Value-at-Risk

Value at Risk (VaR) is a technique for analyzing portfolio market risk based on a known—or at least posited—return model. VaR has dreadful limitations, both as a model and as a practice of risk managers in real-world applications. We talk about these at length in Chapters 10 and 11. And VaR has come under ferocious, and to a large extent justified, attack from many quarters. But it is nonetheless an extremely important technique, for a number of reasons. It is worth laying these out carefully, lest the poor reputation of VaR lead to the neglect of an important analytical tool:

- Uniquely among risk measures, VaR provides a single number summarizing the risk of even large and complex portfolios. That number may be inaccurate. But in situations that require at least an order-of-magnitude estimate of the potential loss of a portfolio, or in which a comparison must be made as to which of two portfolios is riskier, it is hard to avoid using some sort of VaR estimate, at least as one among other risk measures.
- Examining VaR, its assumptions, the different ways to compute it, and its pitfalls, is an excellent way to understand the issues involved in risk measurement. We study VaR techniques not because we believe in them or in any particular distributional hypothesis literally, but so that we can learn to ask the right questions.
- VaR is a reasonably accurate guide to the "main body" risks of most portfolios, that is, losses that are large, but nonetheless routine. There are, of course, important exceptions to the information value of VaR even as a general indicator of riskiness, such as option portfolios and dynamic strategies. We discuss these cases in Chapters 4 and 13. But that VaR is a poor guide to extreme losses and for some types of portfolios does not diminish its value as a measure of overall risk in many cases.
- As we see in Chapter 13, there are risk measures related to VaR that give insights into the contributions of different positions to the total risk

of a portfolio. These measures can be of value in the process of portfolio construction.

- ▪ Criticism of VaR is often bundled with criticism of modeling assumptions, such as normally distributed returns, that are not inherent in VaR, but are made in specific implementations of VaR. As we will see below, these are distinct issues and should be distinguished in critical review.
- ▪ Finally, VaR was developed because there was need for it. Before VaR, there were risk measures that pertained to equity portfolios, or to fixed income, and so on, but no way to summarize the risk of a portfolio containing different asset classes. As first banks and then other types of financial firms became more heavily involved in trading activities, VaR was an early response to this gap.

3.1   DEFINITION OF VALUE-AT-RISK

Let's start by defining VaR formally. We denote portfolios by x. We can think of x as a vector, with elements representing the amounts of different securities or risk factors, or as a scalar, representing the exposure of a single position to a single risk factor. For most of the rest of this chapter, we stick to the simplest case: We assume the portfolio consists of a single long position, x shares or par amount of an asset, or units of exposure to a single risk factor, with an initial price of S_t. Section 3.4 below discusses short positions.

We have already been using S_t interchangeably for risk factors and simple asset prices that are captured by the standard model and its near relations. Later in this and subsequent chapters, we discuss some of the issues around *mapping*, that is, deciding how to represent a risky position with specific observable data. We will extend the discussion to portfolios and to positions that are more difficult to represent in Chapters 4 and 5.

The value of the position at time t is

$$V_t = xS_t$$

The current, time-t value of the position is a known value V_t, but the future value $V_{t+\tau}$ is a random variable. The dollar return or *mark-to-market* (MTM) profit or loss (P&L) $V_{t+\tau} - V_t$ is also a random variable:

$$V_{t+\tau} - V_t = x(S_{t+\tau} - S_t)$$

VaR is, in essence, a quantile of the distribution of this P&L. That is, VaR is a loss level that will only be exceeded with a given probability in a given time frame. If we know the distribution of returns on S_t, we can compute the VaR of the position. In measuring VaR, we hold the portfolio, that is, x, constant, and let the risk factor or asset price fluctuate.

As noted in the introduction to this chapter, the way the term "VaR" is used can be misleading. VaR is often thought of as a particular model and procedure for estimating portfolio return quantiles. We describe this standard model in detail in a moment. Right now, we are presenting a more general definition of VaR as a quantile. It is not tied to a particular model of risk factor returns, or to a particular approach to estimating quantiles or other characteristics of the portfolio return.

As a quantile of $V_{t+\tau} - V_t$, VaR is the expected worst case loss on a portfolio with a specific confidence level, denoted α, over a specific holding period; one minus the confidence level is the probability of a loss greater than or equal to the VaR over the specified time period:

$$\text{VaR}_t(\alpha, \tau)(x) \equiv -(V^* - V_t) \quad \text{s.t.} \quad 1 - \alpha = \mathbf{P}\left[V_{t+\tau} - V_t \leq V^* - V_t\right]$$

The minus sign in this definition serves to turn a loss into a positive number, the way VaR is typically expressed. A negative VaR means that the worst P&L expected at the given confidence level is a profit, not a loss.

We can also express the P&L in terms of returns. Let $r_{t,\tau}$ denote the τ-period logarithmic return on the risk factor at time t. To simplify notation, we set the compounding interval of the return equal to the time horizon of the VaR. Thus, for example, if we are computing a one-day VaR, we will express the return at a daily rate. The return is a random variable, related to the P&L $V_{t+\tau} - V_t$, as we saw in the previous chapter, by

$$V_{t+\tau} - V_t = x(S_{t+\tau} - S_t) = xS_t\left(\frac{S_{t+\tau}}{S_t} - 1\right) = xS_t(e^{r_{t,\tau}} - 1)$$

That is, P&L is equal to the initial value of the position times a proportional shock to the risk factor.

From this definition, we can derive r^*, the risk factor return corresponding to the VaR:

$$V^* = xS_t e^{r^*}$$

We'll call V^*, the position value at which a loss equal to the VaR is realized, the *VaR scenario*, and r^*, the logarithmic change in S_t at which the VaR

scenario occurs, the *VaR shock*. The VaR shock, in other words, is a quantile of the return distribution. The VaR itself is

$$\text{VaR}_t(\alpha, \tau)(x) = -(V^* - V_t) = -xS_t(e^{r^*} - 1) = xS_t(1 - e^{r^*}) \qquad (3.1)$$

Note that, in the definition thus far, we haven't made any assumptions about the distribution of risk factor returns. A common misconception about VaR is that it always assumes log returns are normally distributed. It is true that in practice, for a number of good and bad reasons, the normal return model is, in fact, usually assumed. But that is not inherent in VaR. VaR can be computed in a number of ways, and can be based on any tractable hypothesis about return distributions; a quantile does not have to be a quantile of the normal distribution.

Figure 3.1 illustrates, with a simple example, a single long position in one unit of AMEX SPDR shares, an ETF that represents a tradable version of the S&P 500 index. The date on which the calculations are made is November 10, 2006, with the closing value of the S&P 500 index at 1376.91. The graph plots the probability distribution of the next-day index level. The one-day
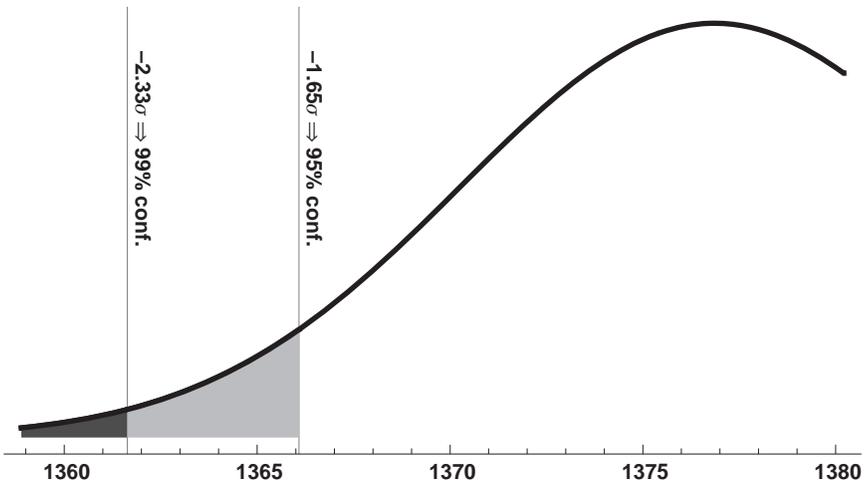

FIGURE 3.1   Definition of VaR
Overnight (1-day) VaR for one unit of the S&P 500 index on November 10, 2006. The drift is set equal to zero, and the volatility is computed using the EWMA technique, described in Section 3.2.2, with 250 days of data and a decay factor if 0.94. The vertical grid lines represent the VaR scenarios corresponding to $\alpha = 0.95$ and 0.99. The VaR is the difference between the index value in the VaR scenario and 1376.91, the closing value of the index.

VaR is the difference between the 1- and 5-percent quantiles of the next-day price, represented by vertical grid lines, and the closing price.

3.1.1 The User-Defined Parameters

Regardless of the distributional hypothesis, the risk manager determines two parameters when computing a VaR. They are set pragmatically, with reference to the type of investor and its activities, or by a financial intermediary's regulatory overseer, rather than being determined by other model elements:

The time horizon or holding period. We will denote this τ, measured in days or years. For a portfolio of short-term positions, it is typically 1 day to 1 month. Longer-term investors, such as individuals and pension funds, will be more interested in a VaR with a time horizon of one year or more. For a given portfolio, VaR for a longer time horizon (say, 1 month) is greater than VaR for short horizon (say, overnight).

When treating periods less than one year, as is typical in VaR calculations, we usually consider only trading days rather than calendar days. This convention is consistent with that used in calculating, say, daily from annual volatilities.

The confidence level (typically 95, 99, or 99.9 percent), that is, what is the probability of the quantile being exceeded? For a given portfolio, VaR at a higher confidence level is greater than VaR for lower confidence interval. This can be seen in Figure 3.1; the worst case loss at a 99 percent confidence level is higher than that at a 95 percent confidence level.

The choice of confidence level is a difficult problem in using VaR. On the one hand, one would like to use VaR at a high confidence level, since it then provides an estimate of the extremely high, unusual and infrequent losses that the investor seeks to avoid. At a lower confidence level, the VaR provides an estimate of "main body" risks, the smaller, more regularly occurring losses that, while interesting, are not as crucial an element of risk management. For example, VaR at a confidence level of 95 percent and with an overnight time horizon provides an estimate of losses that can be expected to occur once a month. Such losses could be considered a recurrent fact of investing life or a cost of doing business, rather than an extreme loss absolutely to be avoided.

On the other hand, in our current state of knowledge, the accuracy of VaR tends to decrease as the confidence level rises. While

one would like to use VaR at a 99.99 percent confidence level, and obtain estimates of the worst loss to be expected on average every 10,000 trading days, or about 40 years, such estimates are likely to be devoid of predictive, and thus practical, value.

3.1.2   Steps in Computing VaR

Having defined VaR, we now describe how to compute it. At this point, more needs to be said about maintained hypotheses concerning return distributions. At an "outline" level, the choices and steps in computing VaR are

Establish a return distribution. In this step, based on some combination of real conviction about the actual distribution and mere convenience, we stipulate a distributional hypothesis, that is, we state what family of distributions the returns are drawn from. We can then estimate the parameters of the distribution. Or we may instead let the data speak more directly for themselves and rely on the quantiles of the returns, without imposing a distributional hypothesis. Either way, a judgement call is required. Nothing in theory or the data will tell the risk manager unambiguously which approach is better.

Choose the horizon and confidence level. This choice will be based both on the type of market participant using the VaR, and on limitations of data and models.

Mapping. This step relates positions to risk factors. It is as important as the distributional hypothesis to the accuracy of the results. To actually compute a VaR, we need return data. For any portfolio, we have to decide what data best represent its future returns. Sometimes this is obvious, for example, a single position in a foreign currency. But most often it is *not* obvious, for reasons ranging from paucity of data to the complexity of the portfolio, and choices must be made.

Compute the VaR. The last step, of course, is to put data and models together and compute the results. Many implementation details that have a material impact on the VaR estimates are decided in this stage. For this reason, software, whether internally developed or obtained from a vendor, is so important, as we see in Chapter 11's discussion of model risk.

Each of these steps requires judgement calls; none of these choices will be dictated by something obvious about the models or in the environment.

3.2   VOLATILITY ESTIMATION

The typical assumption made in VaR estimation is that logarithmic returns are normally distributed, generally with a mean of zero. The key parameter to be estimated therefore is the volatility. There are a number of techniques for estimating volatility. Most are based on historical return data, but one can also use implied volatility, which we discuss in Chapters 4, 5 and 10.

Criteria for a "good" volatility estimator include:

- We want to capture the true behavior of risk factor returns. Recall that in the geometric Brownian motion model, the volatility for an asset is a constant, never changing over time. This is far from characteristic of real-world asset returns.
- We also want a technique that is easy to implement for a wide range of risk factors.
- The approach should be extensible to portfolios with many risk factors of different kinds.

3.2.1   Short-Term Conditional Volatility Estimation

Next, we describe the standard approach to estimating a VaR. We do so in the following steps:

- Posit a statistical model for the joint distribution of risk factor returns.
- Estimate the parameters of the model.
- Estimate the portfolio return distribution.

If we posit that logarithmic returns are jointly normally distributed, risk factors follow a geometric Brownian motion. As we saw in Chapter 2, the level of the risk factor is then a random variable that can be described by Equation (2.2), reproduced here:

$$S_{t+\tau} = S_0 e^{\mu dt + \sigma dW_t}$$

We look at several ways to use historical return data to estimate the volatility. As we do so, bear in mind that VaR is oriented towards short-term risks. It is harder to forecast short-term expected return than short-term volatility. In fact, short-term volatility forecasts are relatively accurate, compared to forecasts of returns, or to longer-term volatility forecasts. Over short time horizons, the drift term μ is likely to make a much smaller contribution to total return than volatility. Practitioners make a virtue of

necessity and typically restrict VaR estimates to a time horizon over which forecasting can have some efficacy.

What does "short-term" mean in this context? There is no fixed standard of the longest time horizon for which a VaR estimate is useful. It depends partly on one's views on the true underlying model to which the lognormal is an approximation, and what that model says about the forecastability of volatility and the unimportance of the drift. The practice is to neutralize the drift, which is likely to be comparatively small. In the geometric Brownian motion model outlined above, we set

$$\mu = -\frac{\sigma^2}{2}$$

so as to zero out the slight acceleration of growth resulting from continuous compounding, and keep the drift over discrete intervals equal to zero.

To see why we can safely ignore the mean, consider the potential range of returns for varying levels of the drift and volatility, and over long and short time intervals, if the asset price follows a geometric Brownian motion. The table below shows the 1 percent quantile of the logarithmic return, for time horizons of one day and one year, and for varying assumptions on the drift and volatility parameters. In other words, for each set of assumed parameters, there is a 1 percent chance that the return will be even lower than that tabulated below:

	$\tau = \frac{1}{252}$ (1 day)			
	$\mu = 0.00$	$\mu = 0.10$	$\mu = 0.20$	$\mu = 0.30$
$\sigma = 0.10$	-0.015	-0.014	-0.014	-0.013
$\sigma = 0.50$	-0.074	-0.073	-0.073	-0.073
	$\tau = 1$ (1 year)			
	$\mu = 0.00$	$\mu = 0.10$	$\mu = 0.20$	$\mu = 0.30$
$\sigma = 0.10$	-0.233	-0.133	-0.033	0.067
$\sigma = 0.50$	-1.163	-1.063	-0.963	-0.863

The comparison is similar to that in Figure 2.5. We see that, for short time horizons, the drift makes very little difference, since it is linear in time ($\mu\tau$) and small for a short time horizon, even for a high value of μ. The potential loss is greater when volatility is high, 50 percent as compared to

10 percent per annum, but the difference, for either volatility, between a drift of zero and a drift of 30 percent per annum is small. In contrast, for a time horizon of one year, the drift can make a substantial difference. At a low volatility of 10 percent, increasing the drift from zero to 30 percent raises the worst-case outcome with a 99 percent confidence level from a loss of 23.3 percent to a gain of 6.7 percent. The upward shift in worst-case outcomes arising from a higher drift is the same at a higher volatility, but stands out less against the backdrop of the higher potential loss the higher volatility makes possible.

With the drift set to zero, we have

$$S_{t+\tau} = S_t e^{\sigma \, dW_t} = S_t e^{r_t}$$

where

$$r_t \sim N\left(-\frac{\sigma^2}{2}\tau, \sigma\sqrt{\tau}\right)$$

For a long position of x units of an exposure to a single risk factor, the P&L

$$V_{t+\tau} - V_t = xS_t\left(e^{r_t} - 1\right)$$

is lognormally distributed. Assume for now that we know the value of the portfolio or asset return volatility σ. In fact, of course, we don't; it has to be estimated.

The P&L is negative when $r_t < 0$. The VaR at a confidence level α is the 1-α quantile of the P&L distribution. This corresponds to a return equal to $z_*\sigma\sqrt{\tau}$,

where α is the confidence level of the VaR, e.g., 0.99 or 0.95,
           z_* is the ordinate of the standard normal distribution at
              which $\Phi(z) = 1 - \alpha$,
           σ is the time-t annual volatility estimate,
           τ is the time horizon of the VaR, measured as a fraction of a year.

Note that for any $\alpha > 0.50$, $z_* < 0$. Figure 3.1 illustrates how VaR is calculated and how it is affected by the confidence interval. The VaR shock is related to volatility by

$$r^* = z_*\sigma\sqrt{\tau}$$

so the α percentile of $V_{t+\tau} - V_0$ is

$$xS_t \left(e^{z_* \sigma \sqrt{\tau}} - 1 \right)$$

and the VaR is estimated as

$$\text{VaR}_t(\alpha, \tau)(x) = -xS_t \left(e^{z_* \sigma \sqrt{\tau}} - 1 \right)$$

The logarithmic change in the asset price at which a loss equal to the VaR is incurred, $z_* \sigma \sqrt{\tau}$, is the VaR shock, and $e^{z_* \sigma \sqrt{\tau}} xS_t$, the position value at which a loss equal to the VaR is incurred, is the VaR scenario.

Example 3.1   At a 99 percent confidence level, we have $1 - \alpha = 0.01$ and $z_* = -2.33$. Suppose time is measured in years, the time horizon of the VaR estimate is one day or overnight, portfolio value is lognormally distributed, and the volatility is estimated at $\sigma = 0.24$ per annum. The daily volatility, using the rule of thumb described in Chapter 2, is then $\sigma \sqrt{\tau} = 0.24 \frac{1}{16} = 0.015$, or 1.5 percent. The VaR shock, finally, is $z_* \sigma \sqrt{\tau} = -2.33 \cdot 0.015$ or approximately -3.5 percent. In other words, the overnight, 99 percent VaR is about 3.5 percent of the initial portfolio value.

Note that for this simple long position, $z_* < 0$. As we will see below, for short positions or for securities whose values decrease as the risk factor increases, we may have $z_* > 0$ at the same confidence level.

In the model as presented so far, the volatility parameter σ is a constant. Volatility, however, varies over time. *Time-varying volatility* is observed in the returns of all financial assets. Volatility changes over time in a certain particular way. Periods of lower and higher volatility are persistent. If the magnitude of recent returns has been high, it is likelier that returns in the near future will also be of large magnitude, whether positive or negative.

The phenomenon, also described as *volatility clustering*, presents a number of challenges in interpreting financial markets. In empirical research, it is difficult to "explain" asset prices by associating them with fundamental factors in the economy via a model. The observed variations in asset return volatility cannot be persuasively linked to variations in fundamental drivers of asset returns. For example, stock market return volatility is a well-documented phenomenon over centuries of available historical price data, but cannot be satisfactorily explained by the behavior of earnings, dividends, economic growth, and other fundamentals, which are far less variable.

From one point of view, this presents an obstacle to validating the theory that financial markets are efficient. The difficulty of explaining asset return volatility via fundamentals has been described as the problem of *excess volatility* and has given rise to a large literature attempting to explain this excess with models of how traders put on and unwind positions over time. However, it is also difficult to find fundamentals-based return models that outperform simple short-term return forecasts such as one based on the random walk, that is, essentially, a coin toss.

We put time-varying volatility in the context of other departures from the standard geometric Brownian motion return model, such as "fat tails," in Chapter 10. Here, we will present techniques for forecasting time-varying volatilities.

Future volatility can be forecast from recent volatility with some accuracy, at least over short horizons, even though returns themselves are hard to predict. We can obtain a current estimate of volatility, using recent data on return behavior, denoted $\hat{\sigma}_t$ to remind us that we are computing an estimate of the *conditional volatility* dependent on information available at time t.

A first approach is to use the sample standard deviation, the average square root of daily return deviations from the mean over some past time period:

$$\hat{\sigma}_t = \sqrt{\frac{1}{T} \sum_{\tau=1}^{T} (r_\tau - \bar{r})^2}$$

where T is the number of return observations in the sample, and $\bar{r}$, the sample mean, is an estimator of μ.

Since we've decided that, for short periods at least, we can safely ignore the mean, we can use the *root mean square* of r_t:

$$\hat{\sigma}_t = \sqrt{\frac{1}{T} \sum_{\tau=1}^{T} r_\tau^2}$$

This statistic is the second sample moment and the maximum likelihood estimator of the standard deviation of a normal, zero-mean random variable. This estimator adapts more or less slowly to changes in return variability, depending on how long an observation interval T is chosen.

3.2.2   The EWMA Model

The sample standard deviation does too little to take the phenomenon of time-varying volatility into account. An alternative approach is the "Risk-Metrics model," which applies an *exponentially weighted moving average* (EWMA) to the return data. A *decay factor* $\lambda \in (0, 1)$ is applied to each observation so that more recent observations are given greater weight in the volatility estimate than observations in the more remote past.

The decay factor is typically close to, but less than unity. As long as the decay factor is less than unity, the resulting volatility estimator is stable. The decay factor is chosen so as to be consistent with the long-term behavior of short-term volatility. For quite short horizons—a few days—a value of about 0.94 has found considerable empirical support. For somewhat longer periods, such as one month, a "slower" decay factor of 0.97 has been recommended, among others, by bank regulators.[1]

The user chooses a number T of daily time series observations $\{r_1, r_2, \ldots, r_T\}$ on risk factor returns. The weight on each return observation is

$$\frac{1 - \lambda}{1 - \lambda^T} \lambda^{T-\tau} \qquad \tau = 1, \ldots, T$$

The weight is composed of two parts:

- $\lambda^{T-\tau}$ is the impact of the decay factor on the $T - t$-th observation. The most recent observation is the last (i.e., sequence number T) in a time series in standard ascending order by date. It has a weight of $\lambda^{T-T} = 1$; as the most recent, it carries a "full weight." For the first, that is, the observation furthest in the past, the value is λ^{T-1}. As $T \to \infty$, this tends to zero.
- $\frac{1-\lambda}{1-\lambda^T}$ "completes the weight" and ensures all the weights add up to unity. From the formula for the sum of a geometrical series, we have

$$\sum_{\tau=1}^{T} \lambda^{T-\tau} = \sum_{\tau=1}^{T} \lambda^{\tau-1} = \frac{1 - \lambda^T}{1 - \lambda}$$

so the sum of the weights is unity.

---

[1]Fleming, Kirby, and Ostdiek (2001) find a decay factor of 0.94 fits a wide range of assets well. The use of VaR is used in setting regulatory capital in some developed countries and is discussed in Chapter 15.
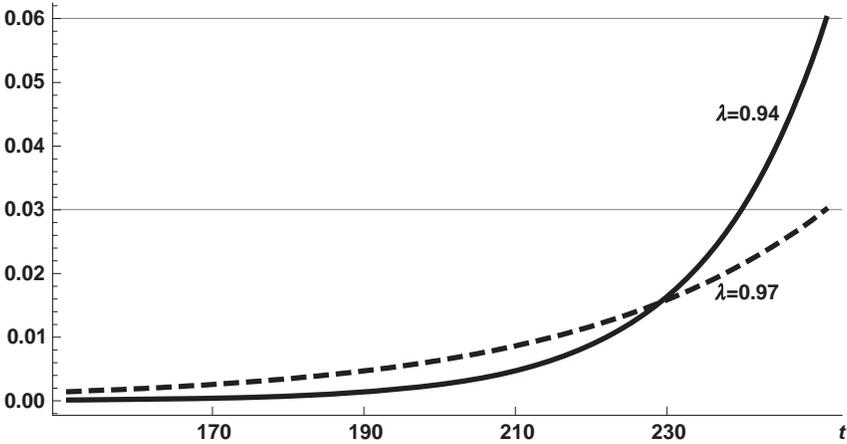
FIGURE 3.2   The EWMA Weighting Scheme
The graph displays the values of the last 100 of $T = 250$ EWMA weights $\frac{1-\lambda}{1-\lambda^T}\lambda^{T-t}$
for $\lambda = 0.94$ and $\lambda = 0.97$.

Figure 3.2 displays the behavior of the weighting scheme. The weights decline smoothly as the return observations recede further into the past. If the decay factor is smaller, more recent observations have greater weight, and past observations are de-emphasized more rapidly.

If you start out with a time series of asset prices, you lose one observation for the return calculations and T observations for the EWMA calculation, for a total of $T + 1$ fewer observations on the EWMA volatility than on the asset prices themselves.

As $T \to \infty$, $\lambda^T \to 0$, so if T is large, the weights are close to $(1 - \lambda)\lambda^{T-\tau}$. The EWMA estimator can then be written

$$\hat{\sigma}_t = \sqrt{(1 - \lambda) \sum_{\tau=1}^{t} \lambda^{t-\tau} r_\tau^2}$$

If T is large, the EWMA estimator can also be written in *recursive form* as

$$\hat{\sigma}_t^2 = \lambda \hat{\sigma}_{t-1}^2 + (1 - \lambda) r_t^2 \tag{3.2}$$

To use the recursive form, you need a starting value $\hat{\sigma}_0$ of the estimator, which is then updated using the return time series to arrive at its current value $\hat{\sigma}_t$. If $\hat{\sigma}_0$ is not terribly far from $\hat{\sigma}_t$, it will rapidly converge to the correct value of $\hat{\sigma}_t$.

The recursive form provides additional intuition about the procedure:

- The smaller is λ, the faster the EWMA estimator adapts to new information.
- The larger the absolute value of the latest return observation, the larger the change in the EWMA estimator. The sign of the return doesn't matter, only its magnitude.
- The EWMA estimator is not that different from the sample moment; it merely adapts faster to new information. But if you believe that volatility is time-varying, you should not use the sample moment.

Example 3.2 (Computing the EWMA Volatility Estimator)   Suppose we wish to find the value of the EWMA volatility estimator for the S&P 500 index on November 10, 2006, using closing values of the index and setting $T = 250$ and $\lambda = 0.94$. We need 251 observations on the index, giving us 250 return observations. The table displays the weights, index values, and returns for the last 11 observations used.

Obs	Date	λ^{T-t}	$\frac{1-\lambda}{1-\lambda^T}\lambda^{T-t}$	S_t	$100r_t$
240	27Oct06	0.539	0.032	1377.34	−0.849
241	30Oct06	0.573	0.034	1377.93	0.043
242	31Oct06	0.610	0.037	1377.94	0.001
243	01Nov06	0.648	0.039	1367.81	−0.738
244	02Nov06	0.690	0.041	1367.34	−0.034
245	03Nov06	0.734	0.044	1364.30	−0.223
246	06Nov06	0.781	0.047	1379.78	1.128
247	07Nov06	0.831	0.050	1382.84	0.222
248	08Nov06	0.884	0.053	1385.72	0.208
249	09Nov06	0.940	0.056	1378.33	−0.535
250	10Nov06	1.000	0.060	1376.91	−0.103

3.2.3   The GARCH Model

The *generalized autoregressive conditionally heteroscedastic* (GARCH) model can be seen as a generalization of EWMA. It places the time series of conditional volatility estimates in the spotlight rather than the return series. The return series plays the role of the randomly generated shocks or innovations to the process that generates the conditional volatility series.

We will look at GARCH(1,1), the simplest version of the model. The returns are assumed to follow a process

$$r_t = \epsilon_t \sigma_t$$

with $\epsilon_t \sim N(0, 1)$ and independent $\forall t$; that is, successive shocks to returns are independent. The conditional volatility process, the equation that describes how the estimate of the current level of volatility is updated with new return information, is

$$\sigma_t^2 = \omega + \alpha \sigma_{t-1}^2 + \beta r_t^2$$

We want this to be a stationary process, that is, we don't want volatility to be able to wander far from ω, since that would admit the possibility of implausibly high or low volatility levels. We therefore impose the assumption $\alpha + \beta \leq 1$.

The conditional volatility process then looks remarkably like the recursive form of the EWMA estimator, and in fact has the same interpretation: it shows how the current estimate of volatility is updated with new information about the magnitude of returns.

Figure 3.3 illustrates the relationship between the GARCH(1,1) and EWMA estimators with estimated volatilities over a 10-year period for
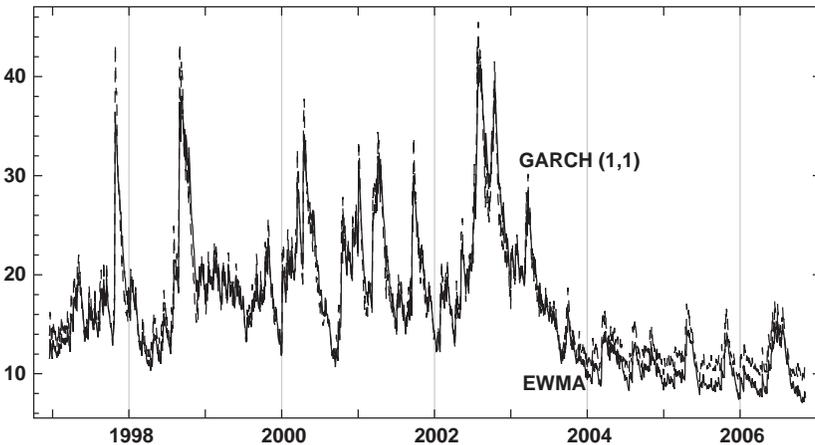

FIGURE 3.3   Comparison of Volatility Estimators
Volatility estimates for the S&P 500 index, January 3, 1996, to November 10, 2006. The GARCH(1,1) volatilities (dotted line) use the following parameter estimates: $\omega = 0.000002$, $\alpha = 0.091$, $\beta = 0.899$. The GARCH parameter estimates are based on the observation range May 16, 1995, to April 29, 2003. The EWMA estimates (solid line) use a decay factor of 0.94 and 250 days of data.
Data source: Bloomberg Financial L.P.

the S&P 500 index. The GARCH(1,1) estimates are those presented as background information for the 2003 Nobel Prize in Economics.[2] The GARCH(1,1) estimates up to April 29, 2003, are fitted values, while those later, are forecasts. The EWMA estimates use a decay factor $\lambda = 0.94$.

We can see, first, that the EWMA and GARCH(1,1) estimates are very close to one another. The main difference between the two is that, when conditional volatility abruptly spikes, the GARCH(1,1) estimate is somewhat more reactive, updating more drastically. Overall, the EWMA estimates mimic GARCH(1,1) closely. The estimates begin to diverge as the forecast period gets further from the end of the observation period on which they were based, but remain reasonably close for a long time, indicating that the GARCH(1,1) parameters are fairly stable and that, with a suitable decay factor, EWMA is a good approximation to GARCH(1,1).

3.3 MODES OF COMPUTATION

Let's look next at how to estimate VaR and how the estimate varies with the time horizon of the VaR. We continue the example of a single long position in the S&P 500 index with an initial value of $1,000,000 as of Nov. 10, 2006. We'll set the time horizon to one day and the confidence level to 99 percent.

3.3.1 Parametric

Parametric VaR relies on algebra rather than simulation to compute an estimate of the VaR. For a portfolio x consisting of a single position, linear in the underlying risk factor, that estimate is

$$\text{VaR}_t(\alpha, \tau)(x) = -xS_t\left(e^{z_*\hat{\sigma}_t\sqrt{\tau}} - 1\right)$$

This expression is the same as that we developed above, only now $\hat{\sigma}_t$ represents an *estimate* of the unobservable volatility parameter.

For our example, using the the RiskMetrics/EWMA estimate $\hat{\sigma}_t = 7.605\%$, we have

$$\begin{aligned}\text{VaR}\left(0.99, \frac{1}{252}\right) &= -1,000,000\left(e^{\frac{-2.33 \cdot 0.07605}{\sqrt{252}}} - 1\right) \\ &= 1,000,000 \cdot (1 - e^{-0.0111}) \\ &= \$11,084\end{aligned}$$

---

[2]See Royal Swedish Academy of Sciences (2003).

Practitioners sometimes apply the approximation $e^r \approx 1 + r$, leading to the VaR estimate

$$\text{VaR}_t(\alpha, \tau) = -z_* \hat{\sigma}_t \sqrt{\tau} x S_t \tag{3.3}$$

This approximation of logarithmic by arithmetic returns (see Chapter 2 for more detail) is particularly common in calculating VaR for a portfolio, as we will see in Chapter 4.

In our example, we have the slightly greater VaR estimate

$$\text{VaR}\left(0.99, \frac{1}{252}\right) = -1{,}000{,}000 \frac{(-2.33) \cdot 0.07605}{\sqrt{252}} = \$11{,}146$$

3.3.2 Monte Carlo Simulation

Our model of returns is

$$r_t = \sqrt{t} \sigma \epsilon_t$$

where ϵ_t is a standard normal random variate. To estimate the VaR via Monte Carlo simulation:

1. Generate I independent draws from a standard normal distribution. We'll label these $\tilde{\epsilon}_1, \ldots, \tilde{\epsilon}_I$.
2. For each of the I draws, we get a single simulation thread for the return by calculating

$$\tilde{r}_i = \sqrt{t} \hat{\sigma} \tilde{\epsilon}_i \qquad i = 1, \ldots, I$$

where $\hat{\sigma}$ is an estimate of the volatility.
3. The corresponding estimate of the asset price is

$$\tilde{S}_i = S_0 e^{\tilde{r}_i} = S_0 e^{\sqrt{t} \hat{\sigma} \tilde{\epsilon}_i} \qquad i = 1, \ldots, I$$

4. For each of the I simulation threads, we now revalue the portfolio, either by multiplying the number of shares x, or equivalently by multiplying the portfolio value $x S_0$ by $e^{\tilde{r}_i} - 1$. This gives us the P&L $\tilde{V}_i - V_0$ of the portfolio in the scenario corresponding to the i-th simulation.
5. Reorder the simulated P&Ls $\tilde{V}_i - V_0$ by size. This gives us the order statistics $\tilde{V}^{(i)} - V_0$, with $\tilde{V}^{(1)} - V_0$ corresponding to the largest loss and $\tilde{V}^{(I)} - V_0$ to the largest profit.

6. Finally, for the stipulated confidence level α, we identify the P&L $i^* = (1-\alpha)I$ corresponding to the VaR. We have

$$\text{VaR}(\alpha, \tau) = -(\tilde{V}_{i^*} - V_0) = -V_0(e^{\tilde{r}_{i^*}} - 1)$$
$$= xS_0(1 - e^{\tilde{r}_{i^*}})$$

If i^* is not an integer, we average the neighboring simulation results.

In the example, as set out in Table 3.1, the 99 percent overnight VaR would be estimated as the tenth worst scenario out of $I = 1,000$ simulation trials, or $10,062. This is fairly close to the parametric VaR; the differences can be accounted for by simulation noise.

The virtues of Monte Carlo are that it can accommodate a lot of troublesome characteristics of models and portfolios, such as multiple risk factors, options, or non-normal returns. We discuss these issues in subsequent chapters. All we need for Monte Carlo to work is a return distribution that we either believe in or are willing to stipulate. The main drawback of Monte

TABLE 3.1 Example of Monte Carlo Simulation

i	$\tilde{r}^{(i)}$	$\tilde{S}^{(i)}$	$\tilde{V}^{(i)} - V_0$
1	−0.01538	1,355.89	−15,263.83
2	−0.01300	1,359.13	−12,913.46
3	−0.01279	1,359.42	−12,705.08
4	−0.01254	1,359.75	−12,465.97
5	−0.01159	1,361.04	−11,524.08
6	−0.01097	1,361.89	−10,911.72
7	−0.01087	1,362.02	−10,814.77
8	−0.01081	1,362.11	−10,750.62
9	−0.01071	1,362.24	−10,655.52
10	−0.01011	1,363.06	−10,061.59
11	−0.01008	1,363.10	−10,032.39
12	−0.00978	1,363.50	−9,735.93
. . .	. . .	. . .	. . .
997	0.01281	1,394.66	12,894.69
998	0.01288	1,394.76	12,964.92
999	0.01355	1,395.70	13,644.34
1,000	0.01597	1,399.07	16,093.82

The example computes VaR via Monte Carlo simulation for a single position in the S&P 500 index with an initial value of $1,000,000, with a time horizon of one day, and a confidence level of 99 percent, as of Nov. 10, 2006. The portfolio's initial value is $1,000,000. The horizontal lines mark the VaR scenario. The $\tilde{V}^{(i)} - V_0$ and $\tilde{r}^{(i)}$ are the order statistics of $\tilde{V}_i$ and $\tilde{r}_i$, and $\tilde{S}^{(i)} \equiv S_0 e^{\tilde{r}^{(i)}}$.

Carlo simulation is that it can be computationally slow or "expensive," and is less of a drawback as computing time cheapens.

3.3.3 Historical Simulation

Historical simulation relies on the "natural experiment" of historical returns to estimate VaR. It does not depend on any distributional hypotheses and does not require the estimation of distribution moments.

To estimate the VaR via historical simulation:

1. Select a historical "look-back" period of I business days. Compute the price return for each day in this interval, for which we'll reuse the notation $\tilde{r}_1, \ldots, \tilde{r}_I$. Each i corresponds to a particular date.
2. The corresponding estimate of the asset price is

$$\tilde{S}_i = S_0 e^{\tilde{r}_i}$$

3. The rest of the procedure is identical to that for Monte Carlo simulation: Order the $\tilde{S}_i$ (or the $\tilde{r}_i$) as in step 5 of the Monte Carlo procedure. The VaR corresponds to the historical return in the $i^* = (1 - \alpha)I$-th order statistic, that is, the i^*-th worst loss

Table 3.2 shows how the data are to be arranged and identifies the VaR shock, VaR scenario, and VaR. The VaR computed using historical simulation is $17,742, or about 60 to 70 percent higher than estimated via Monte Carlo or parametrically.

Let's compare the two simulation approaches in more detail. Figure 3.4 displays the histograms of simulation results for the Monte Carlo and historical simulations. We can read it in conjunction with the tabular display of the extreme simulation results. The VaR shocks in the two simulation approaches are quite different, but the behavior of the most extreme returns, in ordered simulation threads numbered 1 to 3 or 998 to 1,000, are even more different from one mode of computation to the other. The historical simulation data show that the equity return behavior is fat-tailed; on the 0.5 percent of days on which the return is most extreme, returns have a magnitude of about 1.5 percent in the Monte Carlo simulations, but about 3.5 percent in the historical simulations.

The results of historical simulation can depend heavily on the observation interval. There is no prescribed amount of history one should take into account. Rather, it is a risk manager choice, but a difficult one; there is no right answer, only trade-offs. If one takes a longer historical period, one is likelier to capture low-probability, large-magnitude returns and obtain a

TABLE 3.2 Example of VaR Computed via Historical Simulation

i	Date	$\tilde{r}^{(i)}$	$\tilde{S}_i$	$\tilde{V}^{(i)} - V_0$
1	24Mar03	−0.0359	1,328.40	−35,231.47
2	24Jan03	−0.0297	1,336.66	−29,233.44
3	10Mar03	−0.0262	1,341.34	−25,829.72
4	19May03	−0.0252	1,342.60	−24,917.93
5	30Jan03	−0.0231	1,345.45	−22,849.28
6	24Sep03	−0.0193	1,350.62	−19,095.65
7	24Feb03	−0.0186	1,351.60	−18,380.75
8	20Jan06	−0.0185	1,351.68	−18,326.28
9	05Jun06	−0.0180	1,352.40	−17,799.75
10	31Mar03	−0.0179	1,352.48	−17,741.75
11	05Aug03	−0.0178	1,352.59	−17,663.46
12	17May06	−0.0170	1,353.72	−16,841.06
. . .	. . .	. . .	. . .	. . .
997	21Mar03	0.0227	1,408.55	22,976.69
998	02Apr03	0.0258	1,412.87	26,115.93
999	13Mar03	0.0339	1,424.35	34,457.03
1,000	17Mar03	0.0348	1,425.69	35,426.69

The example computes VaR via historical simulation for a single position in the S&P 500 index with an initial value of USD 1,000,000, with a time horizon of one day and a confidence level of 90 percent, as of Nov. 10, 2006. The portfolio's initial value is USD 1,000,000. The horizontal lines mark the VaR scenarios. The $\tilde{V}^{(i)} - V_0$ and $\tilde{r}^{(i)}$ are the order statistics of $\tilde{V}_i$ and $\tilde{r}_i$, and $\tilde{S}^{(i)} \equiv S_0 e^{\tilde{r}^{(i)}}$.
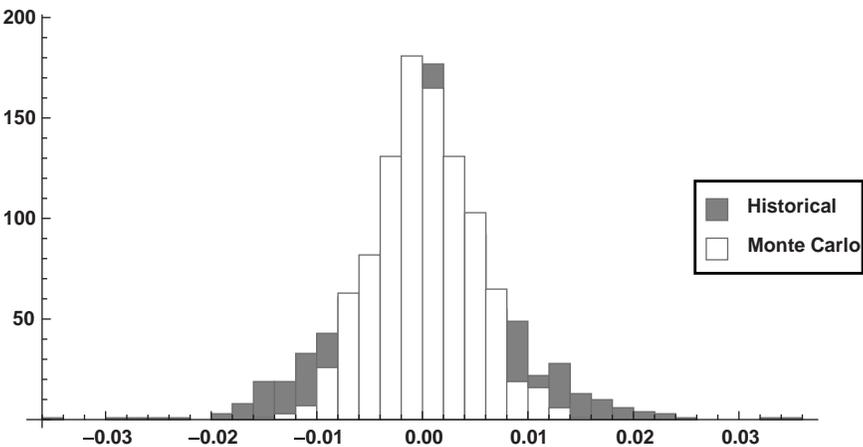

FIGURE 3.4 Comparison of Simulation Approaches
Histograms of Monte Carlo and historical simulated return results displayed in Tables 3.1 and 3.2.

more accurate estimate of the tail risks in the portfolio. But, as noted in discussing short-term conditional volatility estimation, estimates based on longer history are less likely to do a good job estimating a short-term VaR. If the focus is on the main-body, rather than tail risks, longer history may be counterproductive.

Much depends on whether the fat-tailed behavior manifests itself in the short or long term. Do large returns appear routinely? Or do the returns behave close to normally most of the time, but subject to infrequent large shocks? We explore these issues further in Chapter 10. Boudoukh, Richardson, and Whitelaw (1997) have proposed a hybrid approach in which the historical data are weighted, as in the EWMA volatility estimation procedure, but then used in historical simulation, rather than in a volatility estimate. A more-recent large-magnitude return will then have more impact on the VaR than a less-recent one.

3.4 SHORT POSITIONS

For a short position, the number of units x of exposure, say, shares borrowed and sold, is negative. The P&L can still be written

$$V_{t+\tau} - V_t = xS_t \left(e^{r_t} - 1\right)$$

but it is negative when $r_t > 0$. The parametric VaR at a confidence level α therefore corresponds to a return equal to $z_*\sigma\sqrt{\tau}$, with z_* now the ordinate of the standard normal distribution at which $\Phi(z) = \alpha$, rather than $\Phi(z) = 1 - \alpha$. In other words, it corresponds to a right-tail rather than left-tail quantile of the normal distribution. The VaR is still estimated as

$$\text{VaR}_t(\alpha, \tau)(x) = xS_t \left(e^{z_*\hat{\sigma}_t\sqrt{\tau}} - 1\right)$$

but we have $z_* > 0$, not $z_* < 0$.

In our example, if we have a short rather than long S&P position, the parametric VaR is computed as

$$\text{VaR}\left(0.99, \frac{1}{252}\right) = -1,000,000(e^{0.0111} - 1) = \$\,11,208$$

which is slightly higher than for a long position; if we use the approximation of Equation (3.3), $-z_*\hat{\sigma}_t\sqrt{\tau}xS_t$, the result is identical to that for a long position.

Short positions provide a good illustration of some inadequacies of VaR as a risk measure. There is no difficulty computing VaR for short positions; in fact, as we have just seen, the results are essentially the same as for a long position. However, there are a number of crucial differences between long and short positions. Long positions have limited "downside," or potential loss, since the market value of the asset can go to zero, but no lower. But short positions have unlimited downside, since prices can rise without limit. Short positions are thus inherently riskier, and this additional risk is not captured by VaR.

No matter which approach is used to compute VaR or estimate volatility, the risk manager has to decide how to represent the risk of a position with a risk factor S_t. He can then determine how many units x of this exposure are in the position. In many cases, it is fairly straightforward to associate positions with risk factors. A spot currency position, for example, should be associated with the corresponding exchange rate, and the number of units is just the number of currency units in the position.

But it is surprising how few cases are actually this straightforward. Most individual securities are exposed to several risk factors. To continue the currency example, currency positions are usually taken via forward foreign exchange, which are exposed to domestic and foreign money-market rates as well as exchange rates. We continue the discussion of mapping in discussing portfolio VaR in Chapter 5.

Even in the simplest cases, mapping a security to a single risk factor neglects the *financing risk* of a position. This, too, is a particularly glaring omission for short positions, as we see in Chapter 12.

3.5   EXPECTED SHORTFALL

VaR provides an estimate of a threshold loss. That is, it is a loss level that is unlikely to be exceeded, with "unlikely" defined by the confidence level. Another measure of potential loss, called *expected shortfall* (ES), *conditional value-at-risk* (CVaR), or *expected tail loss*, is the expected value of the loss, given that the threshold is exceeded. In this section, we define this risk measure and discuss why it is useful.

Expected shortfall is a conditional expected value. If x is a well-behaved continuous random variable with density function $f(x)$, the conditional expected value of x given that $x \leq x^*$ is

$$\frac{\int_{-\infty}^{x^*} x f(x) dx}{P[x \leq x^*]}$$

To understand this expression, let's compare its numerator, $\int_{-\infty}^{x^*} xf(x)dx$, to the very similar expression $\int_{-\infty}^{\infty} xf(x)dx$ for the mean of x. We can see that the numerator gives the expected value, not over the whole range of x, but only of values at or below x^*. In other words, it represents the average of low values of x, multiplied by the correspondingly low probabilities of occurring.

The denominator is the probability of some value below x^* occurring. In Figure 3.1, this is illustrated for the value of a long position as the shaded areas in the left tail of the density. To obtain the conditional expected value of this tail event, we apply its definition as the ratio of the expected value $\int_{-\infty}^{x^*} xf(x)dx$ to its probability.

Similarly, the formal expression for ES is also composed of two parts. The expected value of the position value, if it ends up at or below the VaR scenario V^*, is $\int_0^{V^*} vf_\tau(v)dv$, where $f_\tau(v)$ is the time-t density function of the position value. Note that the value of a long position can't go lower than zero. By definition, the probability of a loss greater than or equal to the VaR is $1 - \alpha$, where α is the confidence level of the VaR. So the conditional expected value of the position is

$$\frac{1}{1-\alpha} \int_0^{V^*} vf_\tau(v)dv$$

The conditional expected loss, treated as a positive number, is just the initial position value less this quantity:

$$V_t - \frac{1}{1-\alpha} \int_0^{V^*} vf_\tau(v)dv$$

Expected shortfall is always larger than the VaR, since it is an average of the VaR and loss levels greater than the VaR.

Example of Expected Shortfall			
Mode	VaR	ES	$\frac{ES}{VaR}$
Parametric	11,083.57	12,686.68	1.145
Monte Carlo	10,061.59	11,806.67	1.173
Historical	17,741.75	22,940.60	1.293

In the standing example of this chapter, the expected shortfall can be computed using any of the modes of computation we have laid out. For the

parametric approach, the expected shortfall is computed as

$$\frac{1}{1-\alpha} x S_t \mathrm{E}\left[1 - e^{z\sigma_t\sqrt{\tau}} | z < z_*\right]$$

For the Monte Carlo and historical simulation approaches, the expected shortfall is computed as the average of $x S_t \left[1 - \exp\left(\tilde{r}^{(i)}\right)\right]$, where the averaging is over the i^* simulated returns that are as bad as or worse than the VaR scenario. In our example, in which the number of simulations is set to $I = 1,000$, we average over the 10 worst simulation threads.

Just as we saw for VaR itself, Monte Carlo simulation gives results for expected shortfall that are not very different from those of the parametric approach assuming normally distributed returns. However, the historical simulation approach is better able to capture the extremes of the distribution, so not only is the VaR larger, but the expected value of outcomes worse than the VaR is even larger relative to the VaR than for approaches relying on the normal distribution.

If we compute VaR parametrically, using the convenient approximation of Equation (3.3), we have a similarly convenient way to compute the expected shortfall; it is equal to

$$\frac{\phi(-\alpha)}{(1-\alpha)z_*} \mathrm{VaR}_t(\alpha, \tau)(x)$$

where $\phi(\cdot)$ represents the standard normal probability density function.

Expected shortfall is useful for a number of reasons, which will be discussed in greater detail below:

- Expected shortfall can be a more useful risk measure than VaR when the standard model is wrong in certain ways, in particular when the return distribution has fat tails (see Chapter 10).
- Expected shortfall via historical simulation can be compared with the VaR to obtain clues as to how far from the standard model the true distribution actually is.
- Expected shortfall has theoretical properties that occasionally have a useful practical application (see the discussion of "coherent" risk measures in Chapter 11).

FURTHER READING

Jorion (2007) and Dowd (2005) are general risk management textbooks with a focus on market risk measurement. Introductions to VaR models are

provided by Duffie and Pan (1997), Morgan Guaranty Trust Company (1996), Mina and Xiao (2001), and Gourieroux and Jasiak (2010). Culp, Miller, and Neves (1998) is an end-user-focused introduction. Second-generation developments in VaR modeling are described in Engle and Manganelli (2001) and Pearson and Smithson (2002). Basel Committee on Banking Supervision (2011) is a literature survey from a regulatory point of view.

See Black (1993) and Merton (1980) on the difficulty of estimating expected return.

Comparative surveys of approaches to forecasting volatility are presented by Brailsford and Faff (1996), Tay and Wallis (2000), Engle and Patton (2001), Brooks and Persand (2003), Andersen, Bollerslev, Christoffersen, and Diebold (2010), and Poon and Granger (2003, 2005). Christoffersen and Diebold (2000) focuses on longer-term volatility forecasts. Schwert (1990*b*) and (1990*a*) present the results of studies of the macroeconomic influences on U.S. stock market volatility using a long historical data set going back to the late 19th century.

Vlab is a web site maintained at New York University Stern School of Business presenting updated volatility and correlation forecasts for a range of assets. It can be accessed at http://vlab.stern.nyu.edu/.