Chapter 2
Constructing a Practical Model

Don't always blindly follow guidance and step-by-step instructions; you might run into something interesting.

(Georg Cantor)

As we saw in the introductory chapter, economic capital is a model-based approach to assess a firm's worst-case capital demand across a broad range of enterprise risks. The three largest elements are generally credit, market, and operational risk. While it is conceptually possible to try to describe some of these key risks in a *joint* manner, it is more typical to handle them separately. It is simply too difficult to construct a defensible simultaneous description of these three major risks. Moreover, estimating each piece separately and then simply adding them together takes a conservative stance on the interactions between these risks.[1]

For the vast majority of banking institutions, credit risk represents the single most important element of economic capital. The principal concern of a lending institution is (almost invariably) extending funds to another entity and ultimately not being paid back.[2] Depending on the size of the loan—and any amounts recovered during bankruptcy proceedings—the loss to the firm can be substantial. A second, related worry, is that the borrowing entity's financial situation worsens, making default (and attendant future losses) more likely. The deterioration in a borrower's creditworthiness also typically generates (unrealized) financial losses. This is because loan (and market) pricing—determined based on the borrower's situation at the time of the loan—is usually inadequate to cover the heightened risk of default. These two dimensions are referred to as default and credit-migration risk, respectively.

Lending and investing activities are nonetheless a banking institution's *daily bread*, so they cannot be simply avoided. Instead, they need to be managed. A critical aspect of lending, of course, is the initial assessment of a firm's financial

---

[1] In particular, it assumes no scope for diversification among these central risks.

[2] This can occur in a few ways: a direct loan, the purchase of a firm's bond, or even investment in a firm's equity. The mechanics of the extension of credit may differ, but the basic risks are the same.

position. This involves capturing all current information and also taking a forward-looking perspective. Such analysis allows the lender to determine if it makes sense to extend the loan in the first place. A second component is credit mitigation; these are features that safeguard the lender in the future. One important tool involves structuring the loan contract to provide recourse to the lending firm in specific situations. Broadly, such contractual features are referred to as loan covenants. These might take the form of limits on the lending firm's financial ratios or optionality in the loan contract permitting early termination and repayment.[3] An additional form of credit mitigation relates to assignment of collateral—in the form of other assets from the borrowing firm—to be collected in the event of default. A final aspect involves guarantees, where another entity steps in to fulfil the obligation in the event original borrower is not capable of meeting it.

Credit mitigation thus includes loan covenants, collateralization, and guarantees. These are standard credit-risk management tools for every lending institution, but some amount of risk always still remains. If the lending institution accepted no risk in a lending transaction, they could hardly expect to earn any return for their efforts.[4] At the same time, there is also competition among lenders in the economy. This mainly occurs along the pricing dimension, but it also naturally extends to the extent and severity of credit-mitigation measures.[5] In short, there are good reasons why credit risk can be mitigated, but not avoided. Managing credit risk must thus also extend beyond credit-risk mitigants. It also needs to be carefully measured; managing and balancing risk in a prudent fashion is dreadfully uncomfortable if you cannot measure it. Explaining how this might reasonably be performed is the principal task of this chapter.

2.1 A Naive, but Informative, Start

Every serious credit-risk economic capital model is complex and multifaceted. Jumping straight into the detail, without a bit of context, would be intimidating and counterproductive. Let's instead begin with a simpler, although admittedly somewhat naive, alternative modelling approach: the independent-default model.[6] As the name clearly indicates, it treats every default event as independent. The corollary of this central assumption is that it also entirely ignores systemic risk.

---

[3] These are only a few examples. Many other types of loan covenants are possible.

[4] As clichéd as it might sound, there is truth to the old saying "no risk, no reward." It also sounds far more serious and erudite in its original latin form, *periculum praemio*.

[5] Faced with two loans of the same price, a borrower will rationally accept the alternative offering the least number of constraints on its activities.

[6] In the spirit of full disclosure, this method represents one of our challenger models computed every day in our economic-capital framework as a comparator to the more complex production methodology.

Table 2.1 *Important credit-risk variables*: This table introduces the three of the most important elements associated with the measurement of credit risk within any modelling venture: the exposure at default, (unconditional) default probability, and loss-given-default.

Variable	Symbol	Acronym	Description
Exposure-at-default	c_i	EAD	Magnitude of the loan at default
(Unconditional) default probability	p_i	PD	Probability of default over a $(T - t)$-year horizon
Loss-given-default	γ_i	LGD	Severity of the loss in the event of default

This lack of realism makes it a poor choice of production model, but its associated simplicity makes it an excellent place to start.

To build this initial model, we require some ingredients. Imagine that one's portfolio consists of lending exposures to I distinct credit obligors.[7] We require three pieces of information about each of these counterparties: the total amount of exposure at play, the likelihood (or probability) of default over a given time horizon, and the amount of recovery in the event of default. These are referred to as the exposure-at-default, (unconditional) probability of default, and loss-given-default; we will denote them—for the ith credit obligor—as c_i, p_i, and γ_i, respectively. Table 2.1 provides a detailed summary of these three central quantities. It is useful, for the less experienced reader, to gain a good familiarity with these objects from the start. They will be employed repeatedly, in a variety of different ways, during each of the following chapters.

In principle, all three elements outlined within Table 2.1 are random variables. This implies that their future values are unknown and their range of possible outcomes is described by some statistical distribution. Understanding this fact is critical, because it will colour important modelling efforts in later discussion. For this introductory discussion, however, we will unrealistically treat them as known, deterministic quantities.

The kernel of the independent-default model is dead simple. The default event associated with the ith credit obligor—which we will denote as $\mathcal{D}_i$—occurs with probability p_i. Survival, or non-incidence of default, naturally arrives with probability $1 - p_i$. Default, in this context, is essentially a Bernoulli trial. Default either occurs or it does not; it is a binary state. We can succinctly (and conveniently) summarize this fact in the form of an indicator variable for each individual obligor,

$$\mathbb{I}_{\mathcal{D}_i} = \begin{cases} 1 : \text{ default occurs during } (t, T] \text{ with probability } p_i \\ 0 : \text{ survival until time } T \text{ with probability } 1 - p_i \end{cases}. \tag{2.1}$$

---

[7] We might also refer to these as risk owners; the idea is that they should represent a collective entity for the determination of a default event. This can become rather involved in practice, but is conceptually easy to understand.

To be clear, we represent the current time as t and the final time-point of our analysis as T. $T - t$ describes the length of risk horizon.[8]

This development directly leads to a description of the portfolio loss as,

$$L = \sum_{i=1}^{I} c_i \gamma_i \mathbb{I}_{\mathcal{D}_i}. \tag{2.2}$$

This short expression might not look like much, but it is actually the launch-site for all portfolio credit-risk models. It includes the exposures of all credit obligors, their loss-given-defaults, and a trigger describing when default occurs (or not). The objective is to characterize the distribution of the overall portfolio loss (L as at time T) through a description of the statistical behaviour of each of the three elements on the right-hand side of Eq. 2.2. Given this loss distribution, we can compute a range of interesting metrics for use in the measurement (and ultimately) management of credit risk.

In the independent-default model, the only source of uncertainty arises from each $\mathbb{I}_{\mathcal{D}_i}$ term defined in Eq. 2.1. Since each $\mathbb{I}_{\mathcal{D}_i}$ is statistically independent, the problem is rather tractable. If we are willing to assume that all credit obligors have the same default probability—that is, $p_i = p$ for all $i = 1, \ldots, N$—we can even find a closed-form solution for the distribution of L.[9] Practically, characterization of the distribution of L is readily determined via simulation methods. We can conceptualize this as a coin-tossing exercise. Imagine that we have an (unfair) coin where the probability of tails is p_i and $1 - p_i$ is heads. Tails represents default and heads is survival. Now extend this to I coins each with this property. If we independently flip all I coins, we can evaluate one realization of the portfolio loss. If we flip this collection of coins thousands (or even millions) of times, we can trace out all the possible permutations and combinations of portfolio credit loss. Equipped with this information, we can proceed to estimate virtually any risk metric we might desire.

This brute-force solution clearly requires an intimidating amount of coin flipping.[10] The effort is completely justified. If our objective is to capture one of the most important risks to a lending institution, we need to answer the following question: how likely is the coincident default of multiple important credit obligors over the next $T - t$ periods of time? The independent-default model—which basically takes the simplest possible route—provides a first incomplete answer to this question. Blessed with simplicity, the independent-default model will nevertheless underestimate this risk. The reason is that we are ignoring any possible default

---

[8] It is often set to one-year, but it is regularly altered for different situations and applications.

[9] A collection of independent, identically distributed Bernoulli trials follows the binomial distribution. See Bolder [7, Chapter 2] for much more detail on this question. A semi-analytic solution is available in the more general case if one is willing to delve into more complicated mathematics. Bolder [7, Chapter 7] touches on this idea.

[10] Thankfully, we can comfortably delegate this daunting task to a computer.

dependence (or correlation) between our lending counterparties. The larger point is that even in the simplest setting, there is a surprising amount of complexity in tracing out the credit-loss distribution. This situation will only get worse as we add more realism to our modelling framework.

> **Colour and Commentary 12** (THE INDEPENDENT-DEFAULT MODEL)*: The simplest approach to the modelling of portfolio credit risk involves a very strong assumption: that the default events of each of one's credit obligors are independent. This is tantamount to assuming that all credit portfolio risk is idiosyncratic or, in other words, there is no systemic credit risk. The logical consequence of such a choice is somewhat ridiculous; it implies that with a sufficiently large and diversified portfolio one could completely eliminate credit risk. As discussed in the previous chapter, this is neither consistent with economic theory nor is it a particularly conservative stance. We can thus conclude that the independent-default model is* not *a serious option. This does not, however, imply that it is not useful. Its simplicity, its use of the bare minimum of inputs, and its (fairly radical) base assumptions make it easy to understand and compute. In other words, not much can go wrong in its calculation. This makes it an excellent choice of challenger model to assist in trouble-shooting, interpreting, and communicating one's production methodology.*

2.2   Mixture and Threshold Models

All production models—in an attempt to capture a higher degree of realism— involve a number of complicated twists and additions. Addressing and organizing this complexity is the main task of this chapter. It is nonetheless useful to quickly sketch out the details of a few competing approaches before we jump into a description of our actual choice. Serious candidates for the modelling of portfolio credit risk fall into *two* main categories: mixture and threshold models.[11] This section will touch on both of these modelling frameworks. To keep it brief and permit us to focus on the core concepts, we'll restrict our attention to the one-factor setting.

---

[11] An excellent overview of these two approaches is found in McNeil et al. [29, Chapter 11].

2.2.1 The Mixture Model

The principal shortcoming of the independent-default model is its failure to capture systemic risk. The question thus becomes: how might we capture this dimension in a portfolio credit-risk model? The mixture-model setting solves this problem through the randomization of the default probability. Instead of using the given value p_i to describe the unconditional default probability, we cleverly replace it with something else that induces default correlation.

The general approach employed by all mixture models is to write each p_i as $p_i(Z)$, where Z is a random variable that is common to all credit obligors in one's portfolio. Z, which affects all variables—albeit in perhaps different ways—is the systemic risk factor. This trick leads to a revised definition of our default indicator from Eq. 2.2 as

$$
\mathbb{I}_{\mathcal{D}_i} = \begin{cases} 1: & \text{default occurs during } (t, T] \text{ with probability } p_i(Z) \\ 0: & \text{survival until time } T \text{ with probability } 1 - p_i(Z) \end{cases} . \quad (2.3)
$$

An entire family of models can be constructed involving various choices of Z and alternative specifications of $p_i(Z)$. One simple, mathematically pleasant, and educational entry point is to set $p_i(Z) \equiv Z$ and $Z \sim \beta(\alpha, \beta)$. In plain English, the unconditional default probability is assumed to follow a beta distribution. The unit-interval support of the beta distribution makes it a natural candidate for a probability and, quite conveniently, also yields closed-form solutions for the loss distribution.[12]

The loss function remains unchanged from Eq. 2.2. Estimation of each mixture model loss distribution can always be performed numerically via stochastic simulation. It involves the same frenzy of coin-flipping as in the independent-default model with one important difference. At each iteration of the simulation model, one draws the common systemic random variable, Z, from a hat.[13] This value then resets the individual probabilities of various individual coin flips for that realization. It is useful to think of Z as a global macroeconomic variable. Each draw of Z thus tells us something about the state of the world. When Z is large and positive, this pushes up the default probabilities for all credit obligors in the portfolio; this would be an adverse economic outcome. A small realization of Z creates the reverse chains of events. In both cases, however, Z is the object inducing default correlation within the model.

Given the value of Z, however, each coin flip is independent. This important property is referred to as conditional independence. It ensures that the model has both a systemic element—via the outcome of Z—and an idiosyncratic component that enters through the coin flip. A particularly interesting, and popular, choice of mixture model illustrates this fact very well. Consider the following default

---

[12] See Bolder [7, Chapter 3] for more details on this so-called beta-binomial mixture model.

[13] Or, if you prefer, it can be drawn from an urn or a bucket or from your favourite scientific computing software.

probability definition for the ith credit obligor:

$$p_i(Z) = p_i\left(\omega_{i0} + \omega_{i1}Z\right), \tag{2.4}$$

where $Z \sim \Gamma(a, b)$ and $\omega_{i0} + \omega_{i1} = 1$. This yields the gamma-Poisson mixture implementation, which is referred to as the CreditRisk+ model in practical applications.[14] Expanding Eq. 2.4, we arrive at a very useful decomposition,

$$p_i(Z) = \underbrace{p_i\omega_{i0}}_{\substack{\text{Idio-}\\\text{syncratic}\\\text{component}}} + \underbrace{p_i\omega_{i1}Z}_{\substack{\text{Systemic}\\\text{component}}}, \tag{2.5}$$

The randomized default probability is broken into both idiosyncratic and systemic pieces. The unconditional default probability as well as the ω_{i0} and ω_{i1} parameters—termed factor loadings—determine the relative importance of these two key aspects of risk.[15] Determining the appropriate values of the ω's is a critical aspect of the practical model implementation.

Mixture or, as they are sometimes called, actuarial models approach the problem from a reduced-form perspective. That is, the default event as an exogenous occurrence. Although it certainly induces default dependence and thereby introduces a systemic-risk element, it is silent on how default actually happens. For some, this is a weakness, for others it is a strength. In reality, it is neither; it is simply a feature of mixture models. In the subsequent section, we introduce a competing approach that attacks this question from another (more structural) angle.

Colour and Commentary 13 (MIXTURE MODELS)*: The family of mixture models represents a first possible step in extending the independent-default model into the realm of systemic risk. The origins of the term mixture model stem from the fact that the binomial structure of the model is* mixed *with another random variable driving the default probability.[a] There are thus two sources of uncertainty: the coin flips themselves and the probabilities of each coin. Both elements are, in the context of each simulation, in flux. Their combination yields a full-blown description of the permutations and combinations of default events incorporating ideas of both idiosyncratic and*

(continued)

---

[14] The model was first suggested and popularized by Wilde [43]. An excellent, and exhaustive, overview of this approach is found in Gundlach and Lehrbass [17].

[15] We can also see how this approach nests the independent-default model. It is readily recovered by setting $\omega_{i0} = 1$. The consequence of the condition $\omega_{i0} + \omega_{i1} = 1$ is that the systemic part falls out of the model.

2.2.2   The Threshold Model

Shortly after the introduction of the celebrated Black and Scholes [4] model, Merton [30] made a central contribution to the study of risk management. In an attempt to identify a general and consistent approach for the pricing of corporate debt, he provided the key notion underlying much of the credit-risk literature. The idea is that a firm practically enters into bankruptcy when its equity is exhausted. Or, in other words, when the value of its assets dip below its liabilities. In the language of the previous chapter, we can think of this as a situation when a firm's capital demand exceeds its supply.

The genius of this observation is that it provides a concrete path for the modelling of portfolio credit risk: one needs to describe the joint distribution of the individual credit-obligor assets in one's credit portfolio. Merton [30] offers a detailed approach—using continuous-time mathematics—to characterize, parametrize, and model this situation. Some years later, Vasicek [40, 41, 42] offered a simplification of the basic structure of Merton [30]'s proposal. Further contributions to the literature have lead to the current family of portfolio credit-risk threshold models.

The additional modelling element stems from a description of the firm's asset values.[16] We thus specify for the ith credit obligor in one's portfolio, the following variable:

$$y_i = \underbrace{a_i Z}_{\substack{\text{Systemic} \\ \text{component}}} + \underbrace{b_i \epsilon_i}_{\substack{\text{Idio-} \\ \text{syncratic} \\ \text{component}}} , \qquad (2.6)$$

where $Z, \epsilon_i \sim \text{i.i.d.} \mathcal{N}(0, 1)$ and $a_i, b_i \in \mathbb{R}$ are selected such that $y_i \sim \mathcal{N}(0, 1)$. We are no longer flipping coins, but we are not completely in uncharted territory.

---

[16] Practically, these can also be viewed as asset returns. Sometimes this notion is generalized to the idea of creditworthiness index or latent state variable. The effect is the same and the details are important; for this introductory discussion, it is probably preferable to keep the logical, asset-related link to the original Merton [30] approach.

Inspection of Eq. 2.6 reveals that our variable, y_i, is the sum of two familiar components: systemic and idiosyncratic risk. A single realization of Z, which captures the systemic component, provides a common effect—experienced in potentially different ways—for all credit counterparties. Again, this can be viewed as a global macroeconomic variable. Each of the firm-specific ϵ_i terms describes the idiosyncratic effect. This explains why there are I of them and they are all independent. Although the mechanics are very different, the ϵ's are conceptually not so far removed from our coin-flipping exercise. Finally, the parameters a_i and b_i (i.e., factor loadings) determine the relative importance of the systemic and idiosyncratic parts.[17]

It still requires a bit more discussion to get to our default event. We wish to describe default as the situation when a firm's assets are less than its liabilities. We view each y_i as a (random) characterization of the firm's assets. Let's introduce the value K_i to represent the liabilities. By extension, using the Merton [30] insight, we can (cheerfully) define the default event as,

$$\mathcal{D}_i \equiv \{y_i \leq K_i\}. \tag{2.7}$$

If, via the realizations of Z and ϵ_i, y_i falls below the value K_i, then default will have occurred. The firm will have exhausted its equity and find itself in bankruptcy. This directly permits us to re-express our default indicator variable—from Eqs. 2.1 and 2.3—in the following form:

$$\mathbb{I}_{\mathcal{D}_i} \equiv \mathbb{I}_{\{y_i \leq K_i\}} = \begin{cases} 1 : & \text{default occurs during } (t, T] \text{ when } y_i \leq K_i \\ 0 : & \text{survival until time } T \text{ if } y_i > K_i \end{cases}. \tag{2.8}$$

Once again, the loss definition from Eq. 2.2 remains the same. We have simply provided an alternative characterization of the default event.

How might we now estimate the associated credit-loss distribution? Again, we use simulation. We begin by drawing the common Z from the standard-normal distribution. This is the systemic piece, which determines the general state of the world. We then draw I independent ϵ_i random variates; once again, from the standard normal distribution. These are the specific, firm idiosyncratic elements. Using Eq. 2.6, we combine these inputs to construct our collection of firm asset values, or y_i's. Comparing these to their liability values permits us to evaluate our indicator variables from Eq. 2.8 and construct a single loss outcome.[18] This process is then repeated—more or less, *ad nauseum*—by a computer program until we have a sufficiently clear view of our loss distribution to estimate risk metrics.

---

[17] These coefficients should be linked back to the CreditRisk+ ω parameters introduced in Eq. 2.5. Do not forget that the a_i and b_i parameters are chosen to force $y_i \sim \mathcal{N}(0, 1)$; this, as we'll see in subsequent discussion, leads to some practical headaches.

[18] This also, of course, requires a sensible value for each threshold value, K_i.

The class of threshold models, in contrast to the mixture-model world, takes a structural approach to the characterization of credit risk. This means that it endogenizes the default event; or rather, default is determined within the model, not outside of it. Again, this does not make it necessarily superior to the mixture-model approach, just different. For multiplicity of perspective, different is good.

> **Colour and Commentary 14** (THRESHOLD MODELS)*: The family of threshold models represents a second possible approach towards adding the systemic element into the independent-default model. The structure is, practically at least, rather different. It begins with the assignment of a latent state variable—conceptually related to the idea of a firm's assets—to each individual credit obligor. This state variable linearly combines two elements: a systemic and an idiosyncratic component. The term threshold model arises from the operation required to determine each individual default event. If a credit obligor's state variable falls below a pre-defined* threshold—*which is motivated by the firm's level of liabilities—then default is triggered. As in the mixture-model setting, there is a common element impacting all obligors; this induces default correlation and creates systemic risk. The independent, firm-specific elements represent idiosyncratic risk. CreditMetrics is a well-known industrial model, introduced by Gupton et al. [18], that uses the threshold-model methodology. A twist upon this approach will form the foundation of the our production credit-risk model.*

2.3 Asset-Return Dynamics

Equipped with this background, we can now proceed to examine—in rather greater detail and mathematical rigour—our specific choice. As previously indicated, the origin of capital modelling at the NIB dates back a few decades. As with most institutions, the framework has moved forward in a gradual—and sometimes uneven—manner. The bank's economic-capital model was initially implemented in, and around, 2004 using an external vendor tool. Some years later, an in-house approach was developed and brought into production. Over the course of time, improvements and adjustments to the basic methodology were introduced. We will refer to this as the *legacy* model. This situation continued until it received, during 2019 and 2020, a serious conceptual and practical overhaul. Following extensive statutory revision, a new implementation involving numerous methodological changes became a necessity. The remaining sections of this chapter seek to help the reader gain a better understanding of these choices, which together can be described as our *current* (or revised) production model.

While our legacy approach can be classified as a multivariate, multi-period Gaussian threshold model, it has a few features that differ from the general class of such models. As such, it is useful and important to review, motivate, and to the extent reasonable and feasible, derive the main elements of the legacy approach. This will provide some context and justification for the recently revised methodology.

NIB has, from the beginning, opted for a threshold model. Use of this approach requires one to introduce, as lightly introduced in the previous section, what is referred to as a latent creditworthiness index. In the parlance of the Merton [30] approach, this would be related to the value of an obligor's assets (or asset return).[19] The generalization to a creditworthiness index is thus sensible, but it naturally requires a mathematical structure. Specification of the mechanics of one's creditworthiness process is thus a *first* step into the construction of a threshold model. We are going to allocate quite a bit of effort into understanding this first step, because it has important implications for the overall development of the legacy (and revised) NIB credit-risk modelling framework.

The characterization of the asset or creditworthiness state variable has customarily been done in an industrial strength manner. The legacy approach does not deviate from this practice. A bit of formality is thus in order. On the probability space, $(\Omega, \mathcal{F}, \mathbb{P})$, a slight generalization of Merton [30]'s approach leads to the following stochastic differential equation (SDE) to describe the ith obligor's latent (i.e., unobserved) creditworthiness process,

$$d\tilde{X}_i(t) = \sum_{j=1}^{J} \tilde{\beta}_{ij} dZ_j(t) + \tilde{\sigma}_i dW_i(t). \tag{2.9}$$

These are the intertemporal dynamics of $\{\tilde{X}_i(t) : i = 1, \ldots, I, \ t \geq 0\}$, which is a multivariate system of creditworthiness indices; there is one for each obligor. Equation 2.9 lies at the heart of the legacy modelling methodology; indeed, some variation on it is, in fact, the cornerstone of a sizable proportion of extant industrial and academic models.

Equation 2.9, despite being only a single expression, requires a fairly dramatic amount of explanation. The actual unpacking process will provide useful insight into the structure of the model. Let us, therefore, try to address each definitional and implicational point in turn:

1. **DRIFT**: The SDE in Eq. 2.9 does *not* have a drift term; as such, each increment of $X_i(t)$ should be centred around zero for all $i = 1, \ldots, I$.[20] One could

---

[19] This latter interpretation makes logical sense in the context of a corporate entity, but can be a bit more complicated for a sovereign, municipality, or government agency.

[20] The actual individual values of the underlying process, $\{X_i(t) : i = 1, \ldots, I, \ t \geq 0\}$, will however depend on the set of starting values for X. If they are all set to zero at inception, then the expected value of each X_i will remain zero across time, although the actual outcomes will practically deviate from this point.

potentially imagine a generic drift term related to the business cycle, but not only would this be difficult to identify, its incorporation would create important challenges for the model's implementation.

2. **SYSTEMIC FACTORS**: The stochastic element in the first term in Eq. 2.9, $\{Z_j(t) : j = 1, \ldots, J, \ t \geq 0\}$, is a correlated system of Brownian motions with instantaneous correlation matrix, $\tilde{S}$. Each individual Z_j is intended to represent a market-related systemic risk factor. In general, this could be related to macroeconomic outcomes, commodity prices, or even volatility. Practically, for the purposes of our model, these factors are considered to be regional- and industrial-related variables.

3. **IDIOSYNCRATIC FACTORS**: The second stochastic term in Eq. 2.9 is a standard scalar Wiener process, $\{W_i(t) : t \geq 0\}$ for $i = 1, \ldots, I$. Each W_i is, by construction, independent of both one another and the correlated systemic risk factors. These elements, therefore, describe the idiosyncratic (i.e., specific) risks associated with each individual obligor.

4. **FACTOR DIMENSIONALITY**: As a practical matter, we have that $I \gg J$. The individual obligors in our portfolio are counted in hundreds, whereas the maximal number of correlated systemic factors is unlikely to exceed a few dozen. This is both conceptually important—that is, there are relatively few sources of systemic risk, but many types of idiosyncratic risk—and also of practical interest since it determines the dimensionality of one's parametrization.

5. **PROPERTIES OF BROWNIAN MOTION**: The presence of the Brownian motions implies time-independent increments and Gaussianity. There is something for everyone. To be more specific—for the probabilist or statistician—the time-increment of a standard Wiener process associated with the ith credit obligor, $W_i(t) - W_i(s)$ with $t > s$, are independent and distributed as $\mathcal{N}(0, t - s)$. Moreover, given a set of time increments, $t > s > v$, then the increments $W_i(t) - W_i(s)$ and $W_i(s) - W_i(v)$ are independent. For an econometrician, this implies that the idiosyncratic factors are independent from both a cross-sectional and time-series perspective. The systemic factors, however, are *not* cross-sectionally independent; their linear dependence is captured by the instantaneous correlation matrix, $\tilde{S}$. Since the system $\{Z_j(t) : j = 1, \ldots, J, \ t \geq 0\}$ is a collection of Brownian motions, non-overlapping systemic increments remain independent over time. For the economist, this implies that—whether from an idiosyncratic or systemic perspective—shocks regarding the creditworthiness of our set of obligors in one period are independent of similar shocks received in other periods. In the real world, for the practitioner, this is quite unlikely to be true. There are probably temporal trends in the evolution of creditworthiness, but this model assumes them away. Even if you believed such trends exist, however, reliably identifying their magnitude and dynamics is likely to be empirically very difficult (or even impossible).

6. **TIME CONTINUITY**: In this general setting, Eq. 2.9 implies that our creditworthiness system has continuous sample paths. While this is an attractive theoretical property, practically we cannot observe the creditworthiness of any entity at each possible instant in time. Economic-capital is typically estimated

in annual increments; in some cases, one might wish to consider monthly credit migration, but this is probably the feasible lower limit of time granularity. The consequence is that it will be necessary to discretize Eq. 2.9 to create a workable implementation.

7. **TIME HOMOGENEITY**: The model parameters, $\tilde{\beta}_{ij}, \tilde{\sigma}_i \in \mathbb{R}$ for all $i = 1, \ldots, I$ and $j = 1, \ldots, J$, determine the instantaneous relative weight (or loading) of each systemic and idiosyncratic risk factor to an individual obligor's creditworthiness. They also, indirectly, determine the relative weights of the systemic and idiosyncratic risk factors for each individual obligor. Since these parameters are not indexed to time, we may safely surmise them to be time homogeneous. In other words, the importance of a given risk factor to a credit counterparty's creditworthiness is assumed to be constant over time. This, as a choice, is probably somewhat dubious; specification of meaningful time-varying parameters is, however, a daunting undertaking.

8. **SYSTEMIC FACTOR LOADINGS**: The systemic $\tilde{\beta}$ parameters and the correlated system of Brownian motions, $\{Z_j(t); j = 1, \ldots, J; t \geq 0\}$, are responsible for the portfolio effects. They do so, however, in different ways. The systemic risk factors describe the current state of regional and sectoral risk. The instantaneous correlation matrix, $\tilde{S}$, captures the dependence between these outcomes.[21] The $\tilde{\beta}$ parameters, as previously suggested, determine the loading of each risk factor onto a given obligor. The actual interdependence between the creditworthiness of each obligor thus depends on the $\tilde{\beta}$ values and the relevant entries in $\tilde{S}$. Were $\tilde{S}$ to be an identity matrix, then all of the cross-obligor correlations would be determined by the $\tilde{\beta}$ parameters. If, however, all of the $\tilde{\beta}$ parameters were set to zero, then systemic risk is removed from the model and creditworthiness can be considered to be entirely idiosyncratic. This would bring us back to the previously discussed independent-default model.[22]

9. **PARAMETER DIMENSIONALITY**: The theoretical number of model parameters is quite unsettling. There are, in principle, $J \times I$ systemic and I idiosyncratic parameter values. For a representative portfolio of 500 obligors and 24 systemic risk factors—values roughly consistent with the current implementation—this amounts to in excess of 10,000 possible model parameters. This is a clearly unmanageable (even laughable) situation. Some form of defensible parametric restrictions will be required to manage the dimensionality needed to simultaneously inform the actual dependence structure between obligor creditworthiness and thus, indirectly, portfolio-level default and credit migration.

---

[21] One could—and many model implementations actually do—orthogonalize this collection of risk factors to create a set of independent systemic risk drivers. In this case, the factor loadings would solely determine the factor correlations.

[22] It is difficult to conceive of placing all of the responsibility upon the correlation matrix, $\tilde{S}$. Setting all $\tilde{\beta}$ parameters to unity (or some other fixed value) does not quite achieve this; instead such an action essentially reduces the model to a single factor structure, which is an equally weighted linear combination of all systemic factors.

10. **MAPPING X_i TO CREDIT MIGRATION**: Each increment is a zero-mean, time-scaled Gaussian outcome. This implies that, over time, the $\tilde{X}$ values will cover the support of the Gaussian distribution. Each $\tilde{X}_i(t)$ can thus, in principle, take values over the interval, $(-\infty, \infty)$. In practice, assuming each $\tilde{X}_i(0)$ is normalized to zero and divided by the instantaneous volatility, actual simulated outcomes will cover the continuum from roughly $(-5, 5)$. There are, however, only q discrete credit-state outcomes. An important element of the model will, therefore, involve creating a link between continuous creditworthiness and discrete credit-state outcomes. By necessity, given their differing scales, this will require a many-to-one mapping between the creditworthiness and credit-state values.

There are clearly many details and insights that can be drawn from this choice of creditworthiness process. Equation 2.9 is thus a critical foundational element of the legacy model, but by itself, it raises more questions than it actually answers. In the following sections, we will make numerous adjustments to Eq. 2.9 with a view towards addressing some of the previous points and, to the extent possible, simplifying and standardizing the development. This journey will bring us, rather better prepared and informed, to our production model.

Colour and Commentary 15 (DISMANTLING THE CONTINUOUS-TIME MACHINERY)*: The legacy approach attempted to make a clear link to the original Merton [30] framework replete with its continuous-time mathematical structure. The first step is the specification of a multivariate stochastic differential equation to describe the latent, creditworthiness state variable for each credit obligor. Indeed, this is a very standard beginning. While theoretically appealing, it is practically rather difficult to use. Most importantly, the incremental complexity does not provide significant, if any, additional value in terms of flexibility and descriptive improvement. As a consequence, it makes logical sense to move as quickly as possible to simplify a number of key aspects of the legacy structure. This constructive approach may, to some readers, seem a bit excessive and painful. There is certainly truth to that view, but this form of exposition was selected because it permits us, from a first-principles pedagogical perspective, to reveal, motivate, and resolve a range of important issues and choices.*

2.3.1 Time Discretization

The continuous-time construction is a bit cumbersome. Using it will require potentially complex mathematics for little pay-off—a situation to be generally

avoided—since the model will be implemented in discrete time.[23] Indeed, it is unlikely that more than one or two steps will be taken. As a consequence, before proceeding further, it makes logical and practical sense to discretize the creditworthiness SDE presented in Eq. 2.9. There are a few alternatives techniques for such an operation, but we opt for the simplest approach: the simple and intuitive Euler-Maruyama method.[24] The first step is to create a partition of our time horizon, $[0, T]$ into a collection of κ sub-intervals:

$$0 = t_0 < t_1 < t_2 < \cdots < t_\kappa = T. \tag{2.10}$$

Generally, the length of each subinterval is equal. That is, $\Delta t = t_k - t_{k-1}$ is constant for all $k = 1, \ldots, \kappa$. Given this partition, we proceed to transform Eq. 2.9 over any arbitrary discrete time interval, $[t_{k-1}, t_k]$, into:

$$X_i(t_k) - X_i(t_{k-1}) = \sum_{j=1}^{J} \tilde{\beta}_{ij} \underbrace{\left(Z_j(t_k) - Z_j(t_{k-1}) \right)}_{\Delta Z_j(k)} + \tilde{\sigma}_i \underbrace{\left(W_i(t_k) - W_i(t_{k-1}) \right)}_{\Delta W_i(k)}, \tag{2.11}$$

$$\Delta X_i(k) = \sum_{j=1}^{J} \tilde{\beta}_{ij} \Delta Z_j(k) + \tilde{\sigma}_i \Delta W_i(k),$$

where the Δ operator is introduced to somewhat reduce the notational burden. Recall, that the variance of each Brownian increment remains Δt. We may, with the following adjustment, slightly simplify our life and take one step closer to more typical threshold models. In particular, we write

$$\Delta X_i(k) = \underbrace{\frac{\sqrt{\Delta t}}{\sqrt{\Delta t}}}_{=1} \underbrace{\left(\sum_{j=1}^{J} \tilde{\beta}_{ij} \Delta Z_j(k) + \tilde{\sigma}_i \Delta W_i(k) \right)}_{\text{Eq. 2.11}}, \tag{2.12}$$

$$= \sum_{j=1}^{J} \underbrace{\tilde{\beta}_{ij} \sqrt{\Delta t}}_{\beta_{ij}} \underbrace{\frac{\Delta Z_j(k)}{\sqrt{\Delta t}}}_{\Delta z_i(k)} + \underbrace{\tilde{\sigma}_i \sqrt{\Delta t}}_{\sigma_i} \underbrace{\frac{\Delta W_i(k)}{\sqrt{\Delta t}}}_{\Delta w_i(k)},$$

for $k = 1, \ldots, \kappa$ where $\Delta z_j(k) \sim \mathcal{N}(0, \Omega_{jj})$ and $\Delta w_i(k) \sim \mathcal{N}(0, 1)$ for all $i = 1, \ldots, I$ and $j = 1, \ldots, J$. The beauty of this trick is that the parameters have also been scaled by the square-root of the common time step so that we do not

---

[23] Continuous-time models are essential in pricing applications where arbitrage relationships need to hold constantly over time. In risk-management settings, such as this one, this property is rarely required permitting us to work freely in discrete time.

[24] See Oksendal [34] for more information on this technique.

need to carry this aspect around with us in our development.[25] For this reason, $\beta_{ij} = \tilde{\beta}_{ij}\sqrt{\Delta t}$ and $\sigma_i = \tilde{\sigma}_i\sqrt{\Delta t}$ thus represent the parametrization—which must be determined empirically—associated with our selected time step. We also replace the instantaneous correlation matrix (i.e., $\tilde{S}$) for the correlated Brownian system, Z, with a non-instantaneous correlation matrix, Ω.

Since we will be taking annual time steps, let's simplify our life somewhat by setting $\Delta t = 1$; while this is not really in the spirit of discretization, this decision is easily relaxed and it rather dramatically eases the notational burden. Finally, since all time steps are assumed to be equal, we may drop the reference to the kth time step leading to the following streamlined, discrete-time representation of Eq. 2.9:

$$\Delta X_i = \sum_{j=1}^{J} \beta_{ij}\Delta z_j + \sigma_i \Delta w_i, \tag{2.13}$$

for $i = 1,\ldots,I$. Depending on the context, of course, it may be useful to reintroduce specific reference to the time step in the notation. One, for example, might wish to include some notion of time dependency in the economic-capital computations.[26]

To summarize, not only is Eq. 2.13 easier to work with—from a mathematical perspective—it finds itself in a more convenient, and realistic, form for the important task of model parametrization. We have basically taken—for better or for worse—a decisive step away from the world of continuous-time mathematics.

2.3.2   Normalization

Equation 2.13 is qualitatively similar to the Vasicek [40, 41, 42] formulation—touched upon in the introductory section—but it is missing a few important elements. More specifically, there is a lack of clarity about the relative importance of the systemic and idiosyncratic elements and the creditworthiness increment, for a given obligor, is *not* standard normal. These two elements are, of course, determined by choices of the β and σ elements. Practically, it is extremely useful, for the interpretation of the model structure and its results, to directly handle these two aspects. It also indirectly solves issues surrounding the determination of the σ parameters.[27]

---

[25] Indeed, from a statistical perspective, the parameters will be consistent with the time steps used in the underlying estimation data.

[26] Consider a financial instrument, with a maturity less than T, that needs to be reinvested. Deposits are one example, where the decision to reinvest would reasonably depend on the creditworthiness index—and associated credit state—in the previous period.

[27] As we'll see in later discussion, there are sensible proxies for the common, systemic sources of risk, but the idiosyncratic element is, by its very definition, much more difficult to manage.

The first stepping stone in this development involves computation of the variance of the systemic element in Eq. 2.13. Undeniably tedious—and frankly is much more naturally expressed in matrix notation—its derivation is both informative and essential. In particular, we have

$$\text{var}\left(\sum_{j=1}^{J}\beta_{ij}\Delta z_j\right) = \mathbb{E}\left(\left(\sum_{j=1}^{J}\beta_{ij}\Delta z_j\right)^2 - \left(\mathbb{E}\left(\sum_{j=1}^{J}\beta_{ij}\Delta z_j\right)\right)^2\right), \quad (2.14)$$

$$= \mathbb{E}\left(\left(\sum_{j=1}^{J}\beta_{ij}\Delta z_j\right)^2 - \left(\sum_{j=1}^{J}\beta_{ij}\underbrace{\mathbb{E}(\Delta z_j)}_{=0}\right)^2\right),$$

$$= \mathbb{E}\left(\left(\sum_{j=1}^{J}\beta_{ij}\Delta z_j\right)\left(\sum_{k=1}^{J}\beta_{ik}\Delta z_k\right)\right),$$

$$= \mathbb{E}\left(\sum_{j=1}^{J}\sum_{k=1}^{J}\beta_{ij}\beta_{ik}\Delta z_j\Delta z_k\right),$$

$$= \sum_{j=1}^{J}\sum_{k=1}^{J}\beta_{ij}\beta_{ik}\text{cov}\left(\Delta z_j, \Delta z_k\right),$$

$$= \sum_{j=1}^{J}\sum_{k=1}^{J}\beta_{ij}\beta_{ik}\Omega_{jk},$$

for $i = 1, \ldots, I$. The takeaway from Eq. 2.14 is that the variability in the systemic contribution to the creditworthiness of the ith obligor is a fairly complicated function of the β choices and the covariance structure of the systemic risk factors.

The standard deviation of the systemic-risk contribution—which is essential for normalization—is now simply defined as

$$\sigma\left(\sum_{j=1}^{J}\beta_{ij}\Delta z_j\right) = \sqrt{\sum_{\ell=1}^{J}\sum_{k=1}^{J}\beta_{i\ell}\beta_{ik}\Omega_{\ell k}}. \quad (2.15)$$

Putting this quantity in our back pocket for the moment, we now rewrite (rather boldly) Eq. 2.13 in the following form,

$$\Delta X_i = \alpha_i\sum_{j=1}^{J}\frac{\beta_{ij}}{\underbrace{\sqrt{\sum_{\ell=1}^{J}\sum_{k=1}^{J}\beta_{i\ell}\beta_{ik}\Omega_{\ell k}}}_{\text{Eq. 2.15}}}\Delta z_j + \sqrt{1-\alpha_i^2}\Delta w_i, \quad (2.16)$$

where $\alpha_i \in (0, 1) \subset \mathbb{R}$ and $\sigma_i = \sqrt{1 - \alpha_i^2}$ for $i = 1, \ldots, I$. This is the *discretized* and *normalized* form of our creditworthiness indicator state variable. Although technically correct, the form of the factor loadings in Eq. 2.16 is in dire need of some simplification. Let's, therefore, define

$$\mathrm{B}_{ij} = \frac{\beta_{ij}}{\sqrt{\displaystyle\sum_{\ell=1}^{J}\sum_{k=1}^{J}\beta_{i\ell}\beta_{ik}\Omega_{\ell k}}} \equiv \frac{\beta_{ij}}{\sigma\left(\displaystyle\sum_{j=1}^{J}\beta_{ij}\Delta z_j\right)}, \tag{2.17}$$

for $i = 1, \ldots, I$ and $j = 1, \ldots, J$. This allows us to restate Eq. 2.16 more succinctly as

$$\Delta X_i = \alpha_i \sum_{j=1}^{J} \mathrm{B}_{ij}\Delta z_j + \sqrt{1 - \alpha_i^2}\,\Delta w_i. \tag{2.18}$$

At first glance, this may not seem to be much in the way of progress. Indeed, it may appear to actually further complicate the creditworthiness process dynamics. Practically, we have introduced *two* important elements:

1. there is an α parameter, for each obligor, that determines the relative importance of the systemic and idiosyncratic contributions to the creditworthiness index—it is broadly referred to as the systemic weight; and
2. the normalized factor loadings (i.e., the B's) work along the specification of the systemic weights (i.e., the α's), ensure that each ΔX_i outcome has unit variance.[28]

Both of these points naturally require demonstration. Indeed, the systemic weights (i.e., the α_i's) have, more or less, fallen from the sky. The easiest place to start is to verify the zero expectation. In particular, for the set of I credit counterparties,

$$\mathbb{E}(\Delta X_i) = \mathbb{E}\left(\underbrace{\alpha_i \sum_{j=1}^{J} \mathrm{B}_{ij}\Delta z_j + \sqrt{1 - \alpha_i^2}\,\Delta w_i}_{\text{Eq. 2.18}}\right) \tag{2.19}$$

$$= \alpha_i \sum_{j=1}^{J} \mathrm{B}_{ij}\underbrace{\mathbb{E}(\Delta z_j)}_{=0} + \sqrt{1 - \alpha_i^2}\,\underbrace{\mathbb{E}(\Delta w_i)}_{=0},$$

$$= 0,$$

---

[28] As we'll see shortly, the cross correlation between any two increments—say ΔX_n and ΔX_m, for example—will be more complicated.

as expected. This is a relatively obvious consequence of the previous structure, but is nonetheless usefully verified.

Let's now move on to the variance of the collection of creditworthiness increments. It has the following form,

$$
\text{var}\left(\Delta X_i\right) = \text{var}\left(\underbrace{\alpha_i \sum_{j=1}^{J} \mathbf{B}_{ij}\Delta z_j + \sqrt{1-\alpha_i^2}\,\Delta w_i}_{\text{Eq. 2.18}}\right) \tag{2.20}
$$

$$
= \underbrace{\text{var}\left(\alpha_i \sum_{j=1}^{J} \mathbf{B}_{ij}\Delta z_j\right) + \text{var}\left(\sqrt{1-\alpha_i^2}\,\Delta w_i\right)}_{\text{By independence}}
$$

$$
= \alpha_i^2 \text{var}\left(\underbrace{\sum_{j=1}^{J} \frac{\beta_{ij}}{\sqrt{\displaystyle\sum_{\ell=1}^{J}\sum_{k=1}^{J}\beta_{i\ell}\beta_{ik}\Omega_{\ell k}}}\Delta z_j}_{\text{Eq. 2.17}}\right) + (1-\alpha_i^2)\underbrace{\text{var}\left(\Delta w_i\right)}_{=1}
$$

$$
= \left(\frac{\alpha_i}{\sqrt{\displaystyle\sum_{\ell=1}^{J}\sum_{k=1}^{J}\beta_{i\ell}\beta_{ik}\Omega_{\ell k}}}\right)^2 \underbrace{\text{var}\left(\sum_{j=1}^{J}\beta_{ij}\Delta z_j\right)}_{\text{Eq. 2.14}} + (1-\alpha_i^2)\underbrace{\text{var}\left(\Delta w_i\right)}_{=1}
$$

$$
= \frac{\cancel{\alpha_i^2}}{\displaystyle\sum_{\ell=1}^{J}\sum_{k=1}^{J}\cancel{\beta_{i\ell}\beta_{ik}\Omega_{\ell k}}}\underbrace{\left(\cancel{\sum_{\ell=1}^{J}\sum_{k=1}^{J}\beta_{i\ell}\beta_{ik}\Omega_{\ell k}}\right)}_{\text{Eq. 2.14}} + 1 - \cancel{\alpha_i^2}
$$

$$
= 1,
$$

as desired. This is a rather more laborious computation, but it clearly illustrates the effectiveness of the normalization in the previous section. Each individual $\Delta X_i \sim \mathcal{N}(0, 1)$; this simple property, as will be demonstrated in a moment, dramatically streamlines the default and credit-migration computations.

Another rather unexciting computation involves identifying the specific form of the cross asset-correlation between the nth and mth arbitrary obligors in one's credit portfolio. Working from first principles, it is derived as[29]

$$\text{cov}\left(\Delta X_n, \Delta X_m\right) = \mathbb{E}\left(\left(\Delta X_n - \underbrace{\mathbb{E}(\Delta X_n)}_{=0}\right)\left(\Delta X_m - \underbrace{\mathbb{E}(\Delta X_m)}_{=0}\right)\right), (2.21)$$

$$= \mathbb{E}\left(\Delta X_n \cdot \Delta X_m\right),$$

$$= \mathbb{E}\left(\left(\alpha_n \sum_{j=1}^{J} \mathrm{B}_{nj}\Delta z_j + \sqrt{1-\alpha_n^2}\Delta w_n\right)\right.$$

$$\left.\times \left(\alpha_m \sum_{k=1}^{J} \mathrm{B}_{mk}\Delta z_k + \sqrt{1-\alpha_m^2}\Delta w_m\right)\right),$$

$$= \underbrace{\mathbb{E}\left(\left(\alpha_n \sum_{j=1}^{J} \mathrm{B}_{nj}\Delta z_j\right)\left(\alpha_m \sum_{k=1}^{J} \mathrm{B}_{mk}\Delta z_k\right)\right)}_{\text{Cross terms vanish by independence}},$$

$$= \alpha_n\alpha_m \mathbb{E}\left(\left(\sum_{j=1}^{J} \mathrm{B}_{nj}\Delta z_j\right)\left(\sum_{k=1}^{J} \mathrm{B}_{mk}\Delta z_k\right)\right),$$

$$= \alpha_n\alpha_m \mathbb{E}\left(\left(\sum_{j=1}^{J} \underbrace{\frac{\beta_{nj}}{\sigma\left(\sum_{j=1}^{J}\beta_{nj}\Delta z_j\right)}}_{\text{Eq. 2.17}}\Delta z_j\right)\right.$$

$$\times \left(\left(\underbrace{\sum_{k=1}^{J} \frac{\beta_{mk}}{\sigma\left(\sum_{j=1}^{J} \beta_{mj}\Delta z_j\right)} \Delta z_k}_{\text{Eq. 2.17}} \right) \right),$$

$$= \left(\frac{\alpha_n}{\sigma\left(\sum_{j=1}^{J} \beta_{nj}\Delta z_j\right)} \right) \left(\frac{\alpha_m}{\sigma\left(\sum_{j=1}^{J} \beta_{mj}\Delta z_j\right)} \right)$$

$$\times \text{cov}\left(\sum_{j=1}^{J} \beta_{nj}\Delta z_j, \sum_{k=1}^{J} \beta_{mk}\Delta z_k \right),$$

$$= \alpha_n \left(\frac{\text{cov}\left(\sum_{j=1}^{J} \beta_{nj}\Delta z_j, \sum_{k=1}^{J} \beta_{mk}\Delta z_k \right)}{\sigma\left(\sum_{j=1}^{J} \beta_{nj}\Delta z_j\right) \sigma\left(\sum_{j=1}^{J} \beta_{mj}\Delta z_j\right)} \right) \alpha_m,$$

$$= \alpha_n \text{corr}\left(\sum_{j=1}^{J} \beta_{nj}\Delta z_j, \sum_{k=1}^{J} \beta_{mk}\Delta z_k \right) \alpha_m,$$

$$= \alpha_n \cdot \rho_{nm} \cdot \alpha_m,$$

where the final step follows from the definition of correlation. We use the succinct expression, ρ_{nm}, to describe the observed linear dependence between the factor-loaded systemic factors of the nth and mth credit obligors. This quantity can be referred to as factor correlation. Moreover, since $\text{var}(\Delta X_i) = 1$ for all choices of i, the covariance and correlation of the creditworthiness increments are equivalent. Practically, this means that

$$\text{cov}\left(\Delta X_n, \Delta X_m\right) = \text{corr}\left(\Delta X_n, \Delta X_m\right), \tag{2.22}$$

$$= \alpha_n \cdot \rho_{nm} \cdot \alpha_m.$$

As a quick sanity check, setting $n = m$, we have that $\rho_{nn} = 1$, by definition. The consequence is that α_n^2 is the contribution to variance from the systemic factor.

When adding the idiosyncratic component of variance, $1 - \alpha_n^2$, we replicate the unit variance result from Eq. 2.20.[30]

The most important conclusion from Eq. 2.22 is the interpretation of the α parameters. The individual α_n terms, which we should get in the habit of referring to as systemic weights, can be viewed as driving the correlation between the nth creditworthiness increment and the systemic factor. An easy way to see this is to consider a classic one-factor model,

$$\Delta X_n = \alpha_n \Delta G + \sqrt{1 - \alpha_n^2} \Delta w_n, \tag{2.23}$$

where ΔG is a single global risk factor and we can assume that β_n is equal to unity. If we compute the correlation of the creditworthiness increment with the single global factor, G, we have

$$\begin{aligned}
\operatorname{cov}(\Delta X_n, \Delta G) &= \operatorname{cov}\left(\alpha_n \Delta G + \sqrt{1 - \alpha_n^2} \Delta w_n, \Delta G\right), \tag{2.24} \\
&= \mathbb{E}\left(\left(\alpha_n \Delta G + \sqrt{1 - \alpha_n^2} \Delta w_n\right) \Delta G\right), \\
&= \alpha_n \underbrace{\mathbb{E}\left(\Delta G^2\right)}_{=1}, \\
&= \alpha_n.
\end{aligned}$$

A similar computation to that found in Eq. 2.21—in the univariate setting— reveals that $\operatorname{cov}(\Delta X_m, \Delta X_n) = \alpha_n \alpha_m$. The idea is analogous in our multivariate case. To handle the additional complexity of the systemic factors, however, the α_n and α_m terms need to load onto the more involved cross-correlation term, ρ_{nm}. We thus have a parsimonious and intuitive interpretation of the pairwise correlation between obligors in our credit portfolio. Moreover, this addition brings our model much closer, from a practical perspective, to the standard threshold-model implementation.

The preceding pages have been a flurry of changing notation and derivations. It is not always easy to follow all of the various steps. Table 2.2 attempts to help by chronicling the evolution of our state variables and coefficients throughout the discretization and normalization processes. Although the final column is basically our end point, there is value in following the steps from our stochastic differential equation—analogous to the one found in the Merton [30] work and more common a few decades ago—and the Vasicek [40, 41, 42] style formulation.

---

[30] Recall that, by independence, the idiosyncratic element impacts the variance measure, but does *not* contribute to covariance.

Table 2.2 *Keeping track of variables and coefficients*: This table helps us keep track of the various state variables, their sub-components, and coefficients as we work from the original stochastic differential equation to the final discretized and normalized form in Eq. 2.18.

Variable	Continuous time	Discretized	Normalized
State-variable increment	$dX_i(t)$	ΔX_i	ΔX_i
Systemic-factor increment	$dZ_j(t)$	Δz_j	Δz_j
Idiosyncratic-factor increment	$dW_i(t)$	Δw_i	Δw_i
Systemic-factor loading	$\tilde{\beta}_{ij}$	β_{ij}	B_{ij}
Idiosyncratic-factor loading	$\tilde{\sigma}_i$	σ_i	$\sqrt{1 - \alpha_i^2}$
Systemic weight	1	1	α_i^2

Colour and Commentary 16 (GAUSSIAN-THRESHOLD MODEL)*: Despite the heavy first step involving the introduction of Eq. 2.9, the situation has improved. Time discretization, factor-level normalization, and the introduction of a proper form for the systemic weights have had a useful effect. We can now clearly see that, at its heart, the legacy model essentially follows a multivariate Gaussian-threshold methodology.[a] Examination of the latent-variable correlation structure also reveals a rich interaction between systemic weights and systemic-factor correlations. Specifying these coefficients will actually bring the model to life. For the remainder of this chapter, we will keep the discussion fairly abstract. Chapter 3 will address the important question of model parametrization.*

---

[a] It is not, however, expressed in its canonical form. For readers interested in more background, Bolder [7, Chapter 4] examines the class of threshold models in significant detail.

2.3.3 A Matrix Formulation

Another awkward aspect of the base formulation is its scalar representation. Given the large potential numbers of correlated and independent risk drivers—along with their loadings—it makes much more sense to place our discretized creditworthiness system into matrix notation. This turns out, unfortunately, to be a bit complicated. Let us begin with the most natural representation

$$\underbrace{\begin{bmatrix} \Delta X_1 \\ \vdots \\ \Delta X_I \end{bmatrix}}_{I \times 1} = \underbrace{\begin{bmatrix} \alpha_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \alpha_I \end{bmatrix}}_{I \times I} \underbrace{\begin{bmatrix} \mathrm{B}_{11} & \cdots & \mathrm{B}_{1J} \\ \vdots & \ddots & \vdots \\ \mathrm{B}_{I1} & \cdots & \mathrm{B}_{IJ} \end{bmatrix}}_{I \times J} \underbrace{\begin{bmatrix} \Delta z_1 \\ \vdots \\ \Delta z_J \end{bmatrix}}_{J \times 1}$$

$$+ \underbrace{\begin{bmatrix} \sqrt{1-\alpha_1^2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sqrt{1-\alpha_I^2} \end{bmatrix}}_{I \times I} \underbrace{\begin{bmatrix} \Delta w_1 \\ \vdots \\ \Delta w_I \end{bmatrix}}_{I \times 1}. \tag{2.25}$$

If we define the base systemic weights as the following diagonal matrix,

$$A = \begin{bmatrix} \alpha_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \alpha_I \end{bmatrix}, \tag{2.26}$$

then it naturally follows that $(I - A^2)^{\frac{1}{2}} \in \mathbb{R}^{I \times I}$ is readily calculated.[31] Assigning reasonable symbols to the remaining matrices and vectors in Eq. 2.25, we can concisely express Eq. 2.19 as

$$\Delta X = AB\Delta z + (I - A^2)^{\frac{1}{2}} \Delta w, \tag{2.27}$$

where $B \in \mathbb{R}^{I \times J}$ is a matrix of normalized factor loadings, $\Delta z \in \mathbb{R}^{J \times 1}$ is a column vector of systemic risk factors, and $\Delta w \in \mathbb{R}^{I \times 1}$ is a column vector of idiosyncratic risk variables. Equation 2.27 is basically a simplified recipe for describing the entire system, $\Delta X \in \mathbb{R}^{I \times 1}$, of latent creditworthiness state variables in a single step. This quantity falls into the *nice-to-have* category, but it is not used very frequently.

While we have Eq. 2.27, let's put it to good use. The asset-return (or creditworthiness index) correlation matrix is of central importance in simulating economic-capital outcomes, trouble-shooting simulation results, and interpreting and communicating the model values. The more we can learn about it, the better our overall understanding. Examination in matrix notation has an important advantage of allowing us to see the entire picture. Let's see where Eq. 2.27 takes us. By construction, we have that $\Delta z \sim \mathcal{N}(0, \Omega)$ and $\Delta w \sim \mathcal{N}(0, I)$. As a consequence, it follows that

$$\text{var}(\Delta X) = \text{var}\left(AB\Delta z + \left(I - A^2\right)^{\frac{1}{2}} \Delta w \right), \tag{2.28}$$

$$= AB \underbrace{\text{var}(\Delta z)}_{\Omega}(AB)^T + \left(I - A^2\right)^{\frac{1}{2}} \underbrace{\text{var}(\Delta w)}_{I} \left(\left(I - A^2\right)^{\frac{1}{2}} \right)^T$$

---

[31] This follows from the fact that A is diagonal, but also that each α_i falls in the unit interval. Any negative α_i would otherwise lead to complex values. It should also be clear, from context, that the identity matrix has dimensions $I \times I$, but we'll underscore that point here to be certain.

$$= AB\Omega B^T A + \left(\left(I - A^2 \right)^{\frac{1}{2}} \right) \left(I - A^2 \right)^{\frac{1}{2}} \right),$$

$$= AB\Omega B^T A + I - \underbrace{AIA}_{A^2},$$

$$= I + A \left(B\Omega B^T - I \right) A,$$

which is an interesting, and somewhat unexpected, form.[32] Although each ΔX_i has unit variance, the full asset-correlation matrix is rather more complicated. In particular, we have that $\Delta X \sim \mathcal{N}\left(0, I + A \left(B\Omega B^T - I \right) A \right)$. All the main characters—systemic weights, normalized factor loadings, and factor correlations—make an appearance. The basic pattern is, upon examination, rather similar to the results from Eq. 2.21.

Equation 2.28, while interesting and occasionally useful, is *not* the most typical mathematical representation of our creditworthiness index. In most calculations, we find ourselves working at the credit obligor level. We typical use the less elegant, but surprisingly helpful, mixed-matrix representation. To get to this point, we first define the following fairly trivial, but convenient, row-vector quantity:

$$\beta_i = \begin{bmatrix} \beta_{i1} & \cdots & \beta_{iJ.} \end{bmatrix}. \tag{2.29}$$

This immediately allows us to restate the awkward double-sum in the factor-loading normalization from Eq. 2.17 as,

$$\sqrt{\beta_i \Omega \beta_i^T} = \sqrt{\sum_{\ell=1}^{J} \sum_{k=1}^{J} \beta_{i\ell} \beta_{ik} \Omega_{\ell k}}. \tag{2.30}$$

The normalized factor loadings thus become

$$\mathbf{B}_i = \begin{bmatrix} \mathbf{B}_{i1} & \cdots & \mathbf{B}_{iJ} \end{bmatrix}, \tag{2.31}$$

$$= \frac{1}{\sqrt{\beta_i \Omega \beta_i^T}} \overbrace{\begin{bmatrix} \beta_{i1} & \cdots & \beta_{iJ} \end{bmatrix}}^{\beta_i}.$$

---

[32] Indeed, it is hard to tell if it makes sense. The corner cases of $A \equiv 0$ and $A \equiv I$—representing the domination of idiosyncratic and systemic risk, respectively—collapse to values of I and $B\Omega B^T$ for Eq. 2.28. This quick sanity check suggests that the formulation, at least for the extremes, is mathematically reasonable.

There is nothing new in this representation, but it is rather easier to read (and manipulate). Using these constructs, for each i, we may now write our discretized creditworthiness index as

$$\Delta X_i = \alpha_i \underbrace{\frac{\beta_i}{\sqrt{\beta_i \Omega \beta_i^T}}}_{B_i} \Delta z + \sqrt{1 - \alpha_i^2} \Delta w_i, \qquad (2.32)$$

$$= \alpha_i B_i \Delta z + \sqrt{1 - \alpha_i^2} \Delta w_i.$$

In short, we have vectorized the interaction between the systemic variables and their factor loadings, but otherwise maintained a scalar structure.[33] Again, there is not much exciting going on here; Eq. 2.32 simply turns out to be both a convenient and parsimonious representation.

> **Colour and Commentary 17** (MIXED MATRIX NOTATION)*: It is entirely possible—and indeed, even advisable—to write our collection of I discretized and normalized creditworthiness indices in matrix notation. The final result is a concise and interesting mathematical representation. There is, however, a problem. Most common mathematical operations, in this setting, tend to occur at the obligor level. This means that the full matrix form is seldom employed. Instead, it is more common to work with the mixed-matrix notation summarized in Eq. 2.32. The interaction between the systemic variables and their factor loadings is vectorized, but otherwise the scalar structure has been maintained. Slightly clumsy, it nevertheless turns out to be an efficient and pragmatic way to represent the individual latent creditworthiness indices. In particular, this layered approach to the problem turns out to be very helpful in managing the parametrization questions addressed in Chap. 3.*

2.3.4 Orthogonalization

Before jumping to the remainder of the proper model construction, we will take a short detour. In the base formulation, the systemic factors are correlated. Abstracting from the systemic weights, this implies that the actual factor structure depends on a complicated combination of the factor loadings and the factor correlation matrix. Many industrial applications are constructed such that the systemic factors are actually orthogonal. This clearly offers some conceptual advantages. The good news

---

[33] With $B_i \in \mathbb{R}^{1 \times J}$ and $\Delta z \in \mathbb{R}^{J \times 1}$, their dot product (i.e., $B_i \Delta z$) happily reduces to a scalar.

is that, with a bit of work, Eq. 2.32 can be readily orthogonalized. In this brief aside, we will examine precisely how this might be done.

The discretized systemic factor correlation matrix of Δz is, by construction, symmetric, real-valued, and positive definite. The consequence is that we can decompose Ω as,

$$\Omega = CDC^T, \tag{2.33}$$

where $D \in \mathbb{R}^{I \times I}$ is a diagonal matrix of positive real-valued eigenvalues. $C \in \mathbb{R}^{I \times I}$ is the collection of orthonormal eigenvectors.[34] The properties of D allow us to immediately rewrite Eq. 2.33 as

$$\Omega = C \overbrace{\underbrace{D^{\frac{1}{2}} D^{\frac{1}{2}}}_{D}}^{\text{Eq. 2.33}} C^T, \tag{2.34}$$

$$= CD^{\frac{1}{2}} I \left(CD^{\frac{1}{2}} \right)^T,$$

where, once again, I denotes the identity matrix.[35]

This might seem like a curiosity, but it is just what we need. Recall that $\Delta z \sim \mathcal{N}(0, \Omega)$. We can actually reconstruct this quantity starting from $\Delta v \sim \mathcal{N}(0, I)$; that is, a vector of I i.i.d. standard normal variates. Indeed, we might simply define

$$\Delta z = CD^{\frac{1}{2}} \Delta v. \tag{2.35}$$

The expectation of Δv is clearly zero, but what about its variance? This is easily verified:

$$\text{var}(\Delta z) = \text{var} \left(CD^{\frac{1}{2}} \Delta v \right), \tag{2.36}$$

$$= CD^{\frac{1}{2}} \underbrace{\text{var}(\Delta v)}_{I} \left(CD^{\frac{1}{2}} \right)^T,$$

$$= CDC^T,$$

$$= \Omega.$$

---

[34] This is a special case of the eigenvalue problem, which is sometimes referred to as the spectral decomposition. Much more background can be found—in increasing levels of theoretical complexity—in Harris and Stocker [21, Section 9.9], Press et al. [35, Chapter 9], and Golub and Loan [15, Chapter 8].

[35] There are, it turns out, other ways to do this. We could define $U = CD^{\frac{1}{2}}$ implying that $\Omega = UU^T$, which is often referred to as the Cholesky decomposition of Ω.

The consequence is that $\Delta z = CD^{\frac{1}{2}}\Delta v \sim \mathcal{N}(0, \Omega)$. The immediate corollary is that by introducing a slight variation on our normalized factor loadings as

$$\breve{B}_i = B_i CD^{\frac{1}{2}}, \tag{2.37}$$

we can restate Eq. 2.32 as,

$$\Delta X_i = \alpha_i \breve{B}_i \Delta v + \sqrt{1 - \alpha_i^2}\,\Delta w_i. \tag{2.38}$$

To avoid more repetitive calculations, we leave it as an exercise for the reader to verify that ΔX_i retains unit variance—that is, $\Delta X_i \sim \mathcal{N}(0, 1)$—under this transformation.[36]

Ultimately, it makes little practical difference if one defines the model using the B_i or $\breve{B}_i$ formulation of the factor loadings. The orthogonalized implementation offers conceptual clarity in the interpretation of the coefficients, while parametrization is slightly easier to manage with the base approach. It is ultimately a question of taste. The current implementation employs the B_i definition, but we are keenly aware that both methods are entirely legitimate.[37]

2.4 The Legacy Model

Our legacy credit-risk economic capital model—used in production until late 2020—was a version of the multivariate Gaussian threshold model. Thus far, we have principally focused on the asset-return, or creditworthiness, state variables. In this section, we will incorporate the remaining elements required to flesh out the full-blown model.

2.4.1 Introducing Default

Each individual default event is defined, in the spirit of the threshold model introduced in the first section, as

$$\mathbb{I}_{\{\Delta X_i \leq K_i\}}, \tag{2.39}$$

---

[36] As a hint, it suffices to show that $\breve{B}_i \breve{B}_i^T = 1$.

[37] In Chap. 11, we will actually exploit this orthogonalized form within some interesting Pillar II regulatory calculations.

where K_i represents an as-yet-undefined default threshold. In words, this implies that if the realization of the creditworthiness index falls below K_i, then the obligor is deemed to have entered the default state. Following the Merton [30] logic, we can think of the K_i value as being somehow related to the firm's liabilities.

The immediate challenge is to determine a sensible value for K_i. While there is a range of possible choices, one option dramatically simplifies the problem and provides an entirely logical answer. The basic idea is to calibrate the expected value of the default event to the probability of default associated with the ith credit obligor. Defining the one-period default probability of the ith credit counterpart as p_i, we have

$$\mathbb{E}\left(\mathbb{I}_{\{\Delta X_i \leq K_i\}}\right) = p_i, \tag{2.40}$$

$$\underbrace{\mathbb{P}\left(\Delta X_i \leq K_i\right)}_{\Delta X_i \sim \mathcal{N}(0,1)} = p_i,$$

$$\Phi\left(K_i\right) = p_i,$$

$$\Phi^{-1}\left(\Phi\left(K_i\right)\right) = \Phi^{-1}\left(p_i\right),$$

$$K_i = \Phi^{-1}\left(p_i\right),$$

where $\Phi(\cdot)$ and $\Phi^{-1}(\cdot)$ denote the cumulative and inverse cumulative standard normal distribution functions, respectively.[38] This allows us to directly restate the default event more concretely as,

$$\mathbb{I}_{\{\Delta X_i \leq \Phi^{-1}(p_i)\}}, \tag{2.41}$$

for $i = 1, \ldots, I$. In turn, this definition permits us to write down a straightforward expression for the total default loss in one's portfolio as,

$$L_{\mathcal{D}} = \sum_{i=1}^{I} c_i \underbrace{(1 - \mathcal{R}_i)}_{\gamma_i} \mathbb{I}_{\{\Delta X_i \leq \Phi^{-1}(p_i)\}}, \tag{2.42}$$

where—drawing from Table 2.1—c_i, $\mathcal{R}_i$, and γ_i denote the ith counterparty's exposure, recovery rate, and loss-given-default, respectively. In a one-period, default-only setting, the distribution of $L_{\mathcal{D}}$ is precisely our modelling object of interest.

The default probability estimate, p_i, is an unconditional estimate of the failure of the ith obligor to meet its credit obligations—this value relates only to each counterpart on an individual basis. The entire justification for the addition of ΔX_i, however, was to induce dependence between these default events. To understand

---

[38] In principle, the periodicity of the default probability should be consistent with the length of the discretized time step ΔX_i.

better how this works, we can examine the probability of a default event conditional upon a given realization of the set of systemic risk factors. This is defined as,

$$\mathbb{P}\left(\Delta X_i \le \Phi^{-1}(p_i)\middle| \Delta Z\right) = \mathbb{P}\left(\underbrace{\alpha_i B_i \Delta z + \sqrt{1-\alpha_i^2}\Delta w_i}_{\text{Eq. 2.32}} \le \Phi^{-1}(p_i)\middle| \Delta z\right), \quad (2.43)$$

$$p_i(\Delta z) = \mathbb{P}\left(\Delta w_i \le \frac{\Phi^{-1}(p_i)-\alpha_i B_i \Delta z}{\sqrt{1-\alpha_i^2}}\middle| \Delta z\right),$$

$$= \Phi\left(\frac{\Phi^{-1}(p_i)-\alpha_i B_i \Delta z}{\sqrt{1-\alpha_i^2}}\right)$$

$$\equiv \boxed{\Phi\left(\frac{\Phi^{-1}(p_i)-\alpha_i \breve{B}_i \Delta v}{\sqrt{1-\alpha_i^2}}\right)},$$

which follows directly from the fact that the idiosyncratic shock, Δw_i, is a standard normal variate. Just for fun, Eq. 2.43 also includes the orthogonalized version of the model. In either case, the consequence is that large negative outcomes of the systemic risk factors will tend, for all obligors, to push the probability of default upwards. This is the source of default dependence. Since different obligors load onto the individual systemic risk factors in varying ways, the actual dependence structure is actually quite complex. Nevertheless, this quantity provides significant insight into the inner workings of our credit-risk model.

> **Colour and Commentary 18** (DEFAULT CONDITIONALITY)*: The threshold model induces default correlation by conditioning each credit ogligor's latent creditworthiness state variable upon a common set of systemic risk factors. A positive systemic risk-factor realization will improve all credit counterparties' default probabilities, whereas a negative draw works in the opposite direction. Given the systemic risk-factor realization, however, all default events are independent. Moreover, each credit obligor—through the specification of the α and B parameters—are potentially impacted in a different way by the systemic values. The conditional default probability, as described in Eq. 2.43, is a rather useful object in understanding the impact of systemic risk-factor outcomes on a given credit obligor's risk profile. As a final point, the form changes only very slightly when using the dependent or orthogonalized formulations of systemic risk factors.*

2.4.2 Stochastic Recovery

The magnitude of the credit loss also depends, as suggested in Eq. 2.42, upon the amount of the total credit exposure *recovered* through the default process. The larger the amount recovered, of course, the lower the default loss. We can think of this recovery amount as taking a value from zero to unity. A recovery value of zero would imply that, in the event of default, one would be in the unfortunate situation of losing one's entire exposure. Conversely, a recovery value of one suggests that, happily, no loss is incurred.[39] The entire amount is recovered. Naturally, these are extremes. Typical values, therefore, fall in the open interval, (0, 1).

For each exposure—similar to the default probability—one thus needs to estimate, determine, or otherwise assign a recovery rate in the event of default. Rather than recovery, which we might refer to as $\mathcal{R}$, it is more common to work with the loss-given-default. This latter quantity is simply $\gamma \equiv 1 - \mathcal{R}$; in essence, therefore, recovery and loss-given-default are opposite sides of the same coin.

The loss-given default associated with a credit obligation depends, again similar to the default probability, upon a complex interaction between a range of factors. The financial strength of the firm plays, of course, an important role. Perhaps equally important, however, are the seniority of the claim and the presence of any external guarantees or collateral attached to the loan or treasury asset. These latter points are intimately connected to the inevitable legal aspects associated with the default of any entity. Related to this point is the type of organization; a corporation, a financial institution, a public-sector entity, or a government. Each of these organizational types will, in principle, have different characteristics impacting the loss-given-default in varying ways. Most organizations, and NIB is no exception, have developed a fairly involved loss-given-default framework describing the underlying process of how these values are determined for each of their individual asset exposures.[40]

The simplest modelling treatment of this quantity would involve assignment of a constant loss-given default—let's assume, as is typically the case, that this value is determined by the firm's loss-given-default framework—to each of the individual exposures within one's portfolio. This is a sensible approach, but it ignores the potential uncertainty in the actual size of the final loss. As previously discussed, a multiplicity of factors exerts an influence on the final loss-given-default outcome.[41] Even the most detailed loss-given-default framework will struggle to determine

---

[39] Naturally, in reality, it is not quite so simple. There are often significant time lags, foregone cash-flows, and additional transaction costs. This is, however, a model and thus its representation is defensibly stylistic.

[40] How precisely this is performed at NIB is beyond the scope of this discussion, which focuses on the modelling dimension. This is not to say that no modelling is involved in LGD specifications, far from it, but this area often touches on rather proprietary dimensions.

[41] Altman et al. [2] provides a useful broad based analysis of the factors driving recovery with a particular focus on the link to default probabilities.

precisely the importance of each underlying driver to a specific credit obligor. Assignment of a single, deterministic loss-given-default value is thus probably overly simplistic. A more complicated, and perhaps more realistic, alternative would be to treat the actual loss-given default as a random variable. Such an approach would certainly complicate the model implementation, but it would directly address this important source of uncertainty. For this reason, the employment of stochastic loss-given-default values has become standard practice in credit-risk models.

How might this work? We can succinctly restate—from Eq. 2.42—the default-loss of our portfolio as,

$$L_{\mathcal{D}} = \sum_{i=1}^{I} c_i \gamma_i \mathbb{I}_{\mathcal{D}_i}, \tag{2.44}$$

where $\mathcal{D}_i \equiv \Delta X_i \leq \Phi^{-1}(p_i)$ and $\gamma_i \sim \mathcal{X}(\bar{\gamma}_i, v_{\gamma_i}, \cdots)$. In actuality, we have done relatively little with this restatement. The ith loss-given-default quantity has merely been specified as a yet-to-be-defined random variable given its first two moments: the mean, $\bar{\gamma}_i$ and the variance, v_{γ_i}. Clearly a choice is required regarding the distribution of each γ_i, but this is not the only choice required. One also needs to decide on the dependence between the default event, $\mathbb{I}_{\mathcal{D}_i}$, and the loss-given-default. Are they somehow related or are they independent? Moreover, it is necessary to decide upon any relationship between the loss-given-default of any two credit obligors, say γ_i and γ_j.

Before exploring possible marginal distributions for each individual loss-given-default parameter, let us first consider questions of dependence. Common practice involves the assumption of independence between the loss-given-default outcome of any two arbitrary credit obligors, i and j. Economic intuition aside, it is difficult to practically imagine how calibration of such dependence would work when examining relatively high credit-quality obligors. The loss-given default outcome comes into play only in the case of default. As default is rare, it only affects a small number of credit obligors—very often none—in any given simulation draw. Imposition of correlation at the individual credit obligor level is unlikely to have any appreciable impact on risk outcomes, quite simply because joint defaults are so rare. This is the same reasoning behind the use of global conditioning variables to induce default correlation; default correlation at the obligor level is relatively ineffective.

It is very common to assume that the loss-given-default is an entirely idiosyncratic affair. Although loss-given-default is a random variable, each outcome is independent of the set of global systemic variables, a firm's default probability, and the loss-given-default events associated with other credit obligors. We will also make this choice in our development. More complex and involved specifications are, indeed, possible.[42] The typical viewpoint leans on the fact that virtually any

---

[42] Frye [11], for example, offers an alternative in the one-factor setting that, like the conditional default probability, depends on the common global macroeconomic variable. Frye [12], conversely,

economic-capital model is already rather complex and parameter selection, in this setting, is extremely difficult. Moreover, it is important that the loss-given-default outcome employed in our economic-capital model is conceptually consistent with the assigned value from one's loss-given-default framework. This is not a deviation from general practice, it is principally a choice driven by a combination of analytic convenience and a lack of informative data.

One useful advantage of the assumption of independence is a relatively straightforward computation of the mean of the portfolio default loss. In particular, working from Eq. 2.44, the mean is given as,

$$\mathbb{E}\left(L_{\mathcal{D}}\right) = \mathbb{E}\left(\sum_{i=1}^{I} c_i \gamma_i \mathbb{I}_{\mathcal{D}_i}\right), \tag{2.45}$$

$$= \sum_{i=1}^{I} \mathbb{E}\left(c_i \delta_i \mathbb{I}_{\mathcal{D}_i}\right),$$

$$= \sum_{i=1}^{I} c_i \underbrace{\mathbb{E}\left(\gamma_i\right) \mathbb{E}\left(\mathbb{I}_{\mathcal{D}_i}\right)}_{\text{By independence}},$$

$$= \sum_{i=1}^{I} c_i \bar{\gamma}_i \underbrace{\mathbb{P}(\mathcal{D}_i)}_{p_i}.$$

The expected loss is thus simply the sum of the product of the exposures, the average loss-given-defaults, and the probabilities of default. This further underscores the central importance of these *three* fundamental quantities, which we introduced in Table 2.1.

The final decision involves a choice of distribution for each random loss-given-default variable. There are many possible choices. One important restriction, however, is that the support of this random variable is confined to the unit interval. A trivial possibility would involve the use of a standard uniform distribution. Such a choice implies that every point in the unit interval has an equal probability. This seems unreasonable and can be rejected out of hand, not least because it implies a common mean loss-given-default outcome of 0.5 for all credit obligors.

Ignoring the standard uniform distribution, there are two main approaches to the selection of a suitable distribution. One can find a distribution with support restricted to a fixed interval or transform a random variable with infinite support into the interval between zero and one. One could, for example, draw a random variable from the normal distribution and then map it to (0, 1) using a logistic or

---

offers another possibility where the loss-given-default is a stylized function of the conditional-default probability.

probit transformation.[43] There are rather fewer distributions with bounded support. The classic choice is the beta distribution.[44] There are numerous other choices that, upon closer inspection, are ultimately special cases of the beta distribution. Another, perhaps less common choice, is the so-called Kumaraswamy distribution; see Kumaraswamy [26] for more detail.

We have, once again, opted for simplicity. Each individual loss-given-default random variable is assumed to follow an independent beta distribution. The parameters of each beta distribution are informed, at least in part, from one's internal loss-given-default framework. The specific parametric choices are discussed, in much more detail, in Chap. 3.[45] The important takeaway, at this point, is that our approach incorporates random recovery in a relatively straightforward—and to the extent possible—conservative fashion.

Colour and Commentary 19 (PARAMETRIZING RECOVERY)*: The management of the recovery dimension in the measurement of credit risk is a complicated task. The amount recovered subsequent to a default event—or, closely related, the notion of loss-given-default—is* not *known in advance: it is a stochastic quantity. Although ultimately random, it nevertheless depends upon a range of factors: an obligor's organizational structure, its financial strength, its legal domicile, and any credit-mitigation measures to name a few important elements. Taking this into account, our methodology assigns a separate marginal distribution—describing the interaction between recovery outcomes and likelihoods—to each credit obligor. These marginal distributions, while important, do not tell the entire story. It is also necessary to specify any dependence between the various recovery processes and the default outcomes. Our production model, consistent with general practice, assumes independence between recovery and default. This enhances model tractability and avoids some rather thorny parametric questions. It is an expedient choice, forced upon us by practical data constraints, but it is important to be aware that it is probably not quite correct.*

---

[43] These, of course, are only two popular transformations of a continuous-valued variable into (0, 1); there are many others.

[44] See Johnson et al. [22, Chapter 25] for significantly more background on this distribution.

[45] This discussion incorporates a unique angle that attempts to capture an important stylized fact about loss-given-default.

2.4.3 Risk Metrics

As highlighted in the previous chapter, the economic capital calculation is the distance between a worst-case measure of risk and the accounting- or valuation-based expected loss. Equation 2.45 provides clear direction for the computation of expected loss. To determine the credit-risk economic capital estimate—and complete the basic structure of our model—we still need a worst-case loss quantity. Since there are multiple possible candidates to measure worst-case losses, a few decisions need to be taken. In this brief section, we'll consider *two* main alternatives to motivate our choice.

Using the Gaussian threshold model outlined in the previous sections, we use simulation to trace out the entire credit loss distribution. This provides all of the ingredients required to actually compute virtually any desired measure of portfolio riskiness. We do not, however, typically consider the entire loss distribution. As rather pessimistic risk-management professionals, we invariably focus our attention on the tail of the loss distribution. This is, after all, where all the bad things happen. The classic measure is referred to as Value-at-Risk or VaR. Originally suggested by Morgan/Reuters [32], within the market-risk setting, it is described mathematically as

$$\mathrm{VaR}_\alpha(L) = \inf\left(x : \mathbb{P}\left(L \leq x\right) \geq 1 - \alpha\right), \qquad (2.46)$$

where $\inf(\cdot)$ denotes the infimum operator.[46] Equation 2.46 depends on *two* parameters: α a threshold for the probability on the right-hand side and the default loss, L. The first is a parameter, whereas the second is a random variable: portfolio default-loss. α is also referred to as the confidence level. Imagine it is 0.99. With this parameter choice, $\mathrm{VaR}_\alpha(L)$ describes the largest default-loss outcome exceeding 99% of all possible losses, but is itself exceeded in 1% of all cases. When computed for default or migration risk, it is often referred to as credit VaR.[47]

With a bit of additional effort, we may simplify our abstract specification of VaR. From Eq. 2.46, the VaR measure, by definition, satisfies the following relation,

$$\int_{\mathrm{VaR}_\alpha(L)}^{\infty} f_L(\ell)d\ell = \alpha, \qquad (2.47)$$

$$1 - \mathbb{P}\left(L \geq \mathrm{VaR}_\alpha(L)\right) = 1 - \alpha,$$

---

[46] The infimum is basically—avoiding lots of mathematical formalism—a set-theoretical generalization of the minimum operator. See Royden [37] for the important details.

[47] An excellent reference for all things related to VaR is Jorion [23].

$$\mathbb{P}\left(L \leq \text{VaR}_\alpha(L)\right) = 1 - \alpha,$$

$$F_L\left(\text{VaR}_\alpha(L)\right) = 1 - \alpha,$$

where $F_L(\cdot)$ denotes the cumulative default-loss distribution function. The natural solution is thus,

$$\text{VaR}_\alpha(L) = F_L^{-1}\left(1 - \alpha\right). \tag{2.48}$$

In words, Eq. 2.48 verifies that $\text{VaR}_\alpha(L)$ is nothing other than the $(1 - \alpha)$-quantile of our portfolio credit-loss distribution. Since we are already in the business of running millions of simulations of our model, quantiles are readily estimated by simply ordering the simulation results. Readily computed, easy-to-use, understand, and communicate, it is no wonder that VaR has become such a popular risk measure.

A few decades prior to the writing of this book, Artzner et al. [3] produced a very useful paper providing, in an axiomatic manner, the main desirable properties of a risk metric. Risk measures that fulfil all properties are referred to as *coherent*. VaR was, unfortunately, found wanting along one important dimension.[48] This has led to a search for alternatives. An increasingly common (and coherent) measure, is defined as the following conditional expectation

$$\mathcal{E}_\alpha(L) = \mathbb{E}\left(L \middle| L \geq \text{VaR}_\alpha(L)\right), \tag{2.49}$$

$$= \frac{1}{1 - \alpha} \int_{\text{VaR}_\alpha(L)}^\infty \ell f_L(\ell) d\ell,$$

where $f_L(\ell)$, as before, represents the default-loss density function. Equation 2.49 describes the expected default-loss given that one finds oneself at, or beyond, the $(1 - \alpha)$-quantile (i.e., $\text{VaR}_\alpha(L)$) level. This quantity is, as a consequence, termed the conditional VaR, the tail VaR, or the expected shortfall. We predominately use the term expected shortfall and refer to it symbolically as $\mathcal{E}_\alpha(L)$ to explicitly include the desired quantile defining the tail of the return distribution.[49] Figure 2.1 provides a visualization of the difference between these two central measures of risk.

There are, of course, other possible choices of risk metric that one might consider. Boyle et al. [8] chronicle the disadvantages of VaR—in terms of coherence— and offer a twist on the idea of expected shortfall. Gilli and Këllezi [13] propose

---

[48] In particular, VaR is not sub-additive. See McNeil et al. [29, Chapter 2.3] for a nice exposition of this point.

[49] Given a closed-form expression for the default-loss density, Eqs. 2.48 and 2.49 can often be manipulated to derive analytic formulae for the VaR and expected-shortfall measures. In practical industrial applications, unfortunately, we rarely find ourselves in this happy situation.
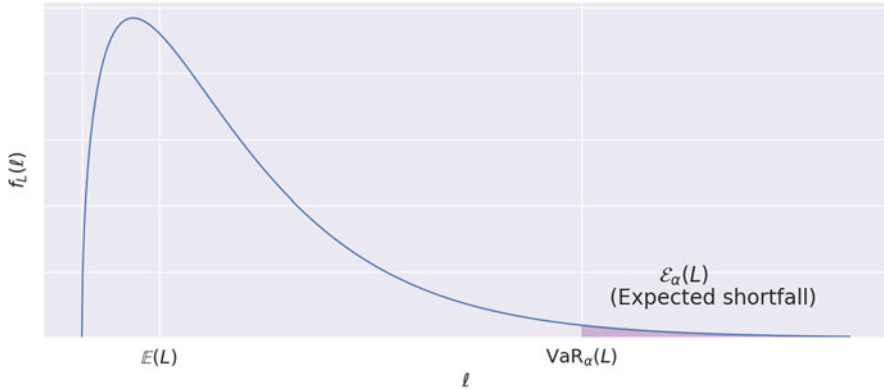
Fig. 2.1 *Visualizing risk metrics*: This graphic, via the illustration of a generic credit-loss density $f_L(\ell)$, illustrates the VaR and expected shortfall measures. VaR is a quantile of the loss distribution, while expected shortfall is the *average* at or beyond a specific quantile.

some ideas from extreme-value theory. Although quite interesting, and potentially powerful, these notions have not yet found popularity in industrial applications. For the moment, therefore, the principal alternatives used in practical circles are the presented VaR and expected-shortfall measures.

Traditionally, like most organizations, NIB has used the VaR measure with varying confidence levels for alternative applications. The production measure in the revised credit-risk economic capital model has nonetheless broken with tradition and moved to expected-shortfall measure. VaR's non-coherence does give some reason for pause, and was certainly a contributing factor in the final decision, but it is not the key factor. A more convincing motive is the inherent conservatism associated with expected shortfall. Instead of looking at losses associated with a fixed quantile, it considers the average losses beyond this point. To put it bluntly, *all* of the really bad outcomes are structurally incorporated into the calculation of expected shortfall. This is, for a risk manager, an attractive perspective.

The final deciding factor in favour of expected shortfall, however, relates to a practical point. A useful, well-functioning, credit-risk economic capital model requires attribution of overall capital to individual loans and treasury instruments. This process is referred to risk attribution; the actual approach will be addressed in the final section of this chapter. VaR-measure risk attribution in the simulation setting turns out to be a messy and noisy affair. Conversely, the same computation applied to expected shortfall is significantly more robust and stable. The central importance of risk attribution, along with the theoretical advantages of expected shortfall, ultimately tipped the scales in favour of this coherent risk measure.

Colour and Commentary 20 (CHOOSING A RISK METRIC)*: A simulation-based, credit-risk model—such as that employed in this development—furnishes the analyst with a full description of the portfolio credit-loss distribution. To transform this object into an estimate of economic capital, a specific choice regarding the description of worst-case losses must be made. While there are many possible risk-metric options to be found in the literature, there are really only* two *main contenders: Value-at-Risk (i.e., VaR) and expected shortfall. Like most organizations, NIB has historically employed the VaR metric in their economic-capital computations. In the most recent revision of our framework, the situation changed. A decision was taken to move to the expected-shortfall measure. Three main reasons motivated this choice. The first is theoretical; in the jargon of Artzner et al. [3] expected shortfall is—unlike VaR—a* coherent *risk measure. The second reason is conceptual. By incorporating all extreme tail observations into its calculation, expected shortfall is both a more complete and conservative description of downside risks. The final reason is practical; simulation-based methods for attribution of risks to individual loans and investments are dramatically more robust for expected shortfall.[a] This combination of sensible reasons feels like sufficient justification for the move.*

_____

[a] The precise details of this calculation follow in the final section of this chapter.

2.5 Extending the Legacy Model

Starting from a generic description of creditworthiness dynamics, we have established that the legacy model is a (non-canonical) multivariate, Gaussian threshold model with stochastic recovery. We have introduced a detailed notation, identified the key aspects of the model and even motivated our choice of risk metric. With the important question of parameter selection relegated to Chap. 3, this might conclude our discussion. As indicated on numerous occasions in previous sections and chapters, however, our 2020 statutory change (quite naturally) prompted reflection on the structure of the credit-risk economic capital model. The reflection lead to action. The *two* central consequences were associated with the choice of threshold-model copula function and the incorporation of credit migration. This section walks through the details, and implications, of these two central elements.

2.5.1 *Changing the Copula*

The legacy credit-risk economic-capital model was founded on the assumption that both the idiosyncratic and systemic risk factors are normally distributed. This choice is, in a general sense, referred to as the Gaussian copula. The Gaussian copula, following the recent financial crisis starting in 2008, has nonetheless come under strong criticism. The source of this disapproval relates to some very strong assumptions about default correlation. A well-cited article—see MacKenzie and Spears [28]—has even made the Gaussian-copula specification rather infamous. The main problem is that it ignores the idea of *tail dependence*. The basic idea is that as we move far enough into the tail of the credit-loss distribution, default becomes independent in the Gaussian copula. This is hardly confidence inspiring. One practical solution, offered in the literature, is to use the so-called *t*-copula; this means that idiosyncratic and systemic risk factors follow a *t* distribution.[50]

The notion of tail dependence is neither particularly well known nor is it extremely straightforward to explain without resort to rather technical arguments.[51] It is nonetheless worth deeper reflection. Tail dependence is conceptually analogous to more commonly used notions of correlation such as linear or rank correlation coefficients. Typically, one begins with two random variables. In our case, let's give these random variables specific names, $\mathbb{I}_{\mathcal{D}_i}$ and $\mathbb{I}_{\mathcal{D}_j}$, naturally representing the default events of two arbitrarily selected credit obligors. Default correlation considers the simple linear dependence between these two random quantities. Tail dependence, conversely, considers the relationship between these two events as we move arbitrarily far into the tail of their joint credit-loss distribution.

Default is, for almost any credit obligor, a rare event. A joint default is worse: it is the coincidence of two rare (i.e., tail) events. The incidence of joint default tells us something about the notion of tail dependence. The Gaussian copula implies, unfortunately, by its very mathematical structure, that as the two events become sufficiently rare, they tend towards independence. In most portfolios, however, it is precisely the multiplicity of default that generates sizable risk scenarios. Since the probability of two independent events is their product, and default is by definition a low-probability outcome, joint defaults are structurally underestimated in the Gaussian setting. Although the notion is quite technical, the impact is easy to understand.

An example provides a useful demonstration of the difference associated with one's choice of copula function. Figure 2.2 provides the results associated with a simple experiment. We examine our credit portfolio for a given date in the spring of 2020. Two credit counterparties were arbitrarily selected. We then investigate the joint default outcomes associated with one million simulations. Using the same

---

[50] Kuhn [25] and Kostadinov [24] both make compelling cases for the benefits of the *t*-distributed threshold implementation.

[51] For the reader interested in a more formal description, please see Bolder [7, Chapter 4] and the multiple references it contains.
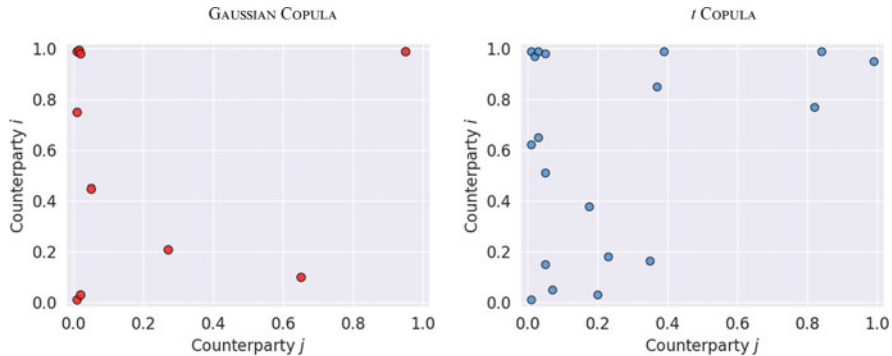
Fig. 2.2 *Joint default*: Joint defaults lie at the heart of default correlation and the conservative description of the worst-case credit-loss outcomes. This figure provides a visual perspective on the differences between the distribution of joint defaults under the Gaussian- and t-copula specifications of the threshold model. In each case, the proportional joint-default losses of two arbitrarily selected obligors are displayed under our competing copula functions.

randomly simulated outcomes *two* copula functions are employed: the presented Gaussian and the t copulas. To be clear, therefore, the only difference relates to the choice of copula function.[52] There are, of course, cases where one credit counterparty defaults, whereas the other does not. Only those situations, to be clear, where both of the obligors experience a default outcome are presented. Generally, of course, there are far fewer incidences of two counterparties defaulting concurrently. In the Gaussian case, there are precisely 10 joint defaults. The t-copula case, conversely, exhibits a twofold increase in the incidence of joint default. This is precisely the impact of (non-zero) tail dependence; a more realistic, and conservative, description of joint default is created.

The results presented in Fig. 2.2 might not seem like a dramatic difference, but expanding this notion across all pairs of credit counterparts across an entire portfolio can, and does, have an important impact on the final risk figures. There is, at the portfolio level given the parameter settings, only about a 20% increase in joint default. At the individual obligor level, however, when joint default does occur, on average it leads to a sizable increase in the number of joint defaults. This, of course, varies from a small reduction in some cases to increases on the order of many multiples of the base number of Gaussian-copula joint defaults. The bottom line, therefore, is that tail dependence matters quite importantly in terms of joint default and, ultimately, practical portfolio-level default correlation. The final consequence is a more realistic and conservative estimate of economic capital.

---

[52] There is also some noise associated with the random draws associated with the recovery outcomes, but this only impacts the magnitude of the loss, not their incidence.

2.5.2 Constructing the t Copula

To construct the t-copula model, the vast majority of the model infrastructure remains unchanged. The creditworthiness process is slightly adjusted as follows,

$$\Delta X_i = \sqrt{\frac{\nu}{W}} \underbrace{\left(\alpha_i B_i \Delta z + \sqrt{1 - \alpha_i^2} \Delta w_i \right)}_{\text{Eq. 2.32}}.$$
(2.50)

where $W \sim \chi^2(\nu)$ is referred to as the mixing variable. In other words, W is a common variable shared across all risk obligors within a given simulation that follows a χ^2 distribution with ν degrees of freedom. Equation 2.50 is a special case of what is generally referred to as a normal-variance mixture model.[53]

While it is not obvious from Eq. 2.50, the $\sqrt{\frac{\nu}{W}}$ terms transforms each y_n into a univariate standard t-distributed random variable.[54] This implies that $\mathbb{E}(y_n) = 0$ and $\text{var}(y_n) = \frac{\nu}{\nu-2}$ for all $n = 1, \ldots, N$. In other words, the marginal distribution of each ΔX_i follows a standard t distribution, while simultaneously moving the joint distribution of the collection of $\{\Delta X_i \ : \ i = 1, \ldots, I\}$ to a multivariate t distribution. The final result is the so-called t-threshold model.

The new moments of ΔX_i merit demonstration. The expected value of Eq. 2.50 is,

$$\mathbb{E}\left(\Delta X_i\right) = \mathbb{E}\left(\sqrt{\frac{\nu}{W}} \left(\alpha_i B_i \Delta z + \sqrt{1 - \alpha_i^2} \Delta w_i \right) \right),$$
(2.51)

$$= \alpha_i \cdot \overbrace{\mathbb{E}\left(\sqrt{\frac{\nu}{W}} \right) B_i \underbrace{\mathbb{E}\left(\Delta z\right)}_{=0}}^{\text{By independence}} + \sqrt{1 - \alpha_i^2} \cdot \overbrace{\mathbb{E}\left(\sqrt{\frac{\nu}{W}} \right) \underbrace{\mathbb{E}(\Delta w_i)}_{=0}}^{\text{By independence}},$$

$$= 0,$$

as expected.

Finding the variance of ΔX_i is a bit more work,

$$\text{var}\left(\Delta X_i\right) = \mathbb{E}\left(\Delta X_i^2 - \underbrace{\mathbb{E}(\Delta X_i)^2}_{=0} \right),$$
(2.52)

---

[53] Quite simply, different choices of W and ν induce alternative copula functions. See Bolder [7, Chapter 4] for much more information on this notion (and many other helpful references).

[54] See Bolder [7, Appendix A] for a detailed description of the construction of the t-distribution.

$$= \mathbb{E}\left(\left(\sqrt{\frac{\nu}{W}}\left(\alpha_i B_i \Delta z + \sqrt{1-\alpha_i^2}\Delta w_i\right)\right)^2\right),$$

$$= \underbrace{\mathbb{E}\left(\frac{\nu}{W}\cdot \alpha_i^2 \cdot B_i \Delta z \Delta z^T B_i^T\right) + \mathbb{E}\left(\frac{\nu}{W}\cdot(1-\alpha_i^2)\cdot \Delta w_i^2\right)}_{\text{Cross terms vanish by independence and zero expectations}},$$

$$= \alpha_i^2 \cdot \nu \cdot \mathbb{E}\left(\frac{1}{W}\right) B_i \underbrace{\mathbb{E}\left(\Delta z \Delta z^T\right)}_{\Omega} B_i^T$$

$$+ (1-\alpha_i^2)\cdot \nu \cdot \mathbb{E}\left(\frac{1}{W}\right)\underbrace{\mathbb{E}\left(\Delta w_i^2\right)}_{=1},$$

$$= \alpha_i^2 \cdot \nu \cdot \mathbb{E}\left(\frac{1}{W}\right)\underbrace{B_i \Omega B_i^T}_{=1} + (1-\alpha_i^2)\cdot \nu \cdot \mathbb{E}\left(\frac{1}{W}\right),$$

$$= \cancel{\alpha_i^2 \cdot \nu \cdot \mathbb{E}\left(\frac{1}{W}\right)} + \nu \cdot \mathbb{E}\left(\frac{1}{W}\right) - \cancel{\alpha_i^2 \cdot \nu \cdot \mathbb{E}\left(\frac{1}{W}\right)},$$

$$= \nu \cdot \mathbb{E}\left(\frac{1}{W}\right),$$

where we need to observe that $B_i \Omega B_i^T = 1$; this is the normalized, factor-loaded variance.[55] Resolving this expression now reduces to evaluating the reciprocal of a chi-squared distribution with ν degrees of freedom. Although tedious, it is readily determined from first principles by solving the following integral,

$$\mathbb{E}\left(\frac{1}{W}\right) = \int_{\mathbb{R}_+}\frac{1}{w}f_W(w)dw, \qquad (2.53)$$

$$= \int_{\mathbb{R}_+}\frac{1}{w}\underbrace{\frac{1}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)}w^{\frac{\nu}{2}-1}e^{-\frac{w}{2}}}_{W\sim\chi^2(\nu)}\,dw,$$

$$= \frac{1}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)}\int_{\mathbb{R}_+}w^{\frac{\nu-2}{2}-1}e^{-\frac{w}{2}}dw,$$

---

[55] It was, in fact, constructed in such a way to be equal to one.

$$= \frac{2^{\frac{\nu-2}{2}}\Gamma\left(\frac{\nu-2}{2}\right)}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)} \int_{\mathbb{R}_+} \overbrace{\frac{1}{2^{\frac{\nu-2}{2}}\Gamma\left(\frac{\nu-2}{2}\right)} w^{\frac{\nu-2}{2}-1} e^{-\frac{w}{2}}}^{\chi^2(\nu-2)}\, dw,$$

$$\underbrace{\qquad\qquad\qquad\qquad}_{=1}$$

$$= \frac{1}{2} \frac{\Gamma\left(\frac{\nu-2}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)},$$

$$= \frac{1}{2} \frac{\left(\frac{\nu-2}{2}-1\right)!}{\left(\frac{\nu}{2}-1\right)!},$$

$$= \frac{1}{2} \frac{1}{\left(\frac{\nu}{2}-1\right)},$$

$$= \frac{1}{\nu-2},$$

which directly implies from Eq. 2.52 that, indeed as we claimed, $\mathrm{var}(\Delta X_i) = \frac{\nu}{\nu-2}$. The final line of Eq. 2.53 follows from two useful facts: $\nu \in \mathbb{N}_+$ and $\Gamma(n) = (n-1)!$.

The marginal systematic and idiosyncratic variable distributions of the t-distributed model remain Gaussian. The joint and marginal distributions of the latent default-state variables (i.e., $\{\Delta X_i \ : \ i = 1, \ldots, I\}$) follow a t-distribution. At the joint distribution level, of course, the covariance—and correlation—between any two arbitrary latent creditworthiness state variables, ΔX_n and ΔX_m, are also slightly transformed.

Recycling our analysis from Eq. 2.21 on page 72 with our new definition in Eq. 2.50, we have

$$\mathrm{cov}(\Delta X_n, \Delta X_m) = \mathbb{E}(\Delta X_n \Delta X_m), \tag{2.54}$$

$$= \mathbb{E}\left(\underbrace{\sqrt{\frac{\nu}{W}}\left(\alpha_n B_n \Delta z + \sqrt{1-\alpha_n^2}\Delta w_n\right)}_{\Delta X_n} \underbrace{\sqrt{\frac{\nu}{W}}\left(\alpha_m B_m \Delta z + \sqrt{1-\alpha_m^2}\Delta w_m\right)}_{\Delta X_m} \right),$$

$$= \mathbb{E}\left(\frac{\nu}{W}\left(\alpha_n \alpha_m B_n \Delta z \Delta z^T B_m^T\right) \right),$$

$$= \nu\mathbb{E}\left(\frac{1}{W}\right) \alpha_n B_n \underbrace{\mathbb{E}(\Delta z \Delta z^T)}_{\Omega} B_m^T \alpha_m,$$

$$= \nu\mathbb{E}\left(\frac{1}{W}\right) \alpha_n \underbrace{B_n \Omega B_m^T}_{\rho_{nm}} \alpha_m,$$

$$= \left(\frac{\nu}{\nu-2}\right) \alpha_n \rho_{nm} \alpha_m,$$

where $\rho_{nm} = \mathrm{B}_n \Omega \mathrm{B}_m^T$ is the normalized, factor-loaded correlation between the nth and mth creditworthiness latent state variables.[56] With the exception of the coefficient involving ν, this looks very similar to the structure found in Eq. 2.21. To get to the correlation coefficient, we need only normalize by the volatility of ΔX_n and ΔX_m as follows:

$$
\begin{aligned}
\rho(\Delta X_n, \Delta X_m) &= \frac{\mathrm{cov}(\Delta X_n, \Delta X_m)}{\sqrt{\mathrm{var}(\Delta X_n)}\sqrt{\mathrm{var}(\Delta X_m)}}, \tag{2.55}\\[2mm]
&= \frac{\left(\frac{\nu}{\nu-2}\right)\alpha_n \rho_{nm}\alpha_m}{\sqrt{\frac{\nu}{\nu-2}}\sqrt{\frac{\nu}{\nu-2}}},\\[2mm]
&= \frac{\left(\cancel{\frac{\nu}{\nu-2}}\right)\alpha_n \rho_{nm}\alpha_m}{\cancel{\frac{\nu}{\nu-2}}},\\[2mm]
&= \alpha_n \rho_{nm}\alpha_m,
\end{aligned}
$$

which happily corresponds exactly with the Gaussian model analogue derived in Eq. 2.22. We may continue to interpret the correlation between two arbitrary latent creditworthiness variables—often referred to as asset correlation—as the product of their respective systemic weight parameters and systemic correlation coefficient. This latter quantity is also sometimes called the factor correlation.

There are a few other small practical differences in the implementation. The default indicator of the ith obligor, $\mathcal{D}_i$, has the same conceptual definition as in the Gaussian case, but the default threshold is somewhat different. Specifically, following the logic from Eq. 2.40, we have

$$
\begin{aligned}
p_i &= \mathbb{E}(\mathbb{I}_{\mathcal{D}_i}), \tag{2.56}\\
&= \mathbb{P}(\mathcal{D}_i),\\
&= \mathbb{P}(\Delta X_n \le K_i),\\
&= F_{\mathcal{T}_\nu}(K_i),
\end{aligned}
$$

implying directly that $K_i = F_{\mathcal{T}_\nu}^{-1}(p_n)$. We will use, for lack of better notation, $F_{\mathcal{T}_\nu}$ and $F_{\mathcal{T}_\nu}^{-1}$ to denote the cumulative and *inverse* cumulative distribution functions of the standard t-distribution with ν degrees of freedom, respectively. This threshold adjustment is necessary and makes logical sense given that we've actually changed the underlying marginal and joint distributions.

---

[56] Because of the normalization, ρ_{nm} is conveniently both the correlation coefficient and covariance.

2.5.3 Default Correlation

Default correlation is a critically important quantity; inducing some degree of correlation between the default of individual credit obligors has been the motivating factor in the construction of the threshold model. To this point, we have seen both systemic factor and asset correlation, but have not examined the precise form of default correlation. We can no longer postpone consideration of this point. Proper treatment begins with the covariance between the nth and mth default events. From first principles, we have

$$\text{cov}\left(\mathbb{I}_{\mathcal{D}_n}, \mathbb{I}_{\mathcal{D}_m}\right) = \mathbb{E}\left(\mathbb{I}_{\mathcal{D}_n}\mathbb{I}_{\mathcal{D}_m}\right) - \mathbb{E}\left(\mathbb{I}_{\mathcal{D}_n}\right)\mathbb{E}\left(\mathbb{I}_{\mathcal{D}_m}\right), \tag{2.57}$$

$$= \mathbb{P}\left(\mathcal{D}_n \cap \mathcal{D}_m\right) - \mathbb{P}\left(\mathcal{D}_n\right)\mathbb{P}\left(\mathcal{D}_m\right).$$

Normalizing the covariance to arrive at the default correlation,

$$\rho\left(\mathbb{I}_{\mathcal{D}_n}, \mathbb{I}_{\mathcal{D}_m}\right) = \frac{\text{cov}\left(\mathbb{I}_{\mathcal{D}_n}, \mathbb{I}_{\mathcal{D}_m}\right)}{\sqrt{\text{var}(\mathbb{I}_{\mathcal{D}_n})}\sqrt{\text{var}(\mathbb{I}_{\mathcal{D}_m})}}, \tag{2.58}$$

$$= \frac{\mathbb{P}\left(\mathcal{D}_n \cap \mathcal{D}_m\right) - \mathbb{P}\left(\mathcal{D}_n\right)\mathbb{P}\left(\mathcal{D}_m\right)}{\sqrt{\mathbb{P}(\mathcal{D}_n)(1 - \mathbb{P}(\mathcal{D}_n))}\sqrt{\mathbb{P}(\mathcal{D}_m)(1 - \mathbb{P}(\mathcal{D}_m))}}.$$

Recalling that the variance of an indicator variable coincides with a Bernoulli trial permits a return to our simpler notation. We thus have

$$\rho\left(\mathbb{I}_{\mathcal{D}_n}, \mathbb{I}_{\mathcal{D}_m}\right) = \frac{\mathbb{P}\left(\mathcal{D}_n \cap \mathcal{D}_m\right) - p_n p_m}{\sqrt{p_n p_m (1 - p_n)(1 - p_m)}}. \tag{2.59}$$

Default correlation thus depends on the unconditional default probabilities as well as the joint probability of default between counterparties n and m. This definition is model independent. The choice of model, however, will determine the form of the joint default probability, $\mathbb{P}\left(\mathcal{D}_n \cap \mathcal{D}_m\right)$.

In the Gaussian threshold model, it should be no surprise that the joint distribution of ΔX_n and ΔX_m is also Gaussian. In this case, $\mathbb{P}\left(\mathcal{D}_n \cap \mathcal{D}_m\right)$ is described by the bivariate normal distribution. It has the following mathematical form,

$$\mathbb{P}\left(\mathcal{D}_n \cap \mathcal{D}_m\right) = \mathbb{P}\left(\Delta X_n \leq \Phi^{-1}(p_n), \Delta X_m \leq \Phi^{-1}(p_m)\right), \tag{2.60}$$

$$= \frac{1}{2\pi\sqrt{1 - (\alpha_n \rho_{nm}\alpha_m)^2}} \int_{-\infty}^{\Phi^{-1}(p_n)}$$

$$\times \int_{-\infty}^{\Phi^{-1}(p_m)} e^{-\frac{(u^2 - 2\alpha_n \rho_{nm}\alpha_m uv + v^2)}{2(1 - (\alpha_n \rho_{nm}\alpha_m)^2)}} \, du dv,$$

$$= \Phi\left(\Phi^{-1}(p_n), \Phi^{-1}(p_m); \alpha_n \rho_{nm}\alpha_m\right).$$

This expression easily permits us—with the help of a good numerical integration library—to directly compute the default correlation between counterparties n and m using Eq. 2.59.

It is helpful to put Eq. 2.60 into matrix form. Define the correlation matrix as,

$$\Omega_{nm} = \begin{bmatrix} 1 & \mathbf{r}_{nm} \\ \mathbf{r}_{nm} & 1 \end{bmatrix}, \tag{2.61}$$

where the asset-correlation coefficient is denoted as $\mathbf{r}_{nm} = \alpha_n \rho_{nm} \alpha_m$ to slightly ease the notational burden. The determinant of Ω_{nm} is given as,

$$\det(\Omega_{nm}) = |\Omega_{nm}| = 1 - \mathbf{r}_{nm}^2, \tag{2.62}$$

and the inverse is simply,

$$\Omega_{nm}^{-1} = \frac{1}{|\Omega_{nm}|} \begin{bmatrix} 1 & -\mathbf{r}_{nm} \\ -\mathbf{r}_{nm} & 1 \end{bmatrix}, \tag{2.63}$$

$$= \begin{bmatrix} \frac{1}{1-\mathbf{r}_{nm}^2} & \frac{-\mathbf{r}_{nm}}{1-\mathbf{r}_{nm}^2} \\ \frac{-\mathbf{r}_{nm}}{1-\mathbf{r}_{nm}^2} & \frac{1}{1-\mathbf{r}_{nm}^2} \end{bmatrix}.$$

Defining $x = \begin{bmatrix} u & v \end{bmatrix}^T$, then

$$x^T \Omega_{nm}^{-1} x = \begin{bmatrix} u & v \end{bmatrix} \begin{bmatrix} \frac{1}{1-\mathbf{r}_{nm}^2} & \frac{-\mathbf{r}_{nm}}{1-\mathbf{r}_{nm}^2} \\ \frac{-\mathbf{r}_{nm}}{1-\mathbf{r}_{nm}^2} & \frac{1}{1-\mathbf{r}_{nm}^2} \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}, \tag{2.64}$$

$$= \frac{u^2 - 2\mathbf{r}_{nm} uv + v^2}{1 - \mathbf{r}_{nm}^2}.$$

Our bivariate Gaussian joint-default probability, from Eq. 2.60, can be more succinctly rewritten as,

$$\Phi\left(\Phi^{-1}(p_n), \Phi^{-1}(p_m); \mathbf{r}_{nm}\right) = \frac{1}{2\pi\sqrt{1 - \mathbf{r}_{nm}^2}} \int_{-\infty}^{\Phi^{-1}(p_n)}$$

$$\times \int_{-\infty}^{\Phi^{-1}(p_m)} e^{-\frac{(u^2 - 2\mathbf{r}_{nm} uv + v^2)}{2(1-\mathbf{r}_{nm}^2)}} \, du dv, \tag{2.65}$$

$$= \frac{1}{2\pi\sqrt{|\Omega_{nm}|}} \int_{-\infty}^{\Phi^{-1}(p_n)} \int_{-\infty}^{\Phi^{-1}(p_m)} e^{-\frac{x^T \Omega_{nm}^{-1} x}{2}} dx.$$

This is referred to as the Gaussian copula. Copula functions essentially describe dependence between random variables. The job of a copula function is to map a

collection of marginals into a joint distribution. This notion—which was originally introduced by Sklar [38]—is the definitive way to describe dependence between random variables.

It is naturally possible to generalize the joint dependence across all of our credit counterparties. Equation 2.65, with I counterparties, where R is the asset correlation matrix between the I counterparties, is then

$$\Phi\left(\Phi^{-1}(p_1), \cdots, \Phi^{-1}(p_I); R\right) = \frac{1}{(2\pi)^{\frac{I}{2}}\sqrt{|R|}} \int_{-\infty}^{\Phi^{-1}(p_1)} \cdots$$
$$\times \int_{-\infty}^{\Phi^{-1}(p_I)} e^{-\frac{x^T R^{-1} x}{2}} dx, \qquad (2.66)$$

where,

$$R = \begin{bmatrix} 1 & \mathbf{r}_{12} & \cdots & \mathbf{r}_{1(I-1)} & \mathbf{r}_{1I} \\ \mathbf{r}_{21} & 1 & \cdots & \mathbf{r}_{2(I-1)} & \mathbf{r}_{2I} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{r}_{I1} & \mathbf{r}_{I2} & \cdots & \mathbf{r}_{I(I-1)} & 1 \end{bmatrix} \qquad (2.67)$$

summarizes the asset-correlation coefficients for every pair of credit obligors in one's portfolio. Equation 2.66 is a function, defined on the unit cube $[0, 1]^I$, transforming a set of marginal distributions into a single joint distribution.

Equation 2.59, as previously mentioned, applies to any choice of threshold model. In the t-threshold setting, however, the joint distribution of ΔX_n and ΔX_m is now assumed to follow a bivariate t-distribution with ν degrees of freedom. Mathematically, such an object is written as

$$\mathbb{P}\left(\mathcal{D}_n \cap \mathcal{D}_m\right) = \mathbb{P}(y_n \leq F_{\mathcal{T}_\nu}^{-1}(p_n), y_m \leq F_{\mathcal{T}_\nu}^{-1}(p_m)), \qquad (2.68)$$
$$= \frac{\Gamma\left(\frac{\nu+d}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)(\nu\pi)^{\frac{d}{2}}|\Omega_{nm}|^{\frac{1}{2}}} \int_{-\infty}^{F_{\mathcal{T}_\nu}^{-1}(p_n)}$$
$$\times \int_{-\infty}^{F_{\mathcal{T}_\nu}^{-1}(p_m)} \left(1 + \frac{x^T \Omega_{nm}^{-1} x}{\nu}\right)^{-\left(\frac{\nu+d}{2}\right)} dx,$$

where $\Gamma(\cdot)$ represents the gamma function,

$$\Gamma(u) = \int_0^\infty x^{u-1} e^{-x} du. \qquad (2.69)$$

and Ω_{nm} is unchanged from the definition in Eq. 2.61, $x \in \mathbb{R}^2$ and $d = 2$. This is the classic, direct form.[57]

An alternative description of the joint default probability, which we provide for completeness and future usage, makes use of the conditional default probability in the t-threshold setting. That is,

$$
\begin{aligned}
p_i(\Delta z, W) &= \mathbb{E}\left(\mathbb{I}_{\mathcal{D}_i} \,\middle|\, \Delta z, W\right), \tag{2.70} \\
&= \mathbb{P}\left(\mathcal{D}_i \,|\, \Delta z, W\right), \\
&= \mathbb{P}\left(\Delta X_i \leq F_{\mathcal{T}_\nu}^{-1}(p_i) \,\middle|\, \Delta z, W\right), \\
&= \mathbb{P}\left(\underbrace{\sqrt{\frac{\nu}{W}}\left(\alpha_i B_i \Delta z + \sqrt{1 - \alpha_i^2}\, \Delta w_i\right)}_{\text{Eq. 2.50}} \leq F_{\mathcal{T}_\nu}^{-1}(p_n) \,\middle|\, \Delta z, W\right), \\
&= \mathbb{P}\left(\Delta w_i \leq \frac{\sqrt{\frac{W}{\nu}} F_{\mathcal{T}_\nu}^{-1}(p_i) - \alpha_i B_i \Delta z}{\sqrt{1 - \alpha_i^2}} \,\middle|\, \Delta z, W\right), \\
&= \Phi\left(\frac{\sqrt{\frac{W}{\nu}} F_{\mathcal{T}_\nu}^{-1}(p_i) - \alpha_i B_i \Delta z}{\sqrt{1 - \alpha_i^2}}\right).
\end{aligned}
$$

To determine the conditional probability of y_n, both the global systematic factor, Δz, and the mixing random variate, W, must be revealed. Our desired joint-default probability, for arbitrarily selected counterparties n and m, is thus determined as,

$$
\begin{aligned}
\mathbb{P}\left(\mathcal{D}_n \cap \mathcal{D}_m\right) &= \mathbb{E}\left(\mathbb{I}_{\mathcal{D}_n} \cap \mathbb{I}_{\mathcal{D}_m}\right), \tag{2.71} \\
&= \underbrace{\mathbb{E}\left(\mathbb{E}\left(\mathbb{I}_{\mathcal{D}_n} \cap \mathbb{I}_{\mathcal{D}_m}\right) \,\middle|\, \Delta z, W\right)}_{\text{By iterated expectations}}, \\
&= \mathbb{E}\left(\underbrace{\mathbb{E}\left(\mathbb{I}_{\mathcal{D}_n} \,\middle|\, \Delta z, W\right) \cdot \mathbb{E}\left(\mathbb{I}_{\mathcal{D}_m} \,\middle|\, \Delta z, W\right)}_{\text{By conditional independence}}\right), \\
&= \mathbb{E}\left(p_n(\Delta z, W) \cdot p_m(\Delta z, W)\right),
\end{aligned}
$$

---

[57] One can loosely think of the gamma function as a continuous-valued extension of the integer-valued factorial function. See Abramovitz and Stegun [1, Chapter 6] for more information on this mathematical object.

$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_{0}^{\infty} p_n(z, w) \cdot p_m(z, w) \cdot f_W(w)$$

$$\cdot \phi_{\Delta z}(z_1, \cdots, z_J) dw dz_1 \cdots dz_J.$$

This approach exploits the conditional independence of the latent state variables to find an alternative, but equivalent, representation of the joint default probability. Since resolution of this expression requires numerical integration, the catching point is the dimensionality of $\Delta z \in \mathbb{R}^J$. If $J = 1$, then Eq. 2.71 reduces to a very manageable two-dimensional integral. If not, then it may very well be computationally impractical to use this trick; in this case, Eq. 2.68 is the appropriate choice for evaluating the joint default probability and, by extension, the pairwise default correlation.

As a final point, the joint-default density under the t-threshold model has the following form

$$F_{\mathcal{T}_\nu}\left(F_{\mathcal{T}_\nu}^{-1}(p_1), \cdots, F_{\mathcal{T}_\nu}^{-1}(p_I); R\right) = \frac{\Gamma\left(\frac{\nu+d}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)(\nu\pi)^{\frac{d}{2}}|R|^{\frac{1}{2}}} \int_{-\infty}^{F_{\mathcal{T}_\nu}^{-1}(p_1)} \cdots \int_{-\infty}^{F_{\mathcal{T}_\nu}^{-1}(p_I)}$$

$$\times \left(1 + \frac{x^T R^{-1} x}{\nu}\right)^{-\left(\frac{\nu+d}{2}\right)} dx, \qquad (2.72)$$

where R remains as in the definition in Eq. 2.67. This copula function, defined on the unit cube $[0, 1]^I$, is the t-distributed analogue of the Gaussian equivalent in Eq. 2.66. The mathematical structure does not provide an enormous amount of insight, but the key difference is that the t-copula has non-zero tail dependence, whereas the Gaussian copula does not. Ultimately, this makes a rather important distinction in the specification of credit-default risk.

Colour and Commentary 21 (*t*-THRESHOLD MODEL): *The great financial crisis, beginning in 2007–2008, exposed some important statistical flaws in the Gaussian threshold model. Most importantly, it has a structural lack of* tail dependence. *In a modelling framework where attention is consistently focused on the extremes of the credit-loss distribution, zero tail dependence moves from a theoretical peculiarity to a serious problem. A practical solution is required to address this serious shortcoming. It turns out that relatively few members of the elliptical family of distributions—those distributions typically best suited for the threshold framework—actually have non-zero tail dependence. The t distribution—induced in this case through a χ^2 mixing variable in the normal-variance mixture setting—fortunately represents one*

(continued)

Colour and Commentary 21 (continued)
*tractable candidate exhibiting non-zero tail dependence. For this reason,
along with the fact that the necessary model adjustments are relatively
modest, we have opted to employ this version of the threshold model. The
driving rationale behind this choice is a desire to simultaneously enhance the
realism and conservatism of our credit-risk economic-capital estimates.*

2.5.4   Modelling Credit Migration

If one assumes that credit losses are experienced only in the event of default, then the
previous framework would be sufficient to describe our methodological approach.
There is nevertheless, as hinted at previously, another dimension to credit risk.
Financial losses are also logically possible in the event that a credit counterparty
is downgraded. Credit deterioration implies that one needs to reassess the relative
likelihood of repayment and, in general, correspondingly write down the value of
one's assets. This occurs even when the attendant obligor continues to properly
service its credit obligation. Improvement in a credit obligor's credit status will,
of course, have a positive valuation impact. The possibility of one's obligors to
move up or down the credit spectrum is referred to, in general, as credit migration.
Default is, in fact, simply a special case of this general behaviour—it considers only
the transition from the current credit state to the default state.

Dealing with this generalization does not dramatically change the model struc-
ture, but it does require some additional overhead. Given q discrete credit-quality
states, or rating categories, we assume that each obligor's credit status follows a
discrete-time Markov chain. This basically means that the only information required
to determine an obligor's credit status in the next period is its current credit status.
This so-called Markov property practically implies that, for a given obligor i,
currently in state m at time t, the probability it finds itself in state n at time $t + 1$ is
written as,

$$p_{mn} = \mathbb{P}\left(\text{Going } \mathbf{to} \ n \,|\, \text{Coming } \mathbf{from} \ m\right), \qquad (2.73)$$
$$= \mathbb{P}\left(S_{i,t+1} = n \,\middle|\, S_{i,t} = m\right),$$

for $n, m = 1, \ldots, q$. $S_{i,t}$ denotes the credit state of the ith obligor at time t.
This generic form describes all obligors, time points, and constellations of starting
and ending points. The driving idea is that a Markov chain characterizes the
transitions between various states over time. These are also referred to as transition
probabilities, of which default probabilities are a special case (i.e., transition directly
into the default state).

The collection of so-called *transition* probabilities, associated with a given Markov chain, are collected into the transition matrix, P. For our q-state process, it is summarized as

$$P = \begin{bmatrix} p_{11} & p_{21} & \cdots & p_{q1} \\ p_{12} & p_{22} & \cdots & p_{q2} \\ \vdots & \vdots & \ddots & \vdots \\ p_{1q} & p_{2q} & \cdots & p_{qq} \end{bmatrix}. \tag{2.74}$$

This notation can, given it is essentially the opposite of what is used in matrix algebra to refer to specific elements, be somewhat confusing. It does, however, have its own useful internal logic. Each row relates to the state of an obligor as of time t. If counterparty i is classified in the second credit state at time t, then only the *second* row is relevant to determining the relative probabilities of transition into the set of q credit states at time $t + 1$. When examining only the default perspective, we ignore all the other elements in this second row and consider only the $p_{q2} = \mathbb{P}\left(S_{i,t+1} = q \mid S_{i,t} = 2\right)$. This is, to be very explicit, the probability that a credit-counterpart currently in state 2 will transition into default (i.e., state q) in the next period. Naturally, if we sum across all the transition probabilities for the next period, the row must sum to unity. That is,

$$\sum_{n=1}^{q} p_{nm} = 1, \tag{2.75}$$

for $m = 1, \ldots, q$.[58] The point is that the rows of the transition matrix, P, must sum to one.[59] Figure 2.3 illustrates—in a schematic manner, again starting from $S_{i,t} = 2$—how the transition probabilities describe the movement from the current credit state to the range of possible state values in the next period. If we wish to capture the full range of possible movements to other credit states, then it will be necessary to make intelligent use of the entire transition matrix.

The transition matrix will inform—as did the default probabilities in the default-only setting—the specific thresholds. Before we examine precisely how, it is preferable to first build the general framework. Clearly, by construction, we have $S_{i,t} \in \{1, \ldots, q\}$ for each $i = 1, \ldots, I$. This portfolio-level credit-obligor knowledge, along with P, is all that is required to simulate the portfolio counterparties' credit states in the next period.[60] The actual computation, however, is somewhat more complicated.

---

[58] In our simple example, $m = 2$.

[59] There are many excellent references on the theory and practice of Markov chains; a few recommended choices include Brémaud [9], Hamilton [20], and Meyn and Tweedie [31].

[60] More information, as we'll see in upcoming discussion, is required to determine the credit losses (or gains) associated with these transitions.
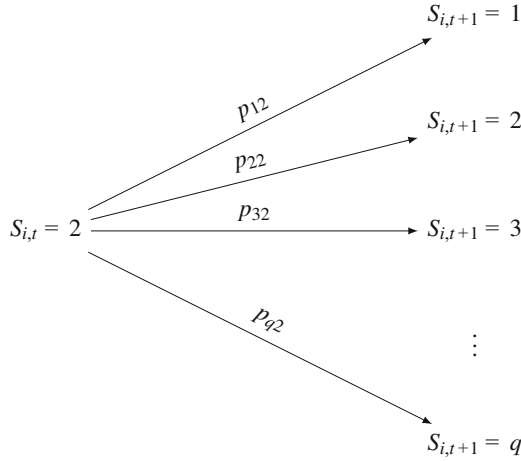
$$S_{i,t+1} = 1$$

$$S_{i,t+1} = 2$$

$$S_{i,t+1} = 3$$

$$S_{i,t} = 2$$

p_{12}

p_{22}

p_{32}

p_{q2}

$\vdots$

$$S_{i,t+1} = q$$

Fig. 2.3 *Transition probabilities*: This schematic illustrates, given the current state of the ith obligor is 2, the range of transition probabilities into the other q states. This is, quite simply, the second row of the transition matrix, P.

In Eq. 2.40, we calibrated the ith default threshold as using an indicator variable. We can generalize this approach somewhat. We need a more flexible definition of the default threshold. Let us, therefore, define $K_{S_{i,t}}(j)$ as the threshold of an obligor in state i as of time t, for moving to state j as of time $t+1$. In the default setting, we only consider the situation of $j = q$; that is, movement from state i to default. The introduction of the transition matrix, however, shows us that transition is possible to any of the individual credit states $j = 1, \ldots, q$. Using this idea, we may redefine the notion of the *default* threshold as follows,

$$\mathbb{E}\left(\mathbb{I}_{\left\{\Delta X_i \in \left[-\infty, K_{S_{i,t}}(q)\right]\right\}}\right) = p_{q,S_{i,t}}, \tag{2.76}$$

$$\mathbb{P}\left(\Delta X_i \in \left[-\infty, K_{S_{i,t}}(q)\right]\right) = p_{q,S_{i,t}},$$

$$\mathbb{P}\left(-\infty \leq \Delta X_i \leq K_{S_{i,t}}(q)\right) = p_{q,S_{i,t}},$$

recalling that $p_{q,S_{i,t}}$ is the probability of default, but also the $(S_{i,t}, q)$ element of the transition represented in Eq. 2.74.[61] This might not seem like much progress, but if we recall that ΔX_i is a continuous, one-dimensional random variable and, rather

---

[61] Don't forget that the transition probability notation is the exact opposite of matrix notation; this often leads to confusion (and occasionally mistakes).

vacuously, $-\infty < K_{S_{i,t}}(q)$, then

$$\mathbb{P}\left(\Delta X_i \leq K_{S_{i,t}}(q)\right) - \mathbb{P}\left(\Delta X_i \leq -\infty\right) = p_{q,S_{i,t}}, \qquad (2.77)$$

$$F_{\Delta X_i}\left(K_{S_{i,t}}(q)\right) - \underbrace{F_{\Delta X_i}\left(-\infty\right)}_{=0} = p_{q,S_{i,t}},$$

$$\Phi\left(K_{S_{i,t}}(q)\right) = p_{q,S_{i,t}},$$

$$K_{S_{i,t}}(q) = \Phi^{-1}(p_{q,S_{i,t}}).$$

While we have not learned anything new relative to the development in Eq. 2.40, we have transformed a threshold into an interval. We also generalized the notation. In particular, we write the default probability using its true identity: a transition probability. That is, p_i is replaced with $p_{q,S_{i,t}}$. The threshold is also not simply linked to the ith obligor, but also includes information on their current credit state, $S_{i,t}$, and the targeted credit-state threshold.

The development summarized in Eq. 2.77, as the reader has certainly noticed, applies to the Gaussian threshold model. The same basic logic applies in the t-threshold setting, but the distributions differ. In this case, the generalized default threshold simply becomes,

$$K_{S_{i,t}}(q) = F_{\mathcal{T}_\nu}^{-1}(p_{q,S_{i,t}}). \qquad (2.78)$$

These generalizations permit us to consider credit-migration for non-default states. Returning to the Gaussian case, a specific example will be helpful. Again, conditioning on the ith credit counterpart currently finding itself in state $S_{i,t}$, we would like to link the (possible) migration to state 3 with the probability of transitioning from state $S_{i,t}$ to 3. This transition probability is simply denoted as $p_{3,S_{i,t}}$. The starting point is inspired by Eq. 2.76,

$$\mathbb{E}\left(\mathbb{I}_{\left\{\Delta X_i \in \left[K_{S_{i,t}}(4), K_{S_{i,t}}(3)\right]\right\}}\right) = p_{3,S_{i,t}}, \qquad (2.79)$$

$$\mathbb{P}\left(K_{S_{i,t}}(4) \leq \Delta X_i \leq K_{S_{i,t}}(3)\right) = p_{3,S_{i,t}},$$

$$\mathbb{P}\left(\Delta X_i \leq K_{S_{i,t}}(3)\right) - \mathbb{P}\left(\Delta X_i \leq K_{S_{i,t}}(4)\right) = p_{3,S_{i,t}},$$

$$\Phi\left(K_{S_{i,t}}(3)\right) - \Phi\left(K_{S_{i,t}}(4)\right) = p_{3,S_{i,t}},$$

since, by construction, $K_{S_{i,t}}(4) < K_{S_{i,t}}(3)$. There does not, however, appear to be any obvious approach to isolate the K terms; this is a situation of a single equation in two unknowns. The solution involves rewriting the right-hand side transition probability, $p_{3,S_{i,t}}$, in an alternative, perhaps initially non-intuitive form.

In particular, we redefine it as,

$$p_{3,S_{i,t}} = \sum_{w=3}^{q} p_{w,S_{i,t}} - \sum_{w=4}^{q} p_{w,S_{i,t}}. \tag{2.80}$$

These are simply the differences in the sum of elements across a given row of the transition matrix, P, with different index starting points. If we plug our definition from Eq. 2.80 into Eq. 2.79, we have that

$$\Phi\left(K_{S_{i,t}}(3)\right) - \Phi\left(K_{S_{i,t}}(4)\right) = \underbrace{\sum_{w=3}^{q} p_{w,S_{i,t}} - \sum_{w=4}^{q} p_{w,S_{i,t}}}_{\text{Eq. 2.80}}. \tag{2.81}$$

If we equate the two corresponding terms on the right- and left-hand sides of Eq. 2.81, then we can identify our thresholds as,

$$K_{S_{i,t}}(3) = \Phi^{-1}\left(\sum_{w=3}^{q} p_{w,S_{i,t}}\right), \tag{2.82}$$

and

$$K_{S_{i,t}}(4) = \Phi^{-1}\left(\sum_{w=4}^{q} p_{w,S_{i,t}}\right). \tag{2.83}$$

More generally, of course, we can conclude that the jth boundary for the ith credit obligor is,

$$K_{S_{i,t}}(j) = \Phi^{-1}\left(\sum_{w=j}^{q} p_{w,S_{i,t}}\right). \tag{2.84}$$

In this manner, the set of finite, real-valued thresholds

$$-\infty < K_{S_{i,t}}(q) < K_{S_{i,t}}(q-1) < \cdots < K_{S_{i,t}}(2) < K_{S_{i,t}}(1) \equiv \infty. \tag{2.85}$$

A bit of reflection reveals that we have created I partitions of the support of a standard normally distributed state variable—each conditioned on the current credit state of the ith counterparty, $S_{i,t}$—informed by the appropriate transition probabilities. Although we have I individual partitions, there are only q unique cases, each linked to a separate row in the transition matrix, P. This represents a large, and messy, number of partitions of the real-number line, $\mathbb{R}$. We can simplify our life somewhat by cumulating, or summing, the transition-matrix entries across each row from right to left. This operation significantly eases working with

quantities like those found on the right-hand side of Eq. 2.80 and leads to the so-called cumulative transition matrix, G, which we define as

$$
G = \begin{bmatrix}
(p_{11} + p_{21} + \cdots + p_{q1}) & (p_{21} + \cdots + p_{q1}) & \cdots & p_{q1} \\
(p_{12} + p_{22} + \cdots + p_{q2}) & (p_{22} + \cdots + p_{q2}) & \cdots & p_{q2} \\
\vdots & \vdots & \ddots & \vdots \\
(p_{1q} + p_{2q} + \cdots + p_{qq}) & (p_{2q} + \cdots + p_{qq}) & \cdots & p_{qq}
\end{bmatrix}, \quad (2.86)
$$

$$
= \begin{bmatrix}
\sum_{w=1}^{q} p_{w1} & \sum_{w=2}^{q} p_{w1} & \cdots & \sum_{w=q}^{q} p_{w1} \\
\sum_{w=1}^{q} p_{w2} & \sum_{w=2}^{q} p_{w2} & \cdots & \sum_{w=q}^{q} p_{w2} \\
\vdots & \vdots & \ddots & \vdots \\
\sum_{w=1}^{q} p_{wq} & \sum_{w=2}^{q} p_{wq} & \cdots & \sum_{w=q}^{q} p_{wq}
\end{bmatrix},
$$

Each one of the values in this cumulative transition matrix, G, lies in the interval, $[0, 1]$—they are, in the end, probabilities. As we saw from the previous development, the cumulative transition matrix is an intermediate step. The next step in determining the appropriate partition requires transforming these probability quantities into the domain of ΔX_i. The entries in G are thus mapped back into the values of the standard normal distribution using the inverse standard normal cumulative distribution function following the basic logic found in Eqs. 2.76 to 2.80. This leads to,

$$
G_{\Phi^{-1}} = \begin{bmatrix}
\Phi^{-1}\left(\sum_{w=1}^{q} p_{w1}\right) & \Phi^{-1}\left(\sum_{w=2}^{q} p_{w1}\right) & \cdots & \Phi^{-1}\left(\sum_{w=q}^{q} p_{w1}\right) \\
\Phi^{-1}\left(\sum_{w=1}^{q} p_{w2}\right) & \Phi^{-1}\left(\sum_{w=2}^{q} p_{w2}\right) & \cdots & \Phi^{-1}\left(\sum_{w=q}^{q} p_{w2}\right) \\
\vdots & \vdots & \ddots & \vdots \\
\Phi^{-1}\left(\sum_{w=1}^{q} p_{wq}\right) & \Phi^{-1}\left(\sum_{w=2}^{q} p_{wq}\right) & \cdots & \Phi^{-1}\left(\sum_{w=q}^{q} p_{wq}\right)
\end{bmatrix}, \quad (2.87)
$$

$$
= \begin{bmatrix}
\Phi^{-1}(g_{11}) & \Phi^{-1}(g_{21}) & \cdots & \Phi^{-1}(g_{q1}) \\
\Phi^{-1}(g_{12}) & \Phi^{-1}(g_{22}) & \cdots & \Phi^{-1}(g_{q2}) \\
\vdots & \vdots & \ddots & \vdots \\
\Phi^{-1}(g_{1q}) & \Phi^{-1}(q_{2q}) & \cdots & \Phi^{-1}(g_{qq})
\end{bmatrix},
$$

$$= \begin{bmatrix} K_1(1) & K_1(2) & \cdots & K_1(q) \\ K_2(1) & K_2(2) & \cdots & K_2(q) \\ \vdots & \vdots & \ddots & \vdots \\ K_q(1) & K_q(2) & \cdots & K_q(q) \end{bmatrix}.$$

The individual entries of the matrix, $G_{\Phi^{-1}}$, thus represent the collection of boundaries, or thresholds, for the determination of credit migration as described in Eq. 2.85. Each row, which captures the current state of the ith counterpart, describes the partition of the ΔX_i space.

The previous development applies to the Gaussian threshold implementation. Virtually identical logic, in the t-threshold approach, brings us to,

$$G_{\mathcal{T}_\nu^{-1}} = \begin{bmatrix} F_{\mathcal{T}_\nu}^{-1}(g_{11}) & F_{\mathcal{T}_\nu}^{-1}(g_{21}) & \cdots & F_{\mathcal{T}_\nu}^{-1}(g_{q1}) \\ F_{\mathcal{T}_\nu}^{-1}(g_{12}) & F_{\mathcal{T}_\nu}^{-1}(g_{22}) & \cdots & F_{\mathcal{T}_\nu}^{-1}(g_{q2}) \\ \vdots & \vdots & \ddots & \vdots \\ F_{\mathcal{T}_\nu}^{-1}(g_{1q}) & F_{\mathcal{T}_\nu}^{-1}(q_{2q}) & \cdots & F_{\mathcal{T}_\nu}^{-1}(g_{qq}) \end{bmatrix}. \tag{2.88}$$

The only difference associated with a change in copula function is the choice of inverse cumulative distribution function to employ. It needs to be consistent with the underlying latent creditworthiness variable definition.

Irrespective of one's choice of copula function, a modicum of computational caution is required. Since the sum of each row of the transition matrix is unity, each element in the first column of G takes the value of one. That the support of the normal distribution is $(-\infty, \infty)$ implies, however, that one would assign a value of ∞ to the elements in the first column of G. This will, of course, inevitably lead to numerical problems. Instead, we assign a value of 10, because the probability of observing an outcome from a standard normal or t distribution behind these values is vanishingly small.[62]

Figure 2.4 provides a visualization of how, for a counterparty currently in credit state $S_{i,t}$, the partition of the ΔX_i space is constructed. We observe the relation to the cumulative transition probabilities and how, stylistically at least, the individual thresholds are placed. The random draw of ΔX_i determines, for the ith counterparty, the location in the interval, $(-\infty, \infty)$. The specific sub-interval that this random draw enters thus determines the ith credit obligor's $t + 1$ credit state. This is a very clever idea. It allows us to map a real-valued variable that might take an uncountable range of values—the creditworthiness index—into a small, finite set of q credit categories.

Figure 2.4 also illustrates how the default outcome, $\Delta X_i \leq K_{S_{i,t}}(q)$, is simply a special case of this more general framework. The size of each individual interval—which is hard to illustrate in a schematic diagram like Fig. 2.4—is directly

---

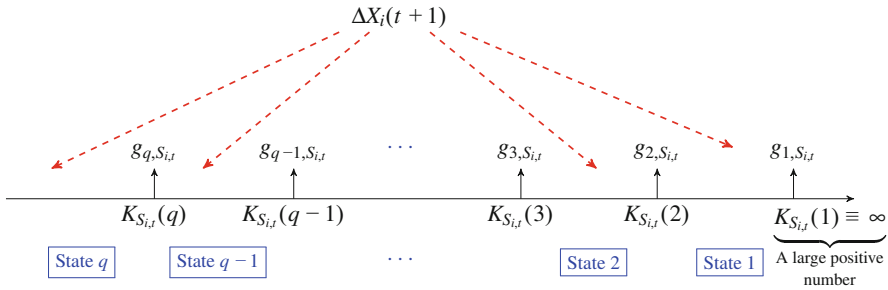[62] Any sufficiently large positive number will, of course, work.

Fig. 2.4 *Partitioning the ΔX_i space*: This figure visually illustrates, for a counterparty currently in credit state $S_{i,t}$, the partition of the $\Delta X_i(t+1)$ space, its relation to the cumulative transition probabilities, and how the individual thresholds are placed. All the necessary information stems directly, or indirectly, from the transition matrix, P.

proportional to the associated probability of transition. Improbable outcomes are assigned sliver thin intervals, while very likely outcomes have significant distance between the two end-points. The idea is to ensure that the final transition outcomes are numerically consistent with a random draw from the distribution of each ΔX_i and that simulated credit-state mappings are driven by the values in our transition matrix, P.

Colour and Commentary 22 (DEFAULT AS SPECIAL CASE)*: A credit-risk model is, quite naturally, centrally concerned with the idea of default. This is, of course, the principal avenue through which loss is experienced by the firm. It is nonetheless, from an analytic perspective, rather useful to think of default as a special case of credit-migration. The default threshold is particularly important, because it involves negotiating repayment and, in some cases, losing one's entire investment. At the same time, movement from one credit state to another (including default) can be viewed a sequence of economically meaningful intervals. Each sub-interval maps the creditworthiness index into a credit state. Default is easy to understand, since it conceptually involves the firm's assets falling below its liabilities. Credit deterioration (or improvement), however, can be seen to follow a similar pattern. The required level of assets associated with these internals are, by contrast, a bit more difficult to define. Ultimately, they are inferred, as the default case, from transition-probability estimates. This generalization forms the central logic of credit-migration modelling.*

Naturally, we need to construct a mathematical formulation of Fig. 2.4. Given the simulated outcome $\Delta X_i(t+1)$ associated with the ith counterparty currently in credit state $S_{i,t}$, we may determine the credit state in the next period, $S_{i,t+1}$ as

follows,

$$
S_{i,t+1} = \begin{cases}
1 & : \ \Delta X_i(t+1) \in \left[K_{S_{i,t}}(2), \, K_{S_{i,t}}(1) \equiv \infty \right] \\
2 & : \ \Delta X_i(t+1) \in \left(K_{S_{i,t}}(3), \, K_{S_{i,t}}(2) \right] \\
& \quad \vdots \\
q-1 & : \ \Delta X_i(t+1) \in \left(K_{S_{i,t}}(q), \, K_{S_{i,t}}(q-1) \right] \\
q & : \ \Delta X_i(t+1) \in \left(-\infty, \, K_{S_{i,t}}(q) \right]
\end{cases}, \qquad (2.89)
$$

for $i = 1, \ldots, I$. This intuitive expression can be further simplified—with a significant computational advantage for practical implementation—by exploiting the monotonicity of the threshold definitions introduced in Eq. 2.85 and assured from the structure in Eqs. 2.87 and 2.88. A bit of reflection reveals that the credit state in the next period is determined by the smallest threshold which exceeds, or is equal to, $\Delta X_i(t+1)$. This is defined as,

$$
S_{i,t+1} = \sum_{\ell=1}^{q} \mathbb{I}_{\left\{ \Delta X_i(t+1) \leq K_{S_{i,t}}(\ell) \right\}}. \qquad (2.90)
$$

While somewhat less intuitive, Eq. 2.90 permits a more parsimonious representation of the credit-state transition process.[63] Not great from a pedagogical perspective, this form saves a significant amount of computational complexity and expense.

2.5.5 The Nuts and Bolts of Credit Migration

The previous section provides a detailed, and workable, description of the link between firm creditworthiness and the underlying credit state associated with each individual obligor in one's portfolio. An important question remains: what happens in the event that a credit counterpart changes credit state? We have already seen the implications—in Eqs. 2.41 and 2.42—of a credit default. One loses some portion of the total exposure; the exact amount depends upon one's assumptions regarding recovery. Considering (non-default) migration, of course, an alternative strategy is required. Practically, we would expect that:

---

[63] Other (probably more clever) computational tricks are possible; one alternative, for example, has a specific structure involving the supremum operator. The important point is to define this condition in an efficient manner consistent with one's definitions.

1. a credit loss (gain) to be recognized in the event of credit deterioration (improvement);
2. the magnitude of the credit loss (or gain) should be consistent with market-valuation effects associated with similarly rated entities;
3. the specific size and interest-rate sensitivity of the obligor's exposures to be incorporated into any credit loss (or gain); and
4. neither gain nor loss is to occur in the event an obligor remains in the same credit state.

A critical aspect of this development is the creation of a link between the value of an obligor's exposures and changes in the market value of these exposures. Given that loan exposures are not (typically) traded in liquid markets, this is something of a challenge. For traded instruments, however, it is rather more straightforward. Any credit-risky security trades at a premium over the lowest-risk borrower—almost always the government—in that currency. The difference between actual bond yields and the government, or Treasury, yield for a similar maturity is referred to as the credit spread.[64] Conceptually, the credit spread is decomposed into two separate components: a general and an idiosyncratic element. The general aspect describes the global risk associated with the entity's credit state or rating. The idiosyncratic element, conversely, is unique to the issue and may incorporate specific risks—and also, more practically, liquidity—associated with its debt claims. For the purposes of credit-migration modelling, a few assumptions are typical. First, only the global or general credit-spread risk is considered.[65] Second, we will assume that spread movements can also be used to model valuation gains and losses for both marketable bonds and non-marketable loans. We further assume—although this is principally for the purposes of simplicity—that the credit spread is constant across the entire maturity spectrum.[66]

A further, perhaps more difficult-to-defend, assumption is the time homogeneity of these credit spreads. Credit spreads clearly vary across time due to investment flows, changes in relative liquidity and variation in aggregate risk preferences. There is also evidence that credit spreads are correlated with the general business cycle. One could clearly attempt to model these dynamics, but the entire credit-risk economic-capital approach seeks to generate long-term, through-the-cycle estimates of the credit-loss distribution. In other words, transition probabilities, correlation parameters, loss-given-default estimates, and assumed credit-spread values are,

---

[64] It is also possible, of course, to define the spread relative to another reference curve. A popular choice, given its liquidity, is the LIBOR-based swap curve. Given ongoing and forthcoming LIBOR-market reform, the plumbing of the swap market is likely to change, but its central role and usefulness for spread computations likely will not.

[65] There is, as an empirical reality, a significant amount of heterogeneity with a given credit rating. Modelling this aspect is complex, noisy, and unlikely to dramatically influence the results. As such, it makes logical sense to exclude it.

[66] Empirically, we do observe a term structure of credit spreads for each credit rating, which varies across the credit spectrum.

$$S_{i,t+1} < S_{i,t} \qquad \textbf{SPREAD NARROWING: } \Delta\mathbb{S}_i(t+1) < 0$$

Upgrade

Staying put

$$S_{i,t} \longrightarrow S_{i,t+1} = S_{i,t} \qquad \textbf{SPREAD UNCHANGED: } \Delta\mathbb{S}_i(t+1) = 0$$

Downgrade

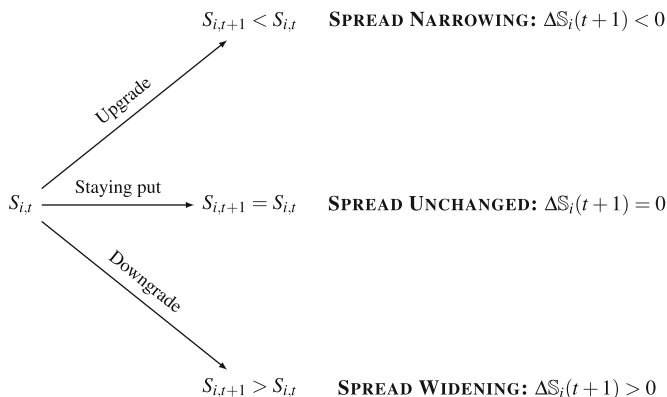$$S_{i,t+1} > S_{i,t} \qquad \textbf{SPREAD WIDENING: } \Delta\mathbb{S}_i(t+1) > 0$$

Fig. 2.5 *Credit-spread cases*: In the credit migration setting, there are three disjoint outcomes: upgrade, downgrade, or staying put. These naturally translate into three associated credit-spread cases: spread narrowing, spread widening, and no change at all. This figure visualizes, in a schematic manner, this situation.

by construction, unconditional.[67] Time homogeneity thus implies that the credit spreads are estimated using long-term, low-frequency data, and are relatively infrequently updated. Much more on this exercise is found in Chap. 3.

To make this more concrete, let us define the credit spread associated with the ith credit counterpart in credit state, $S_{i,t}$ as $\mathbb{S}_{S_{i,t}}$ for $i = 1, \ldots, I$. In fact, there are only $q - 1$ actual credit values; we ignore the default state, q, because it does logically require a spread and the loss implications are fundamentally different. For any time step, therefore, the spread change for a given counterparty can be written as,

$$\Delta\mathbb{S}_i(t+1) = \mathbb{S}_{S_{i,t+1}} - \mathbb{S}_{S_{i,t}}. \tag{2.91}$$

Recall that $S_{i,t} = 1$ describes the highest level of credit quality, $S_{i,t} = 20$ is the lowest, and $S_{i,t} = 21$ denotes default.[68] If $\mathbb{S}_{S_{i,t+1}} < \mathbb{S}_{S_{i,t}}$ (i.e., $\Delta\mathbb{S}_i(t+1) > 0$) then we have a spread widening associated with credit deterioration (i.e., downgrade); this further implies that $S_{i,t+1} > S_{i,t}$. Conversely, $\mathbb{S}_{S_{i,t+1}} > \mathbb{S}_{S_{i,t}}$ ((i.e., $\Delta\mathbb{S}_i(t+1) < 0$) represents a spread tightening with credit improvement (i.e., upgrade), $S_{i,t+1} < S_{i,t}$. Naturally, $\mathbb{S}_{S_{i,t+1}} = \mathbb{S}_{S_{i,t}}$ (i.e., $\Delta\mathbb{S}_i(t+1) = 0$) can only occur when there is no change in underlying credit state. There are thus three disjoint cases for each credit obligor migration in every possible state of the world: downgrade, upgrade, and staying put. Figure 2.5 provides a visualization of these three credit-migration outcomes and the associated credit-spread implications.

---

[67] A number of regulatory mandated buffers attempt to incorporate the, potentially important, consequences of the business cycle for one's economic-capital estimates. These were discussed, in greater length and detail, in Chap. 1.

[68] Other firms will naturally follow differing scales, but the idea is the same.

The definition in Eq. 2.91 permits a straightforward expression for the migration loss associated with the ith credit obligor's migration,

$$- D_{m,i} \cdot \underbrace{\left(\mathbb{S}_{S_{i,t+1}} - \mathbb{S}_{S_{i,t}} \right)}_{\Delta \mathbb{S}_i(t+1)} \cdot c_i, \qquad (2.92)$$

where $D_{m,i}$ and c_i denote the modified spread duration and exposure of the ith credit obligor, respectively. The modified spread duration measure is generally defined as,

$$D_{m,i} = \frac{1}{V_{i,t}} \frac{\partial V_{i,t}}{\partial S_{i,t}}, \qquad (2.93)$$

where $V_{i,t}$ represents the value of the ith credit obligation. While relatively easy to find or compute for market-traded instruments, this value is slightly more difficult to source for non-marketable loan contracts. Our approach to this computation will be discussed in the following chapter.[69]

Given the credit state associated with the ith counterpart in the start and end of each period—which is provided using the ideas from the previous section—determining the credit-spread movement involves looking up one of $q - 1$ values. A simple example may be useful. Imagine that the ith counterparty has a € 15 million exposure with an average modified duration of 4.75 years. At time t, it was in credit state 2 with a (fictitious) credit spread of 100 basis points. In period $t + 1$, obligor ith moved to credit state 3, which has a (equally invented) 120 basis-point spread. From Eq. 2.92, the credit loss amounts to

$$- D_{m,i} \cdot \left(\mathbb{S}_{S_{i,t+1}} - \mathbb{S}_{S_{i,t}} \right) \cdot c_i = -4.75 \cdot (1.20\% - 1.00\%) \cdot (€ \, 15,000,000), \qquad (2.94)$$

$$= -0.95\% \cdot (€ \, 15,000,000),$$

$$= -€ \, 142,500.$$

We thus estimate the credit loss to be approximately € 142,500. Use of the modified duration implies a good first-order approximation of the credit loss and permits a parsimonious, computationally efficient implementation.

The total credit-migration loss is thus,

$$L_{\mathcal{M}} = \sum_{i=1}^{I} D_{m,i} \cdot \Delta \mathbb{S}_i(t+1) \cdot c_i, \qquad (2.95)$$

---

[69] See Bolder [6, Chapter 3] for much more detail on fixed-income instrument sensitivities in general and spread duration in particular.

or, rather, the sum across all credit migrations. We have dispensed with the negative sign in Eq. 2.95 because, as in the default setting, we treat loss as a positive quantity. This is simply convention, but we need to be cautious to treat default and migration losses in a consistent manner. As with the default case, many of the contributions to the sum in Eq. 2.95 will be zero. Transition, however, is typically significantly more probable than default, so relatively speaking, a higher proportion will be non-zero. The overall migration effect—depending, of course, on the overall size of the credit spreads—should involve a smaller magnitude; moreover, there is always the possibility of credit gains to offset loss outcomes. Thus, while the exact size of the migration effect will depend on the portfolio composition, it is generally expected to be somewhat smaller than default losses.[70]

Combining the default and migration losses into a single expression, total credit losses can be defined as,

$$L = L_{\mathcal{D}} + L_{\mathcal{M}}, \tag{2.96}$$

$$= \sum_{i=1}^{I} c_i \left(\underbrace{\gamma_i \mathbb{I}_{\left\{ \Delta X_i(t+1) \leq K_{S_{i,t}}(q) \right\}}}_{\text{Default}} - \underbrace{D_{m,i} \Delta \mathbb{S}_i(t+1) \mathbb{I}_{\left\{ \Delta X_i(t+1) > K_{S_{i,t}}(q) \right\}}}_{\text{Migration}} \right).$$

This representation describes default and credit migration as disjoint, or mutually exclusive, events.[71] This is useful, because although default is a special case of migration, the loss implications are substantially different and need to be modelled separately. The incorporation of migration losses to one's credit-risk modelling framework adds a significant amount of richness to the results.

Both default and migration losses—as highlighted in Eq. 2.96—depend on the systemic and idiosyncratic elements embedded in ΔX_i. This computation continues to be performed with stochastic simulation. We conceptualized the default simulation as pulling systemic variables from hats and then flipping very large numbers of unfair coins. The incorporation of credit migration slightly changes the game. Default is a binary event, whereas migration is multifaceted. While the general anatomy of the simulation approach remains the same, it might help to imagine moving from coins to a dice. With both default and migration, we still draw our (common) systemic factors from a hat. We then proceed to use this outcome

---

[70] The level of confidence also plays a role. A very high-level of confidence will tend to bring out the relatively rare incidence of multiple defaults. Lower levels of confidence may, by this logic, lead to dominance of migration risk in the final economic-capital values.

[71] More specifically, by construction, we have that

$$\left\{ \Delta X_i(t+1) \leq K_{S_{i,t}}(q) \right\} \cap \left\{ \Delta X_i(t+1) > K_{S_{i,t}}(q) \right\} = \emptyset. \tag{2.97}$$

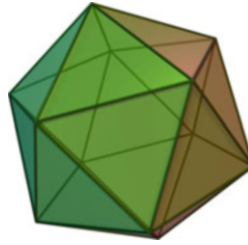DEFAULT: COIN TOSS   MIGRATION: MULTI-SIDED DIE

Fig. 2.6 *Adding migration*: Default, as a binary event, is well conceptualized as an (unfair) coin toss. Migration does not quite work in this setting. It is easier to think about migration as the roll of a large die with many uneven sides where the outcome determines the next period's credit state.

to construct a die with $q + 1$ sides. If the die falls on side 1, the obligor moves to the first credit state; landing on side 2 implies a transition to credit state 2 and so on.[72] Such multi-sided dice actually do exist, but we would need q of them; one corresponding to each row of our transition matrix. Moreover, some of the sides would be incredibly small.[73] Figure 2.6 provides a brief visualization of this important practical twist on the model implementation.

Colour and Commentary 23 (THE IMPORTANCE OF MIGRATION)*: Each firm, when wrestling with the computation of credit-risk economic capital, needs to reflect on the inclusion of migration risk. Historically, our economic-capital computations did not consider the profit-and-loss consequences of credit downgrade or upgrade; only default was considered. During the run-up to statutory change, a decision was taken to incorporate this dimension and capitalize for migration risk. The reasoning was twofold. First of all, our portfolio is, by both construction and mandate, of relatively high credit quality. The risk of default is, for most loan exposures, correspondingly low. Risk of credit deterioration is not.[a] There is a consequent danger of missing a (potentially) important risk dimension. A second rationale arises indirectly through credit impairments or loan-loss provisioning. The loan-loss provision will, in the event of a material deterioration of an obligor's credit quality, lead to an increase in associated impairments. Since this increase will flow through to the profit-and-loss statement, credit-migration has an indirect financial-statement impact.[b] Credit migration has thus been incorporated into our economic capital model with a view towards ensuring greater consistency*

---

[72] Default occurs by rolling $q + 1$ on our die.

[73] So small, in fact, that physically our die idea would never actually work.

Colour and Commentary 23 (continued)
of loan-pricing, creating better incentives for loan-origination, and helping to recognize the potential profit-and-loss implications associated with loan-obligor downgrades. In short, the general view was that incorporation of credit migration provides a more complete *credit-risk characterization. Other firms, depending on their specific situation and objectives, may very well arrive at a different decision.*

[a] Although, it should be stressed, the financial consequences of downgrade are typically rather smaller for downgrade than default.
[b] We return to the interesting question of computing loan impairments in Chap. 9.

2.6   Risk Attribution

This final discussion point in this chapter relates to the notion of risk attribution. Management of economic capital requires us to understand how various sub-components of our portfolio contribute to overall capital demand. This is necessary for loan pricing and capital budgeting. To a certain extent, it also proves helpful in loan-impairment calculations.[74] More simply, it facilitates a better understanding of how risks are distributed throughout one's portfolio. The ability to perform such central analysis requires an allocation, or attribution, of economic capital to individual loans and financial investments. Easy to motivate and understand, it turns out to be a surprisingly mathematically complicated venture. The complexity is, in this case, unavoidable. The simple reason is that, in a practical setting, an economic-capital framework will not get very far without the ability to attribute risks to various aspects of one's portfolio.

In most cases, although not always, risk attribution reduces to an application of Euler's theorem for homogeneous functions. The popularity of this method stems from the fact that it is not only mathematically plausible, but it also ensures both existence and uniqueness. In other words, there are many possible ways one might decompose an aggregate risk metric down to the instrument level, but the standard approach is the most natural and defensible.

To employ the Euler method, we need our risk measure to be a homogeneous function of the portfolio weights or exposures. Homogeneity is a natural feature of a coherent risk measure—this is one of Artzner et al. [3]'s desirable axiomatic properties—and essentially amounts to the idea that if one doubles one's portfolio exposure, all else equal, one doubles the risk.[75] It is hard to argue against such logic.

[74] To see precisely how, please take a look at Chap. 9.

[75] Mathematically, $f(x)$ is termed a homogeneous function of the first order if,

$$f(\lambda x) = \lambda f(x), \tag{2.98}$$

Assuming that one's risk measure, call it $\Xi(c)$, is a first-order homogeneous function of the obligor exposures, $c \in \mathbb{R}^I$, the Euler's result implies that

$$\Xi(c) = \sum_{i=1}^{I} \Xi(c_i), \tag{2.99}$$

$$= \sum_{i=1}^{I} c_i \frac{\partial \Xi(c)}{\partial c_i}.$$

The partial derivative, $\dfrac{\partial \Xi(c)}{\partial c_i}$, is termed the marginal risk associated with Ξ. Conceptually, it is the change in risk associated with an infinitesimally small movement in the underlying exposure of the ith credit counterpart. Both Value-at-Risk (VaR) and expected shortfall, like volatility, are first-order homogeneous risk measures ensuring that Euler decomposition can be employed.[76] Equation 2.99 is an equality; given homogeneity, it must hold. Consequently, one need only identify the marginal risk—that is, partial derivatives—and one has a sensible risk decomposition. In some cases, finding the partial derivatives is straightforward, while in others it can be quite challenging. There are *two* broad approaches to obtaining these partial derivatives in the credit-risk setting: numerical approximation or use of analytical tricks. We'd prefer to use the latter, but sadly we are stuck managing the former.

Despite the usefulness of this general approach to the question of risk attribution, it is not a panacea. Caution is required. The presence of partial derivatives should be, if not a red flag, then a reason for increased attention. By its very construction, a partial derivative is a marginal quantity. In other words, it represents the reasonable approximation of a change in the overall risk for a *small* movement in the exposure of a given contract, or region, or industry. When one considers large changes, however, all bets are off. Zhang and Rachev [44, Page 8] offer a nice overview of the key questions and some of the existing associated research. The proposed risk attribution makes logical and mathematical sense for our purposes, but it important to keep in mind that it is a subtle quantity.

2.6.1 The Simplest Case

The most straightforward setting arises where the partial derivatives can be written in closed form. A good example, which sadly is unavailable to us, is the parametric

---

for any choice of $\lambda \in \mathbb{R}$. See Loomis and Sternberg [27] for more discussion on homogeneous functions.

[76] Bolder [6, Chapter 11] provides, from a market-risk perspective, rather more background on the origins and defensibility of this approach.

market-risk measure. Let's briefly examine the logic. Imagine a portfolio with a vector of K risk-factor weights $\zeta \in \mathbb{R}^K$ along with variance-covariance matrix, Ω. Assuming zero return and abstracting from the (thorny) question of length of time horizon, the parametric VaR, at confidence level α for our imaginary portfolio is typically written as,

$$\text{VaR}_\alpha(L) = \Phi^{-1}(1 - \alpha)\sqrt{\zeta^T \Omega \zeta}. \tag{2.100}$$

The VaR, as we saw previously, is simply the appropriate quantile of the return distribution; given the assumption of Gaussianity, it reduces to a multiple of the portfolio-return variance, $\sqrt{\zeta^T \Omega \zeta}$. Similar expressions are available for expected shortfall. We can, without loss of generality, assume this to be the economic-capital, since subtracting the expected loss (a constant) changes nothing.

The hard work in the risk-factor decomposition of $\text{VaR}_\alpha(L)$ is to compute the marginal value-at-risk. As seen in Eq. 2.99, this is merely the gradient vector of partial derivatives of the VaR measure to the risk-factor exposures summarized in ζ. Mathematically, it has the following form,

$$\frac{\partial \text{VaR}_\alpha(L)}{\partial \zeta} = \Phi^{-1}(1 - \alpha)\frac{\partial \sqrt{\zeta^T \Omega \zeta}}{\partial \zeta}. \tag{2.101}$$

Let's see if we can quickly derive the result in Eq. 2.101 and demonstrate its reasonableness as a risk-attribution method. Directly computing $\frac{\partial \sqrt{\zeta^T \Omega \zeta}}{\partial \zeta}$ is not so straightforward. Instead, define the portfolio volatility as,

$$\sigma_p(\zeta) = \sqrt{\zeta^T \Omega \zeta}. \tag{2.102}$$

Working with the square-root sign is annoying, so let's work with the square of the portfolio volatility (or, rather, variance),

$$\frac{\partial \left(\sigma_p(\zeta)^2\right)}{\partial \zeta} = 2\sigma_p(\zeta)\frac{\partial \sigma_p(\zeta)}{\partial \zeta}, \tag{2.103}$$

$$\frac{\partial \sigma_p(\zeta)}{\partial \zeta} = \frac{1}{2\sigma_p(\zeta)}\frac{\partial \left(\sigma_p(\zeta)^2\right)}{\partial \zeta},$$

$$\frac{\partial \sqrt{\zeta^T \Omega \zeta}}{\partial \zeta} = \frac{1}{2\sigma_p(\zeta)}\frac{\partial \left(\sqrt{\zeta^T \Omega \zeta}^2\right)}{\partial \zeta},$$

$$= \frac{1}{2\sigma_p(\zeta)}\frac{\partial \left(\zeta^T \Omega \zeta\right)}{\partial \zeta},$$

$$= \frac{\Omega \zeta}{\sqrt{\zeta^T \Omega \zeta}},$$

which is a vector in $\mathbb{R}^K$. Plugging this result into Eq. 2.101, we have our desired result,

$$\frac{\partial \text{VaR}_\alpha(L)}{\partial \zeta} = \Phi^{-1}(1-\alpha)\underbrace{\frac{\Omega\zeta}{\sqrt{\zeta^T\Omega\zeta}}}_{\text{Eq. 2.103}}. \tag{2.104}$$

An element-by-element multiplication of these marginal value-at-risk values and the vector, ζ, provides the contribution of each risk-factor to the overall risk. That is,

$$\text{Contribution of }i\text{th risk factor to VaR}_\alpha(L) = \zeta_i \cdot \frac{\partial \text{VaR}_\alpha(L)}{\partial \zeta_i}, \tag{2.105}$$

where ζ_i refers to the ith element of the risk-weight vector, ζ. Does the sum of these individual risk-factor contributions actually lead us to the overall risk-measure value? Consider the dot product of the sensitivity vector, ζ, and the marginal value-at-risk gradient,

$$\zeta^T\frac{\partial \text{VaR}_\alpha(L)}{\partial \zeta} = \zeta^T\underbrace{\left(\Phi^{-1}(1-\alpha)\frac{\Omega\zeta}{\sqrt{\zeta^T\Omega\zeta}}\right)}_{\text{Eq. 2.104}}, \tag{2.106}$$

$$= \Phi^{-1}(1-\alpha)\frac{\zeta^T\Omega\zeta}{\sqrt{\zeta^T\Omega\zeta}},$$

$$= \underbrace{\Phi^{-1}(1-\alpha)\sqrt{\zeta^T\Omega\zeta}}_{\text{Eq. 2.100}},$$

$$= \text{VaR}_\alpha(L).$$

In a few short lines, we have found a sensible and practical approach for the allocation of VaR to individual market-risk factors. Regrettably, while promising and intuitive, this expedient path is not available to us in the credit-risk setting. The threshold and mixture models used for credit-risk measurement do not generally lend themselves to parametric forms, which in turn forces us to rely upon simulation methods for risk-measure estimation.

2.6.2   An Important Relationship

If the partial derivatives cannot be computed analytically, then their numerical computation would appear to be a reasonable solution. Estimating numerical derivatives in a simulation setting, however, is a bit daunting. There is, happily,

a surprising relationship which facilitates the computation of the marginal-risk figures.[77] Its statement requires a bit of background. Let us, as before, define the total default credit loss as the following sum over the I counterparties in the credit portfolio,

$$L = \sum_{i=1}^{I} \underbrace{c_i \gamma_i \mathbb{I}_{\mathcal{D}_i}}_{X_i}, \tag{2.107}$$

where $\mathbb{I}_{\mathcal{D}_i}$, γ_i, and c_i represent our now familiar default-event, loss-given-default, and exposure quantities, respectively. We can use our Gaussian or t-threshold definition of the default event if you like, or imagine something else entirely. Consider Eq. 2.107 to be a general statement.

We now introduce the key relationship. Under positive homogeneity conditions, the VaR measure for the portfolio loss, L, can be written as the sum of the marginal value-at-risk values. That is,

$$\text{Contribution of } i\text{th counterparty to } \text{VaR}_\alpha(L) = \left. \frac{\partial \text{VaR}_\alpha(L + hX_i)}{\partial h} \right|_{h=0} \tag{2.108}$$

$$= \mathbb{E}\left(X_i \,\middle|\, L = \text{VaR}_\alpha(L) \right),$$

where h is a small positive constant. It is also possible to show that,

$$\text{VaR}_\alpha(L) = \sum_{i=1}^{I} \mathbb{E}\left(X_i \,\middle|\, L = \text{VaR}_\alpha(L) \right). \tag{2.109}$$

That is a lot to swallow and does not appear to be particularly believable at first glance. In words, Eq. 2.109 implies that the loss contribution associated with each of the I counterparties—where the total loss exactly coincides with the VaR measure—is equivalent to their overall contribution. Computationally, therefore, one need only use a Monte-Carlo engine to simulate the loss distribution and examine the set of losses at the VaR outcome to determine their individual contributions.

An analogous result is also available for the expected shortfall. In particular,

$$\text{Contribution of } i\text{th counterparty to } \mathcal{E}_\alpha(L) = \left. \frac{\partial \mathcal{E}_\alpha(L + hX_i)}{\partial h} \right|_{h=0}, \tag{2.110}$$

$$= \mathbb{E}\left(X_i \,\middle|\, L \geq \text{VaR}_\alpha(L) \right).$$

---

[77] See Tasche [39], Gourieroux et al. [16], and Rau-Bredow [36] for much more detail.

It can also be analogously shown that

$$\mathcal{E}_\alpha(L) = \sum_{i=1}^{I} \mathbb{E}\left(X_i \,\bigg|\, L \geq \text{VaR}_\alpha(L)\right). \tag{2.111}$$

The difference in the expected-shortfall case—which is not unimportant—is that one needs to compute the average losses greater than or equal to the VaR for the decomposition of this measure. It thus also readily lends itself to extraction from one's Monte-Carlo engine. Derivation of this result is far from trivial, but it is outlined in detail in Bolder [7, Chapter 7]. The reader unfamiliar with these results—and not entirely convinced by their statement without proof—is recommended to invest a bit of time in reviewing their mathematical logic and justification.

2.6.3   The Computational Path

Despite the powerful relationship between conditional expectation and partial derivatives introduced in the previous section, evaluation of the conditional expectation in the case of VaR, from Eq. 2.109, is not terribly straightforward. Algorithms do exist. In the most direct manner, one proceeds in two steps. First, given M simulations, one orders the loss vectors for each iteration in ascending order. Second, using the choice of α, one identifies $\text{VaR}_\alpha(L)$ and then selects the *single* loss vector, L_m, consistent with this VaR estimate. This is, by definition, a singleton set. The losses associated with each instrument—or dimension of interest—are extracted from the loss vector. That is, according to Eq. 2.109, each contribution is equal to $\mathbb{E}\left(L_i \,|\, L = \text{VaR}_\alpha(L)\right)$. It basically comes for free from one's simulation engine. Naturally, there is a catch. This is an incredibly *noisy* estimator of the true risk attributions. As bluntly, but very accurately, stated by Glasserman [14],

> each contribution depends on the probability of a rare event conditional on an even rarer event.

An alternative is to repeat our M-iteration simulation a large number of times and average the resulting portfolio-loss vectors. While this basically works, a depressingly large number of repetitions is required to achieve convergence. In short, the computational expense is exorbitant.

Hallerbach [19] offers an alternative.[78] It is an approximation that begins from the M simulations. Given a small positive number ϵ, the following set is defined

$$\mathcal{L}_{\text{VaR}_\alpha}(\epsilon) = \left[\text{VaR}_\alpha(L) - \epsilon, \, \text{VaR}_\alpha(L) + \epsilon\right]. \tag{2.112}$$

---

[78] This approach, described recently in Mossessian and Vieli [33], appears to still be in broad use in industrial applications.

Each element of $\mathcal{L}_{\mathrm{VaR}_\alpha}(\epsilon)$ is thus a portfolio-loss vector in the neighbourhood of one's VaR estimate. The average across all elements in the portfolio-loss vectors of $\mathcal{L}_{\mathrm{VaR}_\alpha}(\epsilon)$ is a *biased* estimator for the overall VaR and the conditional expectations found in Eq. 2.109. Defining M^* as the number of elements in $\mathcal{L}_{\mathrm{VaR}_\alpha}(\epsilon)$, our estimate of our desired conditional expectations—or equivalently marginal VaR values—is

$$\mathbb{E}\left(L_i \,\middle|\, L = \mathrm{VaR}_\alpha(L)\right) \approx \frac{1}{M^*} \sum_{\ell \in \mathcal{L}_{\mathrm{VaR}_\alpha}(\epsilon)} \ell_i. \tag{2.113}$$

for $i = 1, \ldots, I$. The trick is to find the value of ϵ that increases the number of loss vectors to sufficiently reduce the noise, but not introduce an unreasonable amount of bias. These effects move, of course, in opposite directions. Reducing bias comes at the cost of increased estimation variance and vice versa.[79]

A challenge of this approach is that the sum of approximated conditional expectations will, in principle, no longer sum to the observed VaR. Hallerbach [19] thus, quite practically, suggests rescaling the individual contributions in a proportionate manner to force them back to the desired VaR estimate. If this sounds slightly *ad hoc*, it is because it is. There are, to be honest, not many alternatives.[80]

One legitimate option is to change the risk measure to the expected shortfall. Approximation of its conditional expectation is much more natural. Again, starting from M simulations, the following set is defined

$$\mathcal{L}_{\mathrm{E}_\alpha} = \left\{ L \geq \mathrm{VaR}_\alpha(L) \right\}. \tag{2.114}$$

In this case, we collect all of the portfolio-loss vectors exceeding the designated VaR level. This set, virtually by definition, contains multiple elements. With $M = 10,000$ and $\alpha = 0.99$, it will include by construction, $(1 - 0.99) \cdot 10,000 = 100$ portfolio-loss vectors. If we reduce α to the 99.97th quantile, of course, this falls to just 3.[81] The point nonetheless remains valid. With a judicious choice of M, we can ensure a sufficient number of elements in $\mathcal{L}_{\mathrm{ES}_\alpha}$. The necessity of introducing ϵ and its attendant bias is thus precluded.

---

[79] While practically feasible, the author's personal experience reveals finding the appropriate balance to be a frustratingly slow exercise.

[80] Fan et al. [10] offer an approach exploiting the multivariate copula functions of one's model dependence structure. In a complex, multiple-step simulation, this is unlikely to be extremely helpful.

[81] Naturally, at the 99.99th quantile with $M = 10,000$ we have that $\mathcal{L}_{\mathrm{ES}_\alpha}$ is a singleton set. The interplay between the number of iterations and one's confidence level is, thus, critically important. More on this point in Chap. 4.

Defining $\tilde{M}$ as the number of elements in $\mathcal{L}_{\mathrm{ES}_\alpha}$, our estimate of our desired conditional expectations—or equivalently marginal expected-shortfall values—is

$$\mathbb{E}\left(L_i \,\middle|\, L \geq \mathrm{VaR}_\alpha(L)\right) \approx \frac{1}{\tilde{M}} \sum_{\ell \in \mathcal{L}_{\mathrm{ES}_\alpha}} \ell_i. \qquad (2.115)$$

for $i = 1, \ldots, I$. This is thus both a sharper and unbiased estimate of the risk contribution associated with our expected-shortfall measure. It is also consistent with findings in the literature of greater robustness of expected-shortfall risk measure and attribution estimates.

2.6.4 A Clever Trick

To summarize the previous discussion, the risk-attribution computation for expected shortfall is significantly more stable than its VaR equivalent. The reason, of course, is that our expected-shortfall estimates are an *average* over the values beyond a certain quantile. The VaR figures, conversely, depend on the values *at* a specific quantile. This latter quantile is harder to pinpoint and subject to substantial simulation noise.

Bluhm et al. [5, Chapter 5], keenly aware of the inherent instability of Monte-Carlo-based VaR risk attribution, offer a possible solution. The idea is to find the quantile that equates one's VaR-based economic-capital estimate to the shortfall. They refer to this as *VaR-matched* expected shortfall. Conceptually, it is straight-forward. One computes, using the predetermined level of confidence α, the usual VaR-based economic capital.[82] We seek, however, a new quantile, let's call it α^*, that equates expected shortfall with $\mathrm{VaR}_\alpha(L)$. Mathematically, we seek the α^* root of the following equation,

$$\mathrm{VaR}_\alpha(L) - \mathcal{E}_{\alpha^*}(L) = 0. \qquad (2.116)$$

In some specialized situations, Eq. 2.116 may have an analytic solution. More pragmatically, however, it is readily solved numerically. α^* can be identified, in fact, as the solution to the following one-dimensional non-linear optimization problem

$$\min_{\alpha^*} \left\| \mathrm{VaR}_\alpha(L) - \mathcal{E}_{\alpha^*}(L) \right\|_p, \qquad (2.117)$$

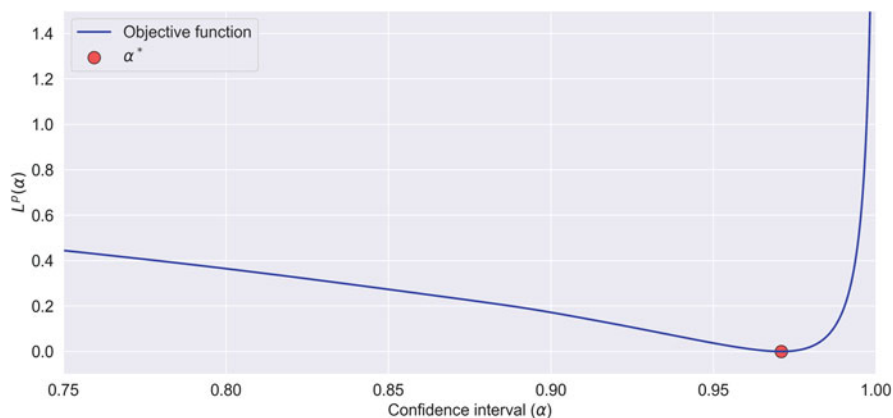[82] Ignoring again, without loss of generality, the constant expected-loss term.

Fig. 2.7 *Equating VaR and expected shortfall*: This graphic highlights the least-squares motivated objective function used to identify the expected-shortfall quantile, α^* that equates the expected shortfall with the 99th quantile VaR estimate.

where $\alpha^* \in (0, 1)$ and $\|\cdot\|_p$ describes the p-norm used to characterize distance. One could, given its low dimensionality, use a brute-force, grid-search algorithm to resolve Eq. 2.117; the ultimate result is conceptually identical.

Figure 2.7 displays the objective function—using the L^2 norm to describe the distance between our two risk measures—for a sample threshold model implementation. The choice of L^2 reduces α^* to a least-squares estimator.[83] $L^p(\alpha)$, as evidenced in Fig. 2.6, demonstrates both a smooth shape and clear minimum highlighted with a red star.

To summarize, an efficient alternative is to identify the α^* equating one's VaR-based economic-capital estimate and the associated expected shortfall. Then one employs this quantity to approximate the VaR-matched expected shortfall contributions. One does not, therefore, directly compute a true VaR risk attribution, but rather accepts Bluhm et al. [5, Chapter 5]'s advice. While numerically defensible, this choice is not without empirical repercussions. Most importantly, a small sleight of hand has been performed. We have quietly switched our risk metric—for risk attribution purposes—from VaR to expected shortfall. While not a criminal act, it is still hard to precisely understand the ultimate consequences.

Despite the cleverness of this trick, there is already a clean logical solution: simply use expected shortfall as one's economic-capital measure. This avoids the need to use mathematical tricks and ensures a structural consistency between one's economic-capital metrics and the associated risk-attribution calculations.[84] Since

---

[83] Other choices of p are, of course, possible, but this option is numerically robust and easily understood.

[84] The remaining issue to keep in mind is the choice of M to ensure adequate numbers for the robust estimation of the risk attributions. This important question is addressed in Chap. 4.

this represents an important change to the economic-capital framework—which is always daunting to revise—it is understandable that one might be reluctant to make a change of this magnitude. Since NIB was already in the midst of a large-scale review of their credit-risk economic capital methodology, it was less difficult. Indeed, this practical risk-attribution advantage—combined with the coherence of expected shortfall and its inherent conservatism—made moving to a new risk metric a relatively easy choice.

Colour and Commentary 24 (RISK ATTRIBUTION)*: The analytic advantages of one's economic-capital framework are severely limited without an ability to attribute overall risk to individual financial instruments. Absent this capability, there is simply no way to understand how capital demand is distributed across one's portfolio. Although essential, in the credit-risk setting this is* not *a trivial task. Useful and tractable semi-analytic methods exist, but are not available in the multivariate setting.[a] We need, therefore, to take recourse to numerical methods. In practice, these methods are much more stable for expected shortfall relative to the VaR metric. Numerous clever tricks exist to improve the situation, but the simplest solution is to use expected shortfall as one's principal economic-capital metric. Accepting this fact—while taking into consideration the other conceptual advantages of the expected-shortfall measure—is precisely what we have done.*

---

[a] Saddlepoint methods, for example, work exceptionally well in low-dimensional cases; see Bolder [7, Chapter 7] for an introduction.

2.7 Wrapping Up

This mathematics-heavy chapter focuses principally on the methodological structure and choices associated with our credit-risk economic-capital model. The entry point is a general, intimidating, and frankly fairly impractical stochastic differential equation describing the creditworthiness (i.e., asset-return) dynamics for the I credit obligors in one's portfolio. Appropriate time discretization, normalization, and notational adjustment reveal the skeleton of a Gaussian threshold model. Some additional machinery—including stochastic recovery—brought us to our legacy implementation. The current economic-capital framework involves two revisions to this basic structure: a new copula function that permits tail dependence and credit migration. This long, and at times winding, road leads us to the final production model.

There are, unfortunately, a multiplicity of moving parts. It is easy to lose the forest for the trees. Table 2.3 thus attempts to help us summarize the key decisions by chronicling the key methodological facts about our model. In one phrase, the

Table 2.3 *Methodology fact sheet*: The underlying table provides, at a glance, a summary of the key methodological choices associated with our credit-risk economic-capital model.

Dimension	Description
Model type	Multivariate threshold model
Time perspective	Discrete time
Horizon	One-year, single step
Copula	t
Credit migration	Yes
Risk metric	Expected shortfall
Confidence level	99.97th quantile
Parameter perspective	Long-term, through-the-cycle approach
Loss-given-default	Random following independent beta distributions.
Transition matrix	20-state NIB master scale

current credit-risk production model is a one-period, multivariate t-threshold model with stochastic recovery and credit migration.

This chapter's objective was to introduce the credit-risk economic capital model. Although these details are absolutely necessary for an understanding of our proposed model, they are *not* sufficient. As every mathematician knows, we need both necessity and sufficiency to be really satisfied.[85] Two additional pieces of the puzzle are required. As a first point, the model does not really come alive until the specific parameters are identified. We would be, in other words, hard-pressed to use the model without clear and concrete coefficients. Model parametrization, which brings many practical challenges and involves numerous decisions, is the focus of the next chapter. The second requirement touches upon model implementation questions. Even if you have all of the mathematics and parameters sorted out, the computations still need to be performed. This task can be categorized as conceptually easy, but practically hard. It is a bit like climbing a mountain. We know that we need only put one foot, and hand, in front of the other. That knowledge, while perhaps comforting in some sense, doesn't necessary ease the difficulty of actually getting up the mountain. Chapter 4 thus acts as our mountain-climbing guide.

References

1. M. Abramovitz and I. A. Stegun. *Handbook of Mathematical Functions*. Dover Publications, New York, 1965.
2. E. I. Altman, B. Brady, A. Resti, and A. Sironi. The link between default and recovery rates: Theory, empirical evidence, and implications. *The Journal of Business*, 78(6):2203–2228, 2005.

---

[85] Even then, one can usually find something to complain about.

3. P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9(3):203–228, July 2001.

4. F. Black and M. S. Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81:637–654, 1973.

5. C. Bluhm, L. Overbeck, and C. Wagner. *An Introduction to Credit Risk Modelling*. Chapman & Hall, CRC Press, Boca Raton, first edition, 2003.

6. D. J. Bolder. *Fixed Income Portfolio Analytics: A Practical Guide to Implementing, Monitoring and Understanding Fixed-Income Portfolios*. Springer, Heidelberg, Germany, 2015.

7. D. J. Bolder. *Credit-Risk Modelling: Theoretical Foundations, Diagnostic Tools, Practical Examples, and Numerical Recipes in Python*. Springer-Nature, Heidelberg, Germany, 2018.

8. P. P. Boyle, M. Hardy, and T. C. Vorst. Life after var. *The Journal of Derivatives*, 13(1):48–55, 2005.

9. P. Brémaud. *Markov Chains: Gibbs Fields, Monte Carlo Simulation and Queues*. Springer Verlag, 175 Fifth Avenue, New York, NY, 1999. Texts in Applied Mathematics.

10. G. Fan, Y. Zeng, and W. K. Wong. Decomposition of portfolio VaR and expected shortfall based on multivariate copula simulation. *International Journal of Management Science*, 7(2): 153–160, 2012.

11. J. Frye. Depressing recoveries. Federal Reserve Bank of Chicago, October 2000.

12. J. Frye. Loss given default as a function of the default rate. Federal Reserve Bank of Chicago, September 2013.

13. M. Gilli and E. Këllezi. An application of extreme value theory for measuring financial risk. *Computational Economics*, 27(1):207–228, 2006.

14. P. Glasserman. Measuring marginal risk contributions in credit portfolios. Risk Measurement Research Program of the FDIC Center for Financial Research, January 2006.

15. G. H. Golub and C. F. V. Loan. *Matrix Computations*. The John Hopkins University Press, Baltimore, Maryland, 2012.

16. C. Gourieroux, J. Laurent, and O. Scaillet. Sensitivity analysis of values at risk. *Journal of Empirical Finance*, 7(3-4):225–245, November 2000.

17. M. Gundlach and F. Lehrbass. *CreditRisk+ in the Banking Industry*. Springer Verlag, Berlin, first edition, 2004.

18. G. M. Gupton, G. Coughlan, and M. Wilson. *CreditMetrics — Technical Document*. Morgan Guaranty Trust Company, New York, 1997.

19. W. G. Hallerbach. Decomposing portfolio value-at-risk: A general analysis. Erasmus University Rotterdam, October 2002.

20. J. D. Hamilton. *Time Series Analysis*. Princeton University Press, Princeton, New Jersey, 1994.

21. J. W. Harris and H. Stocker. *Handbook of Mathematics and Computational Science*. Springer Verlag, New York, New York, 1998.

22. N. L. Johnson, S. Kotz, and N. Balakrishnan. *Continuous Univariate Distributions: Volume 2*. John Wiley and Sons, New York, NY, 1997. Wiley Series in Probability and Statistics.

23. P. Jorion. Risk Management Lessons from Long-Term Capital Management. *European Financial Management*, 6:277–300, 2000.

24. K. Kostadinov. Tail approximation for credit-risk portfolios with heavy-tailed risk factors. Zentrum Mathematik, Technische Universität München, December 2005.

25. G. Kuhn. Tails of credit default portfolios. Zentrum Mathematik, Technische Universität München, December 2005.

26. P. Kumaraswamy. A generalized probability density function for double-bounded random processes. *Journal of Hydrology*, 46(1):79–88, 1980.

27. L. H. Loomis and S. Sternberg. *Advanced Calculus*. World Scientific Publishing, Inc., Hackensack, New Jersey, 2014.

28. D. MacKenzie and T. Spears. "The Formula That Killed Wall Street?" The Gaussian Copula and the Material Cultures of Modelling. University of Edinburgh, June 2012.

29. A. J. McNeil, R. Frey, and P. Embrechts. *Quantitative Risk Management: Concepts, Tools and Techniques*. Princeton University Press, New Jersey, 2015.

30. R. Merton. On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance*, 29:449–470, 1974.
31. S. P. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability*. Springer-Verlag, 175 Fifth Avenue, New York, NY, 1993.
32. J. Morgan/Reuters. *RiskMetrics — Technical Document*. Morgan Guaranty Trust Company, New York, 1996.
33. D. Mossessian and V. Vieli. Decomposing portfolio risk using Monte Carlo estimators. FactSet Research Systems Inc., 2017.
34. B. K. Oksendal. *Stochastic Differential Equations*. Springer Verlag, Berlin, fourth edition, 1995.
35. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, Trumpington Street, Cambridge, second edition, 1992.
36. H. Rau-Bredow. Value-at-risk, expected shortfall, and marginal risk contribution. In G. Szegö, editor, *Risk Measures for the 21st Century*. John Wiley & Sons, 2004.
37. H. L. Royden. *Real Analysis*. Prentice-Hall Inc., Englewood Cliffs, New Jersey, 1988.
38. A. Sklar. Fonctions de répartition à *n* dimensions et leurs marges. *Publications de l'Institut de Statistique de L'Université de Paris*, 8:229–231, 1959.
39. D. Tasche. Risk contributions and performance measurement. Zentrum Mathematik, Technische Universität München, February 2000.
40. O. A. Vasicek. Probability of loss on loan distribution. KMV Corporation, February 1987.
41. O. A. Vasicek. Limiting loan loss probability distribution. KMV Corporation, August 1991.
42. O. A. Vasicek. The distribution of loan portfolio value. *Risk*, (12):160–162, December 2002.
43. T. Wilde. CreditRisk+: A credit risk management framework. Credit Suisse First Boston, January 1997.
44. Y. Zhang and S. Rachev. Risk Attribution and Portfolio Performance Measurement: An Overview. University of California, Santa Barbara, September 2004.