use the Clayton copula and its survival copula as lower and upper limits, respectively.

Suppose, for example, that we set high thresholds at $\boldsymbol{u} = (u_1, u_2)'$, so that $P(X_1 > u_1) \approx P(X_2 > u_2)$ and both probabilities are small. For the conditional distribution of (X_1, X_2) over the threshold $\boldsymbol{u}$ we could assume a model of the form

$$P(\boldsymbol{X} \leqslant \boldsymbol{x} \mid \boldsymbol{X} > \boldsymbol{u}) \approx \hat{C}_\theta^{\mathrm{Cl}}(G_{\xi_1,\beta_1}(x_1 - u_1), G_{\xi_2,\beta_2}(x_2 - u_2)), \quad \boldsymbol{x} > \boldsymbol{u},$$

where $\hat{C}_\theta^{\mathrm{Cl}}$ is the Clayton survival copula and G_{ξ_j,β_j} denotes a GPD, as defined in 5.16. Inference about the model parameters $(\theta, \xi_1, \beta_1, \xi_2, \beta_2)$ would be based on the exceedance data above the thresholds and would use the methods discussed in Section 7.5.

Similarly, for a vector of low thresholds $\boldsymbol{u}$ satisfying $P(X_1 \leqslant u_1) \approx P(X_2 \leqslant u_2)$ with both these probabilities small, we could approximate the conditional distribution of (X_1, X_2) below the threshold $\boldsymbol{u}$ by a model of the form

$$P(\boldsymbol{X} \leqslant \boldsymbol{x} \mid \boldsymbol{X} < \boldsymbol{u}) \approx C_\theta^{\mathrm{Cl}}(\bar{G}_{\xi_1,\beta_1}(u_1 - x_1), \bar{G}_{\xi_2,\beta_2}(u_2 - x_2)), \quad \boldsymbol{x} < \boldsymbol{u},$$

where C_θ^{Cl} is the Clayton copula and $\bar{G}_{\xi_j,\beta_j}$ denotes a GPD survival function. Inference about the model parameters would be based on the data below the thresholds and would use the methods of Section 7.5.

Note and Comments

The GPD-based tail model (16.27) and inference for censored data using a likelihood of the form (16.28) have been studied by Ledford and Tawn (1996), although the derivation of the model uses somewhat different asymptotic reasoning based on a characterization of multivariate domains of attraction of MEV distributions with unit Fréchet margins found in Resnick (2008). The authors of the former paper concentrate on the model with Gumbel (logistic) dependence structure and discuss, in particular, testing for asymptotic independence of extremes. Likelihood inference is non-problematic (the problem being essentially regular) when $\theta > 0$ and $\xi_j > -\frac{1}{2}$, but testing for independence of extremes $\theta = 1$ is not quite so straightforward since this is a boundary point of the parameter space. This case is possibly more interesting in environmental applications than in financial ones, where we tend to expect dependence of extreme values.

A related bivariate GPD model is presented in Smith, Tawn and Coles (1997). In our notation they essentially consider a model of the form

$$\bar{F}(x_1, \ldots, x_d) = 1 + \ln C_0(\mathrm{e}^{\tilde{F}(x_1)-1}, \ldots, \mathrm{e}^{\tilde{F}(x_d)-1}), \quad \boldsymbol{x} \geqslant \boldsymbol{k},$$

where C_0 is an extreme value copula. This model is also discussed in Smith (1994) and Ledford and Tawn (1996); it is pointed out that $\bar{F}$ does not reduce to a product of marginal distributions in the case when C_0 is the independence copula, unlike the model in (16.27).

Another style of statistical model for multivariate extremes is based on the point process theory of multivariate extremes developed in de Haan (1985), de Haan and Resnick (1977) and Resnick (2008). Statistical models using this theory are found in Coles and Tawn (1991) and Joe, Smith and Weissman (1992); see also the texts of Joe (1997) and Coles (2001). New approaches to modelling multivariate extremes can be found in Heffernan and Tawn (2004) and Balkema and Embrechts (2007); the latter paper considers applications to stress testing high-dimensional portfolios in finance.

Limiting threshold copulas are studied in Juri and Wüthrich (2002, 2003). In the latter paper it is demonstrated that the Clayton copula is an attractor for the threshold copulas of a wide class of Archimedean copulas; moreover, a version of our Theorem 16.26 is proved. Limiting threshold copulas for the t copula are investigated in Demarta and McNeil (2005). The usefulness of Clayton's copula and the Clayton survival copula for describing the dependence in the tails of bivariate financial return data was confirmed in a large-scale empirical study of high-frequency exchange-rate returns by Breymann, Dias and Embrechts (2003).

17

Dynamic Portfolio Credit Risk Models and Counterparty Risk

This chapter is concerned with dynamic reduced-form models of portfolio credit risk. It is also the natural place for an analysis of counterparty credit risk for over-the-counter (OTC) credit derivatives, since this risk can only be satisfactorily modelled in the framework of dynamic models.

In Section 17.1 we give an informal introduction to the subject of dynamic models in which we explain why certain risk-management tasks for portfolios of credit derivatives cannot be properly handled in the copula framework of Chapter 12; we also give an overview of the different types of dynamic model used in portfolio credit risk.

A detailed analysis of counterparty credit risk management is presented in Section 17.2, while the remainder of the chapter focusses on two different approaches to dynamic modelling. In Section 17.3 we consider dynamic models with conditionally independent default times, and in Section 17.4 we treat credit risk models with incomplete information.

We use a number of concepts and techniques from continuous-time finance and build on material in Chapters 10 and 12. Particular prerequisites for reading this chapter are Sections 10.5 and 10.6 on pricing single-name credit derivatives in models with stochastic hazard rates and Section 12.3 on CDO pricing in factor copula models.

17.1 Dynamic Portfolio Credit Risk Models

17.1.1 Why Dynamic Models of Portfolio Credit Risk?

In the copula models of portfolio credit risk in Chapter 12, the joint distribution of the default times was specified directly. However, the evolution of this distribution over time, for instance in reaction to new economic information, was not modelled. While the copula approach is sufficient for computing the prices of many credit products such as index swaps or CDO tranches at a given point in time t, it is not possible to say anything about the dynamics of these prices over time. Certain important tasks in the risk management of credit derivative portfolios cannot therefore be handled properly in the copula framework, as we now discuss.

To begin with, in a copula model it is not possible to price options on credit-risky instruments such as options on a basket of corporate bonds. This is an issue of

practical relevance, since the task of pricing an option on a credit-risky instrument arises in the management of counterparty credit risk management, particularly in the computation of a so-called *credit value adjustment* (CVA) for a credit derivative. Consider, for instance, a protection buyer who has entered into a CDS contract with some protection seller. Suppose now that the protection seller defaults during the lifetime of the CDS contract and that the credit spread of the reference entity has gone up. In that case the protection buyer suffers a loss, since in order to renew his protection he has to enter into a new CDS at a higher spread. In order to account for this loss the protection buyer should make an adjustment to the value of the CDS. We will show in Section 17.2 that, roughly speaking, this adjustment takes the form of an option on the value of the future cash flows of the CDS with maturity date equal to the default time of the protection seller.

Moreover, in copula models it is not possible to derive model-based dynamic hedging strategies. For this reason risk managers often resort to sensitivity-based hedging strategies that are similar to the use of duration-based immunization for the risk management of bond portfolios. This is not entirely satisfactory, since it is known from markets for other types of derivatives that hedging strategies that lack a proper theoretical foundation may perform poorly (see Notes and Comments for references). On the theoretical side, the non-existence of model-based hedging strategies implies that pricing results in copula models are not supported by hedging arguments, so that an important insight of modern derivative asset analysis is ignored in standard CDO pricing. Of course, dynamic trading strategies are substantially more difficult to implement in credit-derivative markets than in equity markets and their performance is less robust with respect to model misspecification. However, in our view this should not serve as an excuse for neglecting the issue of dynamic modelling and dynamic hedging altogether. A number of recent papers on the hedging of portfolio credit derivatives are listed in Notes and Comments.

17.1.2 Classes of Reduced-Form Models of Portfolio Credit Risk

The properties of a reduced-form model of portfolio credit risk are essentially determined by the *intensities* of the default times. Different models can be classified according to the mathematical structure given to these intensities. The concept of the intensity of a random default time was encountered in Definition 10.15 and will be explained in more detail in Section 17.3.1.

The simplest reduced-form portfolio models are models with *conditionally independent defaults* (see Section 17.3). These models are a straightforward multi-firm extension of models with doubly stochastic default times. Default times are modelled as conditionally independent given the realization of some observable economic background process $(\boldsymbol{\Psi}_t)$. The default intensities are adapted to the filtration generated by $(\boldsymbol{\Psi}_t)$, and dependence between defaults is generated by the mutual dependence of the default intensities on the common background process. An important special case is the class of models with Markov-modulated intensities where $\lambda_{t,i} = \lambda_i(\boldsymbol{\Psi}_t)$ and $(\boldsymbol{\Psi}_t)$ is Markovian.

In Section 17.4 we consider reduced-form models with *incomplete information*. We consider a set-up where the default times are independent given a common factor. In the simplest case this factor is simply a random variable V, as in the factor copula models studied in Section 12.2.2. Assume for the moment that V is observable. Then defaults are independent and the default intensities have the simple form $\tilde{\lambda}_{t,i} = \gamma_i(V, t)$ for suitable functions γ_i. We assume, however, that the factor is not directly observable. Instead investors are only able to observe the default history and, at most, an auxiliary process representing noisy observations of V. In this context it will be shown that the default intensities are computed by projection. Denoting the default intensity of firm i at time t by $\lambda_{t,i}$ and the information available to investors at t by the σ-field $\mathcal{G}_t$, we have that

$$\lambda_{t,i} = E(\gamma_i(V, t) \mid \mathcal{G}_t). \tag{17.1}$$

Moreover, prices of many credit derivatives are given by similar conditional expectations. We will see that Bayesian updating and, more generally, stochastic filtering techniques can be employed in the evaluation of expressions of the form (17.1) and in the analysis of the model in general.

A further important model class comprises models with *interacting intensities*. We will not present these in any detail for reasons of space but important contributions to the literature are listed in Notes and Comments. In these models the impact of the default of one firm on the default intensities of surviving firms is specified explicitly. For instance, we might assume that, after a default event, default intensities increase by 10% of their pre-default value. This interaction among intensities provides an alternative mechanism for creating dependence between default events. In formal terms, models with interacting intensities are constructed as Markov chains on the set of all possible default states of the portfolio. The theory of Markov chains therefore plays an important role in their analysis.

A common feature of models with interacting default intensities and models with incomplete information is the presence of *default contagion*. This means that the default intensity of a surviving firm jumps (usually upwards) given the information that some other firm has defaulted. As a consequence, the credit spread of surviving firms increases when default events occur in the portfolio. Default contagion can arise via different channels. On the one hand, contagious effects might be due to direct economic links between firms, such as a close business relationship or a strong borrower–lender relationship. For instance, the default probability of a bank is likely to increase if one of its major borrowers defaults. Broadly speaking, this channel of default interaction is linked to counterparty risk. On the other hand, changes in the conditional default probability of non-defaulted firms can be caused by information effects; investors might revise their estimate of the financial health of non-defaulted firms in light of the news that a particular firm has defaulted. This is known as *information-based default contagion*. Note that models with incomplete information generate information-based default contagion by design. At a default event the conditional distribution of the factor V given the investor information is updated; by (17.1) this leads to a jump in the default intensity $\lambda_{t,i}$. There is in

fact substantial evidence for contagion effects. A good example is provided by the default of the investment bank Lehman Brothers in autumn 2008; the default event combined with the general nervousness caused by the worsening financial crisis sent credit spreads to unprecedentedly high levels.

We can also distinguish between *bottom-up* and *top-down* models of portfolio credit risk. This distinction relates to the quantities that are modelled and cuts through all types of reduced-form portfolio credit risk models including the copula models of Chapter 12.

The fundamental objects that are modelled in a bottom-up model are the default indicator processes of the individual firms in the portfolio under consideration; the dynamics of the portfolio loss are then derived from these. In this approach it is possible to price portfolio products consistently with observed single-name CDS spreads and to derive hedging strategies for portfolio products that use single-name CDSs as hedging instruments. These are obvious advantages of this model class. However, there are also some drawbacks related to tractability: in the bottom-up approach we have to keep track of all default-indicator processes and possibly also background processes driving the model. This typically leads to substantial computational challenges in pricing and model calibration, particularly if the portfolio size is fairly large.

In top-down models, on the other hand, the portfolio loss process is modelled directly, without reference to the individual default indicator processes. This obviously drastically reduces the dimensionality of the resulting models. It can be argued that top-down models are sufficient for the pricing of index derivatives, since the payoff of these contracts depends only on the portfolio loss. However, in this model class the information contained in single-name spreads is neglected for pricing purposes, and it is not possible to study the hedging of portfolio derivatives with single-name CDSs.

There is no obvious and universally valid answer to the question of which model class should be preferred; in Notes and Comments we provide a few references in which this issue is discussed further. In our own analysis we concentrate on bottom-up models.

Notes and Comments

The limitations of static copula models are discussed in a number of research papers; a particularly readable contribution is Shreve (2009). An interesting collection of papers that deal with portfolio credit risk models "after copulas" is found in Lipton and Rennie (2008). Dynamic hedging strategies for portfolio credit derivatives are studied by Frey and Backhaus (2010), Laurent, Cousin and Fermanian (2011) and Cont and Kan (2011), among others; an earlier contribution is Bielecki, Jeanblanc and Rutkowski (2004). A detailed mathematical analysis of hedging errors for equity and currency derivatives is given in El Karoui, Jeanblanc-Picqué and Shreve (1998).

There is a rich literature on models with interacting intensities. Bottom-up models are considered by Davis and Lo (2001), Jarrow and Yu (2001), Yu (2007), Frey and Backhaus (2008) and Herbertsson (2008). Top-down models with interacting

intensities include the contributions by Arnsdorf and Halperin (2009) and Cont and Minca (2013). Moreover, there are top-down models where the dynamics of the whole "surface" of CDO tranche spreads—that is the dynamics of CDO spreads for all maturities and attachment points—are modelled directly: see, for example, Ehlers and Schönbucher (2009), Sidenius, Piterbarg and Andersen (2008) and Filipović, Overbeck and Schmidt (2011). The modelling philosophy of these three papers is akin to the well-known HJM models for the term structure of interest rates. A general discussion of the pros and cons of bottom-up and top-down models can be found in Bielecki, Crépey, and Jeanblanc (2010) (see also Giesecke, Goldberg and Ding 2011).

Credit risk models with explicitly specified interactions between default intensities are conceptually close to network models and to models for interacting particle systems developed in statistical physics. Föllmer (1994) contains an inspiring discussion of the relevance of these ideas to financial modelling; the link to credit risk is explored by Giesecke and Weber (2004, 2006) and Horst (2007). Network models are frequently used for the study of systemic risk in financial networks, an issue that has become highly relevant in the aftermath of the financial crisis of 2007–9. Interesting contributions in this rapidly growing field include the papers by Eisenberg and Noe (2001), Elsinger, Lehar and Summer (2006), Gai and Kapadia (2010), Upper (2011) and Amini, Cont and Minca (2012).

There are some empirical papers on default contagion. Jarrow and Yu (2001) provide anecdotal evidence for counterparty-risk-related contagion in small portfolios. In contrast, Lando and Nielsen (2010) find no strong empirical evidence for default contagion in historical default patterns.

Other work has tested the impact of the default or spread widening of a given firm on the credit spreads or stock returns of other firms: see Collin-Dufresne, Goldstein and Helwege (2010) or Lang and Stulz (1992). The evidence in favour of this type of default contagion is quite strong. For instance, Collin-Dufresne, Goldstein and Helwege (2010) found that, even after controlling for other macroeconomic variables influencing bond returns, the returns of large corporate bond indices in months where one or more large firms experienced a significant widening in credit spreads (above 200 basis points) were significantly lower than the returns of these indices in other months; this is clear evidence supporting contagion.

17.2 Counterparty Credit Risk Management

A substantial proportion of all derivative transactions are carried out OTC, so that counterparty credit risk is a key issue for financial institutions. The management of counterparty risk poses a number of challenges. To begin with, a financial institution needs to measure (in close to real time) its counterparty risk exposure to its various trading partners. Moreover, counterparty risk needs to be taken into account in the pricing of derivative contracts, which leads to the issue of computing credit value adjustments. Finally, financial institutions and major corporations apply various risk-mitigation strategies in order to control and reduce their counterparty risk exposure. In particular, many OTC derivative transactions are now *collateralized*.

Consider a derivative transaction such as a CDS contract between two parties—the protection seller S and the protection buyer B—and suppose that the deal is collateralized. If the value of the CDS is negative for, say, S, then S passes cash or securities (the collateral) to B. If S defaults before the maturity of the underlying CDS and if the value of the CDS at the default time τ_S of S is positive for B, the protection buyer is permitted to liquidate the collateral in order to reduce the loss due to the default of S; excess collateral must be returned to S. Most collateralization agreements are symmetric so that the roles of S and B can be exchanged when the value of the underlying CDS changes its sign.

In this section we study quantitative aspects of counterparty risk management. In Section 17.2.1 we introduce the general form of credit value adjustments for uncollateralized derivative transactions and we discuss various simplifications that are used in practice. In Section 17.2.2 we consider the case of collateralized transactions. For concreteness, we discuss value adjustments and collateralization strategies for a single-name CDS, but our arguments apply to other contracts as well.

17.2.1 Uncollateralized Value Adjustments for a CDS

We begin with an analysis of the form of credit value adjustments for an uncollateralized single-name CDS contract on some reference entity R. We work on a filtered probability space $(\Omega, \mathcal{G}, (\mathcal{G}_t), Q)$, where Q denotes the risk-neutral measure used for pricing derivatives and where the filtration $(\mathcal{G}_t)$ represents the information available to investors. Our notation is as follows: the default times of the protection seller S, the protection buyer B and the reference entity R are denoted by the $(\mathcal{G}_t)$ stopping times τ_S, τ_B and τ_R; δ^R, δ^S and δ^B are the losses given default (LGDs) of the contracting parties; $T_1 = \min\{\tau_R, \tau_S, \tau_B\}$ denotes the first default time; $\xi_1 \in \{R, S, B\}$ gives the identity of the firm that defaults first. We assume that δ^R, δ^S and δ^B are constant; for a discussion of the calibration of these parameters in the context of counterparty credit risk, we refer to Gregory (2012).

The CDS contract referencing R has premium payment dates $t_1 < \cdots < t_N = T$, where t_1 is greater than the current time t and a fixed spread x. The default-free short rate is given by the $(\mathcal{G}_t)$-adapted process (r_t). In discounting future cash flows it will be convenient to use the abbreviation

$$D(s_1, s_2) = \exp\left(-\int_{s_1}^{s_2} r_u \, \mathrm{d}u\right), \quad 0 \leqslant s_1 \leqslant s_2 \leqslant T.$$

The promised cash flow of a protection buyer position in the CDS between two time points $s_1 < s_2$, discounted back to time s_1, will be denoted by $\Pi(s_1, s_2)$. Ignoring for simplicity accrued premium payments, we therefore have

$$\Pi(s_1, s_2) = \int_{s_1}^{s_2} D(s_1, u)\delta^R \, \mathrm{d}Y_{R,u} - x \sum_{t_n \in (s_1, s_2]} D(s_1, t_n)(1 - Y_{R,t_n}). \tag{17.2}$$

The first term on the right-hand side of (17.2) represents the discounted default payment, and the second term corresponds to the discounted premium payment.

The value at some stopping time $\tau \geqslant t$ of the promised cash-flow stream for B is then given by

$$V_\tau := E^Q(\Pi(\tau, T) \mid \mathcal{G}_\tau); \tag{17.3}$$

sometimes we will call (V_t) the counterparty-risk-free CDS price. The discounted cash flows that are made or received by B over the period $(s_1, s_2]$ (the real cash flows) are denoted by $\Pi^{\text{real}}(s_1, s_2)$. Note that Π and Π^{real} are in general different as S or B might default before the maturity date T of the transaction.

In order to describe Π^{real} we distinguish the following scenarios.

- If $T_1 > T$ or if $T_1 \leqslant T$ and $\xi_1 = R$, that is if both S and B survive until the maturity date of the CDS, the actual and promised cash-flow streams coincide, so that $\Pi^{\text{real}}(\cdot, T) = \Pi(\cdot, T)$.
- Consider next the scenario where $T_1 < T$ and $\xi_1 = S$, that is the protection seller defaults first and this occurs before the maturity date of the CDS. In that case, prior to T_1, actual and promised cash flows coincide. At T_1, if the counterparty-risk-free CDS price V_{T_1} is positive, B is entitled to charge a *close-out amount* to S in order to settle the contract. Following the literature we assume that this close-out amount is given by V_{T_1}. However, S is typically unable to pay the close-out amount in full, and B receives only a recovery payment of size $(1 - \delta^S)V_{T_1}$. If, on the other hand, V_{T_1} is negative, B has to pay the full amount $|V_{T_1}|$ to S. Using the notation $x^+ = \max\{x, 0\}$ and $x^- = -\min\{x, 0\}$, the actual cash flows on the set $\{T_1 < T\} \cap \{\xi_1 = S\}$ are given by

$$\Pi^{\text{real}}(t, T) = \Pi(t, T_1) + D(t, T_1)((1 - \delta^S)V_{T_1}^+ - V_{T_1}^-). \tag{17.4}$$

- Finally, consider the scenario where $T_1 < T$ and $\xi_1 = B$. If $V_{T_1} > 0$, S has to pay the full amount V_{T_1} to B; if $V_{T_1} < 0$, the protection buyer makes a recovery payment of size $(1 - \delta^B)|V_{T_1}|$ to S. Thus, on the set $\{T_1 < T\} \cap \{\xi_1 = B\}$ we have

$$\Pi^{\text{real}}(t, T) = \Pi(t, T_1) + D(t, T_1)(V_{T_1}^+ - (1 - \delta^B)V_{T_1}^-). \tag{17.5}$$

The correct value of the protection-buyer position in the CDS in the presence of counterparty risk is given by $E^Q(\Pi^{\text{real}}(t, T) \mid \mathcal{G}_t)$. For $t < T_1$ the difference

$$\text{BCVA}_t := E^Q(\Pi(t, T) \mid \mathcal{G}_t) - E^Q(\Pi^{\text{real}}(t, T) \mid \mathcal{G}_t) \tag{17.6}$$

is known as the *bilateral credit value adjustment* (BCVA) at time t. Note that $E^Q(\Pi^{\text{real}}(t, T) \mid \mathcal{G}_t) = V_t - \text{BCVA}_t$: that is, BCVA_t is the adjustment that needs to be made to the counterparty-risk-free CDS price in order to obtain the value of the cash-flow stream Π^{real}. The term *bilateral* refers to the fact that the value adjustment takes the possibility of the default of both contracting parties, B and S, into account. By definition, the bilateral value adjustment is symmetric in the sense that the value adjustment computed from the viewpoint of the protection seller at time

t is given by $-\mathrm{BCVA}_t$; this is obvious since the cash-flow stream received by S is exactly the negative of the cash-flow stream received by B. This contrasts with so-called unilateral value adjustments where each party neglects the possibility of its own default in computing the adjustment to the value of the CDS.

The next proposition gives a more succinct expression for the BCVA.

Proposition 17.1. *For $t < T_1$ we have that* $\mathrm{BCVA}_t = \mathrm{CVA}_t - \mathrm{DVA}_t$, *where*

$$\mathrm{CVA}_t = E^Q(I_{\{T_1 \leqslant T\}} I_{\{\xi_1 = S\}} D(t, T_1) \delta^S V_{T_1}^+ \mid \mathcal{G}_t), \tag{17.7}$$

$$\mathrm{DVA}_t = E^Q(I_{\{T_1 \leqslant T\}} I_{\{\xi_1 = B\}} D(t, T_1) \delta^B V_{T_1}^- \mid \mathcal{G}_t). \tag{17.8}$$

Comments. The CVA in (17.7) reflects the potential loss incurred by B due to a premature default of S; the *debt value adjustment*, or DVA, in (17.8) reflects the potential loss incurred by S due to a premature default of B. A similar formula obviously holds for other products; the only part that needs to be adapted is the definition of the promised cash-flow stream in (17.2).

Accounting rules require that both CVA and DVA have to be taken into account if an instrument is valued via mark-to-market accounting techniques. Note, however, that the use of DVA is somewhat controversial for the following reason: a decrease in the credit quality of B leads to an increase in the probability that B defaults first and hence to a larger DVA term. If both CVA and DVA are taken into account in the valuation of the CDS, this would be reported as a profit for B. It is not clear, however, how B could turn this accounting profit into an actual cash flow for its shareholders.

Proposition 17.1 shows that the problem of computing the BCVA amounts to computing the price of a call option and a put option on (V_t) with strike $K = 0$ and random maturity date T_1. The computation of the value adjustments is therefore more involved than the pricing of the CDS itself, and a dynamic portfolio credit model is needed to compute the value adjustment in a consistent way. The actual computation of value adjustments depends on the structure of the underlying credit model. For further information we refer to Sections 17.3.3 and 17.4.4.

Proof of Proposition 17.1. For $t < T_1 \leqslant s$ it holds that

$$D(t, s) = D(t, T_1) D(T_1, s).$$

Hence, on the set $\{T_1 \leqslant T\}$ we may write $\Pi(t, T) = \Pi(t, T_1) + D(t, T_1)\Pi(T_1, T)$. This yields

$$\Pi(t, T) - \Pi^{\mathrm{real}}(t, T) = I_{\{T_1 \leqslant T\}} D(t, T_1)(\Pi(T_1, T) - I_{\{\xi_1 = S\}}((1 - \delta^S)V_{T_1}^+ - V_{T_1}^-) - I_{\{\xi_1 = B\}}(V_{T_1}^+ - (1 - \delta^B)V_{T_1}^-)).$$

By iterated conditional expectations it follows that

$$E^Q(\Pi(t, T) - \Pi^{\mathrm{real}}(t, T)) = E^Q(E^Q(\Pi(t, T) - \Pi^{\mathrm{real}}(t, T) \mid \mathcal{G}_{T_1})). \tag{17.9}$$

We concentrate on the inner conditional expectation. Since $D(t,T_1)$ and the events $\{T_1 \leqslant T\}$, $\{\xi_1 = S\}$ and $\{\xi_1 = B\}$ are $\mathcal{G}_{T_1}$-measurable, we obtain

$$E^Q(\Pi(t,T) - \Pi^{\text{real}}(t,T) \mid \mathcal{G}_{T_1}) \tag{17.10a}$$

$$= I_{\{T_1 \leqslant T\}} I_{\{\xi_1 = S\}} D(t,T_1) E^Q(\Pi(T_1,T) - ((1-\delta^S)V_{T_1}^+ - V_{T_1}^-) \mid \mathcal{G}_{T_1}) \tag{17.10b}$$

$$+ I_{\{T_1 \leqslant T\}} I_{\{\xi_1 = B\}} D(t,T_1) E^Q(\Pi(T_1,T) - (V_{T_1}^+ - (1-\delta^B)V_{T_1}^-) \mid \mathcal{G}_{T_1}). \tag{17.10c}$$

Now, by the definition of (V_t) we have that $E^Q(\Pi(T_1,T) \mid \mathcal{G}_{T_1}) = V_{T_1}$. Moreover, we use the decomposition $V_{T_1} = V_{T_1}^+ - V_{T_1}^-$, where $V_{T_1}^+$ and $V_{T_1}^-$ are $\mathcal{G}_{T_1}$ measurable. Hence (17.10 *b*) equals $I_{\{T_1 \leqslant T\}} I_{\{\xi_1 = S\}} D(t,T_1)\delta^S V_{T_1}^+$ and, similarly, (17.10 *c*) equals $-I_{\{T_1 \leqslant T\}} I_{\{\xi_1 = B\}} \delta^B V_{T_1}^-$. Putting these together, (17.10 *a*) is equal to

$$I_{\{T_1 \leqslant T\}} D(t,T_1)(I_{\{\xi_1 = S\}}\delta^S V_{T_1}^+ - I_{\{\xi_1 = B\}} D(t,T_1)\delta^B V_{T_1}^-),$$

and substituting this into (17.9) gives the result. □

Simplified value adjustments and wrong-way risk. In order to simplify the computation of value adjustments, it is often assumed that $(Y_{t,S})$, $(Y_{t,B})$ and the counterparty-risk free CDS price (V_t) are independent stochastic processes and that the risk-free interest rate is deterministic. We now explain how the value adjustment formulas (17.7) and (17.8) simplify under this independence assumption. For simplicity we consider the case $t = 0$. Denote by $\bar{F}_S(t)$, $\bar{F}_B(t)$, $f_S(t)$ and $f_B(t)$ the survival functions and densities of τ_S and τ_B. Since $\{\xi_1 = S\} = \{\tau_S < \tau_B\} \cap \{\tau_S < \tau_R\}$ and since $V_{T_1} = V_{\tau_S}$ on $\{\xi_1 = S\}$, we obtain

$$\begin{aligned}\text{CVA} = \text{CVA}_0 &= E^Q(I_{\{\tau_S \leqslant T\}} I_{\{\tau_S < \tau_B\}} I_{\{\tau_S < \tau_R\}} D(0,\tau_S)\delta^S V_{\tau_S}^+) \\ &= \delta^S \int_0^T E^Q(I_{\{\tau_S < \tau_B\}} D(0,\tau_S) V_{\tau_S}^+ \mid \tau_S = t) f_S(t)\,\mathrm{d}t.\end{aligned}$$

In the last line we have used the fact that δ^S is deterministic and we have used the identity $V_s \equiv 0$ on $\{\tau_R \leqslant s\}$, which allows the indicator $I_{\{\tau_S < \tau_R\}}$ to be dropped. The independence of the processes $(Y_{t,S})$, $(Y_{t,B})$, (V_t) and the fact that interest rates are deterministic imply that

$$\begin{aligned}E^Q(I_{\{\tau_S < \tau_B\}} D(0,\tau_S) V_{\tau_S}^+ \mid \tau_S = t) &= E^Q(I_{\{t < \tau_B\}} D(0,t) V_t^+) \\ &= \bar{F}_B(t) D(0,t) E^Q(V_t^+).\end{aligned}$$

Hence $\text{CVA}^{\text{indep}}$, the credit value adjustment at $t = 0$ under the independence assumption, is given by

$$\text{CVA}^{\text{indep}} = \text{CVA}_0^{\text{indep}} = \delta^S \int_0^T \bar{F}_B(t) D(0,t) E^Q(V_t^+) f_S(t)\,\mathrm{d}t, \tag{17.11}$$

and, by a similar argument, the debt value adjustment under independence is

$$\text{DVA}^{\text{indep}} = \text{DVA}_0^{\text{indep}} = \delta^B \int_0^T \bar{F}_S(t) D(0,t) E^Q(V_t^-) f_B(t)\,\mathrm{d}t. \tag{17.12}$$

Note that formulas (17.11) and (17.12) are much easier to evaluate than the "correct" expressions (17.7) and (17.8). In particular, we only need to determine the marginal distribution of τ_S and τ_B and the so-called *expected exposures* $E^Q(V_t^+)$ and $E^Q(V_t^-)$. On the other hand, the assumption that (V_t), $(Y_{t,S})$ and $(Y_{t,B})$ are independent is often difficult to justify, and the simplified adjustments can be misleading. Consider, for instance, the case where S, R and B are financial institutions and suppose that $T_1 < T$ and $\xi_1 = S$. In that case it is quite likely that τ_S falls in a time period where financial institutions face adverse conditions so that the credit spread of the reference entity at τ_S and, hence, the market value V_{τ_S} of the CDS referencing R are comparatively high. We therefore expect that

$$E^Q(V_{\tau_S}^+ \mid \tau_S = t) > E^Q(V_t^+),$$

so that $\text{CVA} > \text{CVA}^{\text{indep}}$. Similarly, we expect that $E^Q(V_{\tau_B}^- \mid \tau_B = t) < E^Q(V_t^-)$, so that $\text{DVA} < \text{DVA}^{\text{indep}}$. The aggregate effect would be that

$$\text{BCVA} > \text{BCVA}^{\text{indep}}$$

in that case. Some numerical results that support this intuition are given in Section 17.4.4. The phenomenon whereby the conditional expected exposure given the default of the counterparty is higher than the unconditional expected exposure is a typical example of an unfavourable dependence between the size of an exposure and the credit quality of the counterparty. In counterparty risk management, such an unfavourable dependence is known as *wrong-way risk* (since the exposure to a counterparty and the credit quality of that party evolve in the "wrong way"). Wrong-way risk is an important issue in counterparty risk management (see, for example, Chapter 15 of Gregory (2012)).

Unilateral credit value adjustments. In a unilateral value adjustment each party neglects the possibility of its own default. The unilateral value adjustment for the protection buyer B is therefore obtained from the formula for the bilateral value adjustment by assuming that $Q(\xi_1 = B) = 0$. This gives

$$\text{UCVA}_t = E^Q(I_{\{\tau_S < T\}} D(t, \tau_S) \delta^S V_{\tau_S}^+ \mid \mathcal{G}_t).$$

An analogous formula holds for the unilateral value adjustment of the protection seller. Unilateral value adjustments avoid the problem that a worsening credit spread of a financial institution leads to an accounting profit. On the other hand, if B and S use unilateral adjustments, they might come to different conclusions about the value of a given deal.

Netting. A further issue that arises in practice is *netting*. Under a legally enforceable netting agreement the market value of all CDS transactions between B and S at T_1 is computed and only the aggregate value is subject to bankruptcy procedures. In particular, perfectly offsetting transactions cancel each other out. Netting can reduce counterparty risk substantially, so netting agreements are widely used in practice. On the other hand, netting substantially increases the computational complexity of CVA and DVA computations, as we now explain. Suppose that there are N transactions

between B and S that fall under a netting agreement and let these be indexed by $n \in \{1, \dots, N\}$. Denote by $(V_{t,n})$ the market value from the point of view of B of the nth transaction. A similar argument to the one used in the proof of Proposition 17.1 implies that

$$\mathrm{CVA}_t = E^Q\left(I_{\{T_1<T\}} I_{\{\xi_1=S\}} D(t, T_1)\delta^S\left(\sum_{n=1}^{N} V_{T_1,n} \right)^+ \,\middle|\, \mathcal{G}_t \right),$$

and a similar formula applies to the debt value adjustment. Hence in the presence of netting agreements the computation of value adjustments amounts to the pricing of an option on the sum of the market value of all transactions covered by the netting agreement. In the case of CDS contracts each would typically refer to a different reference entity, so we have to consider $n + 2$ different default times. This is in general a much more difficult problem than pricing the options individually.

17.2.2 Collateralized Value Adjustments for a CDS

In this section we introduce popular collateralization strategies and analyse qualitatively the impact of collateralization on credit value adjustments for a CDS. To keep things simple we assume that the collateral is posted in the form of cash and earns the risk-free rate of interest. Many collateralization arrangements used in practice are of this form, but arrangements where other securities are used as collateral can also be found.

Details of the collateralization arrangement for an OTC CDS transaction are fixed in the credit support annex of the transaction. Roughly speaking, the procedure works as follows. At $t_0 = 0$ a collateral account is opened. Let C_t denote the cash balance in the account at time t. Here $C_t > 0$ means that S has posted the collateral and that B is the collateral taker, whereas $C_t < 0$ means that B has posted the collateral and that S is the collateral taker. The collateral position is updated at discrete time points $t_1, \dots, t_N \leqslant T$. At t_1 the collateral taker pays interest on the collateral, and the cash balance C_{t_1} is adjusted in reaction to changes in the price of the underlying CDS over $(t_0, t_1]$. This procedure continues up to the maturity of the CDS or until the first default occurs. If $T_1 > T$ or if $T_1 < T$ and $\xi_1 = R$, the collateral account is closed at the "natural end" of the contract, so $C_t \equiv 0$ for $t \geqslant T_1 \wedge T$. If there is an early default of B or S—that is, if $T_1 \leqslant T$ and $\xi_1 \in \{B, S\}$—the collateral is used to reduce the loss of the collateral taker and any remaining collateral is returned.

An issue arising in this context is *rehypothecation*. The collateral taker typically has unrestricted access to the posted collateral; in particular, the funds can be used as collateral in other OTC derivative transactions. It is therefore possible to have a situation in which the collateral taker defaults and a part of the collateral that should be returned is missing. To keep things simple we ignore this issue and assume that, in the case of a default of the collateral taker, the collateral is always returned in full to the other party. We refer to Notes and Comments for references regarding credit value adjustments in the presence of rehypothecation.

Collateralization strategies. We describe the cash balance in the collateral account by a $(\mathcal{G}_t)$-adapted process (C_t), the so-called collateralization strategy. For convenience we allow for strategies where the collateral account is changed continuously and not just at predetermined rebalancing dates. Recall that the counterparty-risk-free CDS price is denoted by (V_t). In practice, most collateralization arrangements take the form of a threshold collateralization strategy. Formally, a *threshold collateralization strategy* with *thresholds* $M_1, M_2 \geqslant 0$, labelled $(C_t^{M_1,M_2})$ for $0 \leqslant t \leqslant T_1 \wedge T$, is given by

$$C_t^{\gamma,M_1,M_2} = (V_t^+ - M_1)I_{\{V_t^+ > M_1\}} - (V_t^- - M_2)I_{\{V_t^- > M_2\}}. \tag{17.13}$$

Under this strategy collateral is posted if V_t^+ (the exposure of B) exceeds the threshold M_1 or if V_t^- (the exposure of S) exceeds the threshold M_2. A threshold strategy is used if B and S want to protect themselves against severe counterparty-risk-related losses, while accepting the possibility of smaller losses in order to simplify the practical management of the collateralization process. For $M_1 = M_2 = 0$ we obtain the special case of *market-value collateralization* with

$$C_t^{\text{market}} = V_t, \quad 0 \leqslant t \leqslant T_1 \wedge T. \tag{17.14}$$

Collateralized value adjustment. Value adjustments for collateralized CDS contracts are largely analogous to the uncollateralized case, so we keep our presentation short. The bilateral collateralized value adjustment BCCVA is the difference between the collateralized credit value adjustment CCVA and the collateralized debt value adjustment CDVA. As before, the CCVA gives the value of the potential loss for B due to an early default of S, whereas the CDVA gives the value of the potential loss for S due to an early default of B.

In order to describe these potential losses we have to consider the payments at an early default. Note that no additional collateral is posted at or after the default of B or S. The amount of collateral available for the settlement of the contract is therefore given by C_{T_1-} (the amount of collateral that has been posted immediately prior to T_1). This distinction matters if the close-out amount (V_t) jumps at T_1, for instance due to contagion effects, or if there is some delay between the last adjustment of the collateral account and the settlement of the positions. We begin with the scenario where the protection seller defaults first. We have to distinguish the cases $V_{T_1} > 0$ and $V_{T_1} < 0$.

- Suppose that $V_{T_1} \geqslant 0$ and that the protection buyer is the collateral taker, that is $C_{T_1-} \geqslant 0$. In that case the collateral is used to reduce the loss of the protection buyer and excess collateral is returned. If C_{T_1-} is smaller than V_{T_1}, the protection buyer incurs a counterparty-risk-related loss of size $\delta^S(V_{T_1} - C_{T_1-})$; if $C_{T_1-} \geqslant V_{T_1}$, the amount of collateral is sufficient to protect B from losses due to counterparty risk. If S is the collateral taker, i.e. if $C_{T_1-} \leqslant 0$, there is no available collateral to protect B and he suffers a loss of size $\delta^S V_{T_1}$.

- Suppose that $V_{T_1} \leqslant 0$. In that case B has no exposure to S, so he does not suffer a loss related to counterparty risk (the fact that he has incurred a loss due to the decrease in the counterparty-risk-free CDS price is irrelevant for the computation of value adjustments for counterparty risk).

Summarizing, the counterparty-risk-related loss of B is given by

$$I_{\{T_1<T\}} I_{\{\xi_1=S\}} \delta^S (V_{T_1}^+ - C_{T_1-}^+)^+.$$

Similarly, S suffers a loss in the scenario where $V_{T_1} \leqslant 0$ and where there is insufficient collateral to settle the contract in full, that is, for $V_{T_1}^- > C_{T_1}^-$. The counterparty-risk-related loss of S is thus given by

$$I_{\{T_1<T\}} I_{\{\xi_1=B\}} \delta^B (V_{T_1}^- - C_{T_1-}^-)^+.$$

Thus BCCVA_t, the *bilateral collateralized credit value adjustment* at time t, is given by

$$\mathrm{BCCVA}_t = \mathrm{CCVA}_t - \mathrm{CDVA}_t, \tag{17.15}$$

where

$$\begin{aligned}
\mathrm{CCVA}_t &:= E(I_{\{T_1<T\}} I_{\{\xi_1=S\}} D(t,T_1) \delta^S (V_{T_1}^+ - C_{T_1-}^+)^+ \mid \mathcal{G}_t), \\
\mathrm{CDVA}_t &:= E(I_{\{T_1<T\}} I_{\{\xi_1=B\}} D(t,T_1) \delta^B (V_{T_1}^- - C_{T_1-}^-)^+ \mid \mathcal{G}_t).
\end{aligned}$$

Without collateralization, i.e. for $C_t \equiv 0$, formula (17.15) reduces to the simpler result of Proposition 17.1.

Performance of market-value collateralization. The sum $\mathrm{CCVA}_t + \mathrm{CDVA}_t$ gives the value in t of the entire counterparty-risk-related loss, and it can therefore be viewed as a measure of the performance of a given collateralization strategy. Here we make the following immediate observation: suppose that market-value collateralization with $C_t^{\mathrm{market}} = V_t$ is used and that the market value of the CDS does not jump at T_1, that is, $V_{T_1} = V_{T_1-}$ almost surely. In that case the formulas for CCVA_t and CDVA_t in (17.15) show that

$$\mathrm{CCVA}_t = \mathrm{CDVA}_t = 0, \quad t \leqslant T_1,$$

so that market-value collateralization works perfectly. If, on the other hand, $|\Delta V_{T_1}| = |V_{T_1} - V_{T_1-}|$ is comparatively large, the performance of market-value collateralization will be not so good. Some numerical results supporting this observation will be presented in Section 17.4.4.

Notes and Comments

The literature on counterparty risk management is growing rapidly, leading to a proliferation of valuation-adjustment acronyms (CVA, DVA, FVA and others). A detailed introduction can be found in the textbooks by Gregory (2012), Cesari et al. (2009) and Brigo, Morini and Pallavicini (2013). A non-technical introduction to the computation of value adjustments for counterparty risk is given in the papers by Hull and White (2012, 2013).

The derivation of the bilateral credit value adjustments in Propositions 17.1 and (17.15) is based on the papers by Brigo and Chourdakis (2009) and Brigo, Capponi and Pallavicini (2014) (see also Frey and Rösler 2014). The last two papers also consider the case of rehypothecation and discuss the actual computation of value adjustments in various portfolio credit risk models. Credit value adjustments in structural credit risk models are studied in Lipton and Sepp (2009). A very general technical analysis of value adjustments is given in Crépey (2012a,b).

17.3 Conditionally Independent Default Times

In this section we discuss models with conditionally independent default times. We begin with a discussion of general mathematical properties; applications and specific examples from the literature are considered in Sections 17.3.2 and 17.3.3.

17.3.1 Definition and Mathematical Properties

Throughout we consider a portfolio of m obligors with default times τ_i and default indicators $Y_{t,i} = I_{\{\tau_i \leqslant t\}}$, $1 \leqslant i \leqslant m$, on a probability space $(\Omega, \mathcal{F}, P)$. The ordered default times are denoted by $0 = T_0 < T_1 < \cdots < T_m$, and $\xi_n \in \{1, \ldots, m\}$ gives the identity of the firm defaulting at time T_n. We introduce the filtrations $(\mathcal{H}_t^i)$, $1 \leqslant i \leqslant m$, and $(\mathcal{H}_t)$ defined by

$$\mathcal{H}_t^i = \sigma(\{Y_{s,i} : s \leqslant t\}) \quad \text{and} \quad \mathcal{H}_t = \mathcal{H}_t^i \vee \cdots \vee \mathcal{H}_t^m. \tag{17.16}$$

$(\mathcal{H}_t^i)$ is the filtration generated by the default observation for obligor i alone; $(\mathcal{H}_t)$ is the filtration generated by default observations for all obligors. Often $(\mathcal{H}_t)$ is called the *default history* of the portfolio or the *internal filtration* generated by the default times $\tau_1, \ldots, \tau_m$. The definition of conditionally independent default times is a straightforward multivariate extension of the notion of doubly stochastic default times from Section 10.5.1. In particular, the distribution of the default times is affected by additional information on top of the default history $(\mathcal{H}_t)$. Formally, we represent this information by a filtration $(\mathcal{F}_t)$ on the underlying probability space. Typically, $(\mathcal{F}_t)$ is generated by some observable background process. The information available to investors is given by the filtration $(\mathcal{G}_t) = (\mathcal{F}_t) \vee (\mathcal{H}_t)$ (see also (10.46)).

Definition 17.2. The default times $\tau_1, \ldots, \tau_m$ are conditionally independent doubly stochastic random times if there are positive, $(\mathcal{F}_t)$-adapted processes $(\gamma_{t,i})$, $1 \leqslant i \leqslant m$, with $\Gamma_{t,i} = \int_0^t \gamma_{s,i}\, \mathrm{d}s$ strictly increasing and finite for every $t > 0$, such that

$$P(\tau_1 > t_1, \ldots, \tau_m > t_m \mid \mathcal{F}_\infty) = \prod_{i=1}^m \exp\left(-\int_0^{t_i} \gamma_{s,i}\, \mathrm{d}s\right). \tag{17.17}$$

Note that the definition implies that each of the τ_i is a doubly stochastic random time with conditional hazard process $(\gamma_{t,i})$ in the sense of Definition 10.10 and that the rvs $\tau_1, \ldots, \tau_m$ are conditionally independent given $\mathcal{F}_\infty$.

Construction and simulation via thresholds. Lemma 17.3 extends Lemma 10.11.

Lemma 17.3. *Let $(\gamma_{t,1}), \dots, (\gamma_{t,m})$ be positive, $(\mathcal{F}_t)$-adapted processes such that $\Gamma_{t,i} := \int_0^t \gamma_{s,i}\,\mathrm{d}s$ is strictly increasing and finite for any $t > 0$. Let $\boldsymbol{X} = (X_1, \dots, X_m)'$ be a vector of independent, standard exponentially distributed rvs independent of $\mathcal{F}_\infty$. Define*

$$\tau_i = \Gamma_i^{\leftarrow}(X_i) = \inf\{t \geqslant 0 \colon \Gamma_{t,i} \geqslant X_i\}.$$

Then $\tau_1, \dots, \tau_m$ are conditionally independent doubly stochastic random times with hazard processes $(\gamma_{t,i})$, $1 \leqslant i \leqslant m$.

Proof. By the definition of τ_i we have $\tau_i > t \iff X_i > \Gamma_{t,i}$. The rvs $\Gamma_{t_i,i}$ are now measurable with respect to $\mathcal{F}_\infty$, whereas the X_i are mutually independent, independent of $\mathcal{F}_\infty$ and standard exponentially distributed. We therefore infer that

$$\begin{aligned} P(\tau_1 > t_1, \dots, \tau_m > t_m \mid \mathcal{F}_\infty) &= P(X_1 > \Gamma_{t_1,1}, \dots, X_m > \Gamma_{t_m,m} \mid \mathcal{F}_\infty) \\ &= \prod_{i=1}^m P(X_i > \Gamma_{t_i,i} \mid \mathcal{F}_\infty) \\ &= \prod_{i=1}^m \mathrm{e}^{-\Gamma_{t_i,i}}, \end{aligned} \tag{17.18}$$

which shows that the τ_i satisfy the conditions of Definition 17.2. □

Lemma 17.3 is the basis for the following simulation algorithm.

Algorithm 17.4 (multivariate threshold simulation).

(1) Generate trajectories for the hazard processes $(\gamma_{t,i})$ for $i = 1, \dots, m$. The same techniques as in the univariate case can be used here.

(2) Generate a vector $\boldsymbol{X}$ of independent standard exponentially distributed rvs (the threshold vector) and set $\tau_i = \Gamma_i^{\leftarrow}(X_i)$, $1 \leqslant i \leqslant m$.

As in the univariate case (see Lemma 10.12), Lemma 17.3 has a converse.

Lemma 17.5. *Let $\tau_1, \dots, \tau_m$ be conditionally independent doubly stochastic random times with $(\mathcal{F}_t)$-conditional hazard processes $(\gamma_{t,i})$. Define a random vector $\boldsymbol{X}$ by setting $X_i = \Gamma_i(\tau_i)$, $1 \leqslant i \leqslant m$. Then $\boldsymbol{X}$ is a vector of independent, standard exponentially distributed rvs that is independent of $\mathcal{F}_\infty$, and $\tau_i = \Gamma_i^{\leftarrow}(X_i)$ almost surely.*

Proof. For $t_1, \dots, t_m \geqslant 0$ the conditional independence of the τ_i implies that

$$P(\Gamma_1(\tau_1) \leqslant t_1, \dots, \Gamma_m(\tau_m) \leqslant t_m \mid \mathcal{F}_\infty) = \prod_{i=1}^m P(\Gamma_i(\tau_i) \leqslant t_i \mid \mathcal{F}_\infty).$$

Moreover, similar reasoning to the univariate case implies that

$$P(\Gamma_i(\tau_i) \leqslant t_i \mid \mathcal{F}_\infty) = P(\tau_i \leqslant \Gamma_i^{\leftarrow}(t_i) \mid \mathcal{F}_\infty) = 1 - \mathrm{e}^{-t_i},$$

which proves that $\boldsymbol{X}$ has the claimed properties. □

Recursive default time simulation. We now describe a second recursive algorithm for simulating conditionally independent default times, which can be more efficient than multivariate threshold simulation. We need the following lemma, which gives properties of the first default time T_1.

Lemma 17.6. *Let $\tau_1, \dots, \tau_m$ be conditionally independent doubly stochastic random times with hazard processes $(\gamma_{t,1}), \dots, (\gamma_{t,m})$. Then T_1 is a doubly stochastic random time with $(\mathcal{F}_t)$-conditional hazard process $\bar{\gamma}_t := \sum_{i=1}^m \gamma_{t,i}$, $t \geqslant 0$.*

Proof. Using the conditional independence of the τ_i we infer that

$$P(T_1 > t \mid \mathcal{F}_\infty) = P(\tau_1 > t, \dots, \tau_m > t \mid \mathcal{F}_\infty) = \prod_{i=1}^m \exp\left(-\int_0^t \gamma_{s,i}\,\mathrm{d}s\right),$$

which is obviously equal to $\exp(-\int_0^t \bar{\gamma}_s\,\mathrm{d}s)$. As this expression is $\mathcal{F}_t$-measurable, the result follows. □

Next we compute the conditional probability of the event $\{\xi_1 = i\}$ given the time T_1 of the first default and full information about the background filtration.

Proposition 17.7. *Under the assumptions of Lemma 17.6 we have*

$$P(\xi_1 = i \mid \mathcal{F}_\infty \vee \sigma(T_1)) = \gamma_i(T_1)/\bar{\gamma}(T_1), \quad i \in \{1, \dots, m\}.$$

Proof. Conditional on $\mathcal{F}_\infty$ the τ_i are independent with deterministic hazard functions $\gamma_i(t)$, so it is sufficient to prove the proposition for independent random times with deterministic hazard functions. Fix some $t > 0$ and note that, since the random vector $(\tau_1, \dots, \tau_m)$ has a joint density, the probability of having more than one default in the interval $(t-h, t]$ is of order $o(h)$. Hence,

$$\begin{aligned} P(\{\xi_1 = i\} \cap \{T_1 \in (t-h, t]\}) & \\ &= P(\{\tau_i \in (t-h, t]\} \cap \{\tau_j > t \text{ for all } j \neq i\}) + o(h) \\ &= P(\tau_i \in (t-h, t]) \prod_{j \neq i} P(\tau_j > t) + o(h) \end{aligned}$$

by the independence of the τ_i. Since $P(\tau_i > t) = \exp(-\int_0^t \gamma_i(s)\,\mathrm{d}s)$, $1 \leqslant i \leqslant m$, the above expression equates to

$$\exp\left(-\int_0^{t-h} \gamma_i(s)\,\mathrm{d}s\right)\left(1 - \exp\left(-\int_{t-h}^t \gamma_i(s)\,\mathrm{d}s\right)\right) \times \prod_{j \neq i} \exp\left(-\int_0^t \gamma_j(s)\,\mathrm{d}s\right) + o(h).$$

Hence,

$$\lim_{h \to 0+} \frac{1}{h} P(\{\xi_1 = i\} \cap \{T_1 \in (t-h, t]\}) = \gamma_i(t) \exp\left(-\int_0^t \bar{\gamma}(s)\,\mathrm{d}s\right).$$

Moreover, by Lemma 17.6,

$$\lim_{h\to 0+} \frac{1}{h} P(T_1 \in (t-h, t]) = \bar{\gamma}(t) \exp\left(-\int_0^t \bar{\gamma}(s)\,\mathrm{d}s\right),$$

so the claim follows from the definition of elementary conditional probability. □

Algorithm 17.8 (recursive default time simulation). This algorithm simulates a realization of the sequence (T_n, ξ_n) up to some maturity date T. In order to formulate the algorithm we introduce the notation A_n for the set of non-defaulted firms immediately after T_n. Formally, we set

$$A_0 := \{1, \dots, m\}, \quad A_n := \{1 \leqslant i \leqslant m \colon Y_i(T_n) = 0\}, \quad n \geqslant 1. \tag{17.19}$$

Moreover we define, for $n \geqslant 0$,

$$\bar{\gamma}_t^{(n)} := \sum_{i \in A_n} \gamma_{t,i}, \quad 0 \leqslant n \leqslant m.$$

The algorithm proceeds in the following steps.

(1) Generate trajectories of the hazard processes $(\gamma_{t,i})$ up to time T and set $n = 0$, $T_0 = 0$ and $\xi_0 = 0$.

(2) Generate the waiting time $T_{n+1} - T_n$ by standard univariate threshold simulation using Algorithm 10.13. For this we use a generalization of Lemma 17.6 that states that for conditionally independent defaults and $n < m$ we have

$$P(T_{n+1} - T_n > t \mid (T_1, \xi_1), \dots, (T_n, \xi_n), \mathcal{F}_\infty) = \exp\left(-\int_{T_n}^{T_n+t} \bar{\gamma}_s^{(n)}\,\mathrm{d}s\right).$$

(3) If $T_{n+1} \geqslant T$, stop. Otherwise use Proposition 17.7 and determine ξ_{n+1} as a realization of a multinomial rv ξ with

$$P(\xi = i) = \frac{\gamma_i(T_{n+1})}{\bar{\gamma}^{(n)}(T_{n+1})}, \quad i \in A_n.$$

(4) If $n + 1 = m$, stop. Otherwise increase n by 1 and return to step (2).

Recursive default time simulation is particularly efficient if we only need to simulate defaults occurring before some maturity date T and if defaults are rare. In that case, $T_n > T$ for n relatively small, so only a few ordered default times need to be simulated. With multivariate threshold simulation, on the other hand, we need to simulate the default times of all firms in the portfolio.

Default intensities. We begin with a general definition of default intensities in reduced-form portfolio credit risk models.

Definition 17.9 (default intensity). Consider a generic filtration $(\mathcal{G}_t)$ such that the default indicator process $(Y_{t,i})$ is $(\mathcal{G}_t)$-adapted. Then a non-negative $(\mathcal{G}_t)$-adapted

right-continuous process $(\lambda_{t,i})$ with $\int_0^t \lambda_{s,i}\,\mathrm{d}s < \infty$ almost surely for all t is called the $(\mathcal{G}_t)$ default intensity of firm i if the process

$$M_{t,i} := Y_{t,i} - \int_0^{t\wedge\tau_i} \lambda_{s,i}\,\mathrm{d}s = Y_{t,i} - \int_0^t (1 - Y_{s,i})\lambda_{s,i}\,\mathrm{d}s$$

is a $(\mathcal{G}_t)$-martingale.

The filtration $(\mathcal{G}_t)$ is an integral part of the definition of $(\lambda_{t,i})$. If the choice of the underlying filtration is clear from the context, we will simply speak of the default intensity of firm i. It is well known from stochastic calculus that the compensator of $(Y_{t,i})$ (the continuous, $(\mathcal{G}_t)$-adapted process $(\Lambda_{t,i})$ such that $Y_{t,i} - \Lambda_{t,i}$ is a $(\mathcal{G}_t)$-martingale) is unique. Since we assumed that (λ_t) is right continuous, it follows that the default intensity $\lambda_{t,i}$ is uniquely defined for $t < \tau_i$.

In the next result we link Definition 17.9 to the informal interpretation of default intensities as instantaneous default probabilities.

Lemma 17.10. *Let $(\lambda_{t,i})$ be the $(\mathcal{G}_t)$ default intensity of firm i and suppose that the process $(\lambda_{t,i})$ is bounded. We then have the equality*

$$I_{\{\tau_i>t\}}\lambda_{t,i} = I_{\{\tau_i>t\}} \lim_{h\to 0^+} \frac{1}{h} P(\tau_i \leqslant t + h \mid \mathcal{G}_t).$$

Proof. Since the firm under consideration is fixed we will simply write τ, Y_t and λ_t for the default time, the default indicator and the default intensity. By the definition of Y_t we have

$$P(\tau \leqslant t + h \mid \mathcal{G}_t) = E(Y_{t+h} \mid \mathcal{G}_t) = Y_t + E(Y_{t+h} - Y_t \mid \mathcal{G}_t).$$

Since $Y_t - \int_0^t (1 - Y_s)\lambda_s\,\mathrm{d}s$ is a $(\mathcal{G}_t)$-martingale, it follows that $E(Y_{t+h} - Y_t \mid \mathcal{G}_t) = E(\int_t^{t+h}(1 - Y_s)\lambda_s\,\mathrm{d}s \mid \mathcal{G}_t)$, and therefore, since $I_{\{\tau>t\}}Y_t = 0$,

$$I_{\{\tau>t\}}P(\tau \leqslant t + h \mid \mathcal{G}_t) = I_{\{\tau>t\}}E\left(\int_t^{t+h} (1 - Y_s)\lambda_s\,\mathrm{d}s \;\middle|\; \mathcal{G}_t \right). \tag{17.20}$$

The right-continuity of the process $((1 - Y_s)\lambda_s)$ implies that

$$\lim_{h\to 0^+} \frac{1}{h} \int_t^{t+h} (1 - Y_s)\lambda_s\,\mathrm{d}s = (1 - Y_t)\lambda_t \quad \text{a.s.}$$

Recalling that (λ_t) is bounded by assumption, we can use (17.20) and the bounded convergence theorem for conditional expectations to deduce that

$$I_{\{\tau>t\}} \lim_{h\to 0^+} \frac{1}{h} P(\tau \leqslant t + h \mid \mathcal{G}_t) = I_{\{\tau>t\}}(1 - Y_t)\lambda_t = I_{\{\tau>t\}}\lambda_t,$$

which proves the claim. □

Remark 17.11. The converse statement is also true; if the derivative

$$\lim_{h\to 0^+} \frac{1}{h} P(\tau \leqslant t + h \mid \mathcal{G}_t)$$

exists almost surely, then τ admits an intensity (under some technical conditions); references are given in Notes and Comments.

Finally, we return to the special case of models with conditionally independent defaults as introduced in Definition 17.2. The following proposition shows that in this model class a default intensity of any given firm i is given by the conditional hazard process of the default time τ_i.

Proposition 17.12. *Let $\tau_1, \dots, \tau_m$ be conditionally independent doubly stochastic random times with $(\mathcal{F}_t)$-conditional hazard processes $(\gamma_{t,1}), \dots, (\gamma_{t,m})$. The processes*

$$M_{t,i} := Y_{t,i} - \int_0^{t\wedge\tau_i} \gamma_{s,i}\, \mathrm{d}s$$

are then $(\mathcal{G}_t)$-martingales.

Proof. The result follows immediately from Proposition 10.14 (which gives the compensator of a doubly stochastic default time in the single-firm case) if we can show that τ_i is a doubly stochastic default time with $(\mathcal{G}_t^{-i})$-conditional hazard process $(\gamma_{t,i})$, where $(\mathcal{G}_t^{-i})$ is the artificial background filtration defined by

$$\mathcal{G}_t^{-i} = \mathcal{F}_t \vee \mathcal{H}_t^1 \cdots \vee \mathcal{H}_t^{i-1} \vee \mathcal{H}_t^{i+1} \vee \cdots \vee \mathcal{H}_t^m, \quad t \geqslant 0. \tag{17.21}$$

According to Definition 10.10 we have to show that

$$P(\tau_i > t \mid \mathcal{G}_\infty^{-i}) = \exp\left(-\int_0^t \gamma_{s,i}\, \mathrm{d}s\right), \quad t \geqslant 0. \tag{17.22}$$

This relationship is quite intuitive; since the τ_i are independent given $\mathcal{F}_\infty$, the default history of obligor $j \neq i$ that is contained in $\mathcal{G}_\infty^{-i}$ but not in $\mathcal{F}_\infty$ has no impact on the conditional default probability of obligor i.

A formal argument is as follows. Using Lemma 17.5, we may assume that there is a vector $\boldsymbol{X}$ of independent, standard exponential rvs, independent of $\mathcal{F}_\infty$, such that for all $1 \leqslant j \leqslant m$ we have $\tau_j = \Gamma_j^{\leftarrow}(X_j)$. Obviously, τ_i is independent of X_j for $j \neq i$, so

$$P(\tau_i > t \mid \mathcal{F}_\infty \vee \sigma(\{X_j : j \neq i\}) = P(\tau_i > t \mid \mathcal{F}_\infty) = \exp\left(-\int_0^t \gamma_{s,i}\, \mathrm{d}s\right). \tag{17.23}$$

On the other hand, if we know X_j and the trajectory $(\gamma_{t,j})_{0\leqslant t\leqslant\infty}$, we can determine the trajectory $(Y_{t,j})_{0\leqslant t\leqslant\infty}$, so $\mathcal{G}_\infty^{-i} \subseteq \mathcal{F}_\infty \vee \sigma(\{X_j : j \neq i\})$ (in fact, the two σ-algebras are equal). Since the right-hand side of (17.23) is measurable with respect to $\mathcal{G}_\infty^{-i}$, equation (17.22) follows from (17.23). □

Remark 17.13 (redundant default information). Suppose that $\tau_1, \dots, \tau_m$ are conditionally independent doubly stochastic random times. Consider a financial product with maturity T whose discounted pay-off H depends only on the evolution of default-free security prices (which we assume to be $(\mathcal{F}_t)$-adapted) and on the default history of a subset $A = \{i_1, \dots, i_k\}$ of the firms in the portfolio and is therefore measurable with respect to the σ-algebra

$$\mathcal{G}_T^A := \mathcal{F}_T \vee \mathcal{H}_T^{i_1} \vee \cdots \vee \mathcal{H}_T^{i_k} \subset \mathcal{G}_T.$$

A typical example would be the default and the premium payment leg of a single-name CDS on a given firm i. In this context an argument similar to the one used to establish (17.22) above shows that

$$E(H \mid \mathcal{G}_t) = E(H \mid \mathcal{G}_t^A);$$

in particular, the default history of the firms that do not belong to A is redundant for computing the price of H. Such a relationship does not hold for more general portfolio credit risk models where the default times are not conditionally independent.

17.3.2 Examples and Applications

In many models from the financial literature with conditionally independent defaults, the hazard rates are modelled as linear combinations of independent affine diffusions, possibly with jumps. A typical model takes the form

$$\gamma_{t,i} = \gamma_{i0} + \sum_{j=1}^{p} \gamma_{ij} \Psi_{t,j}^{\text{syst}} + \Psi_{t,i}^{\text{id}}, \quad 1 \leqslant i \leqslant m, \tag{17.24}$$

where $(\Psi_{t,j}^{\text{syst}})$, $1 \leqslant j \leqslant p$, and $(\Psi_{t,i}^{\text{id}})$, $1 \leqslant i \leqslant m$, are independent CIR square-root diffusions as in (10.68) or, more generally, basic affine jump diffusions as in (10.75); the factor weights γ_{ij}, $0 \leqslant j \leqslant p$, are non-negative constants.

Writing $\boldsymbol{\Psi}_t^{\text{syst}} = (\Psi_{t,1}^{\text{syst}}, \ldots, \Psi_{t,p}^{\text{syst}})'$, it is obvious that $(\boldsymbol{\Psi}_t^{\text{syst}})$ represents common or systematic factors, while the $(\Psi_{t,i}^{\text{id}})$ processes represent idiosyncratic factors affecting only the hazard rate of obligor i. Note that the weight attached to the idiosyncratic factor can be incorporated into the parameters of the dynamics of $(\Psi_{t,i}^{\text{id}})$ and does not need to feature explicitly. Throughout this section we assume that the background filtration is generated by $(\boldsymbol{\Psi}_t^{\text{syst}})$ and $(\Psi_{t,i}^{\text{id}})$, $1 \leqslant i \leqslant m$. In practical applications of the model, the current value of these processes is derived from observed prices of defaultable bonds.

We now consider some examples proposed in the literature. Duffee (1999) has estimated a model of the form (17.24) with $p = 2$; in his model all factor processes are assumed to follow CIR square-root diffusions, so that their dynamics are characterized by the parameter triplet $(\kappa, \bar{\theta}, \sigma)$ (see equation (10.68)). In Duffee's model, $(\boldsymbol{\Psi}_t^{\text{syst}})$ represents factors driving the default-free short rate; the parameters of these processes are estimated from treasury data. The factor weights γ_{ij} and the parameters of (Ψ_t^{id}), on the other hand, are estimated from corporate bond prices.

In their influential case study on CDO pricing, Duffie and Gârleanu (2001) use basic affine jump diffusion processes of the form (10.75) to model the factors driving the hazard rates. Jumps in (γ_t) represent shocks that increase the default probability of a firm. They consider a homogeneous model with one systematic factor where $\gamma_{t,i} = \Psi_t^{\text{syst}} + \Psi_{t,i}^{\text{id}}$, $1 \leqslant i \leqslant m$, and they assume that the speed of mean reversion κ, the volatility σ and the mean jump size μ are identical for (Ψ_t^{syst}) and $(\Psi_{t,i}^{\text{id}})$. It is straightforward to show that this implies that the sum $\gamma_{t,i} = \Psi_t^{\text{syst}} + \Psi_{t,i}^{\text{id}}$ also follows a basic affine jump diffusion with parameters κ, $\bar{\theta}^{\text{syst}} + \bar{\theta}^{\text{id}}$, σ, $(l^0)^{\text{syst}} + (l^0)^{\text{id}}$ and μ; the parameters of $(\gamma_{t,i})$ used in Duffie and Gârleanu (2001) can be found in the row labelled "base case" in Table 17.1.

Table 17.1. Parameter sets for the Duffie–Gârleanu model used in Figure 17.1.

Parameter set	κ	$\bar{\theta}$	σ	l^0	μ
Pure diffusion	0.6	0.0505	0.141	0	0
Base case	0.6	0.02	0.141	0.2	0.1
High jump intensity	0.6	0.0018	0.141	0.32	0.1

Pricing single-name credit products. As discussed in Remark 17.13, in the framework of conditionally independent defaults the pricing formulas obtained in a single-firm model remain valid in the portfolio context. Moreover, when the hazard processes are specified as in (17.24), most computations can be reduced to one-dimensional problems involving affine processes, to which the results of Section 10.6 apply. As a simple example we consider the computation of the conditional survival probability of obligor i. By Remark 17.13 and Theorem 10.19 it follows that

$$P(\tau_i > T \mid \mathcal{G}_t) = P(\tau_i > T \mid \mathcal{G}_t^i) = I_{\{\tau_i > t\}} E\left(\exp\left(-\int_t^T \gamma_{s,i} \, \mathrm{d}s \right) \,\middle|\, \mathcal{F}_t \right).$$

For hazard processes of the form (17.24) this equals

$$I_{\{\tau_i > t\}} \mathrm{e}^{-\gamma_{i0}(T-t)} E\left(\exp\left(-\int_t^T \Psi_{s,i}^{\mathrm{id}} \, \mathrm{d}s \right) \,\middle|\, \mathcal{F}_t \right) \times \prod_{j=1}^{p} E\left(\exp\left(-\int_t^T \Psi_{s,j}^{\mathrm{syst}} \, \mathrm{d}s \right) \,\middle|\, \mathcal{F}_t \right). \quad (17.25)$$

Each of the conditional expectations in (17.25) can now be computed using the results for one-dimensional affine models from Section 10.6. More general models, where hazard rates are given by a general multivariate affine process (and not simply by a linear combination of independent one-dimensional affine processes) can be dealt with using the general affine-model technology developed by Duffie, Pan and Singleton (2000).

The implied one-period model and computation of the loss distribution. The conditional independence assumption and the factor structure (17.24) of the hazard processes have interesting implications for the distribution of the default indicators at a fixed time point.

For simplicity we suppose that there are no idiosyncratic factors $(\Psi_{t,i}^{\mathrm{id}})$ in the model. We fix $T > 0$ and consider the random vector $\boldsymbol{Y}_T = (Y_{T,1}, \ldots, Y_{T,m})'$. For $\boldsymbol{y} \in \{0, 1\}^m$ we can compute that

$$\begin{aligned} P(\boldsymbol{Y}_T = \boldsymbol{y}) &= E(P(\boldsymbol{Y}_T = \boldsymbol{y} \mid \mathcal{F}_\infty)) \\ &= E\Bigg(\prod_{j:\, y_j = 1} P(\tau_j \leqslant T \mid \mathcal{F}_\infty) \prod_{j:\, y_j = 0} P(\tau_j > T \mid \mathcal{F}_\infty) \Bigg). \end{aligned}$$

From (17.24) we also know that

$$P(\tau_i \leqslant T \mid \mathcal{F}_\infty) = 1 - \exp\left(-\gamma_{i0}T - \sum_{j=1}^{p} \gamma_{ij} \int_0^T \Psi_{s,j}^{\text{syst}}\, \mathrm{d}s\right). \tag{17.26}$$

This argument shows that $\boldsymbol{Y}_T$ follows a Bernoulli mixture model with p-factor structure as in Definition 11.5. The factor vector is given by

$$\boldsymbol{\Psi} := \left(\int_0^T \Psi_{s,1}^{\text{syst}}\, \mathrm{d}s, \ldots, \int_0^T \Psi_{s,p}^{\text{syst}}\, \mathrm{d}s\right)'$$

and the conditional default probabilities $p_i(\boldsymbol{\Psi})$ are given in (17.26).

For practical purposes, such as CDO pricing, we need to be able to evaluate the distribution of the portfolio loss $L_T = \sum_{i=1}^m \delta_i Y_{T,i}$ in this model. As explained in Section 11.2.1 it is useful to be able to compute the Laplace–Stieltjes transform $\hat{F}_{L_T}(\lambda) = E(\mathrm{e}^{-\lambda L_T})$ since the distribution of L_T can then be determined using inversion techniques for transforms.

In Section 11.2.5 we noted that it is quite common in practice to approximate Bernoulli mixture models with Poisson mixture models, which are often more tractable. We will derive the Laplace–Stieltjes transform for an approximating Poisson mixture model and see that it consists of terms that can computed using results for affine processes in Section 10.6.3. More precisely, we replace L_T by the rv $\tilde{L}_T := \sum_{i=1}^m \delta_i \tilde{Y}_{T,i}$, and we assume that, conditional on $\mathcal{F}_\infty$, the rvs $\tilde{Y}_{T,1}, \ldots, \tilde{Y}_{T,m}$ are independent, Poisson-distributed rvs with parameters

$$\Gamma_{T,i} = \gamma_{i0}T + \sum_{j=1}^{p} \gamma_{ij} \int_0^T \Psi_{s,j}^{\text{syst}}\, \mathrm{d}s, \quad 1 \leqslant i \leqslant m. \tag{17.27}$$

We also assume that the losses given default $\delta_1, \ldots, \delta_m$ are deterministic and we use the fact that the Laplace–Stieltjes transform of a generic Poisson rv $N \sim \mathrm{Poi}(\Gamma)$ is given by $\hat{F}_N(\lambda) = E(\mathrm{e}^{-\lambda N}) = \exp(\Gamma(\mathrm{e}^{-\lambda} - 1))$. Using the conditional independence of $\tilde{Y}_{T,1}, \ldots, \tilde{Y}_{T,m}$ it follows that

$$\begin{aligned} E(\mathrm{e}^{-\lambda \tilde{L}_T} \mid \mathcal{F}_\infty) &= \prod_{i=1}^m E(\mathrm{e}^{-\lambda\delta_i \tilde{Y}_{T,i}} \mid \mathcal{F}_\infty) = \prod_{i=1}^m \exp(\Gamma_{T,i}(\mathrm{e}^{-\lambda\delta_i} - 1)) \\ &= \exp\left(\sum_{i=1}^m (\mathrm{e}^{-\lambda\delta_i} - 1)\Gamma_{T,i}\right). \end{aligned}$$

Using the definition of $\Gamma_{T,i}$ in (17.27) and the independence of the systematic risk-factor processes $(\Psi_{t,1}^{\text{syst}}), \ldots, (\Psi_{t,p}^{\text{syst}})$, we obtain

$$\begin{aligned} \hat{F}_{\tilde{L}_T}(\lambda) &= E(E(\mathrm{e}^{-\lambda \tilde{L}_T} \mid \mathcal{F}_\infty)) \\ &= \exp\left\{\sum_{i=1}^m (\mathrm{e}^{-\lambda\delta_i} - 1)\gamma_{i0}T\right\} \\ &\quad \times \prod_{j=1}^p E\left(\exp\left\{\left(\sum_{i=1}^m (\mathrm{e}^{-\lambda\delta_i} - 1)\gamma_{ij}\right)\int_0^T \Psi_{s,j}^{\text{syst}}\, \mathrm{d}s\right\}\right). \end{aligned} \tag{17.28}$$

Expression (17.28) can be computed using the results for one-dimensional affine models in Section 10.6.3. For further refinements and references to Laplace inversion methods we refer to Di Graziano and Rogers (2009).

Default correlation. As we have seen in Chapter 10, default correlations (defined as the correlations $\rho(Y_{T,i}, Y_{T,j})$, $i \neq j$, between default indicators) are closely related to the heaviness of the tail of the credit loss distribution. In computing default correlations in models with conditionally independent defaults it is more convenient to work with the *survival indicator* $1 - Y_{T,i}$. By the definition of linear correlation we have

$$\begin{aligned}\rho(Y_{T,i}, Y_{T,j}) &= \rho(1 - Y_{T,i}, 1 - Y_{T,j}) \\ &= \frac{P(\tau_i > T, \tau_j > T) - \bar{F}_i(T)\bar{F}_j(T)}{(\bar{F}_i(T)(1 - \bar{F}_i(T)))^{1/2}(\bar{F}_j(T)(1 - \bar{F}_j(T)))^{1/2}}. \end{aligned} \tag{17.29}$$

For models with hazard rates as in (17.24), the computation of the survival probabilities $\bar{F}_i(T)$ has been discussed above. For the joint survival probability we obtain, using conditional independence,

$$\begin{aligned}P(\tau_i > T, \tau_j > T) &= E(P(\tau_i > T, \tau_j > T \mid \mathcal{F}_\infty)) \\ &= E(P(\tau_i > T \mid \mathcal{F}_\infty)P(\tau_j > T \mid \mathcal{F}_\infty)) \\ &= E\left(\exp\left(-\int_0^T (\gamma_{s,i} + \gamma_{s,j})\,\mathrm{d}s\right)\right). \end{aligned} \tag{17.30}$$

For hazard rates of the form (17.24), expression (17.30) can be decomposed—using a similar approach to the decomposition in (17.25)—into terms that can be evaluated using our results on one-dimensional affine models.

It is often claimed that the default correlation values that can be attained in models with conditionally independent defaults are too low compared with empirical default correlations (see, for example, Hull and White 2001; Schönbucher and Schubert 2001). Since default correlations do have a significant impact on the loss distribution generated by a model, we discuss this issue further. As a concrete example we use the Duffie–Gârleanu model and assume that there are no idiosyncratic factor processes (Ψ_t^{id}) and that all risk is systematic. As discussed above, in that case the default indicator vector $\boldsymbol{Y}_T$ follows an exchangeable Bernoulli mixture model with mixing variable $\tilde{Q}$ given by $1 - \exp(-\int_0^T \Psi_s^{\mathrm{syst}}\,\mathrm{d}s)$.

We have seen in Section 11.2.2 that in exchangeable Bernoulli mixture models every default correlation $\rho \in (0, 1)$ can be obtained by choosing the variance of the mixing variable sufficiently high. It follows that in the Duffie–Gârleanu model high levels of default correlation can be obtained if the variance of the rv $\Gamma_T := \int_0^T \Psi_s^{\mathrm{syst}}\,\mathrm{d}s$ is sufficiently high. A high variance of Γ_T can be obtained by choosing a high value for the volatility σ of the diffusion part of $(\boldsymbol{\Psi}_t^{\mathrm{syst}})$ or by choosing a high value for the mean of the jump-size distribution μ or for the jump intensity l^0. A high value for σ translates into very volatile day-to-day fluctuations of credit spreads, which might contradict the behaviour of real bond-price data. This

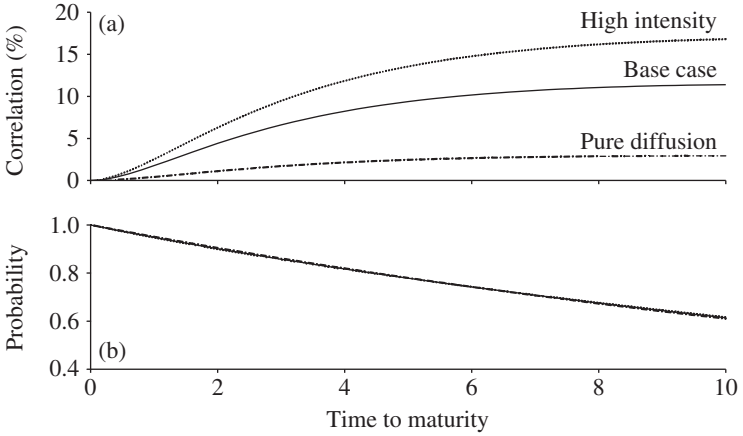

Figure 17.1. Both plots relate to the Duffie–Gârleanu model with $(\Psi_t^{\mathrm{id}}) \equiv 0$ and different parameter sets for (Ψ_t^{syst}), which are given in Table 17.1. (a) The default correlations for varying time to maturity. We see that by increasing the intensity of jumps in (Z_t), the default correlation is increased substantially. Part (b) shows that the survival probabilities for the three parameter sets are essentially equal, so that the differences in default correlations are solely due to the impact of the dynamics of (Ψ_t^{syst}) on the dependence structure of the default times.

shows that it might be difficult to generate very high levels of default correlation in models where hazard rates follow pure diffusion processes.

Instead, in the Duffie–Gârleanu model we can increase the frequency or size of the jumps in the hazard process by increasing l^0 or μ. This is a very effective mechanism for generating default correlation, as is shown in Figure 17.1. In fact, this additional flexibility in modelling default correlations is an important motivation for considering affine jump diffusions instead of the simpler CIR diffusion models.

These qualitative findings obviously carry over to other models with conditionally independent defaults. Summing up, we conclude that it is certainly possible to generate high levels of default correlation in models with conditionally independent defaults; however, the required models for the hazard processes can become quite complex.

17.3.3 Credit Value Adjustments

Finally, we turn to the computation of uncollateralized credit value adjustments for credit default swaps in models with conditionally independent default times. We first recall the notation of Section 17.2.1: τ_S, τ_B and τ_R are the default times of the protection seller, the protection buyer and the reference entity in a given CDS contract; τ_S, τ_B and τ_R are conditionally independent under the risk-neutral pricing measure Q with $(\mathcal{F}_t)$-conditional hazard processes γ^S, γ^B and γ^R; the short rate is given by some $(\mathcal{F}_t)$-adapted process (r_t); recovery rates are deterministic; $T_1 := \tau_S \wedge \tau_B \wedge \tau_R$ is the time of the first default; $D(0,t) = \exp(-\int_0^t r_u \, \mathrm{d}u)$.

As in Section 17.2.1 we denote by $V_t = E^Q(\Pi(t,T) \mid \mathcal{G}_t)$ the net present value of the promised cash flows of the protection-buyer position in the CDS. It was shown in Section 10.5.3 that $V_t = I_{\{\tau_R > t\}} \tilde{V}_t$, where the $(\mathcal{F}_t)$-adapted process $\tilde{V}_t$ is given

by

$$\tilde{V}_t = \delta^R E^Q\bigg(\int_t^T \gamma_s^R \exp\bigg(-\int_t^s (r_u + \gamma_u^R)\,\mathrm{d}u \bigg)\,\mathrm{d}s \;\bigg|\; \mathcal{F}_t \bigg) - x E^Q\bigg(\sum_{\{t_n > t\}} (t_n - t_{n-1}) \exp\bigg(-\int_t^s (r_u + \gamma_u^R)\,\mathrm{d}u \bigg) \;\bigg|\; \mathcal{F}_t \bigg), \quad (17.31)$$

and the time points t_n represent the premium payment dates.

A general formula. For simplicity we consider the case $t = 0$. According to Proposition 17.1, the bilateral uncollateralized value adjustment (for the protection buyer) is given by

$$\begin{aligned} \mathrm{BCVA} &= \mathrm{CVA} - \mathrm{DVA} \\ &= E^Q(I_{\{T_1<T\}} I_{\{\xi=S\}} D(0,T_1)\delta^S(\tilde{V}_{T_1})^+) \\ &\quad - E^Q(I_{\{T_1<T\}} I_{\{\xi=B\}} D(0,T_1)\delta^B(\tilde{V}_{T_1})^-). \end{aligned} \quad (17.32)$$

Here we have used the fact that $V_{T_1} = \tilde{V}_{T_1}$ on $\{\xi = S\}$ or $\{\xi = B\}$. We concentrate on the CVA term in (17.32). Recall from Proposition 17.7 that

$$Q(\xi = S \mid \sigma(T_1) \vee \mathcal{F}_\infty) = \frac{\gamma_{T_1}^S}{\gamma_{T_1}^S + \gamma_{T_1}^B + \gamma_{T_1}^R}.$$

By double conditioning we therefore obtain that

$$\begin{aligned} \mathrm{CVA} &= E^Q(E^Q(I_{\{T_1<T\}} I_{\{\xi=S\}} D(0,T_1)\delta^S(\tilde{V}_{T_1})^+ \mid \sigma(T_1) \vee \mathcal{F}_\infty)) \\ &= E^Q(I_{\{T_1<T\}} D(0,T_1)\delta^S(\tilde{V}_{T_1})^+ Q(\xi = S \mid \sigma(T_1) \vee \mathcal{F}_\infty)) \\ &= E^Q\bigg(I_{\{T_1<T\}} D(0,T_1)\delta^S(\tilde{V}_{T_1})^+ \frac{\gamma_{T_1}^S}{\gamma_{T_1}^S + \gamma_{T_1}^B + \gamma_{T_1}^R} \bigg). \end{aligned}$$

In the terminology of Section 10.5.2 this is a typical payment-at-default claim with payment at the stopping time T_1. By Lemma 17.6, T_1 is a doubly stochastic random time with hazard rate $\gamma_t^S + \gamma_t^B + \gamma_t^R$. Applying the third identity in Theorem 10.19 gives

$$\mathrm{CVA} = E^Q\bigg(\int_0^T \gamma_t^S \delta^S(\tilde{V}_t)^+ \exp\bigg(-\int_0^t (r_s + \gamma_s^S + \gamma_s^B + \gamma_s^R)\,\mathrm{d}s \bigg)\,\mathrm{d}t \bigg). \quad (17.33)$$

A similar computation for the DVA term in (17.32) yields

$$\mathrm{DVA} = E^Q\bigg(\int_0^T \gamma_t^B \delta^B(\tilde{V}_t)^- \exp\bigg(-\int_0^t (r_s + \gamma_s^S + \gamma_s^B + \gamma_s^R)\,\mathrm{d}s \bigg)\,\mathrm{d}t \bigg). \quad (17.34)$$

More explicit results. In order to evaluate (17.33) and (17.34) we need to make more specific assumptions about the form of the hazard rates. Here we derive a PDE for the value adjustments in a one-factor CIR model. More precisely, we assume that γ^S, γ^B, γ^R and the short rate r are affine functions of a one-dimensional CIR process Ψ with parameters κ, θ and σ. Therefore,

$$\gamma_t^R = \gamma^R(t, \Psi_t) = a^R \Psi_t + b^R \tag{17.35}$$

for $a^R > 0, b^R \geqslant 0$, and similarly for γ^B, γ^S and r. In that case, the pre-default value $\tilde{V}_t$ of the CDS is given by a function $\tilde{v}(t, \Psi)$ that can be easily computed using the affine-model techniques introduced in Section 10.6. Moreover, the Feynman–Kac formula (see Lemma 10.21) gives

$$E^Q\left(\int_t^T \gamma_s^S \delta^S \tilde{V}_s^+ \exp\left(-\int_t^T (\gamma_u^S + \gamma_u^B + \gamma_u^R + r_u)\, \mathrm{d}u \right) \,\middle|\, \mathcal{F}_t \right) = h^S(t, \Psi_t),$$

where the function h^S solves the PDE

$$h_t^S + (\kappa - \theta\psi) h_\psi^S + \tfrac{1}{2}\sigma^2 \psi h_{\psi\psi}^S + \gamma^S \delta^S \tilde{v}^+ = (r + \gamma^S + \gamma^B + \gamma^R) h^S. \tag{17.36}$$

Here the arguments (t, ψ) have been omitted for ease of notation. The corresponding PDE for the buyer adjustment is

$$h_t^B + (\kappa - \theta\psi) h_\psi^B + \tfrac{1}{2}\sigma^2 \psi h_{\psi\psi}^B + \gamma^B \delta^B \tilde{v}^- = (r + \gamma^S + \gamma^B + \gamma^R) h^B, \tag{17.37}$$

again with terminal value $h^B(T, \psi) = 0$. Both PDEs can be solved numerically.

Summarizing we obtain, in the special case of the one-factor CIR model, that

$$\mathrm{BCVA} = h^S(t, \Psi_t) - h^B(t, \Psi_t). \tag{17.38}$$

Notes and Comments

For an alternative textbook-level treatment of models with conditionally independent defaults, see, for example, Chapter 9 of Bielecki and Rutkowski (2002). The simulation of conditionally independent default times is discussed in Duffie and Singleton (2003). The existence of default intensities is discussed in Aven (1985) and in Section 2 of Janson, M'Baye and Protter (2011). Both sources contain results that can be viewed as partial converses to Lemma 17.10. A model with conditionally independent defaults where the factor Ψ follows a finite-state Markov chain is considered by Di Graziano and Rogers (2009).

Further empirical work on affine models for credit portfolios includes that of Duffee (1999) and Driessen (2005). We refer the reader to Chapter 10 of Filipović (2009) for further information on affine processes. The empirical estimation of models with conditionally independent defaults and various related issues are discussed in Duffie (2011).

17.4 Credit Risk Models with Incomplete Information

In this section we are concerned with reduced-form portfolio credit risk models under incomplete information. We consider models where the default times are independent given the realization of some common factor; in principle, this factor could be a stochastic process (Ψ_t), as in the models with conditionally independent defaults considered in Section 17.3, but for expository purposes we concentrate on the case where the factor is simply a one-dimensional random variable V. We assume that V is not directly observable by investors. Rather, their information set consists of the default history and—in the more advanced versions of the model—additional noisy observations of the factor represented by an auxiliary σ-algebra.

It will turn out that portfolio credit risk models with incomplete information have a number of attractive features. To begin with, they are able to generate rich dynamics of credit spreads incorporating both default contagion and credit-spread volatility effects. Moreover, the pricing of typical credit derivatives can be carried out using a natural and fairly efficient two-step procedure.

Our presentation starts with a general discussion of credit risk and incomplete information in Section 17.4.1. In Section 17.4.2 we consider simple models in which investors observe only the default history; the extension to a richer information set and applications to counterparty risk management are discussed in Sections 17.4.3 and 17.4.4.

17.4.1 Credit Risk and Incomplete Information

Throughout this section we use the following set-up. We work on a filtered probability space $(\Omega, \mathcal{F}, (\tilde{\mathcal{G}}_t), Q)$, where Q is the risk-neutral measure used for pricing derivatives and $(\tilde{\mathcal{G}}_t)$ is the global filtration, so all processes introduced will be $(\tilde{\mathcal{G}}_t)$-adapted. We consider a portfolio of m obligors with default times $\tau_1, \dots, \tau_m$ and default indicator processes $(Y_{t,i})$ given by $Y_{t,i} = I_{\{\tau_i \leqslant t\}}$, $1 \leqslant i \leqslant m$. We assume that there is a random variable V on $(\Omega, \mathcal{F}, Q)$ such that the rvs $\tau_1, \dots, \tau_m$ are conditionally independent given V, with conditional survival functions of the form

$$\bar{F}_{\tau_i|V}(t \mid v) = \exp\left(-\int_0^t \gamma_i(v, s)\, \mathrm{d}s\right) \tag{17.39}$$

for continuous functions $\gamma_i : \mathbb{R} \times [0, \infty) \to (0, \infty)$, such that $\int_0^t \gamma_i(v, s)\, \mathrm{d}s < \infty$ for all t and v. Note that this implies that $\gamma_i(v, t)$ is the conditional hazard function, and the conditional density $f_{\tau_i|V}(t \mid v)$ satisfies

$$f_{\tau_i|V}(t \mid v) = \gamma_i(v, t) \exp\left(-\int_0^t \gamma_i(v, s)\, \mathrm{d}s\right). \tag{17.40}$$

We consider two cases for the distribution of the factor V. In the first case V is an absolutely continuous rv with density $g_V(v)$, as in the factor copula models considered in Section 12.2.2. In the second case V is a discrete random variable with values in the set $S^V := \{v_1, \dots, v_K\}$ and probability mass function $\pi_k = Q(V = v_k)$, $1 \leqslant k \leqslant K$, as in the class of implied copula models considered in

Section 12.3.3. Finally, we always take the loss given default δ_i of the firms in the portfolio and the default-free interest rate $r(t)$ to be deterministic.

The full-information filtration is given by $\tilde{\mathcal{G}}_t = \mathcal{H}_t \vee \sigma(V) \vee \sigma(B_s : s \leqslant t)$, $t > 0$. Here, $(\mathcal{H}_t)$ is the default history of the portfolio and (B_t) is a Brownian motion independent of V and of the τ_i. (The process (B_t) is used in Section 17.4.3 as a building block in modelling the noisy information available to investors.) If considered under the full-information filtration $(\tilde{\mathcal{G}}_t)$, the model is therefore a model with conditionally independent defaults as discussed in Section 17.3. However, as mentioned before, the rv V is not observable by investors. We therefore describe the information available to investors by a filtration $(\mathcal{G}_t)$ with $\mathcal{G}_t \subset \tilde{\mathcal{G}}_t$, $t \geqslant 0$, and refer to $(\mathcal{G}_t)$ as the *investor filtration*. We assume throughout that investors are able to observe the default history of the portfolio, i.e. $\mathcal{H}_t \subset \mathcal{G}_t$ for all $t \geqslant 0$. On the other hand, V is not $\mathcal{G}_t$ measurable for all $t \geqslant 0$.

Pricing of credit derivatives. Consider a single-name CDS, a CDS index or a CDO tranche on the firms in the portfolio (or a subgroup thereof), and denote by $\Pi(t, T)$ the corresponding stream of discounted future cash flows. In the setting of this section, the price of this cash-flow stream at time t is given by $V_t = E^Q(\Pi(t, T) \mid \mathcal{G}_t)$. Note that V_t is given by a conditional expectation with respect to $\mathcal{G}_t$, the information actually available to investors at time t.

The price V_t can be computed in two steps. By iterated conditional expectations we obtain that

$$V_t = E^Q(E^Q(\Pi(t, T) \mid \tilde{\mathcal{G}}_t) \mid \mathcal{G}_t). \tag{17.41}$$

In the first step we consider the inner conditional expectation. In order to evaluate this expectation we need to determine the $\tilde{\mathcal{G}}_t$-conditional distribution of the non-defaulted firms. Denote by $\{i_1, \ldots, i_\ell\} \subseteq \{1, \ldots, m\}$ the identity of these firms. Since the Brownian motion (B_t) is independent of V and the default times τ_i, we find, for arbitrary $t_1, \ldots, t_\ell \geqslant t$, that

$$Q(\tau_{i_1} > t_1, \ldots, \tau_{i_\ell} > t_\ell \mid \tilde{\mathcal{G}}_t) = Q(\tau_{i_1} > t_1, \ldots, \tau_{i_\ell} > t_\ell \mid \mathcal{H}_t \vee \sigma(V)).$$

Since the τ_i are independent given V, it follows that

$$\begin{aligned} Q(\tau_{i_1} > t_1, \ldots, \tau_{i_\ell} > t_\ell \mid \mathcal{H}_t \vee \sigma(V)) &= \prod_{l=1}^{\ell} Q(\tau_{i_l} > t_l \mid \mathcal{H}_t^{i_l} \vee \sigma(V)) \\ &= \prod_{l=1}^{\ell} I_{\{\tau_{i_l} > t\}} \frac{Q(\tau_{i_l} > t_l \mid \sigma(V))}{Q(\tau_{i_l} > t \mid \sigma(V))} \\ &= \prod_{l=1}^{\ell} I_{\{\tau_{i_l} > t\}} \exp\left(- \int_t^{t_l} \gamma_{i_l}(V, s)\, \mathrm{d}s \right). \end{aligned}$$

The firms in the portfolio are therefore conditionally independent given $\tilde{\mathcal{G}}_t$, with conditional survival probabilities

$$Q(\tau_i > s \mid \tilde{\mathcal{G}}_t) = I_{\{\tau_i > t\}} \exp\left(- \int_t^s \gamma_i(V, s)\, \mathrm{d}s \right), \quad s \geqslant t. \tag{17.42}$$

As in Section 12.3.1 it follows that the value $E^Q(\Pi(t,T) \mid \tilde{\mathcal{G}}_t)$ (the price with respect to the larger σ-algebra $\tilde{\mathcal{G}}_t$) depends only on the time t, the realized value of the factor V and the realized value of the default indicator vector $\boldsymbol{Y}_t = (Y_{t,1}, \ldots, Y_{t,m})'$ at time t. In summary, this means that $E^Q(\Pi(t,T) \mid \tilde{\mathcal{G}}_t) = h(t, V, \boldsymbol{Y}_t)$ for a suitable function $h(\cdot)$. For instance, in the case of a protection-buyer position in a single-name CDS on firm i with premium payment dates $t_1, \ldots, t_N > t$ and spread x, we have that

$$h(t,v,\boldsymbol{y}) = (1-y_i)\bigg(\delta_i \int_t^{t_N} \gamma_i(v,s) \exp\bigg(-\int_t^s r(u) + \gamma_i(v,u)\,\mathrm{d}u\bigg)\,\mathrm{d}s \\ - x\sum_{n=1}^{N}(t_n - t_{n-1}) \exp\bigg(-\int_t^{t_n} r(s) + \gamma_i(v,s)\,\mathrm{d}s\bigg)\bigg).$$

For CDO tranches, the function $h(\cdot)$ can be computed using the analytics for factor credit risk models developed in Section 12.3.1.

Using (17.41) we infer that the price of the cash-flow stream at time t is given by

$$V_t = E^Q(h(t,V,\boldsymbol{Y}_t) \mid \mathcal{G}_t) = \int_{\mathbb{R}} h(t,v,\boldsymbol{Y}_t) g_{V|\mathcal{G}_t}(v)\,\mathrm{d}v, \tag{17.43}$$

where $g_{V|\mathcal{G}_t}(v)$ represents the conditional density of V given $\mathcal{G}_t$. If V is discrete, the last integral is replaced by the sum

$$\sum_{k=1}^{K} h(t, v_k, \boldsymbol{Y}_t) Q(V = v_k \mid \mathcal{G}_t).$$

The pricing formulas therefore have a similar structure to those in the static factor models considered in Section 12.3, but the unconditional density $g_V(v)$ and the unconditional probability mass function $\boldsymbol{\pi}$ have to be replaced by the conditional density $g_{V|\mathcal{G}_t}(v)$ and the conditional probabilities $\pi_t^k := Q(V = v_k \mid \mathcal{G}_t)$, $1 \leqslant k \leqslant K$. The computation of these quantities is discussed below.

Default intensities. A similar two-step procedure applies to the computation of default intensities. Since defaults are conditionally independent given V, we can apply Proposition 17.12 to infer that the default intensity of firm i with respect to the global filtration $(\tilde{\mathcal{G}}_t)$ satisfies $\tilde{\lambda}_{t,i} = \gamma_i(V,t)$. In order to compute the intensity of τ_i with respect to the investor filtration $(\mathcal{G}_t)$, we use the following general result.

Lemma 17.14. *Consider two filtrations $(\mathcal{G}_t)$ and $(\tilde{\mathcal{G}}_t)$ such that $\mathcal{G}_t \subset \tilde{G}_t$ for all $t \geqslant 0$ and some random time τ such that the corresponding indicator process $Y_t = I_{\{\tau \leqslant t\}}$ is $(\mathcal{G}_t)$-adapted. Suppose that τ admits the $(\tilde{G}_t)$-intensity $\tilde{\lambda}_t$. Then the $(\mathcal{G}_t)$ default intensity is given by the right-continuous version of the process $\lambda_t = E(\tilde{\lambda}_t \mid \mathcal{G}_t)$.*

Lemma 17.14 is a special case of Theorem 14 in Chapter 2 of Brémaud (1981) and we refer to that source for a proof. Applying Lemma 17.14 we conclude that $(\lambda_{t,i})$, the $(\mathcal{G}_t)$ default intensity of firm i, is given by

$$\lambda_{t,i} = E(\gamma_i(V,t) \mid \mathcal{G}_t) = \int_{\mathbb{R}} \gamma_i(v,t) g_{V|\mathcal{G}_t}(v)\,\mathrm{d}v; \tag{17.44}$$

in the case where V is discrete, the last integral is replaced by the sum

$$\sum_{k=1}^{K} \gamma_i(v_k, t) Q(V = v_k \mid \mathcal{G}_t).$$

17.4.2 Pure Default Information

In this section we consider the case where the only information available to investors is the default history of the portfolio. This is modelled by setting $(\mathcal{G}_t) = (\mathcal{H}_t)$.

Recursive computation of $g_{V|\mathcal{H}_t}$. We concentrate on the case where V has a density; the discrete case can be handled by analogous arguments. We will compute the conditional density $g_{V|\mathcal{H}_t}$ on the sets $\{t < T_1\}$ (before the first default) and $\{T_1 \leqslant t < T_2\}$ (after the first default and before the second).

Let $A \subseteq \mathbb{R}$ be a measurable set. We first note that on the set $\{t < T_1\}$ we have $Q(V \in A \mid \mathcal{H}_t) = Q(V \in A \mid T_1 > t)$. Now

$$Q(V \in A \mid T_1 > t) = \frac{Q(\{V \in A\} \cap \{T_1 > t\})}{Q(T_1 > t)} \tag{17.45}$$

and, using the conditional independence of defaults given V, the numerator may be calculated to be

$$\begin{aligned} Q(\{V \in A\} \cap \{T_1 > t\}) &= \int_A Q(T_1 > t \mid V = v) g_V(v)\, \mathrm{d}v \\ &= \int_A \prod_{i=1}^{m} Q(\tau_i > t \mid V = v) g_V(v)\, \mathrm{d}v \\ &= \int_A \exp\left(-\int_0^t \sum_{i=1}^{m} \gamma_i(v, s)\, \mathrm{d}s\right) g_V(v)\, \mathrm{d}v. \end{aligned}$$

It follows from (17.45) that on $\{t < T_1\}$ the conditional df $G_{V|\mathcal{H}_t}(v)$ is given by

$$G_V(v \mid \mathcal{H}_t) = \frac{1}{Q(T_1 > t)} \int_{-\infty}^{v} \exp\left(-\int_0^t \sum_{i=1}^{m} \gamma_i(w, s)\, \mathrm{d}s\right) g_V(w)\, \mathrm{d}w,$$

and hence the conditional density $g_{V|\mathcal{H}_t}(v)$ satisfies

$$g_{V|\mathcal{H}_t}(v) \propto \exp\left(-\int_0^t \sum_{i=1}^{m} \gamma_i(v, s)\, \mathrm{d}s\right) g_V(v),$$

where $\propto$ stands for "is proportional to". The constant of proportionality is of course given by $(Q(T_1 > t))^{-1}$, but this quantity is irrelevant for our subsequent analysis. To simplify notation we define the quantity

$$\bar{\gamma}(v, s) = \sum_{i=1}^{m} (1 - Y_{s,i}) \gamma_i(v, s), \tag{17.46}$$

and with this notation we have that on $\{t < T_1\}$

$$g_{V|\mathcal{H}_t}(v) \propto \exp\left(-\int_0^t \bar{\gamma}(v, s)\, \mathrm{d}s\right) g_V(v). \tag{17.47}$$

Next we calculate $g_{V|\mathcal{H}_t}(v)$ on the set $\{T_1 \leqslant t < T_2\}$. As an intermediate step we derive $g_{V|\tau_j}(v \mid t)$, the conditional density of V given the default time τ_j of an arbitrary firm j in the portfolio. Applying the conditional density formula (see, for instance, (6.2)) to the conditional density $f_{\tau_j|V}(t \mid v)$ given in (17.40), we obtain

$$g_{V|\tau_j}(v \mid t) \propto f_{\tau_j|V}(t \mid v) g_V(v) = \gamma_j(v,t) \exp\left(-\int_0^t \gamma_j(v,s)\,\mathrm{d}s\right) g_V(v). \tag{17.48}$$

The available information at t consists of the rvs T_1 and ξ_1 (the default time and identity ξ_1 of the firm defaulting first) and the event $B_t := \{\tau_i > t,\ i \neq \xi_1\}$ (the knowledge that the other firms have not yet defaulted at t). We can therefore write $Q(V \in A \mid \mathcal{H}_t) = Q(V \in A \mid B_t, T_1, \xi_1)$ on $\{T_1 \leqslant t < T_2\}$ and we note that

$$Q(V \in A \mid B_t, T_1, \xi_1) = \frac{Q(\{V \in A\} \cap B_t \mid T_1, \xi_1)}{Q(B_t \mid T_1, \xi_1)} \propto Q(\{V \in A\} \cap B_t \mid T_1, \xi_1).$$

Now we calculate that

$$Q(\{V \in A\} \cap B_t \mid T_1, \xi_1) = \int_A \prod_{i \neq \xi_1} \exp\left(-\int_0^t \gamma_i(v,s)\,\mathrm{d}s\right) g_{V|\tau_{\xi_1}}(v \mid T_1)\,\mathrm{d}v,$$

and we can use (17.48) to infer that this probability is proportional to

$$\int_A \gamma_{\xi_1}(v, T_1) \exp\left(-\int_0^{T_1} \sum_{i=1}^m \gamma(v,s)\,\mathrm{d}s\right) \exp\left(-\int_{T_1}^t \sum_{i \neq \xi_1} \gamma(v,s)\,\mathrm{d}s\right) g_V(v)\,\mathrm{d}v.$$

Using the notation (17.46) we find that on $\{T_1 \leqslant t < T_2\}$

$$g_{V|\mathcal{H}_t}(v) \propto \gamma_{\xi_1}(v, T_1) \exp\left(-\int_0^t \bar{\gamma}(v,s)\,\mathrm{d}s\right) g_V(v). \tag{17.49}$$

If we compare (17.47) and (17.49), we can see the impact of the additional default information at T_1 on the conditional distribution of V. At T_1 the "a posteriori density" $g_{V|\mathcal{H}_{T_1}}(v)$ is proportional to the product of the hazard rate $\gamma_{\xi_1}(v, T_1)$ and the "a priori density" $g_{V|\mathcal{H}_t}(v)$ just before T_1. In the following two examples this result will be used to derive explicit expressions for information-driven contagion effects.

The above analysis is a relatively simple example of a *stochastic filtering problem*. In general, in a filtering problem, we consider a stochastic process (Ψ_t) (a random variable is a special case) and a filtration $(\mathcal{G}_t)$ such that (Ψ_t) is not $(\mathcal{G}_t)$-adapted. We attempt to estimate the conditional distribution of Ψ_t for $t \geqslant 0$ given the σ-algebra $\mathcal{G}_t$ in a recursive way.

Example 17.15 (Clayton copula model). For factor copula models with a Clayton survival copula, the conditional density $g_{V|\mathcal{H}_t}(v)$ and the default intensities $(\lambda_{t,i})$ can be computed explicitly. The Clayton copula model with parameter $\theta > 0$ is an LT-Archimedean copula model (see Example 12.5) with $V \sim \mathrm{Ga}(1/\theta, 1)$. We denote the Laplace–Stieltjes transform of the distribution of V and its functional inverse by $\hat{G}_V$ and $\hat{G}_V^{-1}$, respectively, and we recall that the density $g(v; \alpha, \beta)$ of the $\mathrm{Ga}(\alpha, \beta)$ distribution satisfies $g(v; \alpha, \beta) \propto v^{\alpha-1}\mathrm{e}^{-\beta v}$ and has mean α/β (see Appendix A.2.4).

As shown in Example 12.5, in threshold copula models with an LT-Archimedean copula the conditional survival function is given by

$$\bar{F}_{\tau_i|V}(t \mid v) = \exp(-v\hat{G}_V^{-1}(\bar{F}_i(t))),$$

where $\bar{F}_i(t)$ denotes the marginal survival function. The conditional hazard function $\gamma_i(v,t)$ is therefore given by

$$\gamma_i(v,t) = -\frac{\mathrm{d}}{\mathrm{d}t}\ln \bar{F}_{\tau_i|V}(t \mid v) = -v\frac{\mathrm{d}}{\mathrm{d}t}\hat{G}_V^{-1}(\bar{F}_i(t)). \tag{17.50}$$

Since $Q(T_1 > t \mid V = v) = \exp(-v\sum_{i=1}^m \hat{G}_V^{-1}(\bar{F}_i(t)))$, we can use (17.47) to obtain

$$g_{V|\mathcal{H}_t}(v) \propto v^{(1/\theta)-1}\exp\left(-v\left(1+\sum_{i=1}^m \hat{G}_V^{-1}(\bar{F}_i(t))\right)\right), \quad t < T_1. \tag{17.51}$$

Hence, for $t < T_1$, the conditional distribution of V given $\mathcal{H}_t$ is again a gamma distribution but now with parameters $\alpha = 1/\theta$ and $\beta = 1 + \sum_{i=1}^m \hat{G}_V^{-1}(\bar{F}_i(t))$. Since the mean of this distribution is α/β, we see that the conditional mean of V given $T_1 > t$ is lower than the unconditional mean of V. This is in line with economic intuition; indeed, the fact that $T_1 > t$ is "good news" for the portfolio.

Next we turn to the updating at $t = T_1$. Using (17.49), (17.50) and (17.51) we find that

$$g_{V|\mathcal{H}_{T_1}} \propto v\left\{v^{(1/\theta)-1}\exp\left(-v\left(1+\sum_{i=1}^m \hat{G}_V^{-1}(\bar{F}_i(T_1))\right)\right)\right\}, \tag{17.52}$$

so that given $\mathcal{H}_{T_1}$, V follows a gamma distribution with parameters $\alpha = 1 + 1/\theta$ and $\beta = 1 + \sum_{i=1}^m \hat{G}_V^{-1}(\bar{F}_i(T_1))$. At T_1 the conditional mean $\mu_{V|\mathcal{H}_t}$ of V jumps upwards. We have the formulas

$$\lim_{t\to T_1^-}\mu_{V|\mathcal{H}_t} = \frac{1/\theta}{1+\sum_{j=1}^m \hat{G}_V^{-1}(\bar{F}_j(T_1))}, \quad \mu_{V|\mathcal{H}_{T_1}} = \frac{1+1/\theta}{1+\sum_{j=1}^m \hat{G}_V^{-1}(\bar{F}_j(T_1))}.$$

Readers familiar with Bayesian statistics may note that the explicit form of the updating rules is due to the fact that the gamma family is a conjugate family for the exponential distribution.

Finally, we consider the $(\mathcal{H}_t)$ default intensity process $(\lambda_{t,i})$. Since $\gamma_i(v,t) \propto v$, we use (17.44) to infer that $\lambda_{t,i}$ is proportional to $E(V \mid \mathcal{H}_t)$, the conditional mean of V given $\mathcal{H}_t$. In particular, $\lambda_{t,i}$ jumps upwards at each successive default time and decreases gradually between defaults. This is illustrated in Figure 17.2, where we consider an explicit example in which, for all i, the marginal survival functions are given by $\bar{F}_i(t) = \mathrm{e}^{-\gamma t}$ for a fixed default rate γ. The parameter θ is chosen to obtain a desired level of default correlation.

Example 17.16 (implied copula model). In this example we take a closer look at the case where V is discrete with state space $S^V = \{v_1, \ldots, v_K\}$ and probability mass function $\boldsymbol{\pi}$. In this way we obtain a dynamic version of the implied copula model from Section 12.3.3.

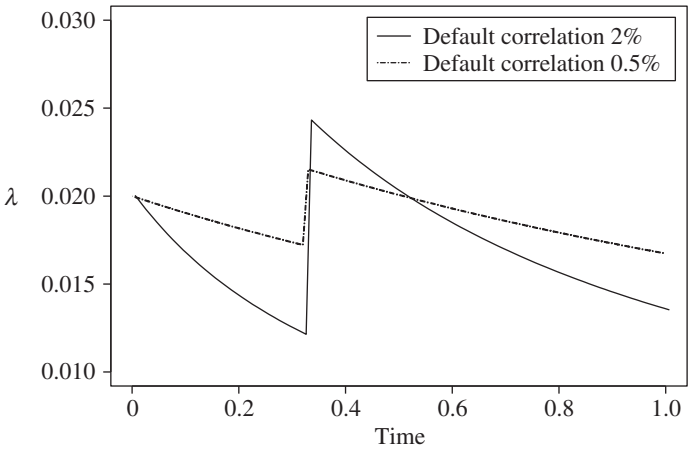

Figure 17.2. Paths of the default intensity (λ_t) in the Clayton copula model, assuming that the first default time T_1 is four months. The parameters are as follows: portfolio size $m = 100$; marginal default rate $\gamma = 0.02$; θ is chosen to give one-year default correlations of 2% and 0.5%. As we expect, a higher default correlation implies a stronger contagion effect.

Using similar arguments to those used to derive (17.47) and (17.49) gives, for $t < T_1$, that

$$Q(V = v_k \mid \mathcal{H}_t) = \frac{\exp(-\int_0^t \bar{\gamma}(v_k, s)\,\mathrm{d}s)\pi_k}{\sum_{l=1}^K \exp(-\int_0^t \bar{\gamma}(v_l, s)\,\mathrm{d}s)\pi_l}, \quad t < T_1; \tag{17.53}$$

at T_1 we have

$$Q(V = v_k \mid \mathcal{H}_{T_1}) \propto \gamma_{\xi_1}(v_k, T_1) \exp\left(-\int_0^{T_1} \bar{\gamma}(v_k, s)\,\mathrm{d}s\right)\pi_k. \tag{17.54}$$

Set $\pi_t^k := Q(V = v_k \mid \mathcal{H}_t)$, $1 \leqslant k \leqslant K$. We now use relations (17.53) and (17.54) to derive a system of stochastic differential equations for the process $\boldsymbol{\pi}_t = (\pi_t^1, \ldots, \pi_t^K)'$, which is driven by time and the default indicator process $(\boldsymbol{Y}_t)$. This helps to understand the dynamics of the default intensities and of credit derivatives.

First we consider the dynamics of the process $(\boldsymbol{\pi}_t)$ between default times. Define $E_t^k := \exp(-\int_0^t \bar{\gamma}(v_k, s)\,\mathrm{d}s)$. With this abbreviation we get, for $t < T_1$, the relationship $\pi_t^k = (\pi_k E_t^k)/(\sum_{l=1}^K \pi_l E_t^l)$, and hence

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\pi_t^k &= \frac{-\pi_k \bar{\gamma}(v_k, t) E_t^k \big(\sum_{l=1}^K \pi_l E_t^l\big) + \pi_k E_t^k \big(\sum_{l=1}^K \pi_l \bar{\gamma}(v_l, t) E_t^l\big)}{(\sum_{l=1}^K \pi_l E_t^l)^2} \\
&= \pi_t^k \left(-\bar{\gamma}(v_k, t) + \frac{\sum_{l=1}^K \pi_l \bar{\gamma}(v_l, t) E_t^l}{\sum_{l=1}^K \pi_l E_t^l}\right) \\
&= \pi_t^k \left(-\bar{\gamma}(v_k, t) + \sum_{l=1}^K \pi_t^l \bar{\gamma}(v_l, t)\right), \quad 1 \leqslant k \leqslant K.
\end{aligned} \tag{17.55}$$

Note that (17.55) is a system of K ordinary differential equations for $(\boldsymbol{\pi}_t)$, so the process $(\boldsymbol{\pi}_t)$ evolves in a deterministic manner up to time T_1. Next we consider the

case $t = T_1$. We will use the general notation $\pi^k_{t-} := \lim_{t \to t^-} \pi^k_t$. From (17.54) we get that

$$\pi^k_{T_1} = \frac{\gamma_{\xi_1}(v_k, T_1) E^k_{T_1} \pi_k}{\sum_{l=1}^K \gamma_{\xi_1}(v_l, T_1) E^l_{T_1} \pi_l} = \frac{\gamma_{\xi_1}(v_k, T_1) \pi^k_{T_1^-}}{\sum_{l=1}^K \gamma_{\xi_1}(v_l, T_1) \pi^l_{T_1^-}}, \tag{17.56}$$

where the second equality follows since $\pi^k_{T_1^-} = \pi_k E^k_{T_1} / (\sum_l \pi_l E^l_{T_1})$. Observe that at $t = T_1$ the process (π^k_t) has a jump of size

$$\pi^k_{t-} \left(\frac{\gamma_{\xi_1}(v_k, t) \pi^k_{t-}}{\sum_{l=1}^K \gamma_{\xi_1}(v_l, t) \pi^l_{t-}} - 1 \right).$$

Combining (17.55) and (17.56) we therefore obtain the following K-dimensional system of stochastic differential equations for the dynamics of $(\boldsymbol{\pi}_t)$ for all t:

$$\begin{aligned} \mathrm{d}\pi^k_t = {} & \pi^k_{t-} \left(-\bar{\gamma}(v_k, t) + \sum_{l=1}^K \pi^l_{t-} \bar{\gamma}(v_l, t) \right) \mathrm{d}t \\ & + \sum_{i=1}^m \pi^k_{t-} \left(\frac{\gamma_i(v_k, t) \pi^k_{t-}}{\sum_{l=1}^K \gamma_i(v_l, t) \pi^l_{t-}} - 1 \right) \mathrm{d}Y_{t,i}, \quad 1 \leqslant k \leqslant K. \end{aligned} \tag{17.57}$$

Recall that $\lambda_{t,i} = \sum_{k=1}^K \pi^k_t \gamma_i(v_k, t)$ is the $(\mathcal{H}_t)$ default intensity of firm i. Define the compensated default indicator processes by

$$M_{t,i} = Y_{t,i} - \int_0^t (1 - Y_{s-,i}) \lambda_{s-,i} \, \mathrm{d}s = Y_{t,i} - \int_0^t (1 - Y_{s-,i}) \sum_{k=1}^K \pi^k_{s-} \gamma_i(v_k, s) \, \mathrm{d}s. \tag{17.58}$$

Next we show that the system (17.57) can be simplified by using the processes $(M_{t,1}), \ldots, (M_{t,m})$ as drivers. This representation will be useful to see the link between our analysis and the more general filtering model considered in Section 17.4.3. To this end, note that the $\mathrm{d}t$ term in (17.57) is equal to

$$\pi^k_{t-} \sum_{i=1}^m \left((1 - Y_{t,i}) \left(\frac{\gamma_i(v_k, t)}{\sum_{l=1}^K \pi^l_{t-} \gamma_i(v_l, t)} - 1 \right) \left(- \sum_{l=1}^K \pi^l_{t-} \gamma_i(v_l, t) \right) \right).$$

Using the compensated default indicator processes as drivers, the dynamics of $(\boldsymbol{\pi}_t)$ can therefore be rewritten in the following simpler form:

$$\mathrm{d}\pi^k_t = \sum_{i=1}^m \pi^k_{t-} \left(\frac{\gamma_i(v_k, t)}{\sum_{l=1}^K \gamma_i(v_l, t) \pi^l_{t-}} - 1 \right) \mathrm{d}M_{t,i}. \tag{17.59}$$

Default intensities and contagion. Equation (17.59) determines the dynamics of the $(\mathcal{H}_t)$ default intensities. As in the Clayton copula model, $(\lambda_{t,i})$ evolves deterministically between default events and jumps at the random default times in the portfolio. It is possible to give an explicit expression for the size of this jump and hence for the contagion effects induced by incomplete information, as we now show.

Consider two firms $i \neq j$. It follows from (17.59) that the jump in the default intensity of firm i at the default time τ_j of firm j is given by

$$\begin{aligned}\lambda_{\tau_j,i} - \lambda_{\tau_j-,i} &= \sum_{k=1}^{K} \gamma_i(v_k, \tau_j)\pi^k_{\tau_j-}\left(\frac{\gamma_j(v_k, \tau_j)}{\sum_{l=1}^{K} \gamma_j(v_l, \tau_j)\pi^l_{\tau_j-}} - 1\right) \\ &= \frac{\mathrm{cov}^{\boldsymbol{\pi}_{\tau_j-}}(\gamma_i, \gamma_j)}{\lambda_{\tau_j-,j}}. \end{aligned} \tag{17.60}$$

Here, $\mathrm{cov}^{\boldsymbol{\pi}}$ denotes the covariance with respect to the probability measure $\boldsymbol{\pi}$ on S^V, and γ_i is shorthand for the random variable $v \mapsto \gamma_i(v, \tau_j)$.

Formula (17.60) makes two very intuitive predictions about default contagion. First, all other things being equal, default contagion increases with increasing correlation of the random variables γ_i and γ_j under $\boldsymbol{\pi}_{\tau_j-}$, i.e. contagion effects are strongest for obligors with similar characteristics. Second, contagion effects are inversely proportional to $\lambda_{\tau_j-,j}$, the default intensity of the defaulting entity. In particular, the default of an entity j with high credit quality and, hence, a low value of $\lambda_{\tau_j-,j}$ has a comparatively large impact on the market, perhaps because the default comes as a surprise to market participants.

17.4.3 Additional Information

In the dynamic version of the implied copula model studied in Example 17.16, default intensities (and hence credit spreads) evolve deterministically between defaults. This unrealistic behaviour is due to the assumption that the investor filtration is the default history $(\mathcal{H}_t)$. With this choice of filtration, significant new information enters the model only at default times. In this section we discuss an extension of the model with a richer investor filtration $(\mathcal{G}_t)$ that is due to Frey and Schmidt (2012).

The set-up. As in Example 17.16 we assume that the factor is a discrete rv V taking values in the set $S^V = \{v_1, \ldots, v_K\}$. The investor filtration $(\mathcal{G}_t)$ is generated by the default history of the firms in the portfolio and—this is the new part—an additional process (Z_t) representing observations of V with additive noise. Formally, we set $(\mathcal{G}_t) = (\mathcal{H}_t) \vee (\mathcal{F}^Z_t)$, where (Z_t) is of the form

$$Z_t = \int_0^t a(V, s)\,\mathrm{d}s + \sigma B_t. \tag{17.61}$$

Here, (B_t) is a Brownian motion, independent of V and the default indicator process $(\boldsymbol{Y}_t)$, $a(\cdot)$ is a function from $S^V \times [0, \infty)$ to $\mathbb{R}$, and σ is a scaling parameter that modulates the effect of the noise.

Equation (17.61) is the standard way of incorporating noisy information about V into a continuous-time stochastic filtering model. To develop more intuition for the model we show how (17.61) arises as the limit of a simpler discrete-time model. Suppose that investors receive noisy information about V at discrete time points $t_k = k\Delta, k = 1, 2, \ldots$, and that this information takes the form $X_{t_k} = a(V, t_k) + \epsilon_k$ for an iid sequence of noise variables ϵ_k, independent of V, with mean 0 and variance

$\sigma_\epsilon^2 > 0$. Now define the scaled cumulative observation process $Z_t^\Delta := \Delta \sum_{t_k \leqslant t} X_{t_k}$ and let $\sigma = \sqrt{\Delta \sigma_\epsilon^2}$. For Δ small we have the approximation

$$Z_t^\Delta = \sum_{t_k \leqslant t} \Delta a(V, t_k) + \Delta \sum_{t_k \leqslant t} \epsilon_k \approx \int_0^t a(V_s, s)\, \mathrm{d}s + \sigma B_t. \tag{17.62}$$

Without loss of generality we set $\sigma = 1$; other values of σ could be incorporated by rescaling the function $a(\cdot)$. In the numerical examples considered below we assume $a(v_k, t) = c \ln \bar{\gamma}(v_k, t)$, where $\bar{\gamma}(v_k, t)$ is as defined in (17.46) and the constant $c \geqslant 0$ models the information content of (Z_t); for $c = 0$, (Z_t) carries no information, whereas, for c large, the state of V can be observed with high precision. The process (Z_t) is an abstract model-building device that represents information contained in security prices; it is not directly linked to observable economic quantities. We come back to this point when we discuss calibration strategies for the model.

Dynamics of $(\boldsymbol{\pi}_t)$. As before we use the notation $\pi_t^k = Q(V = v_k \mid \mathcal{G}_t)$, $1 \leqslant k \leqslant K$, and $\boldsymbol{\pi}_t = (\pi_t^1, \ldots, \pi_t^K)'$. A crucial part of the analysis of Frey and Schmidt (2012) is the derivation of a stochastic differential equation for the dynamics of the process $(\boldsymbol{\pi}_t)$. We begin by introducing the processes that drive this equation. According to Lemma 17.14, the $(\mathcal{G}_t)$ default intensity of firm i is given by $\lambda_{t,i} = \sum_{l=1}^K \pi_t^k \gamma_i(v_k, t)$, and the compensated default indicator processes are given by the martingales $(M_{t,i})$ introduced in (17.58). Moreover, we define the process

$$W_t = Z_t - \int_0^t \sum_{k=1}^K \pi_s^k a(v_k, s)\, \mathrm{d}s, \quad t \geqslant 0. \tag{17.63}$$

In intuitive terms we have the relationship

$$E(\mathrm{d}Z_t \mid \mathcal{G}_t) = E(a(v_k, t)\, \mathrm{d}t \mid \mathcal{G}_t) = \sum_{k=1}^K \pi_t^k a(v_k, t)\, \mathrm{d}t.$$

Hence, $\mathrm{d}W_t = \mathrm{d}Z_t - E(\mathrm{d}Z_t \mid \mathcal{G}_t)\, \mathrm{d}t$, so the increment $\mathrm{d}W_t$ represents the unpredictable part of the new information $\mathrm{d}Z_t$. For this reason, (W_t) is called the *innovations process* in the literature on stochastic filtering. It is well known that (W_t) is a $(\mathcal{G}_t)$-Brownian motion (for a formal proof see, for example, Bain and Crisan (2009) or Davis and Marcus (1981)).

The following result from Frey and Schmidt (2012) generalizes equation (17.59).

Proposition 17.17. *The process* $(\boldsymbol{\pi}_t) = (\pi_t^1, \ldots, \pi_t^K)'$ *solves the SDE system*

$$\mathrm{d}\pi_t^k = \sum_{i=1}^m \pi_{t-}^k \left(\frac{\gamma_i(v_k, t)}{\sum_{l=1}^K \gamma_i(v_l, t) \pi_{t-}^l} - 1 \right) \mathrm{d}M_{t,i} + \pi_{t-}^k \left(a(v_k, t) - \sum_{l=1}^K \pi_t^l a(v_l, t) \right) \mathrm{d}W_t, \tag{17.64}$$

$1 \leqslant k \leqslant K$, *with initial condition* $\boldsymbol{\pi}_0 = \boldsymbol{\pi}$.

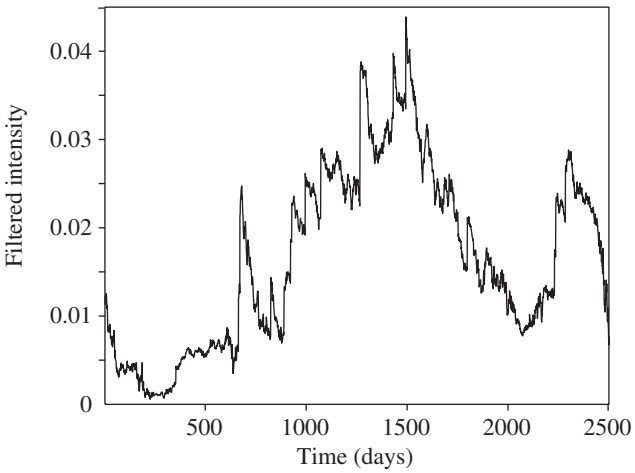

Figure 17.3. A simulated path for the default intensity in the model with incomplete information where additional information is modelled by the noisy observation process (Z_t). The graph was created using the set-up of Example 12.7.

In stochastic filtering, equation (17.64) is known as the *Kushner–Stratonovich equation*. We omit the formal proof of the proposition but try to give some intuitive explanation for equation (17.64). The jump part of this equation has the same form as in equation (17.59). This term represents the impact of default information on $\boldsymbol{\pi}_t$ and it is responsible for contagion effects due to defaults. Next we consider the diffusion part. Define random variables $\tilde{a}$ and $I_k \colon S^V \to \mathbb{R}$ by $\tilde{a}(v) = a(v,t)$ and $I_k(v) = I_{\{v=v_k\}}$. The coefficient of the diffusion part can then be written in the form

$$E^{\boldsymbol{\pi}}(I_k, \tilde{a}) - E^{\boldsymbol{\pi}}(I_k)E^{\boldsymbol{\pi}}(\tilde{a}) = \operatorname{cov}^{\boldsymbol{\pi}}(I_k, \tilde{a}).$$

It follows that a positive increment $\mathrm{d}W_t$ of the innovations process leads to an increase in π_t^k if the rvs I_k and $\tilde{a}$ are positively correlated under the measure $\boldsymbol{\pi}_t$.

The Kushner–Stratonovich equation (17.64) lends itself to simulation. In particular, the equation can be used to generate trajectories of $(\boldsymbol{\pi}_t)$ and of the default intensities $\lambda_{t,i} = \sum_{l=1}^{K} \pi_t^k \gamma_i(v_k, t)$, $1 \leqslant i \leqslant m$. Details are given in Frey and Schmidt (2011).

The extension of the investor information to the larger information set $(\mathcal{G}_t)$ leads to rich and realistic dynamics for default intensities incorporating both random fluctuations of credit spreads between defaults and default contagion. This is illustrated in Figure 17.3, where we plot a typical simulated trajectory of the default intensity. The fluctuation of the intensity between defaults as well as the contagion effects at default times (e.g. around $t = 600$) can be clearly seen.

Pricing of credit derivatives. All commonly encountered credit-risky instruments, credit derivatives and credit value adjustments can be categorized into the following two classes.

Basic credit products: this class comprises products where the cash-flow stream depends on the default history of the underlying portfolio and is therefore $(\mathcal{H}_t)$-adapted. Examples are corporate bonds, CDSs and CDOs.

Options on traded credit products: this class contains derivatives whose pay-off depends on the future *market value* of basic credit products. Examples include options on corporate bonds, options on CDS indices, and credit value adjustments for credit derivatives, as discussed in Section 17.2.

The pricing methodology for these product classes differs and we discuss them separately. We begin with basic credit products. Denote the associated stream of discounted future cash flows by the $\mathcal{H}_T$-measurable rv $\Pi(t, T)$. Then, using risk-neutral pricing the price at time t of the basic credit product is given by $V_t = E^Q(\Pi(t, T) \mid \mathcal{G}_t)$. Recall that $\mathcal{G}_t = \mathcal{H}_t \vee \mathcal{F}_t^Z$. Using a similar approach to the derivation of (17.43) we obtain

$$V_t = \sum_{k=1}^{K} h(t, v_k, \boldsymbol{Y}_t)\pi_t^k, \tag{17.65}$$

where $h(t, V, \boldsymbol{Y}_t) = E^Q(\Pi(t, T) \mid \tilde{\mathcal{G}}_t)$, the hypothetical value of the claim for known V. Note that V_t depends only on $\boldsymbol{Y}_t$, $\boldsymbol{\pi}_t$ and the function $h(\cdot)$; the precise form of the function $a(\cdot)$ in (17.61) is irrelevant for the pricing of basic credit products.

Next we consider options on traded credit products. Assume that N basic credit products—such as CDSs, CDO tranches or index swaps—are traded on the market, and denote their ex-dividend prices at time t by $V_{t,1}, \ldots, V_{t,N}$. The pay-off of an option on a traded credit product then takes the form $I_{\{\tilde{T} \leqslant T\}} g(\boldsymbol{Y}_{\tilde{T}}, V_{\tilde{T},1}, \ldots, V_{\tilde{T},N})$, where $\tilde{T}$ is a $(\mathcal{G}_t)$ stopping time at which the pay-out is made and T is the maturity of the contract. A prime example is the credit value adjustment for a CDS. Denote by (V_t^{CDS}) the market value of the counterparty-risk-free CDS. It was shown in Proposition 17.1 that the credit value adjustment is an option on the CDS with $\tilde{T} = T_1$ and pay-off

$$g(\boldsymbol{Y}_{T_1}, V_{T_1}^{\mathrm{CDS}}) = (1 - Y_{T_1}^R)(1 - Y_{T_1}^B)(V_{T_1}^{\mathrm{CDS}})^+.$$

Another example of an option on a traded credit product would be an option on a CDS index, as discussed in Frey and Schmidt (2011).

It can be shown that the process $(\boldsymbol{Y}_t, \boldsymbol{\pi}_t)$ is a Markov process in the investor filtration so that the price at time t of an option on a traded credit product is a function of Y_t and π_t of the form

$$E\left(\exp\left(-\int_t^{\tilde{T}} r(s)\,\mathrm{d}s\right) I_{\{\tilde{T} \leqslant T\}} g(\boldsymbol{Y}_{\tilde{T}}, V_{\tilde{T},1}, \ldots, V_{\tilde{T},N}) \,\middle|\, \mathcal{G}_t\right) = I_{\{\tilde{T} > t\}} g(t, \boldsymbol{Y}_t, \boldsymbol{\pi}_t). \tag{17.66}$$

The actual evaluation of the function g is usually based on Monte Carlo methods. Note that for an option on traded credit products the function g will in general depend on the entire dynamics of the process $(\boldsymbol{Y}_t, \boldsymbol{\pi}_t)$, as we will see in the analysis of credit value adjustments in the next section.

Calibration. Calibration methodologies are a crucial part of the model of Frey and Schmidt (2012) for the following reason. Recall that we view the process (Z_t) generating the filtration $\mathbb{G}$ as an abstract model-building device that is not directly linked to observable quantities. Consequently, the process $(\boldsymbol{\pi}_t)$ is not directly observable by investors. On the other hand, pricing formulas in our model depend on the values of $\boldsymbol{Y}_t$ and $\boldsymbol{\pi}_t$. Since pricing formulas need to be evaluated in terms of publicly available information, a key point in the application of the model is therefore to determine the realization of $\boldsymbol{\pi}_t$ at time t from prices of basic traded credit products observed at that date.

Suppose that N basic credit products are traded at time t at market prices p_n^*, $1 \leqslant n \leqslant N$, and denote by $h_n(t, V, \boldsymbol{Y}_t)$ the value of contract n in the artificial model where V is known. We then need to find some probability vector $\boldsymbol{\pi} = (\pi^1, \ldots, \pi^K)'$ such that the model prices $V_t^n(\boldsymbol{\pi}) = \sum_{k=1}^K \pi^k h_n(t, v_k, \boldsymbol{Y}_t)$ are close to the observed prices p_n^* for all n. The reader will note that this problem is similar to the calibration problem arising in the implied copula framework of Section 12.3.3, and we refer to that section for details of algorithms and numerical results.

17.4.4 Collateralized Credit Value Adjustments and Contagion Effects

We now study the impact of different price dynamics on the size of credit value adjustments and on the performance of collateralization strategies for a single-name CDS. We are particularly interested in the influence of contagion. To see that contagion might be relevant for the performance of collateralization strategies, consider the scenario in which the protection seller defaults before the maturity of the CDS. In such a case contagion might lead to a substantial increase in the credit spread of the reference entity (the firm on which the CDS is written) and hence in turn to a much higher replacement value for the CDS. In standard collateralization strategies this is taken into account in a fairly crude way, and the amount of collateral posted before the default may well be insufficient to replace the CDS.

Our exposition is based on Frey and Rösler (2014). We use the model with incomplete information from the previous section for our analysis. Slightly extending the set-up of that section, we assume that the factor is a finite-state Markov chain (Ψ_t) (details are not relevant for our discussion here). We consider two versions of the model that differ with respect to the amount of information that is available to investors. In the version with full information it is assumed that the process (Ψ_t) is observable, so there are no contagion effects. In the version with incomplete information investors observe only the process $Z_t = \int_0^t a(\Psi_s)\,\mathrm{d}s + B_t$ for $a(\psi) = c \ln \bar{\gamma}(\psi)$ (and, of course, the default history). As we have seen before, under incomplete information there is default contagion caused by the updating of the conditional distribution of (Ψ_t) at default times.

In order to compute credit value adjustments and to measure the performance of collateralization strategies, we use the bilateral collateralized credit value adjustment (BCCVA) introduced in Section 17.2.2. The actual computation of credit value adjustments is mostly carried out using Monte Carlo simulation. The numerical experiments that follow are based on a Markov chain (Ψ_t) with $K = 8$ states

Table 17.2. Results of model calibration for the case study on credit value adjustments.

State	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8
π_0	0.0810	0.0000	0.2831	0.0548	0.0000	0.0000	0.0000	0.5811
γ_B	0.0000	0.0010	0.0027	0.0040	0.0050	0.0059	0.0091	0.0195
γ_R	0.0031	0.0669	0.1187	0.1482	0.1687	0.1855	0.2393	0.3668
γ_S	0.0007	0.0245	0.0482	0.0627	0.0732	0.0818	0.1108	0.1840

$v_1, \dots, v_8$, where v_1 is the best state (lowest default probabilities for all firms considered) and v_8 is the worst state. In order to calibrate the model we assume that the protection buyer B has a credit spread of 50 bp, the reference entity R has a credit spread of 1000 bp, and the protection seller S has a credit spread of 500 bp, so that B is of far better credit quality than S; moreover, the default correlations of the three firms are fixed to be $\rho_{BR} = 2.0\%$, $\rho_{BS} = 1.5\%$ and $\rho_{RS} = 5\%$. The results of the calibration exercise are given in Table 17.2. Note that the default intensities at any fixed time t are comonotonic random variables.

Now we present the results of the simulation study. We begin with an analysis of the performance of popular collateralization strategies, for which we refer the reader to Section 17.2.2. The market value of the counterparty-risk-free CDS referencing R will be denoted by (V_t^{CDS}). We compare market-value collateralization where $C_t^{\mathrm{market}} = V_t^{\mathrm{CDS}}$, $t \leqslant T$, and threshold collateralization where

$$C_t^{M_1,M_2} = ((V_t^{\mathrm{CDS}})^+ - M_1) I_{\{(V_t^{\mathrm{CDS}})^+ > M_1\}} - ((V_t^{\mathrm{CDS}})^- - M_2) I_{\{(V_t^{\mathrm{CDS}})^- > M_2\}}$$

for thresholds M_1 and M_2. Note that market-value collateralization can be viewed as threshold collateralization with $M_1 = M_2 = 0$.

Numerical results illustrating the performance of these collateralization strategies are given in Table 17.3. We see that market value collateralization is very effective in the model with complete information. The performance of threshold collateralization is also satisfactory, as can be seen by comparing the CCVA for a small positive threshold with the CCVA for the uncollateralized case. The CDVA is quite low for all collateralization strategies, since in the chosen example B is of far better credit quality than S, so $Q(\xi_1 = B)$ is very small. Under incomplete information, on the other hand, the performance of market-value and threshold collateralization is not entirely satisfactory. In particular, even for $M_1 = M_2 = 0$ the CCVA is quite high compared with the case of full information. The main reason for this is the fact that, because of the contagion effects, threshold collateralization systematically underestimates the market value of the CDS at T_1, which leads to losses for the protection buyer. Note that the joint distribution of the default times is the same in the two versions of the model, so the differences in the sizes of the value adjustments and in the performance of the collateralization strategies can be attributed to the different dynamics of credit spreads (contagion or no contagion) in the two model variants. Our case study therefore clearly shows that the dynamics of credit spreads matters in the management of counterparty credit risk.

Finally, we show that for the given parameter values there is clear evidence for *wrong-way risk*. To demonstrate this we compare the correct value adjustments to

Table 17.3. Value adjustments in the model with complete information and in the model with incomplete information under threshold collateralization and market-value collateralization ($M_1 = M_2 = 0$). The nominal of the CDS is normalized to 1; all numbers are in basis points. In the last row we also report the value adjustment corresponding to the simplified value adjustment formulas (17.11) and (17.12).

	Full information			Partial information		
Threshold	CCVA	CDVA	BCCVA	CCVA	CDVA	BCCVA
$M_1 = M_2 = 0$	0	0	0	35	0	35
$M_1 = M_2 = 0.02$	16	0	15	45	0	45
$M_1 = M_2 = 0.05$	38	1	37	60	0	60
No collateralization with						
(i) correct formula	93	1	92	83	1	82
(ii) simplified formula	68	6	62	54	4	49

the value adjustments computed via the simplified formulas (17.11) and (17.12), both in the case of no collateralization. The results are given in the last two rows of Table 17.3. It turns out that in both versions of the model the value adjustments computed via the correct formula are substantially larger than the adjustments computed from the simplified formulas. This suggests that in situations where there is a non-negligible default correlation between the protection seller and the reference entity, the simplified formulas may not be appropriate.

Notes and Comments

There is a large literature on credit risk models with incomplete information. Kusuoka (1999), Duffie and Lando (2001), Giesecke and Goldberg (2004), Jarrow and Protter (2004), Coculescu, Geman and Jeanblanc (2008), Frey and Schmidt (2009), Cetin (2012) and Frey, Rösler and Lu (2014) all consider structural models in the spirit of the Merton model, where the values of assets and/or liabilities are not directly observable. The last three of these papers use stochastic filtering techniques to study structural models under incomplete information.

Turning to reduced-form models with incomplete information, the models of Schönbucher (2004) and Collin-Dufresne, Goldstein and Helwege (2010) are similar to the model considered in Section 17.4.2. Both papers point out that the successive updating of the conditional distribution of the unobserved factor in reaction to incoming default observations generates information-driven default contagion. Duffie et al. (2009) assume that the unobservable factor (Ψ_t) is given by an Ornstein–Uhlenbeck process. Their paper contains interesting empirical results; in particular, the analysis provides strong support for the assertion that an unobservable stochastic process driving default intensities (a so-called *dynamic frailty*) is needed on top of observable covariates in order to explain the clustering of defaults in historical data. The link between stochastic filtering and reduced-form models is explored extensively in Frey and Runggaldier (2010) and Frey and Schmidt (2012). Our analysis

in Section 17.4.3 is largely based on these papers. The numerical results on the performance of collateralization strategies are due to Frey and Rösler (2014). A survey of credit risk modelling under incomplete information can be found in Frey and Schmidt (2011).