9

Applications in Risk Management

This chapter discusses applications of Monte Carlo simulation to risk management. It addresses the problem of measuring the risk in a portfolio of assets, rather than computing the prices of individual securities. Simulation is useful in estimating the profit and loss distribution of a portfolio and thus in computing risk measures that summarize this distribution. We give particular attention to the problem of estimating the probability of large losses, which entails simulation of rare but significant events. We separate the problems of measuring market risk and credit risk because different types of models are used in the two domains.

There is less consensus in risk management around choices of models and computational methods than there is in derivatives pricing. And while simulation is widely used in the practice of risk management, research on ways of improving this application of simulation remains limited. This chapter emphasizes a small number of specific techniques for specific problems in the broad area of risk management.

9.1 Loss Probabilities and Value-at-Risk

9.1.1 Background

A prerequisite to managing market risk is measuring market risk, especially the risk of large losses. For the large and complex portfolios of assets held by large financial institutions, this presents a significant challenge. Some of the obstacles to risk measurement are administrative — creating an accurate, centralized database of a firm's positions spanning multiple markets and asset classes, for example — others are statistical and computational. Any method for measuring market risk must address two questions in particular:

- What statistical model accurately yet conveniently describes the movements in the individual sources of risk and co-movements of multiple sources of risk affecting a portfolio?

- How does the value of a portfolio change in response to changes in the underlying sources of risk?

The first of these questions asks for the joint distribution of changes in risk factors — the exchange rates, interest rates, equity, and commodity prices to which a portfolio may be exposed. The second asks for a mapping from risk factors to portfolio value. Once both elements are specified, the distribution of portfolio profit and loss is in principle determined, as is then any risk measure that summarizes this distribution.

Addressing these two questions inevitably involves balancing the complexity required by the first with the tractability required by the second. The multivariate normal, for example, has known deficiencies as a model of market prices but is widely used because of its many convenient properties. Our focus is more on the computational issues raised by the second question than the statistical issues raised by the first. It is nevertheless appropriate to mention two of the most salient features of the distribution of changes in market prices and rates: they are typically heavy-tailed, and their co-movements are at best imperfectly described by their correlations. The literature documenting evidence of heavy tails is too extensive to summarize — an early reference is Mandelbrot [246]; Campbell, Lo, and MacKinlay [74] and Embrechts, Klüppelberg, and Mikosch [111] provide more recent accounts. Shortcomings of correlation and merits of alternative measures of dependence in financial data are discussed by, among others, Embrechts, McNeil, and Straumann [112], Longin and Solnik [240], and Mashal and Zeevi [255]. We revisit these issues in Section 9.3, but mostly work with simpler models.

To describe in more detail the problems we consider, we introduce some notation:

$$
\begin{aligned}
S &= \text{vector of } m \text{ market prices and rates;}\\
\Delta t &= \text{risk-measurement horizon;}\\
\Delta S &= \text{change in } S \text{ over interval } \Delta t;\\
V(S,t) &= \text{portfolio value at time } t \text{ and market prices } S;\\
L &= \text{loss over interval } \Delta t\\
&= -\Delta V = V(S,t) - V(S+\Delta S, t+\Delta t);\\
F_L(x) &= P(L < x), \text{ the distribution of } L.
\end{aligned}
$$

The number m of relevant risk factors could be very large, potentially reaching the hundreds or thousands. In bank supervision the interval Δt is usually quite short, with regulatory agencies requiring measurement over a two-week horizon, and this is the setting we have in mind. The two-week horizon is often interpreted as the time that might be required to unwind complex positions in the case of an adverse market move. In other areas of market risk, such as asset-liability management for pension funds and insurance companies, the relevant time horizon is far longer and requires a richer framework.

The notation above reflects some implicit simplifying assumptions. We consider only the net loss over the horizon Δt, ignoring for example the maximum and minimum portfolio value within the horizon. We ignore the dynamics of the market prices, subsuming all details about the evolution of S in the vector of changes ΔS. And we assume that the composition of the portfolio remains fixed, though the value of its components may change in response to the market movement ΔS and the passage of time Δt, which may bring assets closer to maturity or expiry.

The portfolio's value-at-risk (VAR) is a percentile of its loss distribution over a fixed horizon Δt. For example, the 99% VAR is a point x_p satisfying

$$1 - F_L(x_p) \equiv P(L > x_p) = p$$

with $p = 0.01$. (For simplicity, we assume throughout that F_L is continuous so that such a point exists; ties can be broken using (2.14).) A quantile provides a simple way of summarizing information about the tail of a distribution, and this particular value is often interpreted as a reasonable worst-case loss level. VAR gained widespread acceptance as a measure of risk in the late 1990s, in large part because of international initiatives in bank supervision; see Jorion [203] for an account of this history. VAR might more accurately be called a measure of capital adequacy than simply a measure of risk. It is used primarily to determine if a bank has sufficient capital to sustain losses from its trading activities.

The widespread adoption of VAR has been accompanied by frequent criticism of VAR as a measure of risk or capital adequacy. Any attempt to summarize a distribution in a single number is open to criticism, but VAR has a particular deficiency stressed by Artzner, Delbaen, Eber, and Heath [19]: combining two portfolios into a single portfolio may result in a VAR that is larger than the sum of the VARs for the two original portfolios. This runs counter to the idea that diversification reduces risk. Many related measures are free of this shortcoming, including the conditional excess $\mathsf{E}[L|L > x]$, calling into question the appropriateness of VAR.

The significance of VAR (and related measures) lies in its focus on the tail of the loss distribution. It emphasizes a probabilistic view of risk, in contrast to the more formulaic accounting perspective traditionally used to gauge capital adequacy. And through this probabilistic view, it calls attention to the importance of co-movements of market risk factors in a portfolio-based approach to risk, in contrast to an earlier "building-block" approach that ignores correlation. (See, for example, Section 4.2 of Crouhy, Galai, and Mark [93].) We therefore focus on the more fundamental issue of measuring the tail of the loss distribution, particularly at large losses — i.e., on finding $P(L > x)$ for large thresholds x. Once these loss probabilities are determined, it is a comparatively simple matter to summarize them using VAR or some other measure.

The relevant loss distribution in risk management is the distribution under the objective probability measure describing observed events rather than the risk-neutral or other martingale measure used as a pricing device. Historical data is thus directly relevant in modeling the distribution of ΔS. One can imagine a nested simulation (alluded to in Example 1.1.3) in which one first generates price-change scenarios ΔS, and then in each scenario simulates paths of underlying assets to revalue the derivative securities in a portfolio. In such a procedure, the first step (sampling ΔS) takes place under the objective probability measure and the second step (sampling paths of underlying assets) ordinarily takes place under the risk-neutral or other risk-adjusted probability measure. There is no logical or theoretical inconsistency in this combined use of the two measures. It is useful to keep the roles of the different probability measures in mind, but we do not stress the distinction in this chapter. Over a short interval Δt, it would be difficult to distinguish the real-world and risk-neutral distributions of ΔS.

9.1.2 Calculating VAR

There are several approaches to calculating or approximating loss probabilities and VAR, each representing some compromise between realism and tractability. How best to make this compromise depends in part on the complexity of the portfolio and on the accuracy required. We discuss some of the principal methods because they are relevant to our treatment of variance reduction in Section 9.2 and because they are of independent interest.

Normal Market, Linear Portfolio

By far the simplest approach to VAR assumes that ΔS has a multivariate normal distribution and that the change in value ΔV (hence also the loss L) is linear in ΔS. This gives L a normal distribution and reduces the problem of calculating loss probabilities and VAR to the comparatively simple task of computing the mean and standard deviation of L.

It is customary to assume that ΔS has mean zero because over a short horizon the mean of each component ΔS_j is negligible compared to its standard deviation, and because mean returns are extremely difficult to estimate from historical data. Suppose then that ΔS has distribution $N(0, \Sigma_S)$ for some covariance matrix Σ_S. Estimation of this covariance matrix is itself a significant challenge; see, for example, the discussion in Alexander [10].

Further suppose that

$$\Delta V = \delta^\top \Delta S, \tag{9.1}$$

for some vector of sensitivities δ. Then $L \sim N(0, \sigma_L^2)$ with $\sigma_L^2 = \delta^\top \Sigma_S \delta$, and the 99% VAR is $2.33\sigma_L$ because $\Phi(2.33) = 0.99$.

One might object to the normal distribution as a model of market movements because it can theoretically produce negative prices and because it is inconsistent with, for example, a lognormal specification of price levels. But all we need to assume is that the change ΔS over the interval $(t, t + \Delta t)$ is *conditionally* normal given the price history up to time t. Given S_j, assuming that ΔS_j is normal is equivalent to assuming that the return $\Delta S_j/S_j$ is normal. For small Δt,

$$S_j(1 + \Delta S_j/S_j) \approx S_j \exp(\Delta S_j/S_j),$$

so the distinction between normal and lognormal turns out to be relatively minor in this setting.

It should also be noted that assuming that ΔS is conditionally normal imposes a much weaker condition than assuming that changes over disjoint intervals of length Δt are i.i.d. normal. In our calculation of the distribution of ΔV based on (9.1), Σ_S is the conditional covariance matrix for changes from t to $t + \Delta t$, given the price history to time t. At different times t, one would ordinarily estimate different covariance matrices. The unconditional distribution of the changes ΔS would then be a mixture of normals and could even be heavy-tailed. This occurs, for example, in GARCH models (see Section 8.4 of Embrechts et al. [111]). Similar ideas are implicit in the discretization methods of Chapter 6: the increments of the Euler scheme (6.2) are conditionally normal at each step, but the distribution of the state can be far from normal after multiple steps.

Delta-Gamma Approximation

The assumption that V is linear in S holds, for example, for a stock portfolio if S is the vector of underlying stock prices. But a portfolio with options has a nonlinear dependence on the prices of underlying assets, and fixed-income securities depend nonlinearly on interest rates. The model in (9.1) is thus not universally applicable.

A simple way to extend (9.1) to capture some nonlinearity is to add a quadratic term. The quadratic produced by Taylor expansion yields the *delta-gamma* approximation

$$\Delta V \approx \frac{\partial V}{\partial t}\Delta t + \delta^\top \Delta S + \tfrac{1}{2}\Delta S^\top \Gamma \Delta S, \tag{9.2}$$

where

$$\delta_i = \frac{\partial V}{\partial S_i}, \quad \Gamma_{ij} = \frac{\partial^2 V}{\partial S_i \partial S_j}$$

are first and second derivatives of V evaluated at $(S(t), t)$. This in turn yields a quadratic approximation to $L = -\Delta V$.

For this approximation to have practical value, the coefficients must be easy to evaluate and finding the distribution of the approximation must be substantially simpler than finding the distribution of L itself. As discussed in Chapter 7, calculating δ and Γ can be difficult; however, these sensitivities are routinely calculated for hedging purposes by individual trading desks and can be aggregated (at the end of the day, for example) for calculation of firmwide risk. This is a somewhat idealized description — for example, many off-diagonal gammas may not be readily available — but is sufficiently close to reality to provide a valid premise for analysis.

If ΔS has a multivariate normal distribution, then finding the distribution of the approximation in (9.2) requires finding the distribution of a quadratic function of normal random variables. This can be done numerically through transform inversion. We detail the derivation of the transform because it is relevant to the techniques we apply in Section 9.2.

Delta-Gamma: Diagonalization

The first step derives a convenient expression for the approximation. As in Section 2.3.3, we can replace the correlated normals $\Delta S \sim N(0, \Sigma_S)$ with independent normals $Z \sim N(0, 1)$ by setting

$$\Delta S = CZ \quad \text{with} \quad CC^\top = \Sigma_S.$$

In terms of Z, the quadratic approximation to $L = -\Delta V$ becomes

$$L \approx a - (C^\top \delta)^\top Z - \tfrac{1}{2} Z^\top (C^\top \Gamma C) Z \tag{9.3}$$

with $a = -(\Delta t)\partial V/\partial t$ deterministic.

It is convenient to choose the matrix C to diagonalize the quadratic term in (9.3), and this can be accomplished as follows. Let $\tilde{C}$ be any square matrix for which $\tilde{C}\tilde{C}^\top = \Sigma_S$, such as the one found by Cholesky factorization. The matrix $-\frac{1}{2}\tilde{C}^\top \Gamma \tilde{C}$ is symmetric and thus admits the representation

$$-\tfrac{1}{2}\tilde{C}^\top \Gamma \tilde{C} = U \Lambda U^\top$$

in which

$$\Lambda = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_m \end{pmatrix}$$

is a diagonal matrix and U is an orthogonal matrix ($UU^\top = I$) whose columns are eigenvectors of $-\frac{1}{2}\tilde{C}^\top \Gamma \tilde{C}$. The λ_j are eigenvalues of this matrix and also of $-\frac{1}{2}\Gamma \Sigma_S$. Now set $C = \tilde{C}U$ and observe that

$$CC^\top = \tilde{C}UU^\top \tilde{C}^\top = \Sigma_S$$

and

$$-\tfrac{1}{2}C^\top \Gamma C = -\tfrac{1}{2}U^\top (\tilde{C}^\top \Gamma \tilde{C})U = U^\top (U \Lambda U^\top) U = \Lambda.$$

Thus, by setting $b = -C^\top \delta$ we can rewrite (9.3) as