specific risk depends on that of general market risk. The Basel rules have a separate charge for specific risk.[1]

To illustrate this decomposition, consider a portfolio of N stocks. We are mapping each stock on a position in the stock market index, which is our primitive risk factor. The return on a stock R_i is regressed on the return on the stock market index R_m, that is, which gives the exposure β_i. In what follows, ignore α, which does not contribute to risk. We assume that the specific risk owing to ϵ is not correlated across stocks or with the market. The relative weight of each stock in the portfolio is given by w_i. Thus the portfolio return is

$$R_i = \alpha_i + \beta_i R_m + \epsilon_i \tag{11.1}$$

$$R_p = \sum_{i=1}^{N} w_i R_i = \sum_{i=1}^{N} w_i \beta_i R_m + \sum_{i=1}^{N} w_i \epsilon_i \tag{11.2}$$

These exposures are aggregated across all the stocks in the portfolio. This gives

$$\beta_p = \sum_{i=1}^{N} w_i \beta_i \tag{11.3}$$

If the portfolio value is W, the mapping on the index is $x = W\beta_p$.

Next, we decompose the variance of R_p in Equation (11.2) and find

$$V(R_p) = (\beta_p^2) V(R_m) + \sum_{i=1}^{N} w_i^2 \sigma_{\epsilon i}^2 \tag{11.4}$$

The first component is the general market risk. The second component is the aggregate of specific risk for the entire portfolio. This decomposition shows that with more detail on the primitive or general-market risk factors, there will be less specific risk for a fixed amount of total risk $V(R_p)$.

As another example, consider a corporate bond portfolio. Bond positions describe the distribution of money flows over time by their amount, timing, and credit quality of the issuer. This creates a continuum of risk factors, going from overnight to long maturities for various credit risks.

In practice, we have to restrict the number of risk factors to a small set. For some portfolios, one risk factor may be sufficient. For others, 15 maturities may be necessary. For portfolios with options, we need to model movements not only in yields but also in their implied volatilities.

Our primitive risk factors could be movements in a set of J government bond yields z_j and in a set of K credit spreads s_k sorted by credit rating. We model the movement in each corporate bond yield dy_i by a movement in z at the closest maturity and in s for the same credit rating. The remaining component is ϵ_i.

The movement in value W then is

$$dW = \sum_{i=1}^{N} \mathrm{DVBP}_i dy_i = \sum_{j=1}^{J} \mathrm{DVBP}_j dz_j + \sum_{k=1}^{K} \mathrm{DVBP}_k ds_k + \sum_{i=1}^{N} \mathrm{DVBP}_i d\epsilon_i \quad (11.5)$$

where DVBP is the total dollar value of a basis point for the associated risk factor. The values for DVBP_j then represent the summation of the DVBP across all individual bonds for each maturity.

This leads to a total risk decomposition of

$$V(dW) = \text{ general risk} + \sum_{i=1}^{N} \mathrm{DVBP}_i^2 V(d\epsilon_i) \quad (11.6)$$

A greater number of general risk factors should create less residual risk. Even so, we need to ascertain the size of the second, specific risk term. In practice, there may not be sufficient history to measure the specific risk of individual bonds, which is why it is often assumed that all issuers within the same risk class have the same risk.

11.2 MAPPING FIXED-INCOME PORTFOLIOS

11.2.1 Mapping Approaches

Once the risk factors have been selected, the question is how to map the portfolio positions into exposures on these risk factors. We can distinguish three mapping systems for fixed-income portfolios: principal, duration, and cash flows. With *principal mapping*, one risk factor is chosen that corresponds to the average portfolio maturity. With *duration mapping*, one risk factor is chosen that corresponds to the portfolio duration. With *cash-flow mapping*, the portfolio cash flows are grouped into maturity buckets. Mapping should preserve the market value of the position. Ideally, it also should preserve its market risk.

TABLE 11-2

Mapping for a Bond Portfolio ($ Millions)

	Cash Flows			Mapping (PV)		
Term (Year)	5-Year	1-Year	Spot Rate	Principal	Duration	Cash Flow
1	$6	$104	4.000%	0.00	0.00	$105.77
2	$6	0	4.618%	0.00	0.00	$5.48
2.733	—	—		—	$200.00	—
3	$6	0	5.192%	$200.00	0.00	$5.15
4	$6	0	5.716%	0.00	0.00	$4.80
5	$106	0	6.112%	0.00	0.00	$78.79
Total				$200.00	$200.00	$200.00

As an example, Table 11-2 describes a two-bond portfolio consisting of a $100 million 5-year 6 percent issue and a $100 million 1-year 4 percent issue. Both issues are selling at par, implying a market value of $200 million. The portfolio has an average maturity of 3 years and a duration of 2.733 years. The table lays out the present value of all portfolio cash flows discounted at the appropriate zero-coupon rate.

Principal mapping considers the timing of redemption payments only. Since the average maturity of this portfolio is 3 years, the VAR can be found from the risk of a 3-year maturity, which is 1.484 percent from Table 11-3. VAR then is $200 X 1.484/100= $2.97 million. The only positive aspect of this method is its simplicity. This approach overstates the true risk because it ignores intervening coupon payments.

The next step in precision is duration mapping. We replace the portfolio by a zero-coupon bond with maturity equal to the duration of the portfolio, which is 2.733 years. We discuss in Appendix 11.A how to allocate the portfolio to the adjoining 2-and 3-year vertices. Table 11-3 shows VARs of 0.987 and 1.484 for these maturities, respectively. Using a linear interpolation, we find a risk of 0.987 + (1.484–0.987) X (2.733–2) = 1.351 percent for this hypothetical zero. With a $200 million portfolio, the duration-based VAR is $200 X 1.351/100 = $2.70 million, slightly less than before.

Finally, the cash-flow mapping method consists of grouping all cash flows on term-structure “vertices” that correspond to maturities for which volatilities are provided. Each cash flow is represented by the present value of the cash payment, discounted at the appropriate zero-coupon rate.

The diversified VAR is computed as

$$\mathrm{VAR} = \alpha\sqrt{x'\Sigma x} = \sqrt{(x \times V)' R (x \times V)} \qquad (11.7)$$

where $V=\alpha\sigma$ is the vector of VAR for zero-coupon bond returns, and *R*is the correlation matrix.

Table 11-4 shows how to compute the portfolio VAR using cash-flow mapping. The second column reports the cash flows *x* from Table 11-2. Note that the current value of $200 million is fully allocated to the five risk factors. The third column presents the product of these cash flows with the risk of each vertex *x* X *V*, which represents the individual VARs.

With perfect correlation across all zeroes, the VAR of the portfolio is

$$\text{Undiversified VAR} = \sum_{i=1}^{N} |x_i| V_i$$

which is $2.63 million. This number is close to the VAR obtained from the duration approximation, which was $2.70 million.

The right side of the table presents the correlation matrix of zeroes for maturities ranging from 1 to 5 years. To obtain the portfolio VAR, we premultiply and postmultiply the matrix by the dollar amounts (*xV*) at each vertex. Taking the square root, we find a diversified VAR measure of $2.57 million.

Note that this is slightly less than the duration VAR of $2.70 million. This difference is due to two factors. First, risk measures are not perfectly linear with maturity, as we have seen in a previous section. Second, correlations are below unity, which reduces risk even further. Thus, of the $130,000 difference in these measures, ($2.70–$2.57 million), $70,000 is due to differences in yield volatility, and ($2.70–$2.63 million), $60,000 is due to imperfect correlations. The last column presents the component VAR using computations as explained in Chapter 7.

TABLE 11-3

Computing VAR from Change in Prices of Zeroes

Term (Year)	Cash Flows	Old Zero Value	Old PV of Flows	Risk (%)	New Zero Value	New PV of Flows
1	$110	0.9615	$105.77	0.4696	0.9570	$105.27
2	$6	0.9136	$5.48	0.9868	0.9046	$5.43
3	$6	0.8591	$5.15	1.4841	0.8463	$5.08
4	$6	0.8006	$4.80	1.9714	0.7848	$4.71
5	$106	0.7433	$78.79	2.4261	0.7252	$76.88
Total			$200.00			$197.37
Loss						$2.63

11.2.2 Stress Test

Table 11-3 presents another approach to VAR that is directly derived from movements in the value of zeroes. This is an example of stress testing. Assume that all zeroes are perfectly correlated. Then we could decrease all zeroes' values by their VAR. For instance, the 1-year zero is worth 0.9615. Given the VAR in Table 11-3 of 0.4696, a 95 percent probability move would be for the zero to fall to 0.9615 X (1–0.4696/100)=0.9570. If all zeroes are perfectly correlated, they should all fall by their respective VAR. This generates a new distribution of present-value factors that can be used to price the portfolio. Table 11-3 shows that the new value is $197.37 million, which is exactly $2.63 million below the original value. This number is exactly the same as the undiversified VAR just computed.

The two approaches illustrate the link between computing VAR through matrix multiplication and through movements in underlying prices. Computing VAR through matrix multiplication is much more direct, however, and more appropriate because it allows nonperfect correlations across different sectors of the yield curve.

TABLE 11-4

Computing the VAR of a $200 Million Bond Portfolio (Monthly VAR at 95 Percent Level)

	PV Cash Flows	Individual VAR	Correlation Matrix *R*					Component VAR
Term (Year)	x	$x \times V$	1Y	2Y	3Y	4Y	5Y	$x\Delta$VAR
1	$105.77	0.4966	1					$0.45
2	$5.48	0.0540	0.897	1				$0.05
3	$5.15	0.0765	0.886	0.991	1			$0.08
4	$4.80	0.0947	0.866	0.976	0.994	1		$0.09
5	$78.79	1.9115	0.855	0.966	0.988	0.998	1	$1.90
Total	$200.00	2.6335						
Undiversified VAR		$2.63						
Diversified VAR								$2.57

11.2.3 Benchmarking a Portfolio

Next, we provide a practical fixed-income example by showing how to compute VAR in relative terms, that is, relative to a performance benchmark. Table 11-5 presents the cash-flow decomposition of the J.P. Morgan U.S. bond index, which has a duration of 4.62 years. Assume that we are trying to benchmark a portfolio of $100 million. Over a monthly horizon, the VAR of the index at the 95 percent confidence level is $1.99 million. This is about equivalent to the risk of a 4-year note.

TABLE 11-5

Benchmarking a $100 Million Bond Index (Monthly Tracking Error VAR at 95 Percent Level)

			Position: Portfolio				
Vertex	Risk (%)	Position: Index ($)	1 ($)	2 ($)	3 ($)	4 ($)	5 ($)
≤1m	0.022	1.05	0.0	0.0	0.0	0.0	84.8
3m	0.065	1.35	0.0	0.0	0.0	0.0	0.0
6m	0.163	2.49	0.0	0.0	0.0	0.0	0.0
1Y	0.470	13.96	0.0	0.0	0.0	59.8	0.0
2Y	0.987	24.83	0.0	0.0	62.6	0.0	0.0
3Y	1.484	15.40	0.0	59.5	0.0	0.0	0.0
4Y	1.971	11.57	38.0	0.0	0.0	0.0	0.0
5Y	2.426	7.62	62.0	0.0	0.0	0.0	0.0
7Y	3.192	6.43	0.0	40.5	0.0	0.0	0.0
9Y	3.913	4.51	0.0	0.0	37.4	0.0	0.0
10Y	4.250	3.34	0.0	0.0	0.0	40.2	0.0
15Y	6.234	3.00	0.0	0.0	0.0	0.0	0.0
20Y	8.146	3.15	0.0	0.0	0.0	0.0	0.0
30Y	11.119	1.31	0.0	0.0	0.0	0.0	15.2
Total		100.00	100.0	100.0	100.0	100.0	100.0
Duration		4.62	4.62	4.62	4.62	4.62	4.62
Absolute VAR		$1.99	$2.25	$2.16	$2.04	$1.94	$1.71
Tracking error VAR		$0.00	$0.43	$0.29	$0.16	$0.20	$0.81

Next, we try to match the index with two bonds. The rightmost columns in the table display the positions of two-bond portfolios with duration matched to that of the index. Since no zero-coupon has a maturity of exactly 4.62 years, the closest portfolio consists of two positions, each in a 4-and a 5-year zero. The respective weights for this portfolio are $38 million and $62 million.

Define the new vector of positions for this portfolio as x and for the index as x_0. The VAR of the deviation relative to the benchmark is

$$\text{Tracking error VAR} = \alpha\sqrt{(x - x_0)'\Sigma(x - x_0)} \qquad (11.8)$$

After performing the necessary calculations, we find that the *tracking error* VAR (TE-VAR) of this duration-hedged portfolio is $0.43 million. Thus the maximum deviation between the index and the portfolio is at most $0.43 million under normal market conditions. This potential shortfall is much less than the $1.99

million absolute risk of the index. The remaining tracking error is due to nonparallel moves in the term structure.

Relative to the original index, the tracking error can be measured in terms of variance reduction, similar to an R^2 in a regression. The variance improvement is which is in line with the explanatory power of the first factor in the variance decomposition detailed in Chapter 8.

$$1-\left(\frac{0.43}{1.99}\right)^2 = 95.4\,\text{percent}$$

Next, we explore the effect of altering the composition of the tracking portfolio. Portfolio 2 widens the bracket of cash flows in years 3 and 7. The TE-VAR is \$0.29 million, which is an improvement over the previous number. Next, portfolio 3 has positions in years 2 and 9. This comes the closest to approximating the cash-flow positions in the index, which has the greatest weight on the 2-year vertex. The TE-VAR is reduced further to \$0.16 million. Portfolio 4 has positions in years 1 and 10. Now the TE-VAR increases to \$0.20 million. This mistracking is even more pronounced for a portfolio consisting of 1-month bills and 30-year zeroes, for which the TE-VAR increases to \$0.81 million.

Among the portfolios considered here, the lowest tracking error is obtained with portfolio 3. Note that the absolute risk of these portfolios is lowest for portfolio 5. As correlations decrease for more distant maturities, we should expect that a duration-matched portfolio should have the lowest absolute risk for the combination of most distant maturities, such as a *barbell* portfolio of cash and a 30-year zero. However, minimizing absolute market risk is not the same as minimizing relative market risk.

This example demonstrates that duration hedging only provides a first approximation to interest-rate risk management. If the goal is to minimize tracking error relative to an index, it is essential to use a fine decomposition of the index by maturity.

11.3 MAPPING LINEAR DERIVATIVES

11.3.1 Forward Contracts

Forward and futures contracts are the simplest types of derivatives. Since their value is linear in the underlying spot rates, their risk can be constructed easily from basic building blocks. Assume, for instance, that we are dealing with a forward contract on a foreign currency. The basic valuation formula can be derived from an arbitrage argument.

To establish notations, define

S_t = spot price of one unit of the underlying cash asset

K = contracted forward price

r = domestic risk-free rate

y = income flow on the asset

τ = time to maturity.

When the asset is a foreign currency, y represents the foreign risk-free rate r^*. We will use these two notations interchangeably. For convenience, we assume that all rates are compounded continuously.

We seek to find the current value of a forward contract f_t to buy one unit of foreign currency at K after time τ. To do this, we consider the fact that investors have two alternatives that are economically equivalent: (1) Buy $e^{-y\tau}$ units of the asset at the price S_t and hold for one period, or (2) enter a forward contract to buy one unit of the asset in one period. Under alternative 1, the investment will grow, with reinvestment of dividend, to exactly one unit of the asset after one period. Under alternative 2, the contract costs f_t upfront, and we need to set aside enough cash to pay K in the future, which is $Ke^{-r\tau}$. After 1 year, the two alternatives lead to the same position, one unit of the asset. Therefore, their initial cost must be identical. This leads to the following valuation formula for outstanding forward contracts:

$$f_t = S_t e^{-y\tau} - Ke^{-r\tau} \tag{11.9}$$

Note that we can repeat the preceding reasoning to find the current forward rate F_t that would set the value of the contract to zero. Setting $K = F_t$ and $f_t = 0$ in Equation (11.9), we have

$$F_t = (S_t e^{-y\tau})e^{r\tau} \tag{11.10}$$

This allows us to rewrite Equation (11.9) as

$$f_t = F_t e^{-r\tau} - Ke^{-r\tau} = (F_t - K)e^{-r\tau} \tag{11.11}$$

In other words, the current value of the forward contract is the present value of the difference between the current forward rate and the locked-in delivery rate. If we are long a forward contract with contracted rate K, we can liquidate the contract by entering a new contract to sell at the current rate F_t. This will lock in a profit of $(F_t - K)$, which we need to discount to the present time to find f_t.

Let us examine the risk of a 1-year forward contract to purchase 100 million euros in exchange for $130.086 million. Table 11-6 displays pricing information for the contract (current spot, forward, and interest rates), risk, and correlations. The first step is to find the market value of the contract. We can use Equation (11.9), accounting for the fact that the quoted interest rates are discretely compounded, as

$$f_t = \$1.2877 \frac{1}{(1+2.2810/100)} - \$1.3009 \frac{1}{(1+3.3304/100)} = \$1.2589 - \$1.2589 = 0$$

TABLE 11-6

Risk and Correlations for Forward Contract Risk Factors (Monthly VAR at 95 Percent Level)

			Correlations		
Risk Factor	**Price or Rate**	**VAR (%)**	**EUR Spot**	**EUR 1Y**	**USD 1Y**
EUR spot	$1.2877	4.5381	1	0.1289	0.0400
Long EUR bill	2.2810%	0.1396	0.1289	1	−0.0583
Short USD bill	3.3304%	0.2121	0.0400	−0.0583	1
EUR forward	$1.3009				

Thus the initial value of the contract is zero. This value, however, may change, creating market risk.

Among the three sources of risk, the volatility of the spot contract is the highest by far, with a 4.54 percent VAR (corresponding to 1.65 standard deviations over a month for a 95 percent confidence level). This is much greater than the 0.14 percent VAR for the EUR 1-year bill or even the 0.21 percent VAR for the USD bill. Thus most of the risk of the forward contract is driven by the cash EUR position.

But risk is also affected by correlations. The positive correlation of 0.13 between the EUR spot and bill positions indicates that when the EUR goes up in value against the dollar, the value of a 1-year EUR investment is likely to appreciate. Therefore, higher values of the EUR are associated with lower EUR interest rates.

This positive correlation increases the risk of the combined position. On the other hand, the position is also short a 1-year USD bill, which is correlated with the other two legs of the transaction. The issue is, what will be the net effect on the risk of the forward contract?

VAR provides an exact answer to this question, which is displayed in Table 11-7. But first we have to compute the positions *x* on each of the three building blocks of the contract. By taking the partial derivative of Equation (11.9) with respect to the risk factors, we have

$$df = \frac{\partial f}{\partial S} dS + \frac{\partial f}{\partial r^*} dr^* + \frac{\partial f}{\partial r} dr = e^{-r^*\tau} dS - Se^{-r^*\tau} \tau dr^* + Ke^{-r\tau} \tau dr \quad (11.12)$$

Here, the building blocks consist of the spot rate and interest rates. Alternatively, we can replace interest rates by the price of bills. Define these as $P = e^{-r\tau}$ and $P^* = e^{-r^*\tau}$. We then replace dr with dP using $dP = (-\tau)e^{-r\tau}\, dr$ and $dP^* = (-\tau)e^{-r^*\tau}\, dr^*$. The risk of the forward contract becomes

TABLE 11-7

Computing VAR for a EUR 100 Million Forward Contract (Monthly VAR at 95 Percent Level)

Position	Present-Value Factor	Cash Flows (CF)	PV of Flows, x	Individual VAR, $\lvert x \rvert V$	Component VAR, $x\Delta$VAR
EUR spot			\$125.89	\$5.713	\$5.704
Long EUR bill	0.977698	EUR100.00	\$125.89	\$0.176	\$0.029
Short USD bill	0.967769	−\$130.09	−\$125.89	\$0.267	\$0.002
Undiversified VAR				\$6.156	
Diversified VAR					\$5.735

$$df = (Se^{-r^*\tau}) \frac{dS}{S} + (Se^{-r^*\tau}) \frac{dP^*}{P^*} - (Ke^{-r\tau}) \frac{dP}{P} \quad (11.13)$$

This shows that the forward position can be separated into three cash flows: (1) a long spot position in EUR, worth EUR 100 million = \$130.09 million in a year, or $(Se^{-r^*\tau})$ = \$125.89 million now, (2) a long position in a EUR investment, also worth \$125.89 million now, and (3) a short position in a USD investment, worth \$130.09 million in a year, or $(Ke^{-r\tau})$ = \$125.89 million now. Thus a position in the forward contract has three building blocks:

Long forward contract = long foreign currency spot + long foreign currency bill + short U.S. dollar bill

Considering only the spot position, the VAR is $125.89 million times the risk of 4.538 percent, which is $5.713 million. To compute the diversified VAR, we use the risk matrix from the data in Table 11-6 and pre-and postmultiply by the vector of positions (PV of flows column in the table). The total VAR for the forward contract is $5.735 million. This number is about the same size as that of the spot contract because exchange-rate volatility dominates the volatility of 1-year bonds.

More generally, the same methodology can be used for long-term currency swaps, which are equivalent to portfolios of forward contracts. For instance, a 10-year contract to pay dollars and receive euros is equivalent to a series of 10 forward contracts to exchange a set amount of dollars into marks. To compute the VAR, the contract must be broken down into a currency-risk component and a string of USD and EUR fixed-income components. As before, the total VAR will be driven primarily by the currency component.

11.3.2 Commodity Forwards

The valuation of forward or futures contracts on commodities is substantially more complex than for financial assets such as currencies, bonds, or stock indices. Such financial assets have a well-defined income flow *y*, which is the foreign interest rate, the coupon payment, or the dividend yield, respectively.

Things are not so simple for commodities, such as metals, agricultural products, or energy products. Most products do not make monetary payments but instead are consumed, thus creating an implied benefit. This flow of benefit, net of storage cost, is loosely called *convenience yield* to represent the benefit from holding the cash product. This convenience yield, however, is not tied to another financial variable, such as the foreign interest rate for currency futures. It is also highly variable, creating its own source of risk.

As a result, the risk measurement of commodity futures uses Equation (11.11) directly, where the main driver of the value of the contract is the current forward price for this commodity. Table 11-8 illustrates the term structure of volatilities for selected energy products and base metals. First, we note that monthly VAR measures are very high, reaching 29 percent for near contracts. In contrast, currency and equity market VARs are typically around 6 percent. Thus commodities are much more volatile than typical financial assets.

TABLE 11-8

Risk of Commodity Contracts (Monthly VAR at 95 Percent Level)

	Energy Products			
Maturity	Natural Gas	Heating Oil	Unleaded Gasoline	Crude Oil–WTI
1 month	28.77	22.07	20.17	19.20
3 months	22.79	20.60	18.29	17.46
6 months	16.01	16.67	16.26	15.87
12 months	12.68	14.61	—	14.05
	Base Metals			
Maturity	Aluminum	Copper	Nickel	Zinc
Cash	11.34	13.09	18.97	13.49
3 months	11.01	12.34	18.41	13.18
15 months	8.99	10.51	15.44	11.95
27 months	7.27	9.57	—	11.59
	Precious Metals			
Maturity	Gold	Silver	Platinum	
Cash	6.18	14.97	7.70	

Second, we observe that volatilities decrease with maturity. The effect is strongest for less storable products such as energy products and less so for base metals. It is actually imperceptible for precious metals, which have low storage costs and no convenience yield. For financial assets, volatilities are driven primarily by spot prices, which implies basically constant volatilities across contract maturities.

Let us now say that we wish to compute the VAR for a 12-month forward position on 1 million barrels of oil priced at $45.2 per barrel. Using a present-value factor of 0.967769, this translates into a current position of $43,743,000.

Differentiating Equation (11.11), we have

$$df = \frac{\partial f}{\partial F} dF = e^{-r\tau} dF = (e^{-r\tau} F) \frac{dF}{F} \qquad (11.14)$$

The term between parentheses therefore represents the exposure. The contract VAR is

VAR = $43,743,000 X 14.05/100 = $6,146,000

In general, the contract cash flows will fall between the maturities of the risk factors, and present values must be apportioned accordingly.

11.3.3 Forward Rate Agreements

Forward rate agreements (FRAs) are forward contracts that allow users to lock in an interest rate at some future date. The buyer of an FRA locks in a borrowing rate; the seller locks in a lending rate. In other words, the "long" receives a payment if the spot rate is above the forward rate.

Define the timing of the short leg as τ_1 and of the long leg as τ_2, both expressed in years. Assume linear compounding for simplicity. The forward rate can be defined as the implied rate that equalizes the return on a τ_2-period investment with a τ_1-period investment rolled over, that is,

$$(1 + R_2\tau_2) = (1 + R_1\tau_1)\,[1 + F_{1,2}(\tau_2 - \tau_1)] \qquad (11.15)$$

For instance, suppose that you sold a 6 × 12 FRA on $100 million. This is equivalent to borrowing $100 million for 6 months and investing the proceeds for 12 months. When the FRA expires in 6 months, assume that the prevailing 6-month spot rate is higher than the locked-in forward rate. The seller then pays the buyer the difference between the spot and forward rates applied to the principal. In effect, this payment offsets the higher return that the investor otherwise would receive, thus guaranteeing a return equal to the forward rate. Therefore, an FRA can be decomposed into two zero-coupon building blocks.

Long 6 X 12 FRA = long 6-month bill + short 12-month bill

Table 11-9 provides a worked-out example. If the 360-day spot rate is 5.8125 percent and the 180-day rate is 5.6250 percent, the forward rate must be such that

$$(1 + F_{1,2} / 2) = \frac{(1 + 5.8125 / 100)}{(1 + 5.6250 / 200)}$$

or F = 5.836 percent. The present value of the notional \$100 million in 6 months is x = \$100/(1 + 5.625/200) = \$97.264 million. This amount is invested for 12 months. In the meantime, what is the risk of this FRA?

Table 11-9 displays the computation of VAR for the FRA. The VARs of 6-and 12-month zeroes are 0.1629 and 0.4696, respectively, with a correlation of 0.8738. Applied to the principal of \$97.26 million, the individual VARs are \$0.158 million and \$0.457 million, which gives an undiversified VAR of \$0.615 million. Fortunately, the correlation substantially lowers the FRA risk. The largest amount the position can lose over a month at the 95 percent level is \$0.327 million.

TABLE 11-9

Computing the VAR of a \$100 Million FRA (Monthly VAR at 95 Percent Level)

Position	PV of Flows, x	Risk (%), V	Correlation Matrix, R		Individual VAR, $\|x\|V$	Component VAR, $x\Delta$VAR
180 days	−\$97.264	0.1629	1	0.8738	\$0.158	−\$0.116
360 days	\$97.264	0.4696	0.8738	1	\$0.457	\$0.444
Undiversified VAR					\$0.615	
Diversified VAR						\$0.327

11.3.4 Interest-Rate Swaps

Interest-rate swaps are the most actively used derivatives. They create exchanges of interest-rate flows from fixed to floating or vice versa. Swaps can be decomposed into two legs, a fixed leg and a floating leg. The fixed leg can be priced as a coupon-paying bond; the floating leg is equivalent to a floating-rate note (FRN).

To illustrate, let us compute the VAR of a \$100 million 5-year interest-rate swap. We enter a dollar swap that pays 6.195 percent annually for 5 years in exchange for floating-rate payments indexed to London Interbank Offer Rate (LIBOR). Initially, we consider a situation where the floating-rate note is about to be reset. Just before the reset period, we know that the coupon will be set at the prevailing market rate. Therefore, the note carries no market risk, and its value can be mapped on cash only. Right after the reset, however, the note becomes similar to a bill with maturity equal to the next reset period.

Interest-rate swaps can be viewed in two different ways: as (1) a combined position in a fixed-rate bond and in a floating-rate bond or (2) a portfolio of forward contracts. We first value the swap as a position in two bonds using risk data from Table 11-4. The analysis is detailed in Table 11-10.

TABLE 11-10
Computing the VAR of a $100 Million Interest-Rate Swap (Monthly VAR at 95 Percent Level)

Term (Year)	Cash Flows Fixed	Cash Flows Float	Spot Rate	PV of Net Cash Flows	Individual VAR	Component VAR
0	$0	+$100		+$100.000	$0	$0
1	−$6.195	$0	5.813%	−$5.855	$0.027	$0.024
2	−$6.195	$0	5.929%	−$5.521	$0.054	$0.053
3	−$6.195	$0	6.034%	−$5.196	$0.077	$0.075
4	−$6.195	$0	6.130%	−$4.883	$0.096	$0.096
5	−$106.195	$0	6.217%	−$78.546	$1.905	$1.905
Total				$0.000		
Undiversified VAR					$2.160	
Diversified VAR						$2.152

The second and third columns lay out the payments on both legs. Assuming that this is an at-the-market swap, that is, that its coupon is equal to prevailing swap rates, the short position in the fixed-rate bond is worth $100 million. Just before reset, the long position in the FRN is also worth $100 million, so the market value of the swap is zero. To clarify the allocation of current values, the FRN is allocated to cash, with a zero maturity. This has no risk.

The next column lists the zero-coupon swap rates for maturities going from 1 to 5 years. The fifth column reports the present value of the net cash flows, fixed minus floating. The last column presents the component VAR, which adds up to a total diversified VAR of $2.152 million. The undiversified VAR is obtained from summing all individual VARs. As usual, the $2.160 million value somewhat overestimates risk.

This swap can be viewed as the sum of five forward contracts, as shown in Table 11-11. The 1-year contract promises payment of $100 million plus the coupon of 6.195 percent; discounted at the spot rate of 5.813 percent, this yields a present value of –$100.36 million. This is in exchange for $100 million now, which has no risk.

The next contract is a 1 X 2 forward contract that promises to pay the principal plus the fixed coupon in 2 years, or –$106.195 million; discounted at the 2-year spot rate, this yields –$94.64 million. This is in exchange for $100 million in 1 year, which is also $94.50 million when discounted at the 1-year spot rate. And so on until the fifth contract, a 4 X 5 forward contract.

Table 11-11 shows the VAR of each contract. The undiversified VAR of $2.401 million is the result of a simple summation of the five VARs. The fully diversified VAR is $2.152 million, exactly the same as in the preceding table. This demonstrates the equivalence of the two approaches.

Finally, we examine the change in risk after the first payment has just been set on the floating-rate leg. The FRN then becomes a 1-year bond initially valued at par but subject to fluctuations in rates. The only change in the pattern of cash flows in Table 11-10 is to add $100 million to the position on year 1 (from –$5.855 to $94.145). The resulting VAR then decreases from $2.152 million to $1.763 million. More generally, the swap's VAR will converge to zero as the swap matures, dipping each time a coupon is set.

TABLE 11-11

An Interest-Rate Swap Viewed as Forward Contracts (Monthly VAR at 95 Percent Level)

	PV of Flows: Contract					
Term (Year)	1	1 × 2	2 × 3	3 × 4	4 × 5	VAR
1	−$100.36	$94.50				
2		−$94.64	$89.11			
3			−$89.08	$83.88		
4				−$83.70	$78.82	
5					−$78.55	
VAR	$0.471	$0.571	$0.488	$0.446	$0.425	
Undiversified VAR						$2.401
Diversified VAR						$2.152

11.4 MAPPING OPTIONS

We now consider the mapping process for nonlinear derivatives, or options.

Obviously, this nonlinearity may create problems for risk measurement systems based on the delta-normal approach, which is fundamentally linear.

To simplify, consider the Black-Scholes (BS) model for European options.[2] The model assumes, in addition to perfect capital markets, that the underlying spot price follows a continuous *geometric brownian motion* with constant volatility $\sigma(dS/S)$. Based on these assumptions, the Black-Scholes (1973) model, as expanded by Merton (1973), gives the value of a European call as

$$c = c(S,K,\tau,r,r^*,\sigma) = Se^{-r^*\tau}N(d_1) - Ke^{-r\tau}N(d_2) \tag{11.16}$$

where $N(d)$ is the cumulative normal distribution function described in Chapter 5 with arguments

$$d_1 = \frac{\ln(Se^{-r^*\tau} / Ke^{-r\tau})}{\sigma\sqrt{\tau}} + \frac{\sigma\sqrt{\tau}}{2}, \qquad d_2 = d_1 - \sigma\sqrt{\tau}$$

where K is now the *exercise price* at which the option holder can, but is not obligated to, buy the asset.

Changes in the value of the option can be approximated by taking partial derivatives, that is,

$$\begin{aligned} dc &= \frac{\partial c}{\partial S}dS + \tfrac{1}{2}\frac{\partial^2 c}{\partial S^2}dS^2 + \frac{\partial c}{\partial r^*}dr^* + \frac{\partial c}{\partial r}dr + \frac{\partial c}{\partial \sigma}d\sigma + \frac{\partial c}{\partial t}dt \\ &= \Delta dS + \tfrac{1}{2}\Gamma dS^2 + \rho^* dr^* + \rho dr + \Lambda d\sigma + \Theta dt \end{aligned} \tag{11.17}$$

The advantage of the BS model is that it leads to closed-form solutions for all these partial derivatives. Table 11-12 gives typical values for 3-month European call options with various exercise prices.

The first partial derivative, or *delta*, is particularly important. For a European call, this is

$$\Delta = e^{-r^*\tau}N(d_1) \tag{11.18}$$

This is related to the cumulative normal density function covered in Chapter 5. Figure 11-2 displays its behavior as a function of the underlying spot price and for various maturities.

The figure shows that delta is not a constant, which may make linear methods inappropriate for measuring the risk of options. Delta increases with the underlying spot price. The relationship becomes more nonlinear for short-term options, for example, with an option maturity of 10 days. Linear methods approximate delta by a constant value over the risk horizon. The quality of this approximation depends on parameter values.

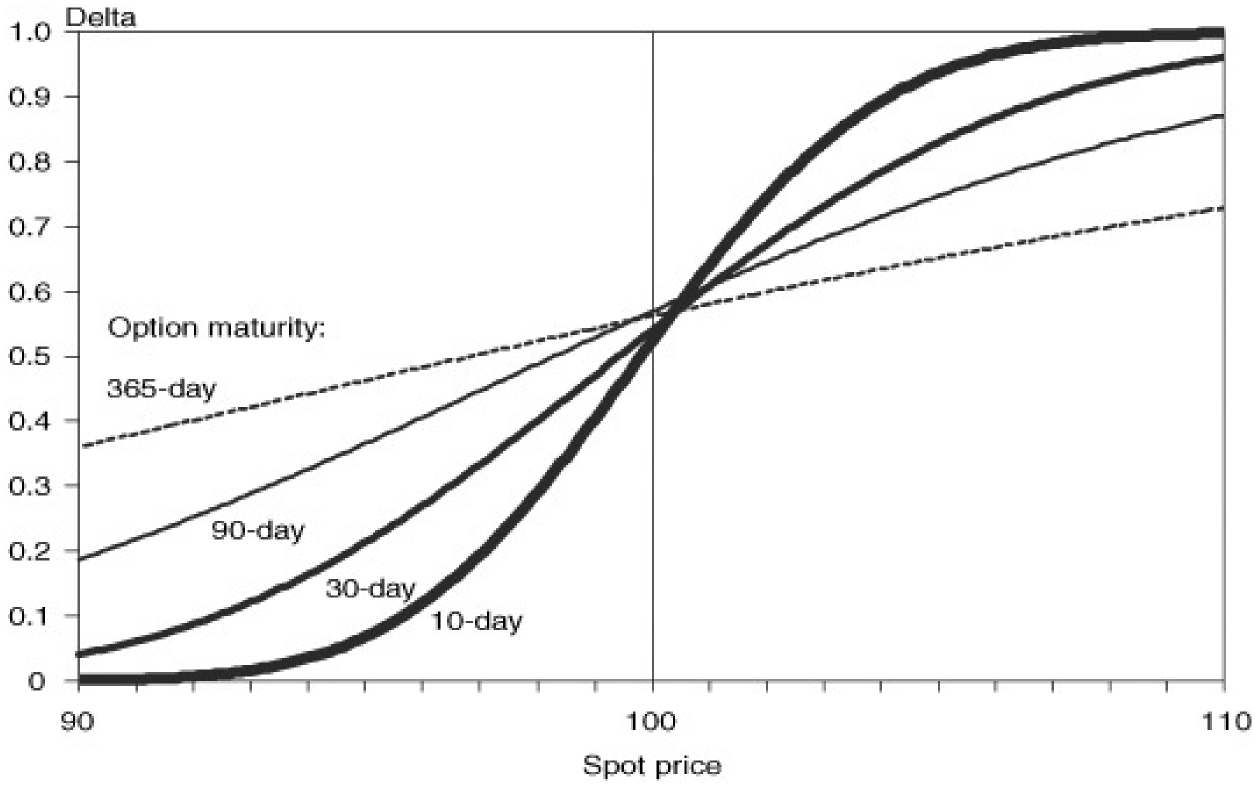
TABLE 11-12

Derivatives for a European Call

Parameters: S = \$100, σ = 20%, r = 5%, r^* = 3%, τ = 3 months					
			Exercise Price		
	Variable	**Unit**	$K = 90$	$K = 100$	$K = 110$
c		*Dollars*	11.01	4.20	1.04
		Change per			
Δ	Spot price	Dollar	0.869	0.536	0.195
Γ	Spot price	Dollar	0.020	0.039	0.028
Λ	Volatility	(% pa)	0.102	0.197	0.138
ρ	Interest rate	(% pa)	0.190	0.123	0.046
ρ^*	Asset yield	(% pa)	−0.217	−0.133	−0.049
θ	Time	Day	−0.014	−0.024	−0.016

FIGURE 11-2

Delta as a function of the risk factor.



For instance, if the risk horizon is 1 day, the worst down move in the spot price is –α *S* σ

$$-\alpha S\sigma\sqrt{T} = -1.645 \times \$100 \times 0.20\sqrt{1/252} = -\$2.08$$

, leading to a worst price of \$97.92. With a 90-day option, delta changes from 0.536 to 0.452 only. With such a small change, the linear effect will dominate the nonlinear effect. Thus linear approximations may be acceptable for options with long maturities when the risk horizon is short.

It is instructive to consider only the linear effects of the spot rate and two interest rates, that is,

$$\begin{aligned} dc &= \Delta dS + \rho^* dr^* + \rho dr \\ &= [e^{-r^*\tau} N(d_1)]dS + [-Se^{-r^*\tau}\tau N(d_1)]dr^* + [Ke^{-r\tau}\tau N(d_2)]dr \\ &= [Se^{-r^*\tau} N(d_1)]\frac{dS}{S} + [Se^{-r^*\tau} N(d_1)]\frac{dP^*}{P^*} - [Ke^{-r\tau} N(d_2)]\frac{dP}{P} \\ &= x_1\frac{dS}{S} + x_2\frac{dP^*}{P^*} + x_3\frac{dP}{P} \end{aligned} \quad (11.19)$$

This formula bears a striking resemblance to that for foreign currency forwards, as in Equation (11.13). The only difference is that the position on the spot foreign currency and on the foreign currency bill $x_1 = x_2$ now involves $N(d_1)$, and the position on the dollar bill x_3 involves $N(d_2)$. In the extreme case, where the option is deep in the money, both $N(d_1)$ and $N(d_2)$ are equal to unity, and the option behaves exactly like a position in a forward contract. In this case, the BS model reduces to $c = Se^{-r^*\tau} - Ke{-}r\tau$, which is indeed the valuation formula for a forward contract, as in Equation (11.9).

Also note that the position on the dollar bill $Ke^{-r\tau}N(d_2)$ is equivalent to $Se^{-r^*\tau}N(d_1) - c = S\Delta\, c$. This shows that the call option is equivalent to a position of Δ in the underlying asset plus a short position of $(\Delta S - c)$ in a dollar bill, that is

$$\text{Long option} = \text{long } \Delta \text{ asset} + \text{short } (\Delta S - c) \text{ bill}$$

For instance, assume that the delta for an at-the-money call option on an asset worth \$100 is $\Delta = 0.536$. The option itself is worth \$4.20. This option is equivalent to a ΔS = \$53.60 position in the underlying asset financed by a loan of $\Delta S - c$ = \$53.60 – \$4.20 = \$49.40.

The next step in the risk measurement process is the aggregation of exposures across the portfolio. Thus all options on the same underlying risk factor are decomposed into their delta equivalents, which are summed across the portfolio. This generalizes to movements in the implied volatility, if necessary. The option portfolio would be characterized by its net *vega*, or Λ. This decomposition also can take into account second-order derivatives using the net *gamma*, or Γ. These exposures can be combined with simulations of the underlying risk factors to generate a risk distribution.

11.5 CONCLUSIONS

Risk measurement at financial institutions is a top-level aggregation problem involving too many positions to be modeled individually. As a result, instruments have to be mapped on a smaller set of primitive risk factors.

Choosing the appropriate set of risk factors, however, is part of the art of risk management. Too many risk factors would be unnecessary, slow, and wasteful. Too few risk factors, in contrast, could create blind spots in the risk measurement system. These issues will be discussed in Chapter 21, where we will discuss limitations of VAR.

The mapping process consists of replacing the current values of all instruments by their exposures on these risk factors. Next, exposures are aggregated across the portfolio to create a net exposure to each risk factor. The risk engine then combines these exposures with the distribution of risk factors to generate a distribution of portfolio values.

For some instruments, the allocation into general-market risk factors is exhaustive. In other words, there is no specific risk left. This is typically the case with derivatives, which are tightly priced in relation to their underlying risk factor. For others positions, such as individual stocks or corporate bonds, there remains some risk, called *specific risk*. In large, well-diversified portfolios, this remaining risk tends to wash away. Otherwise, specific risk needs to be taken into account.

APPENDIX 11.A
Assigning Weights to Vertices

The duration-mapping example in this chapter showed that, in general, cash flows fall between the selected vertices.[3] In our example, the portfolio consists of one cash flow with maturity of D_p =2.7325 years and present value of $200 million. The question is, how should we allocate the $200 million to the adjoining vertices in a way that best represents the risk of the original investment?

A simple method consists of allocating funds according to *duration matching*. Define x as the weight on the first vertex and D_1, D_2 as the duration of the first and second vertices. The portfolio duration D_p will be matched if

$$xD_1 + (1 - x)D_2 = D_p \qquad (11.20)$$

or $x = (D_2 – D_p)/(D_2 – D_1)$. In our case, x = (3 – 2.7325)/(3 – 2) = 0.2675, which leads to an amount of $200 X 0.2675 = $53.49 million on the first vertex. The balance of $146.51 million is allocated to the 3-year vertex.

Unfortunately, this approach may not create a portfolio with the same risk as the original portfolio. The second method aims at *variance matching*. Define σ_1 and σ_2 as the respective volatilities and ρ as the correlation. The portfolio variance is which we set equal to the variance of the zero-coupon bond falling between the two vertices. By linear interpolation of the price volatilities for 2- and 3-year zeroes, the portfolio volatility is σ_p = 1.351 percent, as we have done before. Therefore, the weight x that maintains the portfolio risk to that of the initial investment is found from solving the quadratic equation

$$V(R_p) = x^2\sigma_1^2 + (1-x)^2\sigma_2^2 + 2x(1-x)\rho\sigma_1\sigma_2 \qquad (11.21)$$

$$(\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2)x^2 + 2(-\sigma_2^2 + \rho\sigma_1\sigma_2)x + (\sigma_2^2 - \sigma_p^2) = 0 \qquad (11.22)$$

The solution to the equation

$$ax^2 + 2bx + c = 0 \text{ is } x = (-b \pm \sqrt{b^2 - ac})/a$$

, which leads to the two roots x_1 =0.2635 and x_2 = 5.2168. We choose the first root, which is between zero and unity. As shown in Table 11-13, this translates into a position of \$52.71 million on the 2-year vertex and \$147.29 million on the 3-year vertex.

In this example, the difference between the two approaches is minor. The VAR from variance matching is \$2.702 million versus \$2.698 million with duration matching. In fact, the duration approximation is exact under two conditions: (1) the correlation coefficient is unity, and (2) the volatility of each vertex is proportional to its duration ($\sigma_1 = \sigma D_1, \sigma_2 = \sigma D_2$, $\rho = 1$). Under these conditions, Equation (11.21) simplifies to

$$\begin{aligned} V(R_p) &= x^2\sigma^2 D_1^2 + (1-x)^2\sigma^2 D_2^2 + 2x(1-x)\sigma^2 D_1 D_2 \\ &= \sigma^2[xD_1 + (1-x)D_2]^2 \end{aligned} \qquad (11.23)$$

which equals $(\sigma D_p)^2$ if $[xD_1\ (1-x)D_2] = D_p$. In other words, duration matching is perfectly appropriate under these conditions. In more general cases, especially if ρ is much lower than 1, the duration approximation will fail to provide a portfolio with the same risk as that of the original portfolio.

TABLE 11-13

Assigning Weights to Vertices

	Variance Matching				Duration Matching	
Term (Year)	VAR (%)	Correlation	Weight	Amount	Weight	Amount
2	0.9868		0.2635	$52.71	0.2675	$53.49
3	1.4841	0.9908	0.7365	$147.29	0.7325	$146.51
2.7325	1.3510					
Total			1.0000	$200.00	1.0000	$200.00
VAR				$2.702		$2.698

QUESTIONS

1. Why do risk managers use mapping instead of using historical data on each of the positions?
2. A risk manager is asked to provide a VAR measure for a portfolio of stocks, including IPOs. Discuss whether these IPO stocks should be ignored in the risk measurement process or not.
3. A portfolio of 100 fixed-income instruments is mapped to six risk factors. If the net exposures are zero, the portfolio has no risk. Discuss.
4. A portfolio of corporate bonds carries four different credit ratings and has maturities of 1, 5, and 10 years. How many general risk factors is the portfolio exposed to?
5. What is the drawback of principal mapping for bonds?
6. What is the main assumption for the duration approximation? What is the implication for the structure of a covariance matrix with different maturities as risk factors?
7. Consider a 3-year zero-coupon bond with a face value of $100 and a yield of 4% (annually compounded). Compute this bond’s modified duration, dollar duration, and DVBP.
8. A portfolio manager evaluates the risk of a two-bond portfolio:

	Price	Modified Duration	Number Held
30-year bond	$100	13.84	5,000
10-year bond	$100	7.44	5,000

We assume that specific risk is negligible and that the volatility of changes in market yields is 29 basis points. Under these conditions, what is the volatility of the portfolio value?

9. A portfolio manager enters a 10-year pay-fixed swap with notional of $100 million. The duration of the fixed leg is 7.44 years, and the floating leg is about to be reset. Assume a flat term structure and an annual volatility of yield changes of 100 basis points. What is the 95 percent VAR over the next month?
10. Now assume that the floating leg has just been reset for payment in a year. Compute the VAR.
11. Is duration hedging an appropriate way to minimize tracking error relative to an index?
12. A portfolio manager holds a $100 million price position in 10-year Treasury notes with a daily volatility of 0.9 percent. The manager can hedge by selling 5-year T-notes with a daily volatility of 0.5 percent and correlation of 0.97. For computation of VAR, assume normal distributions and a 95 percent confidence level. Based on this information, what amount did the manager sell, and what was the resulting VAR?
13. Market risk can be defined in absolute or relative terms. Can a portfolio have a positive return yet have relative risk? Give an example.
14. A U.S. exporter sold forward Y125 million at the 7-month forward rate of Y124.27/$. Immediately after the deal is signed, the spot rate moves from Y125 to Y130/$. Dollar and yen rates are still at 6 and 5 percent, respectively. What is the gain/loss of the contract?
15. What features of cash and futures prices tend to make hedging possible?
16. In a foreign-currency futures contract, how is basis risk created?
17. When is basis risk greatest in general?

18. Assume that the spot rate for the euro against the U.S. dollar is $1.05 (i.e., 1 euro buys 1.05 dollars). A U.S. bank pays 5.5 percent compounded annually for 1 year for a dollar deposit, and a German bank pays 2.5 percent compounded annually for 1 year for a euro deposit. The forward exchange rate is set on the contract at $1.06. What is the current value of this forward contract to buy one euro 1 year from now?
19. A U.S. exporter anticipates receiving 1 million British pounds in 3 months. This is hedged with a short position in BP futures expiring in 6 months. The initial spot and futures prices are $1.5000 and $1.4703. At the time the hedge is lifted, the respective prices are $1.4000 and $1.3861. Ignoring the daily marking to market, what are the total proceeds to the exporter?
20. What is the major risk factor for a forward currency position?
21. Explain why the spot price for natural gas has, or should have, greater volatility than for gold.
22. A trader has a long position in at-the-money calls on $1 million worth of an underlying stock with volatility of 20 percent. Roughly, what is the daily VAR at the 95 percent confidence level?
23. Discuss whether the delta VAR computed in the preceding question is likely to be more appropriate if the maturity of the option is 3 months or 5 days.

CHAPTER 12
Monte Carlo Methods

Deus ex machina.

Wall Street is often compared to a casino. The analogy is appropriate in one respect: Securities firms commonly use simulation techniques, known as *Monte Carlo methods*, to value complex derivatives and to measure risk. Simulation methods approximate the behavior of financial prices by using computer simulations to generate random price paths.

These methods are used to simulate a variety of different scenarios for the portfolio value on the target date. These scenarios can be generated in a random fashion (as in Monte Carlo simulation) or from historical data (as in historical simulation) or in other, more systematic ways. The portfolio value at risk (VAR) then can be read off directly from the distribution of simulated portfolio values.

Because of its flexibility, the simulation method is by far the most powerful approach to VAR. It potentially can account for a wide range of risks, including price risk, volatility risk, and complex interactions such as described by copulas in Chapter 8. Simulations can account for nonlinear exposures and complex pricing patterns. In principle, simulations can be extended to longer horizons, which is important for credit risk measurement and to more complex models of expected returns. Also, it can be used for operational risk measurement, as well as integrated risk management.

This approach, however, involves costly investments in intellectual and systems development. It also requires substantially more computing power than simpler methods. VAR measures using Monte Carlo methods often require hours to run. Time requirements, however, are being whittled down by advances in computers and faster valuation methods.

This chapter shows how simulation methods can be used to uncover VAR. The first section presents the rationale for Monte Carlo simulations. Section 12.2 introduces a simple case with just one random variable. Section 12.3 then discusses the tradeoff between speed and accuracy. The case with many sources of risk is discussed in Section 12.4. Next, Sections 12.5 and 12.6 turn to newer methods, such as deterministic simulations. The choice of models is reviewed in Section 12.7.

12.1 WHY MONTE CARLO SIMULATIONS?

The basic concept behind the Monte Carlo approach is to simulate repeatedly a random process for the financial variable of interest covering a wide range of possible situations. These variables are drawn from prespecified probability distributions that are assumed to be known, including the analytical function and its parameters. Thus simulations recreate the entire distribution of portfolio values, from which VAR can be derived.

Monte Carlo simulations were developed initially as a technique of statistical sampling to find solutions to integration problems, as shown in Box 12-1. For instance, take the problem of numerical integration of a function with many variables. A straightforward method is to perform the integration by computing the area under the curve using a number of evenly spaced samples from the function. In general, this works very well for functions of one variable. For functions with many variables, however, this method quickly becomes inefficient. With two variables, a 10 X 10 grid requires 100 points. With 100 variables, the grid requires 10^{100} points, which is too many to compute. This problem is called the *curse of dimensionality*.

BOX 12-1

MONTE CARLO SIMULATIONS

Numerical simulations were first used by atom bomb scientists at Los Alamos in 1942 to crack problems that could not be solved by conventional means. Stanislaw Ulam, a Polish mathematician, is usually credited with inventing the Monte Carlo method while working at the Los Alamos laboratory.

While there, Ulam suggested that numerical simulations could be used to evaluate complicated mathematical integrals that arise in the theory of nuclear chain reactions. This suggestion led to the more formal development of Monte Carlo methods by John Von Neumann, Nicholas Metropolis, and others.

In his autobiography, *Adventures of a Mathematician*, Ulam recollects that the method was named in honor of his uncle, who was a gambler. The name *Monte Carlo* was derived from the name of a famous casino established in 1862 in the south of France (actually, in Monaco). What better way to evoke random draws, roulette, and games of chance?

Monte Carlo simulation instead provides an approximate solution to the problem that is much faster. Instead of systematically covering all values in the multidimensional space, it generates K random samples for the vector of variables. By the *central limit theorem*, this method generates estimates whose standard error decreases at the rate of $1/\sqrt{K}$, which does not depend on the size of the sample space. Thus the method does not suffer from the curse of dimensionality.

12.2 SIMULATIONS WITH ONE RANDOM VARIABLE

12.2.1 Simulating a Price Path

We first concentrate on a simple case with just one random variable. The first, and most crucial, step in the simulation consists of choosing a particular stochastic model for the behavior of prices. A commonly used model is the *geometric brownian motion* (GBM) *model*, which underlies much of options pricing theory. The model assumes that innovations in the asset price are uncorrelated over time and that small movements in prices can be described by

$$dS_t = \mu_t S_t dt + \sigma_t S_t dz \tag{12.1}$$

where dz is a random variable distributed normally with mean zero and variance dt. This variable drives the random shocks to the price and does not depend on past information. It is *brownian*in the sense that its variance decreases continuously with the time interval, $V(dz) = dt$. This rules out processes with sudden jumps, for instance. The process is also *geometric*because all parameters are scaled by the current price S_t.

The parameters μ_t and σ_t represent the instantaneous drift and volatility at time t, which can evolve over time. For simplicity, we will assume in what follows that these parameters are constant over time. But since μ_t and σ_t can be functions of past variables, it would be easy to simulate time variation in the variances as in a GARCH process, for example.

In practice, the process with infinitesimally small increments dt can be approximated by discrete moves of size Δt. Define t as the present time, T as the target time, and $\tau = T - t$ as the (VAR) horizon. To generate a series of random variables S_{t+i} over the interval τ, we first chop up τ into n increments, with $\Delta t =$

τ/n.[1]

Integrating dS/S over a finite interval, we have approximately

$$\Delta S_t = S_{t-1}\,(\mu\Delta t + \sigma\epsilon\sqrt{\Delta t}) \qquad (12.2)$$

where ∈ is now a standard normal random variable, that is, with mean zero and unit variance. We can verify that this process generates a mean $E(\Delta S/S) = \mu\Delta t$, which grows with time, as does the variance $V(\Delta S/S) = \sigma^2\Delta t$.

To simulate the price path for S, we start from S_t and generate a sequence of epsilons (∈'s) *for i* 1,2, . . ., n. Then S_{t+1} is set at $S_{t+1} = S_t =$ $S_{t+1} = S_t + S_t(\mu\Delta t + \sigma\epsilon_1\sqrt{\Delta t})$, S_{t+2} is similarly computed from $S_{t+1} + S_{t+1}(\mu\Delta t + \sigma\epsilon_2\sqrt{\Delta t})$, and so on for future values, until the target horizon is reached, at which point the price is $S_{t+n} = S_T$.

Table 12-1 illustrates a simulation of a process with a drift μ of zero and volatility σ of 10 percent over the total interval. The initial price is $100, and the interval is cut into 100 steps. Therefore, the local volatility is $0.10 \times \sqrt{1/100} = 0.01$.

TABLE 12-1

Simulating a Price Path

Step i	Previous Price S_{t+i-1}	Random Variable ϵ_i	Increment ΔS	Current Price S_{t+i}
1	100.00	0.199	0.00199	100.20
2	100.20	1.665	0.01665	101.87
3	101.87	−0.445	−0.00446	101.41
4	101.41	−0.667	−0.00668	100.74
⋮	⋮	⋮	⋮	⋮
100	92.47	1.153	−0.0153	91.06

The second column starts with the initial price. The next column displays the realization of a standard normal variable. With no drift, the increment in the following column is simply (∈ X 0.01). Finally, the last column computes the current price from the previous price and the increment. The values at each point are conditional on the simulated values at the previous point. The process is repeated until the final price of $91.06 is reached at the 100th step.

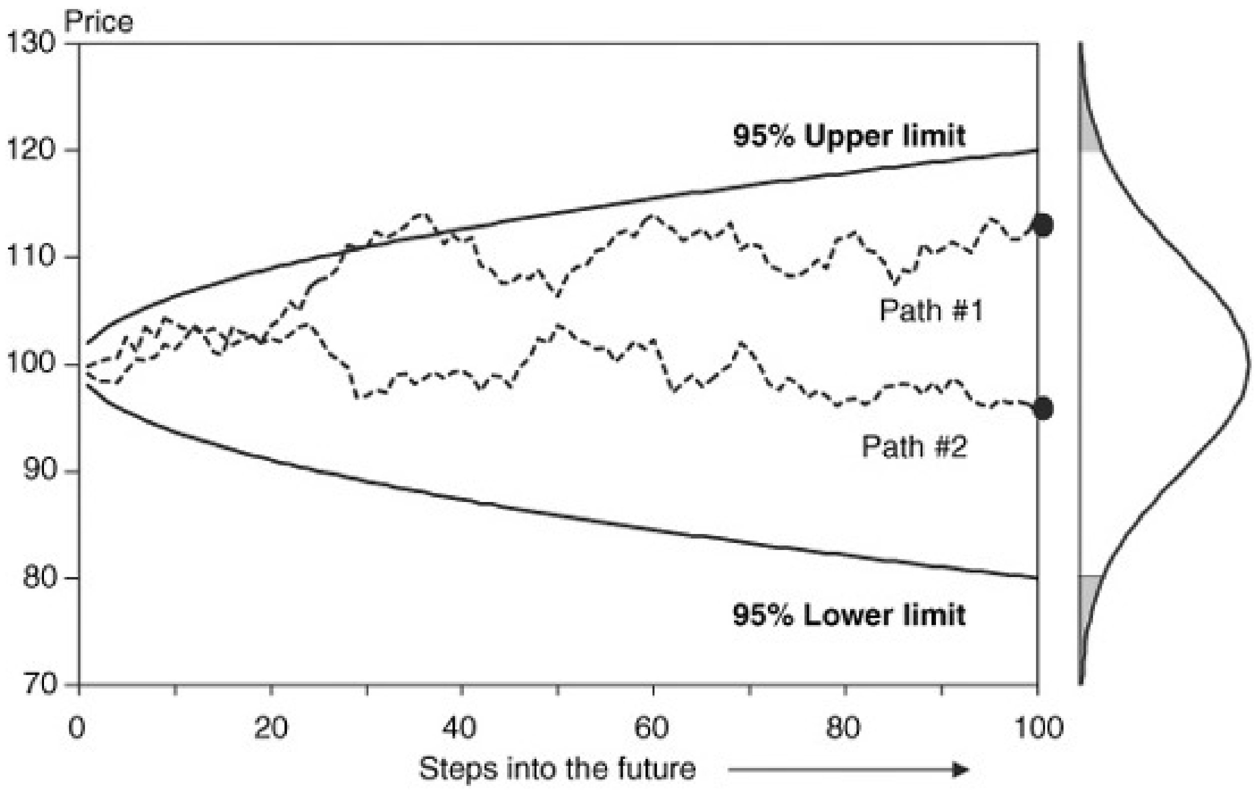
Figure 12-1 presents two price paths, each leading to a different ending price.

Given these assumptions, the ending price must follow a normal distribution with mean of \$100 and standard deviation of \$10.[2] This distribution is illustrated on the right side of the figure, along with 95 percent confidence bands, corresponding to two standard deviation intervals.

But the distribution also is known at any intermediate point. The figure displays 95 percent confidence bands that increase with the square root of time until they reach ±2 X 10 percent. In this simple model, risk can be computed at any point up to the target horizon.

FIGURE 12-1

Simulating price paths.



12.2.2 Creating Random Numbers

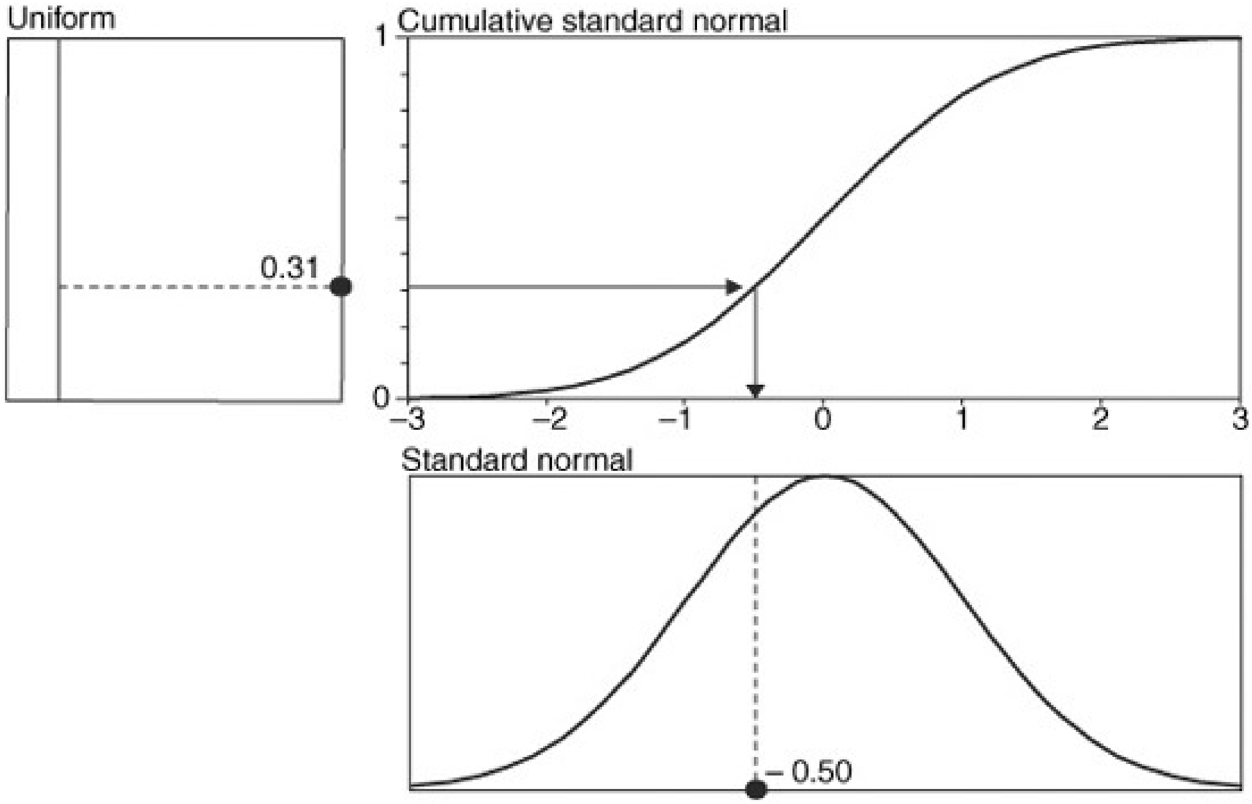
Monte Carlo simulations are based on random draws ∈ from a variable with the desired probability distribution. The numerical analysis usually proceeds in two steps.

The first building block for a random-number generator is a uniform distribution over the interval [0,1] that produces a random variable *x*. More properly speaking, these numbers are “pseudo” random because they are generated from an algorithm using a predefined rule. Starting from the same “seed” number, the sequence can be repeated at will.

The next step is to transform the uniform random number x into the desired distribution through the inverse cumulative probability distribution function (pdf). Take the normal distribution. By definition, the cumulative pdf $N(y)$ is always between 0 and 1. Therefore, to generate a normally distributed random variable, we compute y such that $x = N(y)$ or $y = N^{-1}(x)$.[3] More generally, any distribution function can be generated as long as the function $N(y)$ can be inverted. Figure 12-2 illustrates this procedure, called the *inverse transform method*.

FIGURE 12-2

Transformation from uniform to normal.



At this point, an important caveat is in order. It seems easy to generate variables that are purely random, but in practice, it is quite difficult. A well-designed algorithm should generate draws that “appear” independent over time. Whether this sequence is truly random is a philosophical issue that we will not address. Good random-number generators must create series that pass all conventional tests of independence. Otherwise, the characteristics of the simulated price process will not obey the underlying model.

Most operating systems, unfortunately, provide a random-number generator that is simple but inaccurate. All algorithms "cycle" after some iterations; that is, they repeat the same sequence of pseudorandom numbers. Good algorithms cycle after billions of draws; bad ones may cycle after a few thousand only.

If the cycle is too short, dependencies will be introduced in the price process solely because of the random-number generator. As a result, the range of possible portfolio values may be incomplete, thus leading to incorrect measures of VAR. This is why it is important to use a good-quality algorithm, such as those found in numerical libraries.

12.2.3 The Bootstrap

An alternative to generating random numbers from a hypothetical distribution is to sample from historical data. Thus we are agnostic about the distribution. For example, suppose that we observe a series of M returns $R = \Delta S/S$, $\{R\} = (R_1 \dots R_M)$, which can be assumed to be i.i.d. random variables drawn from an unknown distribution. The historical simulation method consists of using this series once to generate pseudoreturns. But this can be extended much further.

The bootstrap estimates this distribution by the empirical distribution of R, assigning equal probability to each realization. The method was proposed initially by Efron (1979) as a nonparametric randomization technique that draws from the observed distribution of the data to model the distribution of a statistic of interest.[4]

The procedure is carried out by sampling from $\{R\}$, with replacement, as many observations as necessary. For instance, assume that we want to generate 100 returns into the future, but we do not want to impose any assumption on the distribution of daily returns. We could project returns by randomly picking one return at a time from the sample over the past $M = 500$ days, with replacement. Define the index choice as $m(1)$, a number between 1 and 500. The selected return then is $R_{m(1)}$, and the simulated next-day return is $S_{t+1} = S_t(1 + R_{m(1)})$. Repeating the operation for a total of 100 draws yields a total of 100 pseudovalues $S_{t+1}, \dots, S_{t+n}$.

An essential advantage of the bootstrap is that it can include fat tails, jumps, or any departure from the normal distribution. For instance, one could include the return for the crash of October 19, 1987, which would never (or nearly never) occur under a normal distribution. The method also accounts for correlations across series because one draw consists of the simultaneous returns for N series, such as stock, bonds, and currency prices.

The bootstrap approach, it should be noted, has limitations. For small sample sizes M, the bootstrapped distribution may be a poor approximation of the actual one. Therefore, it is important to have access to sufficient data points. The other drawback of the bootstrap is that is relies heavily on the assumption that returns are independent. By resampling at random, any pattern of time variation is broken.

The bootstrap, however, can accommodate some time variation in parameters as long as we are willing to take a stand on the model. For instance, the bootstrap can be applied to the normalized residuals of a GARCH process, that is,

$$\epsilon_t = \frac{r_t}{\sigma_t}$$

where r_t is the actual return, and σ_t is the conditional standard deviation from the estimated GARCH process. To recreate pseudoreturns, one then would first sample from the historical distribution of $\in$ and then reconstruct the conditional variance and pseudoreturns.

12.2.4 Computing VAR

Once a price path has been simulated, we can build the portfolio distribution at the end of the selected horizon. The simulation is carried out by the following steps:

1. Choose a stochastic process and parameters.
2. Generate a pseudosequence of variables $\in_1, \in_2, \ldots, \in_n$, from which prices are computed as $S_{t+1}, S_{t+2}, \ldots, S_{t+n}$.
3. Calculate the value of the asset (or portfolio) $F_{t+n} = F_{\mathrm{T}}$ under this particular sequence of prices at the target horizon.
4. Repeat steps 2 and 3 as many times as necessary, say, $K = 10{,}000$.

This process creates a distribution of values $F_T^1, \ldots, F_T^{10{,}000}$. We can sort the observations and tabulate the expected value $E(F_T)$ and the quantile $Q(F_T, c)$, which is the value exceeded in c times 10,000 replications. VAR relative to the mean then is

$$\mathrm{VAR}(c,T) = E(F_T) - Q(F_T,c) \qquad (12.3)$$

12.2.5 Risk Management and Pricing Methods

It is interesting to note that Monte Carlo methods in finance were proposed originally in the context of options valuation.[5] Simulations are particularly useful to evaluate options that have no closed-form solution. Under the risk-neutral valuation method, Monte Carlo simulation consists of the following steps:

1. Choose a process with a drift equal to the risk-free rate, that is, with $\mu = r$ in Equation (12.1).
2. Simulate prices to the horizon S_T.
3. Calculate the payoff of the derivative at maturity T, $F(S_T)$.
4. Repeat these steps as often as needed.

The current value of the derivative is obtained from discounting at the risk-free rate and averaging across all experiments, that is, where the expectation indicates averaging, and the asterisk is a reminder that the price paths are under *risk neutrality*, that is, both changing the expected return and the discount rate to the risk-free rate.

$$f_t = E^*[e^{-r\tau}F(S_T)] \tag{12.4}$$

This method is quite general and can be applied to options that have price-dependent paths (such as look-back options or average-rate options) or strange payoffs at expiration (such as nonlinear functions of the ending price). Its main drawback is that it cannot price options accurately when the holder can exercise early. Also, the distribution of prices must be finely measured to price options with sharp discontinuities, such as binary options, which pay a fixed amount if the price ends up above or below the strike price. With large “holes” in the price distributions, the payoffs on combinations of binary options simply could not appear in the final portfolio distribution. Thus highly complex payoffs can be handled with increased precision.

Monte Carlo methods also allow users to measure vega risk, or exposure to changes in volatility. All that is required is to repeat the simulation with the same sequence of $\in$ values but with another value for σ. The change in the value of the asset is owing solely to the change in the volatility measures vega risk.

Overall, the methods for pricing derivatives and measuring risk have much in common. Pricing, however, requires *risk-neutral* distributions, whereas risk measurement requires *physical*, or *objective*, distributions.

12.3 SPEED VERSUS ACCURACY

The main drawback of Monte Carlo (MC) methods is their computational time requirements. Consider, for instance, a portfolio exposed to one risk factor only. Say that we require 10,000 replications of this risk factor for acceptable accuracy. If the portfolio contains 1000 assets to be priced using full valuations, we will need 10 million valuations.

If, in addition, the portfolio contains complex instruments, such as mortgages or exotic options, whose valuation itself requires a simulation, measuring risk at a target date then requires "a simulation within a simulation":

- For valuation (i.e., from the VAR horizon to the maturity of the instrument)
- For risk management (i.e., from the present time to the VAR horizon)

Without shortcuts, the number of required simulations soon can reach astronomical values. This is why the industry is busily developing methods to cut down the number of simulations without too much loss in accuracy.

12.3.1 Accuracy

Simulations inevitably generate *sampling variability*, or variations in estimator values, owing to the limited number of replications. More replications lead to more precise estimates but take longer to estimate. Define K as the number of *replications*, or pseudorandom trials. To choose K, it is useful to assess the tradeoff between precision and the number of replications.

FIGURE 12-3

Convergence to true distribution.

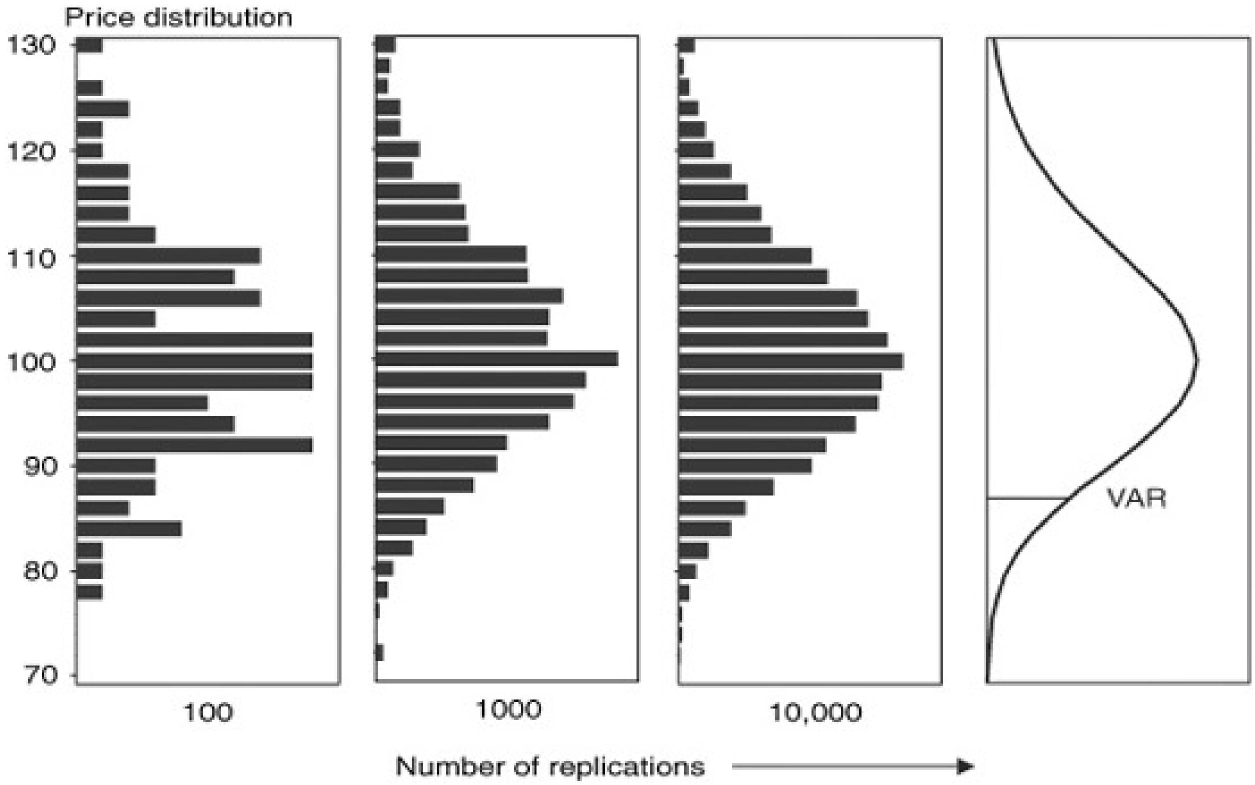


Figure 12-3 illustrates the convergence of the empirical distribution in Figure 12-1 toward the true one. With $K = 100$, the histogram representing the distribution of the ending price is quite irregular. The histogram becomes smoother with 1000 replications, even more so with 10,000 replications, and eventually should converge to the continuous distribution in the right panel. One advantage of the Monte Carlo method is that the user can evaluate the increase in accuracy directly as the number of replications increases.

If the underlying process is normal, the empirical distribution must converge to a normal distribution. In this situation, Monte Carlo analysis should yield exactly the same result as the delta-normal method: The VAR estimated from the sample quantile must converge to the value of $\alpha\sigma$.

Any deviation must be due to sampling variation. Assuming no other source of error, this effect can be measured by the asymptotic standard error for the sample quantile reported in Chapter 5, using K as the sample size. A simple method to assess accuracy is to repeat the simulations multiple times, say, $M = 1000$, and take the standard error of the estimated quantiles across the M experiments. This is illustrated in Table 12-2, which describes the results of 1000 simulation runs on a standard normal distribution with an increasing number of replications.

TABLE 12-2

Convergence Statistics for Risk Measures

		Standard Error			
		Replications			
Left Tail	Expected Quantile	100	500	1000	10,000
1%	−2.326	0.409	0.170	0.119	0.037
5%	−1.645	0.216	0.092	0.066	0.021
10%	−1.282	0.170	0.075	0.052	0.017
Std.dev.	1.000	0.069	0.032	0.022	0.007

The table shows that for a 99 percent VAR with 100 replications, the standard error of the estimate around –2.326 is 0.409, rather high. In our sample of 1000 runs, the VAR estimate ranged from –4.17 to –1.53. This dispersion is rather disturbing. To increase VAR precision by a factor of 10, we need to increase the number of replications by a factor of 100, for a total of 10,000. Indeed, the first line shows that this decreases the standard error to 0.037, which is approximately 0.409 divided by 10. Note that, in contrast, the standard deviation is estimated much more precisely because it uses data from the entire distribution.

We also could report the standard error in relative terms, defined as the ratio of the standard error to the expected value of the risk measure. For example, banks typically report their 99 percent VAR using about 500 days. From Table 12-2, this leads to a relative error in VAR of around 0.170/2.326 = 7.3 percent.

The relative error depends on the number of replications, as well as on the shape of the distribution, as shown in Table 12-3.[6] The table shows that the error is higher for left-skewed distributions and conversely lower for right-skewed distributions. This is so because the longer the left tail, the less precise is the VAR estimate.

TABLE 12-3

Relative Error in 99 Percent VAR for Various Distributions

		Relative Error (Percent)			
		Replications			
Distribution	Skewness	100	500	1000	10,000
Normal	0.00	17.6	7.3	5.1	1.5
Right skew	0.76	9.3	4.2	3.0	0.9
Left skew	−0.76	23.4	9.2	6.3	1.9

Alternatively, we could search for the number of replications required to measure VAR with a relative error of 1 percent. For the normal distribution, we need more than 20,000 replications to make sure that the relative error in the first row is below 1 percent.

12.3.2 Acceleration Methods

This led to a search for methods to accelerate computations. One of the earliest, and easiest, is the *antithetic variable technique*, which consists of changing the sign of all the random samples ϵ. This method, which is appropriate when the original distribution is symmetric, creates twice the number of replications for the risk factors at little additional cost. We still need, however, twice the original number of full valuations on the target date.

This approach can be applied to the historical simulation method, where we can add a vector of historical price changes with the sign reversed. This is also useful to eliminate the effect of trends in the recent historical data.

Another useful tool is the *control variates technique*. We are trying to estimate VAR, a function of the data sample. Call this $V(X)$. Assume now that the function can be approximated by another function, such as a quadratic approximation $V^Q(X)$, for which we have a closed-form solution v^Q.[7]

For any sample, the error then is known to be $V^Q(X) - v^Q$ for the quadratic approximation. If this error is highly correlated with the sampling error in $V(X)$, the control variate estimator can be taken as

$$V_{CV} = V(X) - [V^Q(X) - v^Q] \tag{12.5}$$

This estimator has much lower variance than the original one when the quadratic function provides a good approximation of the true function.

The most effective acceleration method is the *importance sampling technique*, which attempts to sample along the paths that are most important to the problem at hand. The idea is that if our goal is to measure a tail quantile accurately, there is no point in doing simulations that will generate observations in the center of the distribution. The method involves shifts in the distribution of random variables. Glasserman *et al.* (2000) show that relative to the usual Monte Carlo method, the variance of VAR estimators can be reduced by a factor of at least 10.

A related application is the *stratified sampling technique*, which can be explained intuitively as follows: assume that we require VAR for a long position in a call option.[8] We are trying to keep the number of replications at $K = 1000$. To increase the accuracy of the VAR estimator, we could partition the simulation region into two zones. As before, we start from a uniform distribution, which then is transformed into a normal distribution for the underlying asset price using the *inverse transform method*.

Define these two zones, or strata, for the uniform distribution as [0.0, 0.1] and [0.1, 1.0]. Thus *stratification* is the process of grouping the data into mutually exclusive and collectively exhaustive regions. Usually, the probabilities of the random number falling in both zones are selected as $p_1 = 10$ percent and $p_2 = 90$ percent, respectively. Now we change these probabilities to 50 percent for both regions. The number of observations now is $K_1 = 500$ for the first region and $K_2 = 500$ for the second. This increases the number of samples for the risk factor in the first, left-tail region.

Estimators for the mean need to be adjusted for the stratification. We weight the estimator for each region by its probability, that is,

$$E(F_T) = p_1 \frac{\sum_{i=1}^{K_1} F_i}{K_1} + p_2 \frac{\sum_{i=1}^{K_2} F_i}{K_2} \tag{12.6}$$

To compute VAR, we simply examine the first region. The 50 percent quantile for the first region, for example, provides an estimator of a 10 X 0.5 = 5 percent left-tail VAR. Because VAR only uses the number of observations in the right region, we do not even need to compute their value, which economizes on the time required for full valuation.

This reflects the general principle that using more information about the portfolio distribution results in more efficient simulations. In general, unfortunately, the payoff function is not known. All is not lost, however. Instead, the simulation can proceed in two passes. The first pass runs a traditional Monte Carlo. The risk manager then examines the region of the risk factors that causes losses around VAR. A second pass then is performed with many more samples from this region.

12.4 SIMULATIONS WITH MULTIPLE VARIABLES

Modern risk measurement applications are large-scale problems because they apply at the highest level of the financial institution. This requires simulations over multiple sources of risk.

12.4.1 From Independent to Correlated Variables

Simulations generate independent random variables that need to be transformed to account for correlations. Define *N* as the number of sources of risk. If the variables are uncorrelated, the randomization can be performed independently for each variable, that is, where the $\in$ values are independent across time period and series $j = 1, \ldots, N$.

$$\Delta S_{j,t} = S_{j,t-1}\,(\mu_j \Delta t + \sigma_j \epsilon_{j,t}\sqrt{\Delta t}) \qquad (12.7)$$

To account for correlations between variables, we start with a set of independent variables η, which then are transformed into the $\in$. In a two-variable setting, we construct where ρ is the correlation coefficient between the variables $\in$. First, we verify that the variance of $\in_2$ is unity, that is,

$$\begin{aligned} \epsilon_1 &= \eta_1 \\ \epsilon_2 &= \rho\eta_1 + (1-\rho^2)^{1/2}\eta_2 \end{aligned} \qquad (12.8)$$

$$V(\epsilon_2) = \rho^2 V(\eta_1) + [(1-\rho^2)^{1/2}]^2 V(\eta_2) = \rho^2 + (1-\rho^2) = 1$$

Then we compute the covariance of the $\in$ *as*

$$\mathrm{cov}(\epsilon_1,\epsilon_2) = \mathrm{cov}[\eta_1, \rho\eta_1 + (1-\rho^2)^{1/2}\eta_2] = \rho\,\mathrm{cov}(\eta_1, \eta_1) = \rho$$

This confirms that the $\in$ variables have correlation of ρ. The question is, how was the transformation in Equation (12.8) chosen?

More generally, suppose that we have a vector of *N* values of $\in$ for which we would like to display some correlation structure $V(\in) = E(\in\in') = R$. We will use *Cholesky factorization*,named after the French mathematician André-Louis Cholesky, to generate correlated variables. Since the matrix *R* is a symmetric real matrix, it can be decomposed into its *Cholesky factors*, this is, where T is a lower triangular matrix with zeros in the upper right corners.

$$R = TT' \tag{12.9}$$

We start with an N vector η that is composed of independent variables all with unit variances. In other words, $V(\eta) = I$, where I is the identity matrix with zeroes everywhere except on the diagonal. Next, construct the variable $\in = T\eta$. Its covariance matrix is $V(\in) = E(\in\in') = E(T\eta\eta'T') = TE(\eta\eta')T'\ TIT' = TT' = R$. Thus we have confirmed that the values of ∈ have the desired correlations.

As an example, consider the two-variable case. The matrix can be decomposed into

$$\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} = \begin{bmatrix} a_{11} & 0 \\ a_{12} & a_{22} \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{bmatrix} = \begin{bmatrix} a_{11}^2 & a_{11}a_{12} \\ a_{11}a_{12} & a_{12}^2 + a_{22}^2 \end{bmatrix}$$

The entries in the right-hand-side matrix must match exactly each entry in the correlation matrix. Because the Cholesky matrix is triangular, the factors can be found by successive substitution by setting

$$\begin{aligned} a_{11}^2 &= 1 \\ a_{11}a_{12} &= \rho \\ a_{12}^2 + a_{22}^2 &= 1 \end{aligned}$$

which yields

$$\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \rho & (1-\rho^2)^{1/2} \end{bmatrix} \begin{bmatrix} 1 & \rho \\ 0 & (1-\rho^2)^{1/2} \end{bmatrix}$$

And indeed, this is how Equation (12.8) was obtained:

$$\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \rho & (1-\rho^2)^{1/2} \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix}$$

This explains how a multivariate set of random variables can be created from simple building blocks consisting of i.i.d. variables. In addition to providing a method to generate correlated variables, however, this approach generates valuable insight into the random number generation process.

12.4.2 Number of Risk Factors

For the decomposition to work, the matrix R must be *positive definite*.

Otherwise, there is no way to transform N independent source of risks into N correlated variables of ∈

As discussed in Chapter 8, this condition can be verified with the *singular value decomposition*. This decomposition of the covariance matrix provides a check that the matrix is well behaved. If any of the eigenvalues is zero or less than zero, the Cholesky decomposition will fail.

When the matrix R is not positive definite, its *determinant* is zero. Intuitively speaking, the determinant d is a measure of the "volume" of a matrix. If d is zero, the dimension of the matrix is less than N. The determinant can be computed easily from the Cholesky decomposition. Since the matrix T has zeros above its diagonal, its determinant reduces to the product of all diagonal coefficients $d_T = \Pi_{i=1}^{N} a_{ii}$. The determinant of the covariance matrix R then is $d = d_T^2$.

In our two-factor example, the matrix is not positive definite if $\rho = 1$. In practice, this implies that the two factors are really the same. The Cholesky decomposition then yields $a_{11} = 1$, $a_{12} = 1$, and $a_{22} = 0$, and the determinant $d = (a_{11}a_{22})^2$ is 0. As a result, the second factor $\eta_2$is never used, and $\in_1$ is always the same as $\in_2$. The second random variable is totally superfluous. In this case, the covariance matrix is not positive definite. Its true dimension, or *rank*, is 1, which means that it has only one meaningful risk factor.

These conditions may seem academic, but unfortunately, they soon become very real with simulations based on a large number of factors. The RiskMetrics covariance matrix, for instance, is routinely nonpositive definite owing to the large number of assets. These problems can arise for a number of reasons. Perhaps this is simply due to the large number of correlations. With $N = 450$, for instance, we have about 100,000 correlations with rounding errors. This also could happen when the effective number of observations T is less than the number of factors N. One drawback of time-varying models of variances is that they put less weight on older observations, thereby reducing the effective sample size. Or the correlations may have been measured over different periods, which may produce inconsistent correlations.[9] Another reason would be that the series are naturally highly correlated (such as the 9-year zero-coupon bond with the adjoining maturities) or that some series were constructed as a linear combination of others (such as a currency basket).

For simulations, this may be a blessing in disguise because fewer number of variables are sufficient. In Chapter 8 we gave the example of 11 bonds for which the covariance matrix could be reduced without much loss of information to two, or perhaps three, principal components. Thus the problem can be solved using a matrix of smaller dimensions, which speeds up the computation considerably. This illustrates that the design of simulation experiments, including the number of risk factors, is critical. As we have seen, however, the choice of the number of risk factors should be related to the trading strategy.

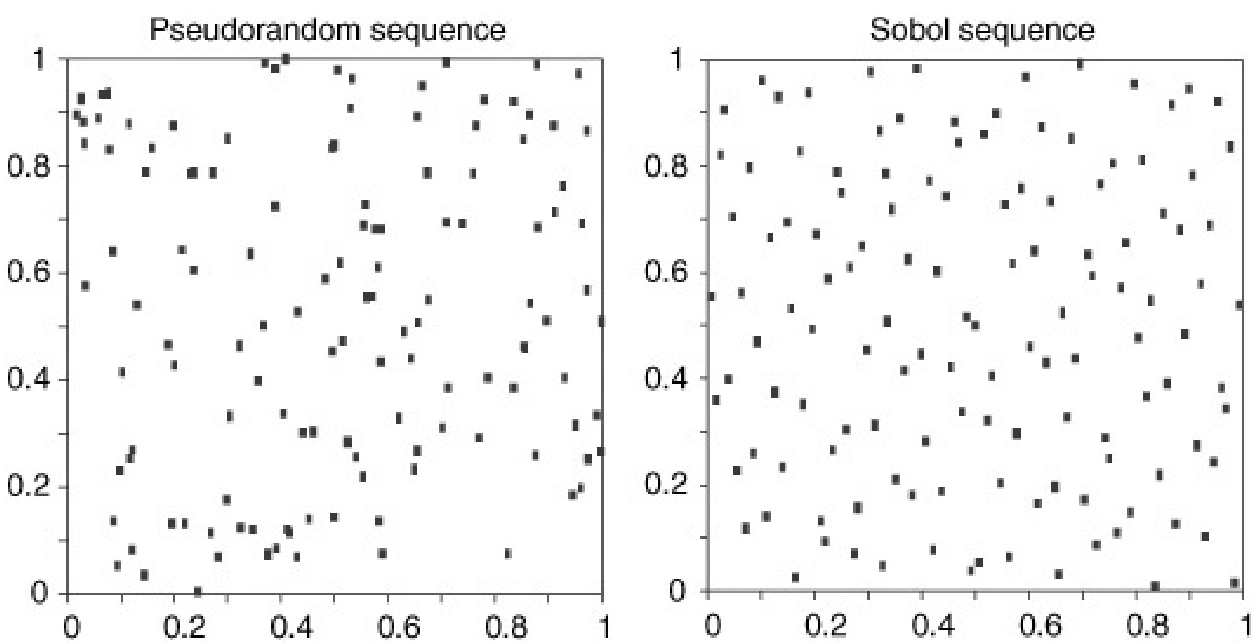
12.5 DETERMINISTIC SIMULATION

Monte Carlo simulation methods generate independent, pseudorandom points that attempt to "fill" an *N*-dimensional space, where *N* is the number of risk factors. The sequence of points does not have to be chosen randomly, however.

Indeed, it is possible to use a *deterministic* scheme that is constructed to provide a more consistent fill to the *N*-space. The choice must account for the sample size, dimensionality of the problem, and possibly the shape of the function being integrated. These deterministic schemes are sometimes called *quasi–Monte Carlo* (QMC), although this is a misnomer because there is nothing random about them. The numbers are not independent but rather are constructed as an ordered sequence of points.

FIGURE 12-4

Comparison of distributions.



To illustrate, Figure 12-4 compares a distribution for two variables only (after all, this is the number of dimensions of a page). The figure shows, on the left, pseudorandom points and, on the right, a deterministic, *low-discrepancy* sequence obtained from a so-called Sobol procedure.[10]

The left graph shows that the points often "clump" in some regions and leave out large areas. These clumps are a waste because they do not contribute more information. The right panel, in contrast, has more uniform coverage. Instead of drawing independent samples, the deterministic scheme systematically fills the space left by the previous numbers in the series.

Quasirandom methods have the desirable property that the standard error shrinks at a faster rate, proportional to close to $1/K$ rather than $1/\sqrt{K}$ for standard simulations. Indeed, a number of authors have shown that deterministic methods provide a noticeable improvement in speed.[11] Papageorgiou and Paskov (1999) compare the computation of VAR for a portfolio exposed to 34 risk factors using 1000 points. They find that the deterministic sequence can be 10 times more accurate than the Monte Carlo method.

One drawback of these methods is that since the draws are not independent, accuracy cannot be assessed easily. For the Monte Carlo method, in contrast, we can construct confidence bands around the estimates. Another issue is that for high-dimensionality problems, some QMC sequences tend to cycle, which leads to decreases in performance. Overall, however, suitably selected QMC methods can provide substantial accelerations in the computations.

12.6 CHOOSING THE MODEL

Simulation methods are most prone to model risk. If the stochastic process chosen for the price is unrealistic, so will be the estimate of VAR. This is why the choice of the underlying process is particularly important.

For example, the geometric brownian motion model in Equation (12.1) adequately describes the behavior of some financial variables, such as stock prices and exchange rates, but certainly not that of fixed-income securities. In the brownian motion models, shocks to the price are never reversed, and prices move as a random walk. This cannot represent the price process for default-free bond prices, which must converge to their face value at expiration.

Another approach is to model the dynamics of interest rates as

$$dr_t = \kappa(\theta - r_t)dt + \sigma r_t^{\gamma} dz_t \tag{12.10}$$

This class of model includes the Vasicek (1977) model when $\gamma = 0$; changes in yields then are normally distributed, which is particularly convenient because this leads to many closed-form solutions. With $\gamma = 0.5$, this is also the Cox, Ingersoll, and Ross (1985) model of the term structure (CIR). With $\gamma = 1$, the model is lognormal.[12]

This process is important because it provides a simple description of the stochastic nature of interest rates that is consistent with the empirical observation that interest rates tend to be mean-reverting. Here, the parameter $k < 1$ defines the speed of mean reversion toward the long-run value θ. Situations where current interest rates are high, such as $r_t > \theta$, imply a negative drift $k(\theta - r_t)$ until rates revert to θ. Conversely, low current rates are associated with positive expected drift. Also note that with $\gamma = 0.5$, the variance of this process is proportional to the level of interest rates; as the interest rate moves toward 0, the variance decreases, so r can never fall below 0. If the horizon is short, however, the trend or mean reversion term will not be important.

Equation (12.10) describes a one-factor model of interest rates that is driven by movements in short-term rates dr_t. In this model, movements in longer-term rates are perfectly correlated with movements in this short-term rate through dz. Therefore, the correlation matrix of zero-coupon bonds consists of ones only. This may be useful to describe the risks of some simple portfolios but certainly not for the leveraged fixed-income portfolios of financial institutions.

For more precision, additional factors can be added. Brennan and Schwartz (1979), for example, proposed a two-factor model with a short and long interest rate modeled as where l is the long rate and the errors have some correlation. Generalizing, one could use the full covariance matrix along some 14 points on the yield curve, as provided by RiskMetrics. In theory, *finer granularity* should result in better risk measures, albeit at the expense of computational time. In all these cases, the Monte Carlo experiment consists of first simulating movements in the driving risk factors and then using the simulated term structure to price the securities at the target date.

$$dr_t = \kappa_1(\theta_1 - r_t)dt + \sigma_1 dz_{1t} \tag{12.11}$$

$$dl_t = \kappa_2(\theta_2 - l_t)dt + \sigma_2 dz_{2t} \tag{12.12}$$

Here is where risk management differs from valuation methods. For short horizons (say, 1 day to 1 month), we could assume that changes in yields are jointly normally distributed. This assumption may be quite sufficient for risk management purposes. Admittedly, it would produce inconsistencies over long horizons (say, beyond a year) because yields could drift in different directions, creating term structures that look unrealistic.[13]

With longer horizons, the drift term in Equation (12.11), for example, becomes increasingly important. To ensure that the two rates cannot move too far away from each other, one could incorporate into the drift of the short rate an *error-correction term* that pushes the short rate down when it is higher than the long rate. For instance, one could set

$$dr_t = \kappa_1[\theta_1 - (r_t - l_t)]dt + \sigma_1 dz_{1t} \tag{12.13}$$

Indeed, much work has been devoted to the analysis of time series that are *cointegrated*, that is, that share a common random component.[14] These error-correction mechanisms can be applied to larger-scale problems, thus making sure that our 14 yields move in a realistic fashion.

But again, over short horizons, the modeling of expected returns is not too important. This also applies to the choice of term structure models, equilibrium models versus arbitrage models. *Equilibrium models* postulate a stochastic process for some risk factors, which generates a term structure. This term structure, however, will not fit exactly the current term structure, which is not satisfactory for fixed-income option traders. They argue that if the model does not even fit current bond prices, it cannot possibly be useful to describe options. This is why *arbitrage models* take today's term structure as an input (instead of output for the equilibrium models) and fit the stochastic process accordingly.

For instance, a one-factor no-arbitrage model is where the function $\theta(t)$ is chosen so that today's bond prices fit the current term structure. This approach has been extended to two-factor Heath-Jarrow-Morton (1992) models, but their estimation is computer-intensive and has been described "at the very boundaries of feasibility."[15] These *arbitrage* models are less useful for risk management, however.

$$dr_t = \theta(t)dt + \sigma dz_t \tag{12.14}$$

For risk management purposes, what matters is to capture the richness in movements in the term structure, not necessarily to price today's instruments to the last decimal point. Thus the "art" of risk management lies in deciding what elements of the model are important.

12.7 CONCLUSIONS

Simulation methods are now used widely for risk management purposes.

Interestingly, these methods can be traced back to the valuation of complex options, except that there is no discounting or risk-neutrality assumption. Thus the investment in intellectual and systems development for derivatives trading can be used readily for computing VAR. No doubt this is why officials at the Fed have stated that derivatives "have had favorable spillover effects on institutions' abilities to manage their total portfolios."

TABLE 12-4

Comparison of VAR Methods

	Valuation Method	
Risk-Factor Distribution	**Local Valuation**	**Full Valuation**
Variance-covariance	Delta-normal	
Historical simulation		Historical path
Deterministic simulation		Full simulation
Monte Carlo simulation	Delta-gamma-MC	Full Monte Carlo

Simulation methods are quite flexible. They can either postulate a stochastic process or resample from historical data. They allow full valuation on the target date. On the downside, they are more prone to model risk owing to the need to prespecify the distribution and are much slower and less transparent than analytical methods. In addition, simulation methods create sampling variation in the measurement of VAR. Greater precision comes at the expense of vastly increasing the number of replications, which slows the process down.

VAR methods are listed in Table 12-4 in order of increasing time requirement. At one extreme is the Monte Carlo method, which requires the most computing time. For the same accuracy, deterministic simulations are faster because they create more systematic coverage of the risk factors. Next is the historical simulation method, which uses recent history in a limited number of simulations. At other extreme is the delta-normal method, which requires no simulation and is very fast. With the ever-decreasing cost of computing power and advances in scientific methods, however, we should expect greater use of simulation methods.

QUESTIONS

1. What is the main assumption for the risk factors underlying the Monte Carlo simulation method?

2. What is the main assumption for the risk factors underlying the historical simulation method?
3. Explain why numerical integration is plagued by the curse of dimensionality and why this is avoided by the Monte Carlo simulation method.
4. Define K as the number of Monte Carlo replications. At what rate does the standard error of estimates decrease?
5. What are the major drawbacks of the Monte Carlo simulation method?
6. Consider an operating system that has a random-number generator with a short cycle, that is, that repeats the same sequence of numbers after a few thousand iterations. Will this lead to inaccuracy in the calculation of VAR? Why?
7. Explain how the inverse transform method could generate draws from a student t distribution.
8. What is an advantage of the bootstrap approach compared with a Monte Carlo simulation based on the normal distribution?
9. If the movements in the risk factors have positive autocorrelation from one day to the next, can we bootstrap on the changes in the risk factors?
10. To compute VAR using a simulation method, which two statistics are required?
11. Can Monte Carlo simulation be adapted to changing volatility?
12. Explain why pricing methods use risk-neutral distributions. Does risk measurement need risk-neutral or physical distributions?
13. A Monte Carlo simulation creates a 99 percent VAR estimate of $10 million with a standard error of $4 million using 1000 replications. How many replications are needed to shrink this standard error to less than $1 million?
14. The relative error in the previous question was 4/10. Would you expect this ratio to be higher or lower for a 95 percent VAR?
15. How many years of daily data do we need to estimate a 99 percent VAR with a precision of 1 percent or better? Would we need more/fewer years for distributions with negative skewness, and why?
16. Explain how stratified sampling could generate more precise estimates

of VAR.

17. A risk manager needs to generate two variables with a correlation of 0.6. Explain how this could be done starting from independent variables. Verify that the final variables have unit volatility.
18. Assume now that the correlation between the two variables is 1. What does this imply in terms of independent risk factors for the simulation?
19. Are sequences of variables in the quasi–Monte Carlo methods independent?
20. Is the geometric brownian motion model a good description of the behavior of fixed-income securities?
21. Explain how equilibrium and no-arbitrage models use the current term structure.

CHAPTER 13
Liquidity Risk

> LTCM then faced severe market liquidity problems when its investments began losing value and the fund attempted to unwind some of its positions.
>
> —*President's Working Group on Financial Markets, 1999*

Traditional value-at-risk (VAR) models assume that the portfolio is "frozen" over the horizon and that market prices represent achievable transaction prices. This marking-to-market approach is adequate to quantify and control risk for an ongoing portfolio but may be more questionable if VAR is supposed to represent the worst loss over a liquidation period.

The question is how VAR can be adapted to deal with liquidity considerations. As we saw in Chapter 1, liquidity risk can be grouped into *funding liquidity risk* and *asset liquidity risk*. Funding liquidity risk arises when financing cannot be maintained owing to creditor or investor demands. The resulting need for cash may require selling assets. Asset liquidity risk arises when a forced liquidation of assets creates unfavorable price movements. Thus liquidity considerations should be viewed in the context of both the assets and the liabilities of the financial institution.

This chapter discusses recent developments that adapt traditional VAR measures to liquidity considerations. Section 13.1 first provides a general introduction to asset and funding liquidity risk.

Next, Section 13.2 attempts to incorporate asset liquidity risk into VAR measures. Immediate liquidation can create losses owing to market impact, which is a drop in the liquidation value relative to mark-to-market prices. Liquidation, however, can take place over many days and should be done so as to balance transactions costs and price risk. Taking both costs and risk into account leads to a measure of liquidity-adjusted VAR (LVAR). Often, however, liquidity is factored into the valuation of positions by decreasing their value by a reserve.

Section 13.3 then discusses measures of funding liquidity risk proposed by the Counterparty Risk Management Policy Group (CRMPG). Even though an institution can have zero traditional VAR, different swap credit terms can generate different cash requirements. The cash liquidity measure is an extension of VAR.

Next, Section 13.4 is devoted to an analysis of the Long-Term Capital Management (LTCM) debacle. LTCM failed because of its lack of diversification combined with its asset and funding liquidity risk, which were due to its sheer size.

Finally, Section 13.5 provides some concluding comments about liquidity risk. Liquidity problems have proved to be crucial in the failure of many financial institutions. Liquidity risk can be factored formally into hybrid VAR measures but only using price impact functions derived from normal market conditions. During episodes of systemic risk, however, liquidity evaporates, invalidating much of this analysis. Thus liquidity risk probably is the weakest spot of market risk management systems.

13.1 DEFINING LIQUIDITY RISK

Table 13-1 displays sources of liquidity risk for a financial institution. Liquidity risk emanates from the liability side, when creditors or investors demand their money back. This usually happens after the institution has incurred or is thought to have incurred losses that could threaten its solvency. The need for cash creates problems on the asset side when the forced liquidation of assets causes transactions losses.

Understanding liquidity risk requires knowledge of several different fields, including *market microstructure*, which is the study of market-clearing mechanisms; *optimal trade execution*, which is the design of strategies to minimize trading costs or to meet some other objective function; and *asset-liability management*, which attempts to match the values of assets and liabilities on balance sheets.

TABLE 13-1
Sources of Liquidity Risk

Assets	**Liabilities**
Size of position	Funding
Price impact for unit trade	Mark to market, haircuts
	Equity
	Investor redemptions

13.1.1 Asset Liquidity Risk

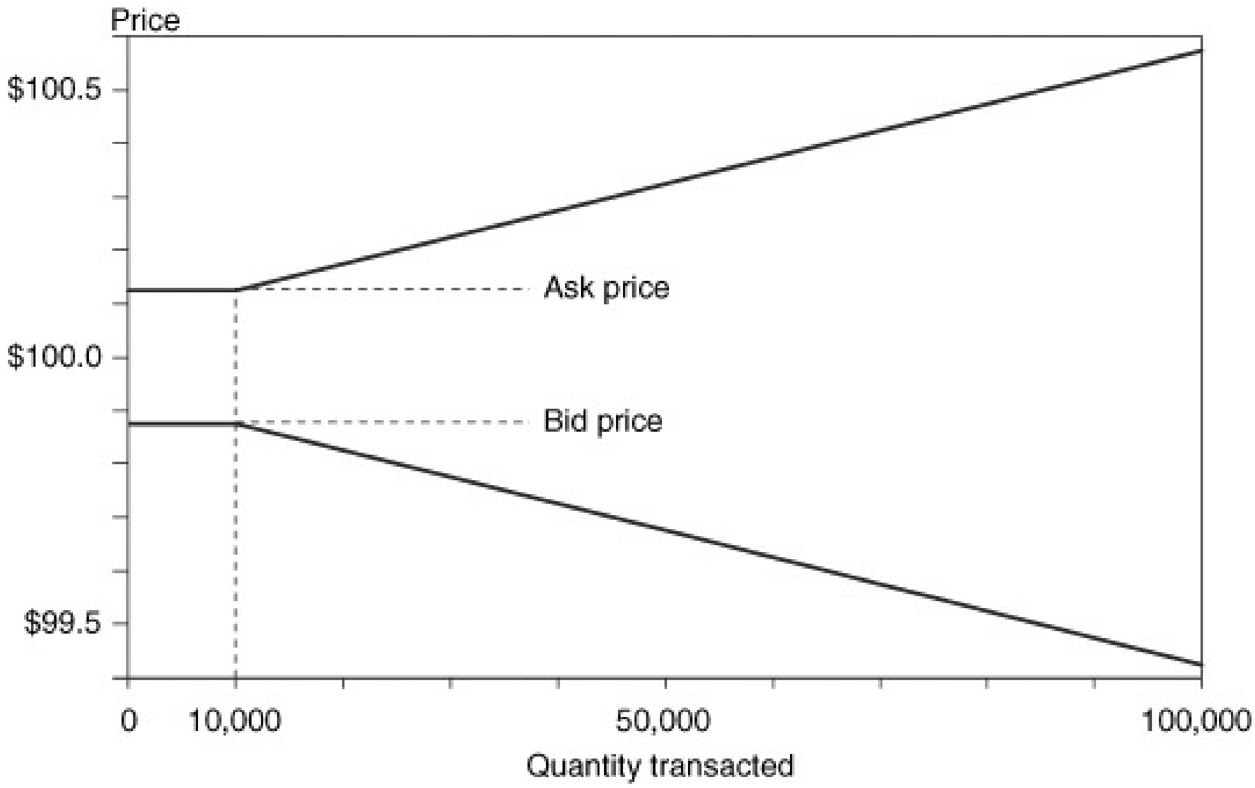
Asset liquidity risk, sometimes called *market/product liquidity risk*, is the risk that the liquidation value of assets may differ significantly from their current mark-to-market values. It is a function of the price impact of trades and the size of the positions.

Asset liquidity can be measured by a price-quantity function. This is also known as the *market-impact* effect. Highly liquid assets, such as major currencies or Treasury bonds, are characterized by *deep markets*, where positions can be offset with very little price impact. *Thin* markets, such as exotic over-the-counter (OTC) derivatives contracts or some emerging market equities, are those where any transaction can quickly affect prices.

This price function is illustrated in Figure 13-1.[1] The starting point is the current *midprice*, which is the average of the bid and ask quotes and can be used to mark the portfolio to market. Here, the *bid-ask spread* is $0.25. Markets with low spreads are said to exhibit *tightness*. In Figure 13-1, this is valid up to some limit, say, 10,000 shares. This is sometimes called the *normal market size*, or *depth*. Relative to midmarket values, the cost of trading is half the spread. This component of trading cost is sometimes called *exogenous* because it does not depend on quantities transacted, as long as these quantities are below the normal market size.

For quantities beyond this point, however, the sale price is a decreasing function of the quantity, reflecting the price pressure required to clear the market. The converse is true for the purchase price. In practice, the position is compared with some metric such as the median daily trading volume. For a widely traded stock such as IBM, for instance, selling 4 percent of the daily trading volume incurs a cost of about 60 basis points. Studies of market microstructure provide empirical evidence on trading costs.

FIGURE 13-1
Price-quantity function.



This relationship is assumed to be linear, although it could take another shape. The slope of the line measures the *market impact*. This varies across assets and, possibly, across time for a given asset. In this example, selling 100,000 shares over 1 day would push the price down from a midmarket value of $100 to about $99.4. Thus, selling this position would incur a liquidation cost of ($100 – $99.4)/$100, or 60 basis points.

This demonstrates that liquidity depends on both the price-impact function and the size of the position. In this example, if the position is below 10,000 shares, then market liquidity is not a major issue. In contrast, if the institution holds a number of shares worth several days of normal trading volume, liquidity should be of primary concern.

In addition to varying across assets, liquidity is also a function of prevailing market conditions. This is more worrying because markets seem to go through regular bouts of liquidity crises. Most notably, liquidity in bond markets dried up during the summer of 1998 as uncertainty about defaults led to a “flight to quality,” that is, increases in the prices of Treasuries relative to those of other bonds. A similar experience occurred during the 1994 bond market debacle, at which time it became quite difficult to deal in Eurobonds or mortgage-backed securities.

Traditionally, asset liquidity risk has been controlled through position limits. The goal of *position limits* is to limit the exposure to a single instrument, even if it provides diversification of market risk, in order to avoid a large market impact in case of forced liquidation.

13.1.2 Funding Liquidity Risk

Cash-flow/funding liquidity risk refers to the inability to meet payment obligations to creditors or investors. This can force unwanted liquidation of the portfolio.

Funding risk arises from the liability side of the balance sheet. Most financial institutions are *leveraged*. Often, this involves posting some collateral (assets) in exchange for cash from a broker. Normally, brokers require collateral that is worth slightly more than the cash loaned, by an amount known as a *haircut*, designed to provide a buffer against decreases in the collateral value. The value of the collateral, however, is constantly *marked to market* by the broker. If this value falls, the broker will require some additional payment, called *variation margin*, to keep the total amount held above the loan value. If the institution does not have enough cash on hand, it will be forced to liquidate some of its other assets.

Brokers also reserve the right of *changes in collateral requirements*, which can create additional cash flow risk. For example, brokers can increase the haircut when markets are more volatile, creating extra demands on cash. Similarly, organized exchanges can change their required margins at will.

Finally, cash-flow liquidity risk also arises owing to *mismatches in the timing of payments*. Even if an institution is perfectly matched in terms of market risk, it may be forced to make a payment on a position without having yet received an offsetting payment on a hedge. Section 13.3 gives examples of such mismatches.

The first line of defense against funding liquidity risk is *cash*. Another may be a *line of credit*, which is a loan arrangement with a bank allowing the customer to borrow up to a prespecified amount.

The institution may be able to meet margin calls by raising funds from another source, such as new debt or a new equity issue. In practice, it may be difficult to raise new funds precisely when the institution is faring badly and needing them most.

Conversely, the institution must evaluate the likelihood of redemptions, or cash requests from debt holders or equity holders. This is most likely to occur when the institution appears most vulnerable, thereby transforming what could be a minor problem into a crisis. It is also important to avoid debt covenants or options that contain “triggers” that would force early redemption of the borrowed funds. Such credit triggers accelerated the fall of Enron, as shown in Box 13-1.

BOX 13-1

ENRON’S CREDIT TRIGGERS: THE BAD AND STUPID

Credit triggers are clauses in financial contracts that allow creditors to demand immediate payments if the credit rating of the borrower falls below some predetermined level.

Enron is now widely viewed as a massive case of accounting fraud. Credit triggers, however, played a role in Enron’s demise. Enron was rated investment-grade until November 28, 2001, when a proposed takeover by Dynergy fell through. On that day, Standard & Poor’s downgraded Enron to speculative-grade, triggering the immediate repayment of almost $4 billion in debt. Unable to pay, Enron filed for bankruptcy on December 2, 2001.

The real cause of Enron’s failure was its poor performance in many business lines, which was hidden through creative off-balance-sheet financing. As early as 1999, Vince Kaminsky, Enron’s risk manager, had railed against these arrangements, which he said had gone from merely “stupid” to fraudulent. His comments, unfortunately, were ignored by top management.

Ostensibly, credit triggers are designed to lower the cost of capital for the issuing company. Because this is an option granted to debt holders, they should be willing to accept a lower interest rate than otherwise. Superficially, such clauses look beneficial because lenders can *put* the obligation back to the borrower.

In practice, however, there have been many cases where credit triggers offered no protection to creditors because they precipitated a default. In such situations, from the viewpoint of borrowers, the cost savings have been swamped by the problems caused by credit triggers.

Credit-rating agencies call these credit triggers “problematic” and “troubling.” As a result, they now examine much more closely the potential effects of credit triggers and take them into account when setting ratings.*

* See Moody's (2001).

Thus liquidity considerations should be viewed in the context of both asset and liabilities. Consider, for instance, *hedge funds*, some of which invest in illiquid assets such as distressed debt. To minimize liquidity risk, such funds impose a longer *lockup period*, or minimum time for investors to keep their funds, and a longer *redemption notice period* for withdrawing funds.

As explained in Chapter 3, *commercial banks* are by their nature susceptible to liquidity risks. They are funded by short-term deposits but can invest in illiquid real estate loans. This setup is fraught with liquidity risk and explains the rationale for deposit insurance, which eliminates the incentives for bank runs.

13.2 ASSESSING ASSET LIQUIDITY RISK

Trading returns are measured typically from midmarket prices. This may be adequate for measuring daily profit and loss (P&L) but may not represent the actual fall in value if a large portfolio were to be liquidated. The question is how to assess potential losses under such conditions.[2] In turn, this can give insights into how to manage this risk.

Traditional adjustments are done on an ad hoc basis. Liquidity risk can be loosely factored into VAR measures by ensuring that the *horizon*is at least greater than an orderly liquidation period. Generally, the same horizon is applied to all asset classes, even though some may be more liquid than others.

Sometimes, longer liquidation periods for some assets are taken into account by artificially increasing the volatility. For instance, one could mix a large position in the dollar/yen with another one in the dollar/Polish zloty, both of which have an annual volatility of 10 percent, by artificially increasing the volatility of the second foreign currency in the VAR computations.

13.2.1 Effect of Bid-Ask Spreads

More formally, one can focus on the various components of liquidation costs. The first and most easily measurable is the quoted bid-ask spread, defined in relative terms, that is,

$$S = \frac{[P(\text{ask}) - P(\text{bid})]}{P(\text{mid})} \tag{13.1}$$

Table 13-2 provides typical spreads. We see that spreads vary from a low of about 0.05 percent for major currencies, large U.S. stocks, and on-therun Treasuries to much higher values when dealing with less liquid currencies, stocks, and bonds. Treasury bills are in a class of their own, with extremely low spreads. These spreads are indicative only because they depend on market conditions. Also, market makers may be willing to trade within the spread.

At this point, it is useful to review briefly the drivers of these spreads. According to market microstructure theory, spreads reflect three different types of costs:

- *Order-processing costs* cover the cost of providing liquidity services and reflect the cost of trading, the volume of transaction, the state of technology, and competition. With fixed operating costs, these order-processing costs should decrease with transaction volumes.

TABLE 13-2

Typical Spreads and Volatility

Asset	Spread (%) (Bid-Ask)	Volatility (%) Daily	Volatility (%) Annual
Currencies			
Major (euro, yen, . . .)	0.02–0.10	0.3–1.0	5–15
Emerging (floating)	0.10–1.00	0.3–1.9	5–30
Bonds			
On-the-run Treasuries	0.03	0.0–0.7	0–11
Off-the-run Treasuries	0.06–0.20	0.0–0.7	0–11
Corporates	0.10–1.00	0.0–0.7	0–11
Treasury bills	0.003–0.02	0.0–0.1	0–1
Stocks			
U.S.	0.05–5.00	1.3–3.8	20–60
Average, NYSE	0.20	1.0	15
Average, all countries	0.40	1.0–1.9	15–30

Note: Author's calculations. Cost of trades excludes broker commissions and fees. See also *Institutional Investor* (November 1999).

- *Asymmetric-information costs* reflect the fact that some orders may come from informed traders, at the expense of market makers who can somewhat protect themselves by increasing the spread.
- *Inventory carrying costs* are due to the cost of maintaining open positions, which increase with higher price volatility, higher interest-rate carrying costs, and lower trading activity or turnover.

If the spread were fixed, one simply could construct a liquidity-adjusted VAR from the traditional VAR by adding a term, that is, where *W* is the initial wealth, or portfolio value. For instance, if we have $1 million invested in a typical stock with a daily volatility of σ = 1 percent and spread of *S* = 0.20 percent, the 1-day LVAR at the 95 percent confidence level would be

$$\mathrm{LVAR} = \mathrm{VAR} + L_I = (W\alpha\sigma) + 1/2(WS) \qquad (13.2)$$

($1,000,000 X 1.645 X 0.01) + 1/2 ($1,000,000 X 0.0020) = $16,450 + $1000 = $17,450

Here, the correction factor is relatively small, accounting for 7 percent of the total.

This adjustment can be repeated for all assets in the portfolio, leading to a series of add-ons, $1/2\Sigma_i|W_i|S_i$. This sequence of positive terms increases linearly with the number of assets, whereas the usual VAR benefits from diversification effects. Thus the relative importance of the correction factor will be greater for large portfolios.

A slightly more general approach is proposed by Bangia *et al.* (1999), who consider the uncertainty in the spread. They characterize the distribution by its mean $\bar{S}$ and standard deviation σ_S. The adjustment considers the worst increase in the spread at some confidence level, that is,

$$\text{LVAR} = \text{VAR} + L_2 = (W\alpha\sigma) + \frac{1}{2}[W(\bar{S} + \alpha'\sigma_s)] \qquad (13.3)$$

This assumes that the worst market loss and increase in spread will occur simultaneously. In general, we observe a positive correlation between volatility and spreads.

At the portfolio level, one theoretically could take into account correlations between spreads. In practice, summing the individual worst spreads provides a conservative measure of the portfolio worst spread.

Typically, σ_s is about half the size of the average spread; for example, σ_s = 0.02 percent against $\bar{S}$ = 0.04 percent for the dollar/euro exchange rate. Relative to a volatility of about 1.0 percent per day, these adjustments are small. Thus transactions costs based on spreads are not very important relative to usual VAR measures.

13.2.2 Incorporating Liquidity in Valuation

If the position is to be sold, the second term in Equation (13.2) represents a certain loss, unlike the volatility term. Assuming that the portfolio is valued using midmarket prices, it represents the loss owing to the liquidation.

Another approach to liquidity is to mark the portfolio to the appropriate bid prices (for long positions) or ask prices (for short positions). In practice, financial institutions generally mark their cash positions to the conservative bid-offer basis.[3] VAR then can be viewed as the worst loss from this value.

Further, financial institutions often apply *reserves*, which are pricing changes in the valuation away from fair value to account for such effects as illiquidity and model risk. Firms deduct this reserve from the fair value of positions to account for the extra time and cost required to close out the position. The reserve amount is often based on judgments about the liquidity of a market. In such cases, there is no need to take liquidity risk into account in VAR because it is already factored into the valuation of positions.

13.2.3 Effect of Price Impact

Although this approach has the merit of considering some transactions costs, it is not totally satisfactory. It only looks at the bid-ask spread component of these costs, which may be appropriate for a small portfolio but certainly not when liquidation can affect market prices. The market-impact factor should be taken into account.

To simplify, let us assume a linear price-quantity function and ignore the spread. For a sale, the price per share is

$$P(q) = P_0(1 - kq) \tag{13.4}$$

Assume that P_0 = \$100 and $k = 0.5 \text{ X } 10^{-7}$. Say that we start with a position of q = 1 million shares of the stock. If we liquidate all at once, the price drop will be P_0kq = \$100 X (0.5 X 10^{-7}) X 1,000,000 = \$5 per share, leading to a total price impact of \$5 million. In contrast, we could decide to work the order through at a constant rate of 200,000 shares over n = 5 days. In the absence of other price movements, the daily price drop will be \$1 per share, leading to a total price impact of \$1 million, much less than before.

Immediate liquidation creates the costs:

$$C_1(W) = q \times [P_0 - P(q)] = q \times (P_0 - P_0 + P_0 kq) = kq^2 P_0 \tag{13.5}$$

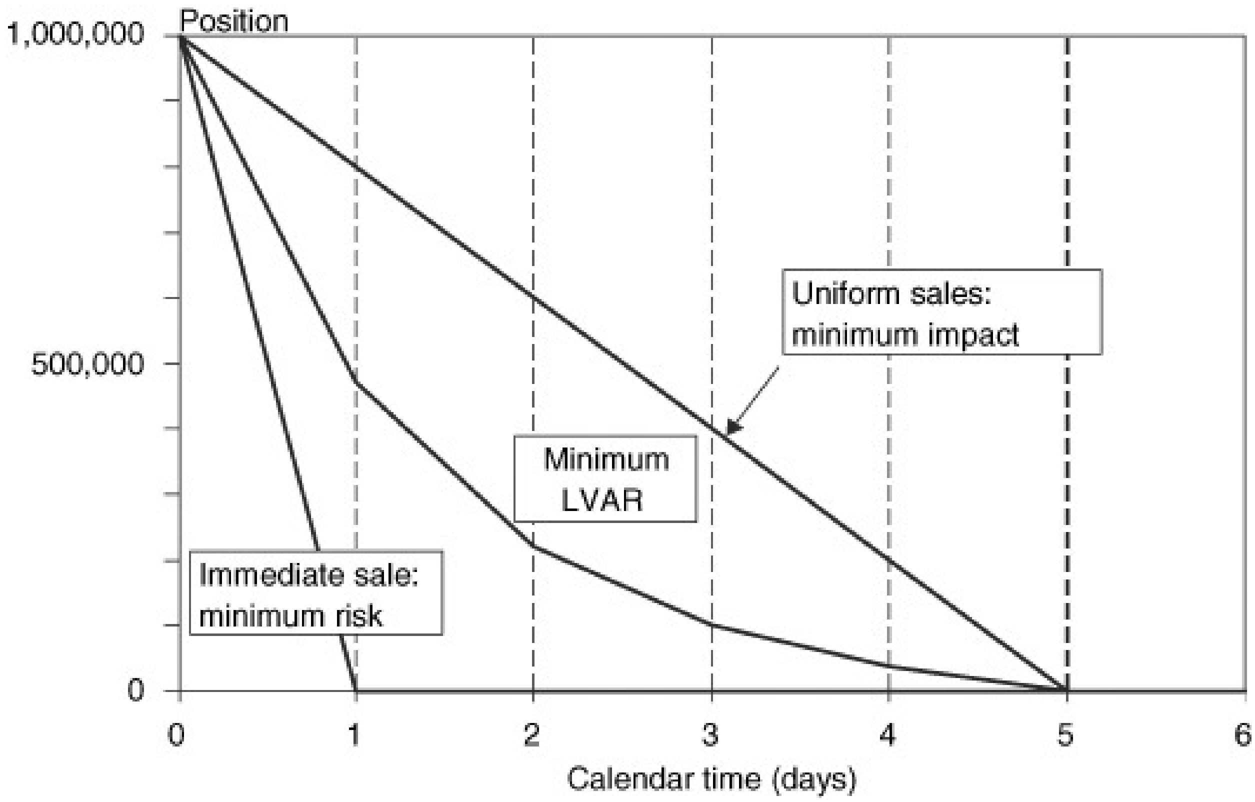
Uniform liquidation creates the costs:

$$C_2(W) = q \times [P_0 - P(q/n)] = q \times (P_0 - P_0 + P_0 kq/n) = k(q^2/n) P_0 \tag{13.6}$$

Because uniform liquidation spreads the price impact over many days, it leads to lower trading costs.

The drawback of liquidating more slowly, however, is that the portfolio remains exposed to price risks over a longer period. The position profiles are compared in Figure 13-2. Under the immediate sale, the position is liquidated before the end of the next day, leading to a high cost but minimum risk. Under the uniform sale, the position is sold off in equalsized lots, leading to low costs but higher volatility.

FIGURE 13-2
Profile of execution strategies.



To analyze the risk profile of these strategies, define σ as the daily volatility of the share price, in dollars. We assume that sales are executed at the close of the business day in one block. Hence, for the immediate sale, the price risk, or variance of wealth, is zero, that is, $V_1(W) = 0$.

For the uniform sale, assume that returns are independent over each day so that the total variance is the sum of the daily variances. The relative positions are defined as $x_0, x_1, \ldots, x_n$. At the end of the first day, the position will have decreased from $x_0 = 1$ to $x_1 = [1 - (1/n)]$. The next day, it will have gone to $[1 - 2(1/n)]$ and so on. With uncorrelated daily returns, the total portfolio variance over n days is the sum of the variance over each day, that is,

$$V_2(W) = \sigma^2 q^2 \left\{ \left(1 - \frac{1}{n}\right)^2 + \left(1 - 2\frac{1}{n}\right)^2 + \cdots + \left[1 - (n-1)\frac{1}{n}\right]^2 \right\} P_0^2 \qquad (13.7)$$

This can be simplified to

$$V_2(W) = \sigma^2 q^2 \left[n\frac{1}{3}\left(1 - \frac{1}{n}\right)\left(1 - \frac{1}{2n}\right) \right] \mathrm{P}_0^2 = \sigma^2 q^2 T * P_0^2 \qquad (13.8)$$

For example, with $n = 5$, the correction factor between brackets is $T^* = 1.20$. Thus the risk of a strategy of uniform liquidation over 5 days is equivalent to the mark-to-market risk of a position held over 1.2 days. It is interesting to note that the 10-day fixed horizon dictated by the Basel Committee is equivalent to a constant liquidation over 31 trading days.

Adding transactions costs leads to a liquidity-adjusted VAR, defined as where α corresponds to the confidence level c. Lawrence and Robinson (1997), for instance, propose choosing n to minimize LVAR.

$$\mathrm{LVAR} = \alpha\sqrt{V(W)} + C(W) \tag{13.9}$$

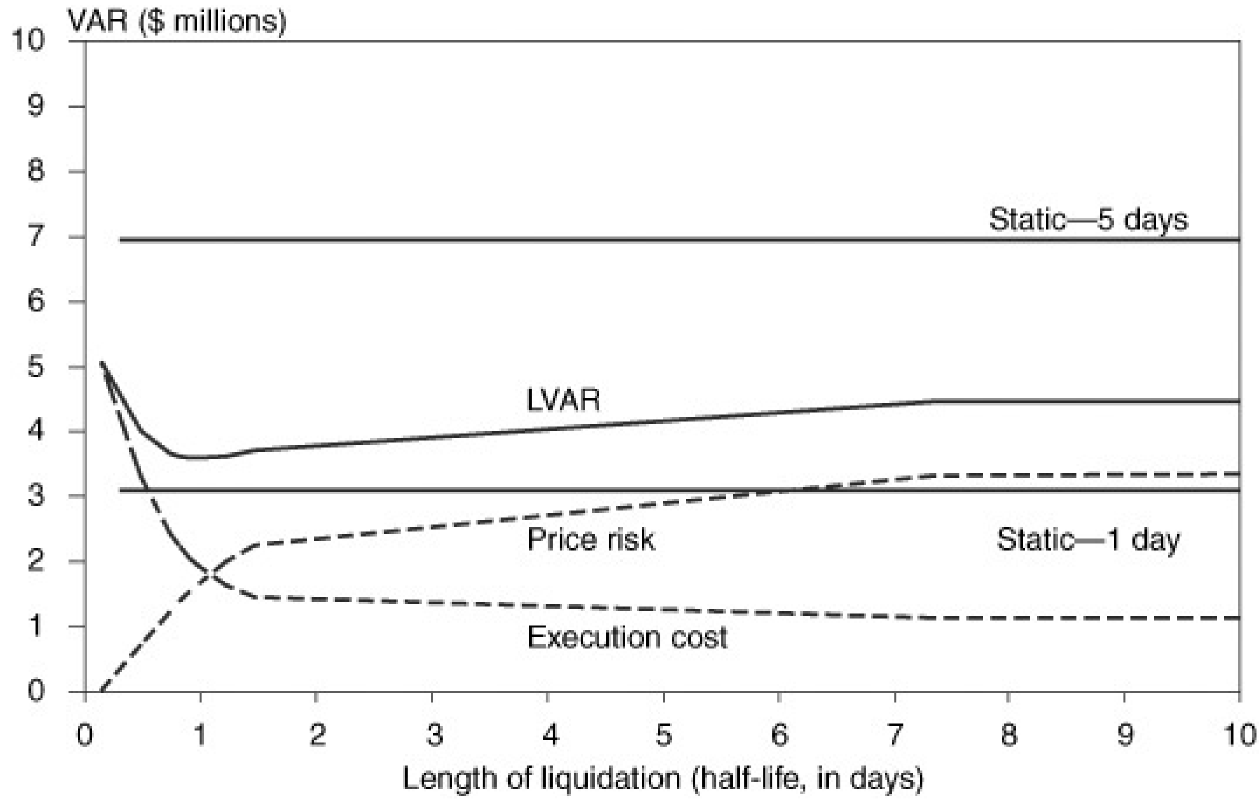
13.2.4 Trading Strategies

Execution strategies need not be limited to these two extreme cases—immediate or uniform liquidation. More generally, we can choose a strategy, or pattern of daily positions that leads to an optimal tradeoff between execution costs and price risk. Almgren and Chriss (2001) provide useful closed-form solutions for efficient execution strategies. Their paper is an important contribution that helped lay the groundwork for *algorithmic trading* on Wall Street.[4]

An optimal trajectory is described in Figure 13-2. This is defined by a set of daily positions $x_0, x_1, x_2, \ldots, x_n$. On the first day, the optimal position drops by more than the uniform sale: $x_0 - x_1 > 1/n$. Intuitively, this is so because it helps lowering price risk over the total horizon. Note that the strategy can be described by its *half-life*, which is the time required to liquidate half the portfolio. In this case, this takes 1 day.

Figure 13-3 compares various VAR measures for different speeds of execution. The "static" 1-and 5-day VARs correspond to the usual mark-to-market VAR measures with 30 percent annual volatility at the

FIGURE 13-3
Liquidity-adjusted VAR.



95 percent level of confidence. Under these conditions, the daily volatility is 1.9 percent, and the 1-day VAR is 1.645 X 0.019 X $100 = $3.1 million for this $100 million portfolio, assuming a normal distribution. Under liquidation, however, we have to account for market impact.

The LVAR measure incorporates the total execution-cost and price-risk components in a consistent fashion. As we extend the length of liquidation, the execution-cost component decreases, but the price-risk component increases. Here, the total LVAR is minimized at a half-life of 1 day. In this case, a 5-day static VAR would provide a conservative measure of liquidation VAR.

The real benefit of this approach is that it draws attention to market-impact effects in portfolio liquidation. It also illustrates that execution strategies should pay close attention to execution costs and price volatility.

Other strategies can be used for liquidation. In the case of stock portfolios, for instance, the portfolio manager could cut the price risk by immediately putting in place a hedge with stock-index futures. In this case, the remaining price risk is “specific” to the security. Orders to sell then could be transmitted so as to minimize their price impact.

13.2.5 Example

In practice, the computational requirements to adjust the conventional VAR numbers are formidable. The method requires a price-quantity function for all securities in the portfolio. Combined with the portfolio position, this yields an estimate of the price impact of a liquidation.

Table 13-3 provides an example of such an analysis, as provided by Morgan Stanley for a four-country $50 million equity portfolio. The data for Switzerland are expanded at the individual-stock level. To estimate the total impact cost, we need information about the historical bid-ask spreads, the median trading volume, and recent volatility. The portfolio relative size then is defined as the number of shares held as a percentage of median trading volume. The total impact cost then is computed as a function of half the bid-ask spread, the price-impact function, and the size of the position.

Here, the total cost of immediate (1-day) liquidation is estimated to be 21.5 basis points. This can be compared with the daily mark-to-market volatility of this portfolio, which is 110 basis points. Using Equation (13.9), if the portfolio were to be liquidated at the end of the next day, the worst LVAR loss at the 95 confidence percent level would be about $50 X (1.645 X 0.011 + 0.0022) = $0.9 million + $0.1 million. This adds up to $1.0 million, most of which is price risk. The relative importance of liquidity no doubt would be much greater for a larger portfolio.

TABLE 13-3

Market Impact-Cost Report

	Portfolio			Cost Analysis			
Asset	Value (US$)	Shares Held	Price	Spread (bp)	Median Volume	Shares/ Volume	Impact Cost (bp)
France	19,300,182	184,063	104.9	19.9		1.3%	18.2
Germany	19,492,570	322,550	60.4	26.1		2.5%	29.3
U.K.	5,860,371	424,373	13.8	20.2		0.6%	17.6
Switzerland	5,351,851	9,355	572.1	12.5		1.1%	9.5
Novartis	2,369,367	1,630	1,453.6	11.7	123,554	1.3%	8.8
Swatch	64,678	400	161.7	32.9	42,559	0.9%	15.5
Nestle	1,752,009	935	1,873.8	6.4	76,004	1.2%	7.3
CS Group	1,165,797	6,390	182.4	22.2	978,168	0.7%	14.1
Total	50,004,974	940,341	53.2	21.6		1.7%	21.5

Source: Morgan Stanley (1999).

13.3 ASSESSING FUNDING LIQUIDITY RISK

Assessing funding liquidity risk involves examining the asset-liability structure of the institution and potential demands on cash and other sources of liquidity. Some lessons are available from the Counterparty Risk Management Policy Group (1999), which was established in the wake of the LTCM near failure to strengthen practices related to the management of market, counterparty credit, and liquidity risk.[5]

The CRMPG proposes to evaluate funding risk by comparing the amount of cash an institution has at hand with to what it could need to meet payment obligations. It defines *cash liquidity* as the ratio of cash equivalent over the potential decline in the value of positions that may create cash-flow needs.

TABLE 13-4
Computing Funding Liquidity Ratio

	Case 1	Case 2
Assets		
Cash	$5	$5
Liabilities		
Equity	$5	$5
Derivatives		
Long 10-year swap	$100, two-way mark to market	$100, unsecured
Short 10-year swap	$100, two-way mark to market	$100, two-way mark to market
Cash equivalent	$5	$5
Funding VAR	$1.1 (1-day)	$3.5 (10-day)
Ratio	4.5	1.4

Suppose that an institution has two swap positions that identically offset each other with two different counterparties. Thus there is no market risk, and the usual VAR is zero. The swaps are structured with different credit terms, however. Table 13-4 summarizes the positions.

In Case 1, each position is a *two-way mark-to-market* swap, also called *bilateral mark-to-market*. Because the two swaps are both marked to market, any cash payment in one swap must be offset by a receipt on the other leg. The only risk is that of a delay in the receipt, say, over 1 day. Assume that the worst move on a $100 million swap at the 99 percent level over 1 day is $1.1 million. Since this is the worst cash need, the funding ratio is $5/$1.1 = 4.5, which indicates sufficient cash coverage.

In Case 2, one of the positions is an unsecured *one-way mark-to-market* swap. Under this arrangement, the institution is required to make payments if the position loses money; it will not, however, receive intermediate payments if the position gains. Because of this asymmetry, the institution is subject to mismatches in the timing of collateral payments if the first swap loses money. We now need to consider a longer horizon, say, 10 days. This gives a VAR of $3.5 million and a funding ratio of 1.4. This seems barely enough to provide protection against funding risk. Thus some of the elements of traditional VAR can be used to compute funding risk, which can be quite different from market risk when the institution is highly leveraged. Box 13-2 illustrates how credit-rating agencies evaluate liquidity risk.

BOX 13-2

HOW RATING AGENCIES ASSESS LIQUIDITY RISK

Liquidity risk is an important component of the risk of a trading operation. Credit-rating agencies do take this risk into account when assessing the credit risk of an institution with a large trading desk.

Standard & Poor's defines *liquidity risk* as the risk that a trading operation's need for cash collateral may exceed its total liquidity resources. Exposure to collateral calls is evaluated under a stress scenario where the institution is downgraded to a speculative rating. Standard & Poor's then determines whether the institution has sufficient dedicated liquidity resources to cover these collateral calls.

The size of the worst collateral calls is estimated by the sum of all positions that have negative market values. This is so because positions with positive values are not subject to margin calls. For instance, if an institution owes $1 million to each of counterparties A and B but is owed $5 million each by counterparties C and D, it may have to post $2 million in the worst-case scenario. This is so because collateral is not transferable. In other words, even if the institution held $10 million from C and D, these funds could not be used to honor margin calls from A and B. When setting its credit rating, Standard & Poor's estimates the probability that the institution would not be able to post $2 million in the worst-case scenario.

13.4 LESSONS FROM LTCM

The story of Long-Term Capital Management (LTCM) provides a number of lessons in liquidity risk. LTCM was founded by John W. Meriwether in 1994, who left Salomon Brothers after the 1991 bond scandal. Meriwether took with him a group of traders and academics and set up a hedge fund that tried to take advantage of "relative value," or "convergence arbitrage" trades, betting on differences in prices, or spreads, among closely related securities.

13.4.1 LTCM's Leverage

Since such strategies tend to generate tiny profits, leverage has to be used to create attractive returns. By December 1997, the total equity in the fund was $5 billion. LTCM's balance sheet was about $125 billion. This represented an astonishing leverage ratio of 25:1. Even more astonishing was the off-balance-sheet position, including swaps, options, and other derivatives, that added up to a notional amount of $1.25 trillion. This represents the total of *gross positions*, measured as the sum of the absolute value of the trade's notional principal amounts.

To give an idea of the magnitude of these positions, the Bank for International Settlements reported a total swap market of $29 trillion in 1998. Hence LTCM's swap positions accounted for 2.4 percent of the global swap market. Many of these trades, however, were offsetting each other, so this notional amount is practically meaningless. What mattered was the net risk of the fund. LTCM, however, failed to appreciate that these gross positions were so large that attempts to liquidate them would provoke large market moves.

13.4.2 LTCM's "Bulletproofing"

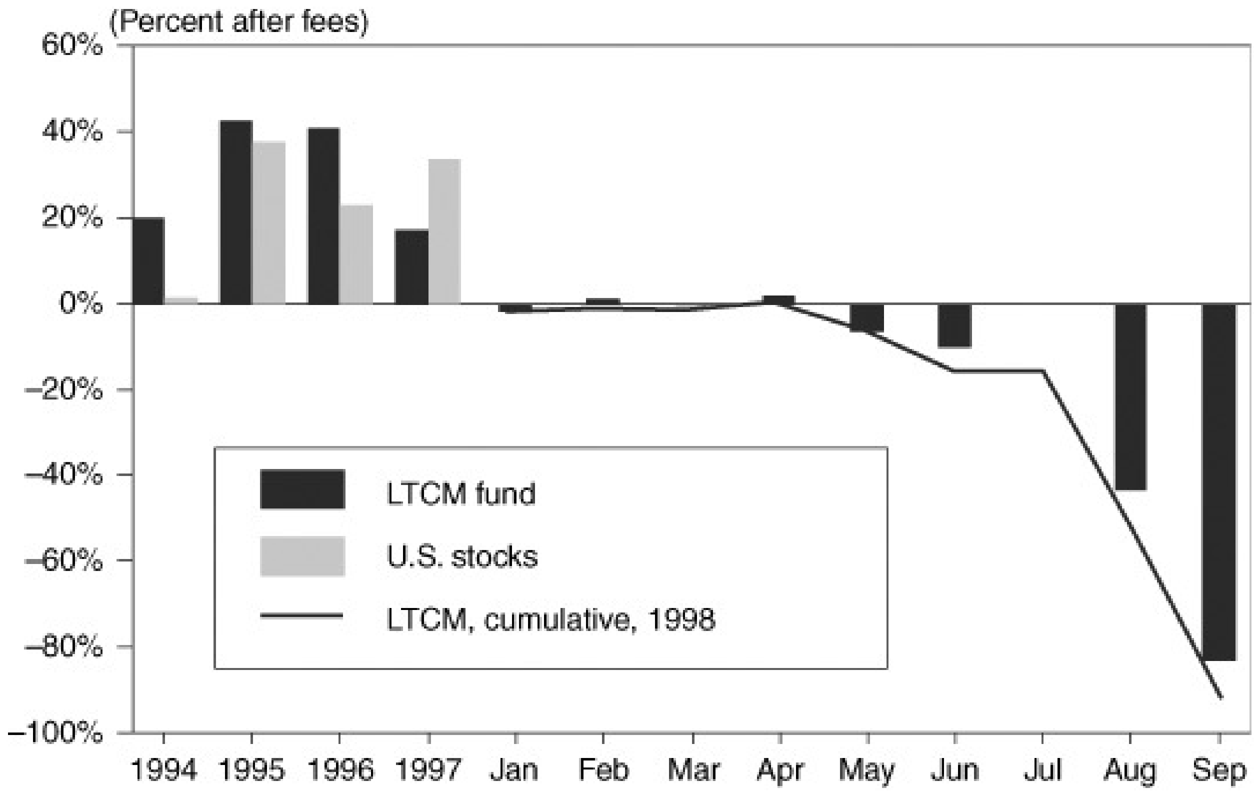
LTCM was able to leverage its balance sheet through sale-repurchase agreements (repos) with commercial and investment banks. Under *repo* agreements, the fund sold some of its assets in exchange for cash and a promise to repurchase them back at a fixed price at some future date. Normally, the value of the assets or collateral exceeds the cash loaned, by an amount known as a *haircut*, which creates a limit to the leverage. LTCM, however, was able to obtain unusually good financing conditions, with next-to-zero haircuts, because it was widely viewed as "safe" by its lenders. In addition, the swaps were subject to two-way marking to market.

On the supply side, LTCM had "bulletproofed" itself against a liquidity squeeze. LTCM initially had required investors to keep their money in the fund for a minimum of 3 years. The purpose of this so-called lockup clause was to avoid forced sales in case of poor performance. LTCM also secured a $900 million credit line from Chase Manhattan and other banks. Even though LTCM had some protection against funding liquidity risk, it was still exposed to market risk and asset-liquidity risk.

13.4.3 LTCM's Downfall

LTCM's strategy profited handsomely from the narrowing of credit spreads during the early years, leading to after-fees returns above 40 percent, as shown in Figure 13-4. Troubles began in May and June of 1998. A downturn in the mortgage-backed securities market led to a 16 percent loss in LTCM's capital. Then came August 17. Russia announced that it was "restructuring" its bond payments—de facto defaulting on its debt. This bombshell led to a reassessment of credit and sovereign risks across all financial markets. Credit spreads, risk premiums, and liquidity spreads jumped up sharply. Stock markets dived. LTCM lost $550 million on August 21 alone.

FIGURE 13-4
LTCM's returns.



By August, the fund had lost 52 percent of its December 31 value. With assets still at $126 billion, the leverage ratio had increased from 28:1 to 55:1. LTCM badly needed new capital. It desperately tried to find new investors, without success.

In September, the portfolio's losses accelerated. Bear Stearns, LTCM's prime broker, faced a large margin call from a losing LTCM T-bond futures position. It then required increased collateral, which depleted the fund's liquid resources.

LTCM now was caught in a squeeze between *funding risk*, as its reserves dwindled, and *asset risk*, as the size of its positions made it impractical to liquidate assets.

A liquidation of the fund would have forced the brokers to sell off tens of billions of dollars of securities and to cover their numerous derivatives trades with LTCM. Because lenders had required next-to-zero haircuts, there was a potential for losses to accrue while the collateral was being liquidated. In credit risk terms, lenders had low current exposure but significant *potential exposure*.

The potential disruption in financial markets was such that the New York Federal Reserve felt compelled to act. On September 23, it organized a bailout of LTCM, encouraging 14 banks to invest \$3.6 billion in return for a 90 percent stake in the firm. These fresh funds came just in time to avoid meltdown. By September 28, the fund's value had dropped to \$400 million only. LTCM investors had lost a whopping 92 percent of their year-to-date investment.

13.4.4 LTCM's Liquidity

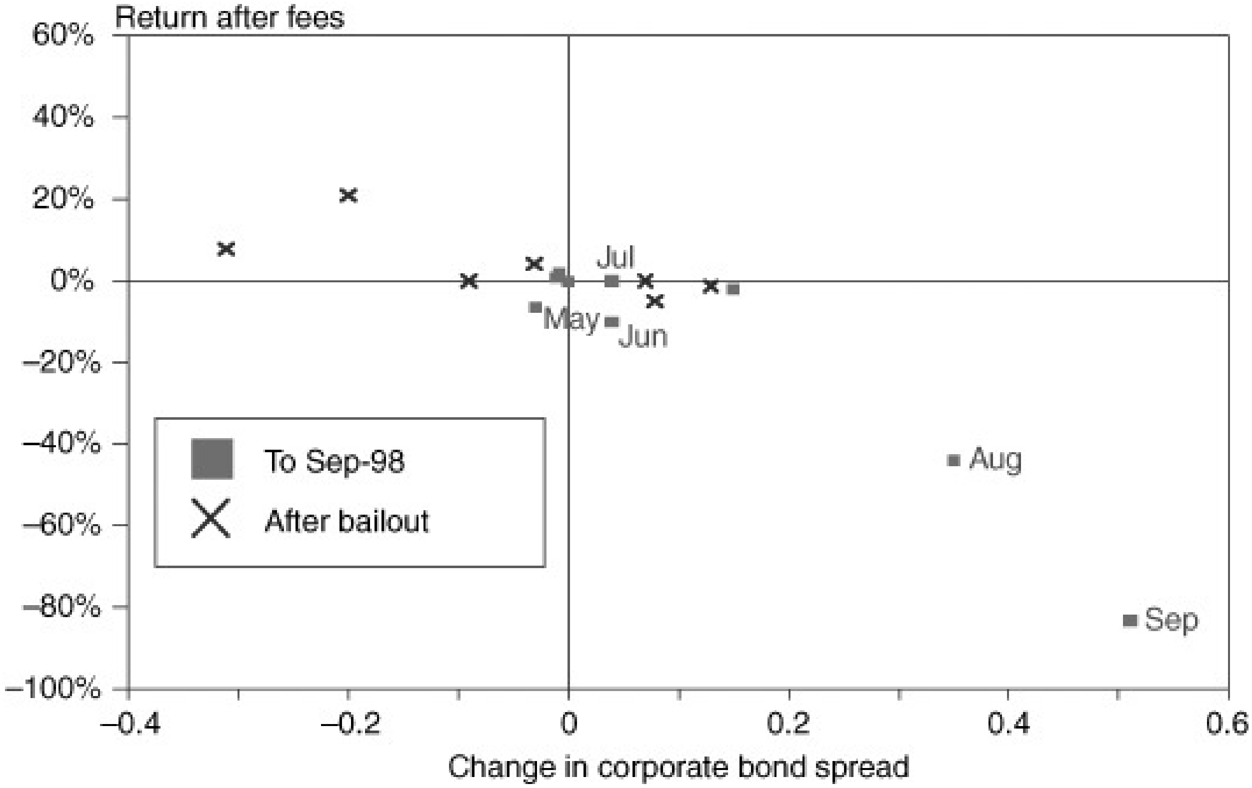
LTCM failed because of its inability to manage its risk. This was due in no small part to the fact that LTCM's trades were rather undiversified. LTCM was reported to have lost about \$1.5 billion from interest-rate swap positions and a similar amount from short positions on equity volatility. As we will show in Chapter 21, this was a result of an ill-fated attempt to manage risk through portfolio optimization.

Table 13-5 describes the exposure of various reported trades to fundamental risk factors. All the trades were exposed to increased market volatility. Most were exposed to increased liquidity risk (which is itself positively correlated with volatility). Many were exposed to increased default risk.

TABLE 13-5
Exposure of LTCM's Portfolio to Risk Factors

	Loss if Risk Factor Increases		
Trade	**Volatility**	**Default**	**Illiquidity**
Long interest-rate swap	Yes	Yes	Yes
Short equity options	Yes		
Long off-the-run/short on-the-run Treasuries			Yes
Long mortgage-backed securities (hedged)	Yes		Yes
Long sovereign debt	Yes	Yes	Yes

FIGURE 13-5
Explaining LTCM's returns.



To illustrate the driving factor behind LTCM's risks, Figure 13-5 plots the monthly returns against monthly changes in credit spreads. The fit is remarkably good, indicating that a single risk factor would explain 90 percent of the variation up to the September bailout. Thus there was little diversification across risk factors.

In addition, LTCM was a victim of both asset and funding liquidity risk. Although it had taken some precautions against withdrawal of funds, it did not foresee that it would be unable to raise new funds as its performance dived. The very size of the fund made it very difficult to organize an orderly portfolio liquidation.

The episode also raised questions about the soundness of the brokers' risk management systems. The brokers lulled themselves into thinking that they were protected because their loans were "fully collateralized." Even so, their loans carried no haircuts and were exposed to the risk that LTCM could default at the same time as the collateral lost value. One of the lessons of this near disaster was to accelerate the integration of credit and market risk management.

13.5 CONCLUSIONS

This chapter has shown how to account for liquidity risk. Traditional VAR models measure the worst change in mark-to-market value over the horizon but do not account for the actual cost of liquidation. These costs depend on the price-impact function, as well as the size of the positions. This leads to a “hybrid” liquidity-adjusted VAR measure that combines price volatility with liquidation costs.

In general, bid-ask spread effects are less important than traditional VAR measures. What matters more are the large price drops owing to liquidating large positions. In normal markets, liquidity effects are fairly predictable. Whether these LVAR measures apply to stressed markets, however, is more doubtful.

An alternative approach is to value positions at the conservative bid-ask quote and even to take a reserve to account for illiquidity. In such cases, there is no need to take liquidity risk into account in VAR because it is already factored into the valuation of positions.

Funding liquidity risk, in contrast, arises when financing for the portfolio cannot be maintained. Here again, VAR can be altered to estimate the risk that a portfolio could run out of cash.

Thus liquidity risk involves the two sides of the balance sheet, assets and liabilities. The greater the liquidation horizon for a portfolio, the greater is the need for extended financing of the portfolio.

The CRMPG recently reviewed the progress made since the original 1999 study in an update dubbed CRMPG II (2005). Many banks have responded that the CRMPG recommendations provided “a useful framework.” CRMPG II reports that institutions now have “a greater focus on liquidity-based adjustments to closeout values and on the interaction of asset liquidity and funding liquidity.” Still, the CRMPG warns that crises will “inevitably occur” and that “investments in risk management systems should continue to be a high priority.”

While LVAR may be somewhat difficult to measure, some rules of thumb are useful. We do know that bid-ask spreads are positively correlated with volatility. A position in illiquid assets will incur greater execution costs as volatility increases. Thus liquidity risk can be mitigated by taking offsetting positions in assets, or businesses, that benefit from increased volatility or have positive vega. Examples are long positions in options and customer trading, which typically benefit when trading volatility and volume spike up.

As with other applications of VAR, the main benefit of this analysis is not so much to come up with one summary risk number but rather to provide a systematic framework for thinking about the interactions among market risk, asset liquidity risk, and funding liquidity risk.

QUESTIONS

1. Define asset and funding liquidity risk.
2. What is a potential problem for the marking-to-market assumption underlying the measurement of VAR if VAR is to measure the worst loss over a liquidation period?
3. Explain how the analysis of market microstructure, or demand and supply curves, is useful to assess liquidity risk.
4. What is the common characteristic of *deep* markets in terms of liquidity risk?
5. Define *normal market sizes*.
6. How is asset liquidity risk controlled?
7. A hedge fund has a position in 1 million shares of a stock whose midprice is $100. The bid-ask spread is $0.40, up to a volume of 100,000. Beyond that, prices fall by $0.50 per share for every 100,000 shares transacted in one day. Compute the loss from the midprice if the entire position is liquidated over 1 day. This should be computed in dollars and in fraction of the initial position value.
8. Repeat with two other scenarios: (a) The sale is spread uniformly over 10 days. (b) The sale is spread over 5 days. Assume that prices are not expected to move.
9. Assuming a daily stock volatility of 1 percent and uncorrelated returns, compute the volatility of holding the original position over 10 days. Then compare the volatility of the three strategies in the previous questions. Ignore intraday risk.
10. What is the tradeoff between liquidating quickly or slowly.
11. Can you explain why hedge funds do not accept new investors after they have reached some size? Would you expect a large or small size for strategies investing in government bonds or high-yield bonds?
12. Some hedge funds have *lockup periods* for their investors, which prevent them from pulling their money within some period. Which type of strategies are more likely to use such clauses: leveraged funds investing

in government bonds or high-yield bonds?

13. Explain how funding liquidity risk can arise for leveraged institutions.
14. Do pension funds, which are not leveraged, face funding liquidity risk?
15. Explain what a haircut is (in the context of liquidity risk).
16. Why do companies issue debt with credit triggers? Do you think these are useful features?
17. Among U.S. stocks, bonds, and Treasury bills, which class of assets has the lowest bid-ask spread?
18. What are sources of bid-ask spreads in market microstructure theory?
19. How is the liquidity-adjusted VAR, LVAR, different from the traditional VAR?
20. Is the relative importance of the liquidity term in LVAR greater or smaller as the number of assets increases in a portfolio?
21. What is cash liquidity, as defined by the Counterparty Risk Management Policy Group?
22. What instruments did LTCM use to leverage its balance sheet? Explain.
23. What were the major risks involved in the LTCM debacle?
24. Reviewing the types of trades done by LTCM, do you think this was a well-diversified fund?

CHAPTER 14
Stress Testing

> This is one of those cases in which the imagination is baffled by the facts.
>
> —*Winston Churchill*

The main purpose of value-at-risk (VAR) measures is to quantify potential losses under “normal” market conditions, where *normal* is defined by the confidence level, typically 99 percent. In principle, increasing the confidence level could uncover progressively larger but less likely losses. In practice, VAR measures based on recent historical data can fail to identify extreme unusual situations that could cause severe losses. This is why VAR methods should be supplemented by a regular program of stress testing. Stress testing is a *nonstatistical* risk measure because it is not associated with a probability statement like VAR.

Stress testing is indeed required by the Basel Committee as one of seven conditions to be satisfied to use internal models. It is also endorsed by the Derivatives Policy Group and by the Group of Thirty. *Stress testing* can be described as a process to identify and manage situations that could cause extraordinary losses. This can be made with a set of tools, including (1) scenario analysis; (2) stressing models, volatilities, and correlations; and (3) policy responses.

Scenario analysis consists of evaluating the portfolio under various extreme but probable states of the world. Typically, these involve large movements in key variables, which requires the application of full-valuation methods. The earliest application of stress tests consisted of sequentially moving key variables by a large amount. This is also called *sensitivity testing*.

This approach, however, ignores correlations, which are crucial to large-scale risk measurement. More generally, scenarios provide a description of the joint movements in financial variables. Scenarios can be *historical*, that is, drawn from historical events, or *prospective*, that is, drawn from plausible economic and political developments. Prospective scenarios are also called *hypothetical*. More recently, the industry has realized that the identification of scenarios should be driven by the particular portfolio at hand. Scenarios that matter are those that could generate extreme losses.

Stress tests are used primarily for understanding the risk profile of a firm.

Increasingly, however, they are also used, in conjunction with VAR, for *capital allocation*. Whenever the stress tests reveal some weakness, management must take steps to manage the identified risks. One solution could be to set aside enough capital to absorb potential large losses. Too often, however, this amount will be cripplingly large, reducing the return on capital. Alternatively, positions can be altered to reduce the exposure. The goal is to ensure that the institution can ride out the turmoil.

Section 14.1 discusses why stress testing is required at all. In theory, extreme losses could be identified by increasing the confidence level of VAR measures. Section 14.2 shows how to use scenarios to generate portfolio losses. Sections 14.3 and 14.4 then examine scenario analysis in great detail. This is no easy matter owing to the large number of risk factors that global financial institutions are exposed to. Next, Section 14.5 turns to stress testing of models and parameters. Section 14.6 then discusses management actions that can be taken in response to stress-test results.

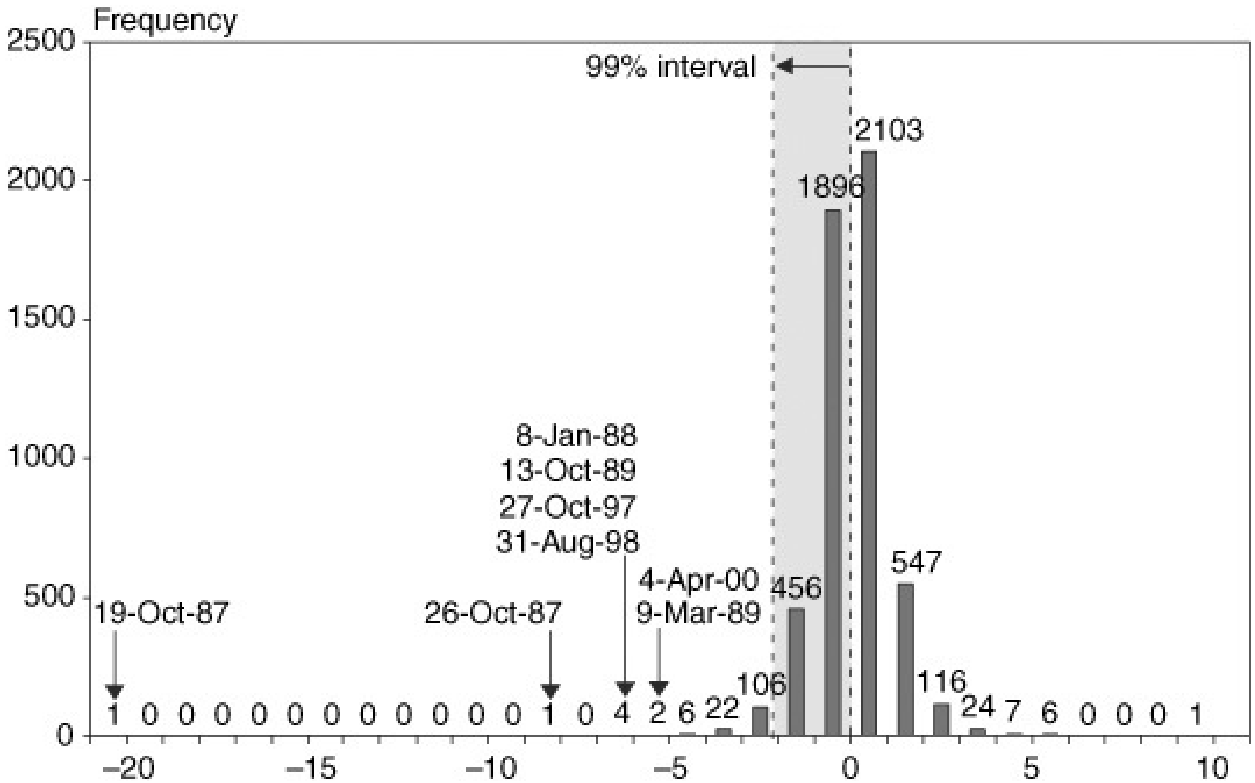
14.1 WHY STRESS TESTING?

Compared with VAR methods, stress testing appears refreshingly simple and intuitive. The first step is scenario analysis, which examines the effect of simulated large movements in key financial variables on the portfolio. Such scenarios have the advantage of linking the loss to a specific event, which is more intuitive to many managers than a draw from a statistical distribution. Owing to its simplicity, this approach actually predates VAR methods.

To understand the need for scenario analysis, consider, for instance, the stock market crash of October 19, 1987. Figure 14-1 displays the distribution of U.S. daily stock returns using data from 1984 to 2004. Over this period, the average volatility was about 1 percent per day. On Monday, October 19, the Standard & Poor’s (S&P) Index lost 20 percent of its value.

FIGURE 14-1

Distribution of daily U.S. stock returns, 1984–2004.



Even if there was some time variation in volatility, this 20 standard deviation event was so far away in the tail that it never have should happened under a normal distribution. The figure also shows that a standard 99 percent VAR interval would have totally missed the magnitude of the actual loss.

More generally, Bookstaber (1997) says that there is

> . . . a general rule of thumb that every financial market experiences one or more daily price moves of 4 standard deviations or more each year. And in any year, there is usually at least one market that has a daily move that is greater than 10 standard deviations.

These observations, however, are an indictment of the distributional assumption rather than VAR itself. In theory, one could fit a better distribution to the data and vary the confidence level so as to cover more and more of the left-tail events. This can be accomplished with historical simulations or, if a smoother distribution is required, through the use of extreme-value theory (EVT). In other words, the generation of a scenario is akin to a particular point in the distribution drawn from historical data. So what is special about stress testing?

The goal of stress testing is to identify unusual scenarios that would not occur under standard VAR models. Berkowitz (2000) classifies these scenarios into the following categories:

1. Simulating shocks that have never occurred or are more likely to occur than historical observation suggests.
2. Simulating shocks that reflect permanent structural breaks or temporarily changed statistical patterns.

Thus one reason to stress test is that VAR measures typically use recent historical data. Stress testing, in contrast, considers situations that are absent from historical data or not well represented but nonetheless likely. Alternatively, stress tests are useful to identify states of the world where historical relationships break down, either temporarily or permanently.

A direct example of the need for stress testing is Niederhoffer's belief, described in Box 14-1, that the market would not drop by more than 5 percent in a day. Indeed, this never happened from 1990 to October 1997. This does not mean that a loss of this magnitude can never happen.

BOX 14-1

VICTOR NIEDERHOFFER: THE EDUCATION OF A SPECULATOR

Victor Niederhoffer outlined his investment philosphy in his book, *Education of a Speculator*, which quickly became a best-seller. An eccentric and brilliant investor, he was a legend of the hedge-fund business. Indeed, he had compiled an outstanding track record—a 32 percent compound annual return since 1982.

Niederhoffer's mission was to "apply science" to the market. Although he had a Ph.D. in business from the University of Chicago, he did not believe in efficient markets and traded on statistical anomalies. He believed, for instance, that the market would never drop by more than 5 percent in a single day. Putting this theory into practice, Niederhoffer sold naked out-of-the-money puts on stock index futures. When the stock market plummeted by 7 percent on October 27, 1997, he was unable to meet margin calls for some $50 million. His brokers liquidated the positions, wiping out his funds.

Apparently, his views were narrowly based on recent history. It is true that the worst loss had been 3 percent in the previous 5-year period. Larger losses do occur once in a while, however. Most notably, the market lost 20 percent on October 19, 1987.

Another illustration is a breakup of a fixed exchange-rate system. In the summer of 1992, it would have been useful to assess potential vulnerabilities in the European monetary system. Indeed, in September 1992, the Italian lira and the British pound abandoned their fixed exchange rates, which led to a disastrous fall in their value. Historical volatilities based on the previous 2 years would have completely missed the possibility of a devaluation. Thus scenario analysis forces risk managers to consider events they otherwise might ignore.

14.2 PRINCIPLES OF SCENARIO ANALYSIS

We now consider the implementation of scenario analysis. Define *s* as a selected scenario. This is constructed as a set of changes in risk factors $\Delta f_{k,s}$ for various k. Based on the new hypothetical risk-factor values, $f_{k,O} + \Delta f_{k,s}$, all the securities in the portfolio are revalued, preferably using a full-valuation method if the portfolio has nonlinear components. The portfolio return then is derived from changes in the portfolio value *V*, which depends on positions and risk factors, that is,

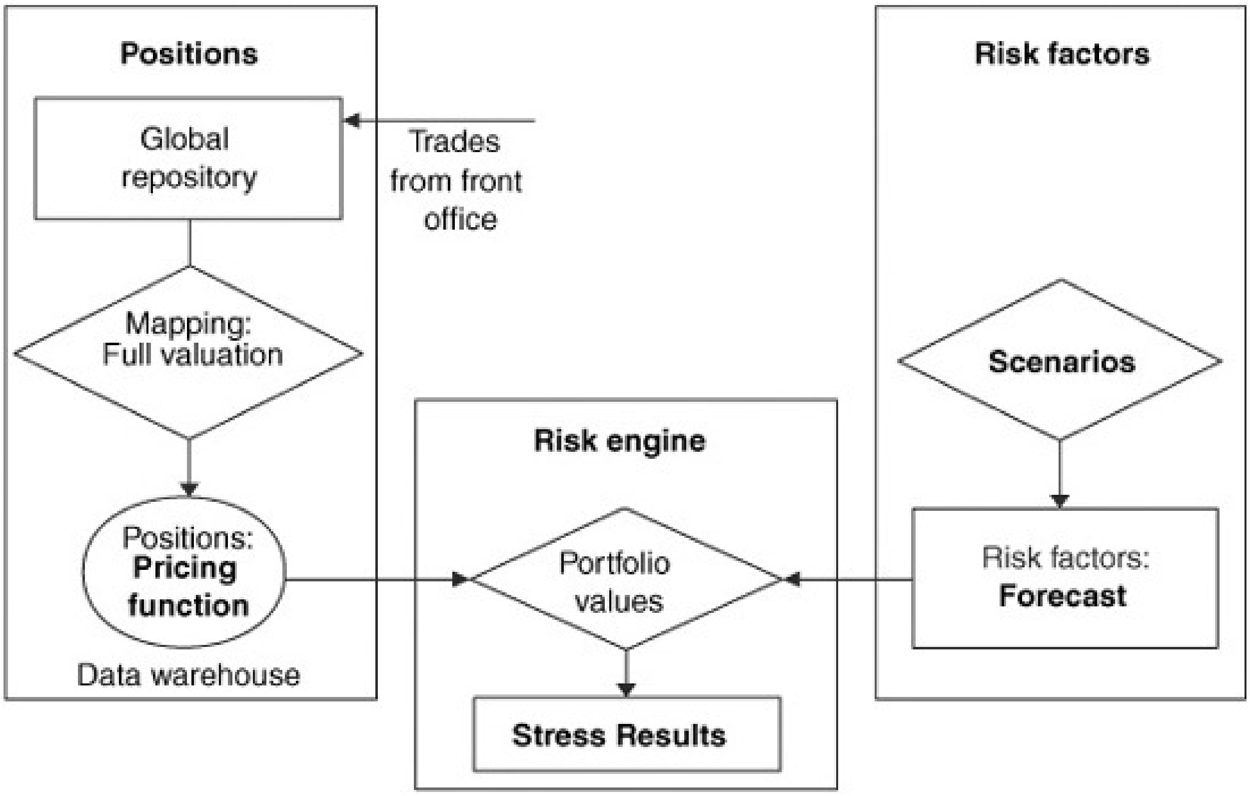
$$R_{p,s} = V_s - V_0 = V(f_{1,0}+\Delta f_{1,s} \ldots, f_{K,0}+\Delta f_{K,s}) - V(f_{1,0} \ldots, f_{K,0}) \qquad (14.1)$$

Note that this is a special case of the historical simulation method. Thus scenario analysis can be implemented easily once a VAR system is in place. Figure 14-2 details the steps involved in this approach. The question is how to generate realistic scenarios.

14.2.1 Portfolio-versus Event-Driven

The generation of scenarios can be either *event-driven* or *portfolio-driven*. In the first case, the scenario is formulated from plausible events that generate movements in the risk factors. In the second case, risk vulnerabilities in the current portfolio are identified first that translate into adverse movements in risk factors. These lead to the generation of scenarios. For instance, institutions that invest in long-term bonds funded by short-term debt are vulnerable to upward movements in the yield curve. It is therefore essential to consider scenarios that reflect such changes.

FIGURE 14-2
Scenario-analysis approach.



14.3 GENERATING UNIDIMENSIONAL SCENARIOS

14.3.1 Sensitivity Tests

The traditional approach to scenario analysis focuses on one variable at a time. For instance, the Derivatives Policy Group (DPG) provides specific guidelines for scenarios. It recommends focusing on a set of specific movements:

1. Parallel yield-curve shifting by ±100 basis points
2. Yield-curve twisting by ±25 basis points
3. Each of the four combinations or yield-curve shifts and twist
4. Implied volatilities changing by ±20 percent of current values
5. Equity index values changing by ±10 percent
6. Currencies moving by ±6 percent for major currencies and ±20 percent for others
7. Swap spreads changing by ±20 basis points

While these movements are quite large for a daily horizon, the DPG's goal was to provide comparable results across institutions in order to assess zones of vulnerabilities. By specifying consistent guidelines, it tried to ensure that all the models used by brokers “possess broadly similar performance.”

These scenarios shock risk factors generally one at a time. The loss in value, scaled by the size of the factor movement, is a *sensitivity* measure. These tests can be run relatively quickly and are intuitive.

This approach is appropriate in situations where the portfolio depends primarily on one source of risk. The Office of Thrift Supervision (OTS), for instance, uses scenario analysis to assess the market risk of savings and loans associations (S&Ls).[1] The OTS requires institutions to estimate what would happen to their economic value under parallel shifts in the yield curve varying from –400 to +400 basis points. The OTS recently has imposed a risk-based capital requirement linked directly to the interest-rate exposure of supervised institutions.

14.3.2 An Example: The SPAN System

The standard portfolio analysis of risk (SPAN) system is a good example of a scenario-based method for measuring portfolio risk. SPAN was introduced in 1988 by the Chicago Mercantile Exchange (CME) to calculate collateral requirements on the basis of overall *portfolio risk* as opposed to position by position. Since its inception, SPAN had become widely used by futures and options exchanges as a mechanism to set margin requirements.

The objective of the SPAN system is to identify movements in portfolio values under a series of scenarios. SPAN then searches for the largest loss that a portfolio may suffer and sets the margin at that level. The SPAN system only aggregates futures and options on the same underlying instrument. It uses full-valuation methods, which is important given that portfolios usually include options.

Consider a portfolio of futures and options on futures involving the dollar/euro exchange rate. SPAN scans the portfolio value over a range of prices and volatilities. These ranges are selected so that they cover a fixed percentage of losses, for example, 99 percent. Contracts have a notional of 125,000 euros. Assume a current price of \$1.05/euro and a

12 percent annual volatility. The value range for the contract is set at the daily VAR, that is,

$$\text{Price range} = 2.33 \times 12 \text{ percent } \sqrt{252}) \times (\text{euro}125{,}000 \times 1.05\$/\text{euro}) = \$2310$$

This is indeed close to the daily margin for an outright futures position, which is around $2500 for this contract. This corresponds to a price range of $0.0176 around the current price of $1.05 per euro. Next, the volatility range is set at 1 percent.

Table 14-1 presents an example of scenario generation. We select scenarios starting from the initial rate plus and minus three equal steps that cover the price range, as well as an up-and-down move for the volatility range. In addition, to provide protection for short positions in deep out-of-the-money options, two scenarios are added with extreme price movements, defined as double the maximum range. Because such price changes are rare, the margin required is 35 percent of the resulting loss.

TABLE 14-1

Example of SPAN Scenario System

Number	Scenario			Gain/Loss	
	Fraction Considered for P&L	Price Scan Expressed in Range	Volatility Scan Expressed in Range	Long Call	Long Futures
1	100%	0	1	$198	$0
2	100%	0	−1	−$188	$0
3	100%	+1/3	1	$395	$767
4	100%	+1/3	−1	−$21	$767
5	100%	−1/3	1	$23	−$767
6	100%	−1/3	−1	−$332	−$767
7	100%	+2/3	1	$615	$1,533
8	100%	+2/3	−1	$170	$1,533
9	100%	−2/3	1	−$132	−$1,533
10	100%	−2/3	−1	−$455	−$1,533
11	100%	+1	1	$858	$2,300
12	100%	+1	−1	$388	$2,300
13	100%	−1	1	−$268	**−$2,300**
14	100%	−1	−1	**−$559**	**−$2,300**
15	35%	+2	0	$517	$1,610
16	35%	−2	0	−$240	−$1,610
Ranges:		$0.0176	1%		

Note: Euro-FX futures and option on futures with notional of 125,000 euros, spot of 1.05\$/euro, strike of \$1.10, 12 percent annual volatility, 90 days to maturity, and interest rate of 5 percent.

Next, the value of each option and futures position is calculated under each scenario, using full valuation. The table presents calculations for two positions only, long one call and long one futures, under each of the 16 scenarios. The result of the computation for each risk scenario is called a *risk-array value*. The set of risk array values for the position is called the *risk array*.

The long-call position would suffer the most under scenario 14, with a large downward move in the futures accompanied by a drop in the volatility. Similarly, the worst loss for a long-futures position also occurs under a large downward move. This analysis is repeated for all options and futures in the portfolio and aggregated across all positions. Finally, the margin is set to the worst portfolio loss under all scenarios.

The SPAN system is a scenario-based approach with full valuation. Its systematic scanning approach is feasible because it considers only two risk factors. The number of combinations, however, soon would become unmanageable for a greater number of factors. This is perhaps the greatest hurdle to systematic scenario analysis.

Another drawback is that the approach essentially places the same probability on most of the scenarios, which ignores correlations between risk factors. And as we have seen, correlations are an essential component of portfolio risk.

14.4 MULTIDIMENSIONAL SCENARIO ANALYSIS

Unidimensional scenarios provide an intuitive understanding of the effect of movements in key variables. The problem is that these scenarios do not account for correlations. This is where multidimensional scenarios are so valuable. The approach consists of (1) positing a state of the world and (2) inferring movements in market variables.

14.4.1 Prospective Scenarios

Prospective scenarios represent *hypothetical* one-off surprises that are analyzed in terms of their repercussions on financial markets. One might want to examine, for instance, the effect of an earthquake in Tokyo, of Korean reunification, of a war in an oil-producing region, or of a major sovereign default. The definition of scenarios should be done with input from top managers, who are most familiar with the firm’s business and extreme events that may affect it.

Let us go back to the example of a scenario analysis for a potential breakup in the exchange-rate mechanism (ERM), evaluated as of summer 1992. The risk manager could hypothesize a 20 percent fall in the value of the Italian lira against the German mark. One could surmise further that if the Italian central bank let the lira float, short-term rates likewise could drop, and the stock market would rally. Beyond the effect on Italian interest rates and equity prices, however, it may not be obvious to come up with plausible movements for other financial variables. The problem is that the portfolio may have large exposures to these other risk factors that remain hidden. Thus this type of subjective scenario analysis is not well suited to large, complex portfolios.

Factor Push Method

Some implementations of stress testing try to account for multidimensionality using a rough two-step procedure. First, push up and down all risk-factor variables individually by, say, $\alpha = 2.33$ standard deviations, and then compute the changes to the portfolio. Second, evaluate a worst-case scenario, where all variables are pushed in the direction that creates the worst loss. For instance, variable 1 could be pushed up by $\alpha\sigma_1$, whereas variable 2 could be pushed down by $\alpha\sigma_2$, and so on.

This approach is very conservative but completely ignores correlations. If variables 1 and 2 are positively correlated, it makes little sense to consider moves in opposite directions. Further, looking at extreme movements may not be appropriate. Some positions such as combinations of long positions in options will lose the most money if the underlying variables do not move at all.

Conditional Scenario Method

There is a systematic method, however, to incorporate correlations across all variables consistently. Let us represent the "key" variables that are subject to some extreme movements as *R**. The other variables are simply represented by *R*. The usual approach to stress testing focuses solely on *R**, setting the other values to zero. Simplifying Equation (14.1) to a linear movement, what we call the *narrow stress loss* (NSL), is $\Sigma_i w_i^* R_i^*$.

To account for multidimensionality, we first regress the *R* variables on the controlled *R** variables, obtaining the conditional forecast from

$$R_j = \alpha_j + \sum_i \beta_{j,i} \mathrm{R}_i^* + \epsilon_j = E(R_j \mid R^*) + \epsilon_j \tag{14.2}$$

This allows us to predict other variables conditional on movements in key variables using information in the covariance matrix. We construct a *predicted stress loss* (PSL) as $\Sigma_i w_i^* R_i^* + \Sigma_j w_j E[R_j | R^*]$. This can be compared with the realized *actual stress loss* (ASL), which is $\Sigma_i w_i^* R_i^* + \Sigma_j w_j R_j$.

Kupiec (1998) illustrates this method with episodes of large moves from 1993 to 1998 using a $1 million portfolio invested in global equity, bond, and currency markets. Table 14-2 presents typical results. For the Philippine peso, for instance, the event was a devaluation, which was a 5.50 standard deviation move. The notional value of the position on this risk factor was $40,700, which led to a narrow stress loss (NSL) of $3070. This number, however, fails to account for other markets, such as Philippine equities, that moved in the opposite direction. Taking this correlation into account, the predicted stress loss (PSL) is much smaller than the NSL, even close to zero. In some other cases, PSL is much worse than NSL.

Interestingly, the table shows that, in all cases, the PSL produces results that are much closer to the ASL than in the simple, narrow model that zeroes out nonkey variables. The conclusion is that the covariance matrix, which is at the core of conditional normal VAR modeling, does provide useful information for stress-testing analysis.

TABLE 14-2

Comparison of Forecast Losses on a $1 Million Portfolio

				Stress Loss		
Key Variable	Period	Event Size (σ)	Position on Key Variable	Narrow	Predicted	Actual
Philippine peso	11 Jul 1997	−5.50	$40,700	−$3,070	$43	$190
Japanese equities	23 Jan 1995	−5.23	$72,120	−$2,700	−$7,730	−$11,700
U.S.equities	27 Oct 1997	−4.93	$136,480	−$6,650	−$5,330	−$5,420
U.K.bonds	29 Dec 1994	−4.84	$122,910	−$2,640	−$3,550	−$3,030
U.S.bonds	20 Feb 1996	−4.86	$122,970	−$1,210	−$7,070	−$10,380

The main drawback of this conditional scenario method is that it relies on correlations derived from the entire sample period. This includes normal periods and *hectic* periods. Should correlations change systematically across these periods, however, the results will differ. For portfolios with long positions only, increases in correlations will increase the worst loss. Banks often reexamine the stress-test results using correlations estimated over hectic periods.[2]

As an example, the correlation between stocks and bonds typically is positive in normal times. In times of stress, however, this correlation often turns negative. When equity markets drop sharply, the demand for Treasury bonds typically increases, reflecting a flight to quality. At the short end of the yield curve, this effect usually is reinforced when the central bank injects liquidity into the financial system, pushing down short-term rates. Thus government bonds are good diversifiers for stocks in times of stress.

A related, useful approach relies on the output from a VAR Monte Carlo analysis or historical simulation. The risk manager could examine the worst losses from the simulation, which specifically accounts for correlations. Such analysis provides valuable insight into the vulnerabilities of a particular portfolio and could guide the construction of scenarios.

14.4.2 Historical Scenarios

Alternatively, scenario analysis can examine historical data to provide examples of joint movements in financial variables. The role of the risk manager is to identify scenarios, such as those listed in Table 14-3, that may be outside the VAR window. Each of these scenarios then will yield a set of joint movements in financial variables that automatically takes correlations into account.

Table 14-3 displays a list of scenarios, both historical and prospective, used by a large group of banks. The largest category of stress tests focuses on interest rates. Historical scenarios include the 1994 bond market crash, the 1997 Asian currency crisis, the LTCM and Russia crises, and the terrorist attack on the World Trade Center, all of which led to global interest-rate shocks. Also common are stress tests involving equities, currencies, commodities, and credit.

TABLE 14-3
Description of Scenarios (Number of Tests Reported by 64 Institutions Surveyed)

Category	Historical	Prospective
Interest rates (173)	1994: Bond market crash (18)	U.S. economic crash (21)
	1997: Asian currency crisis (22)	Global economic crash (11)
	1998: LTCM, Russia crises (26)	Inflation pickup (8)
	2001: World Trade Center (30)	
Equities (86)	1994: Black Monday (83)	Geopolitical unrest (5)
	1997: Asian currency crisis (22)	Terrorist attack (5)
	2000: IT bubble bursting (6)	
	2001: World Trade Center (30)	
Currencies (56)	1992: EMS crisis (8)	Pegged-currencies breakdown (7)
	1997: Asian currency crisis (22)	
	1998: LTCM, Russia crises (26)	
Commodities (22)		Oil-price jump (11)
		Unrest in Middle East (5)
Credit (104)	1997: Asian currency crisis (22)	Emerging market crash (10)
	1998: LTCM, Russia crises (26)	Euro area economic crash (7)
	2001: World Trade Center (30)	Global economic crash (6)
Property (19)		Various episodes

Source: Basel Committee (2005b) survey of financial institutions.

TABLE 14-4
October 1987 Market Crash: Change in Market Variables

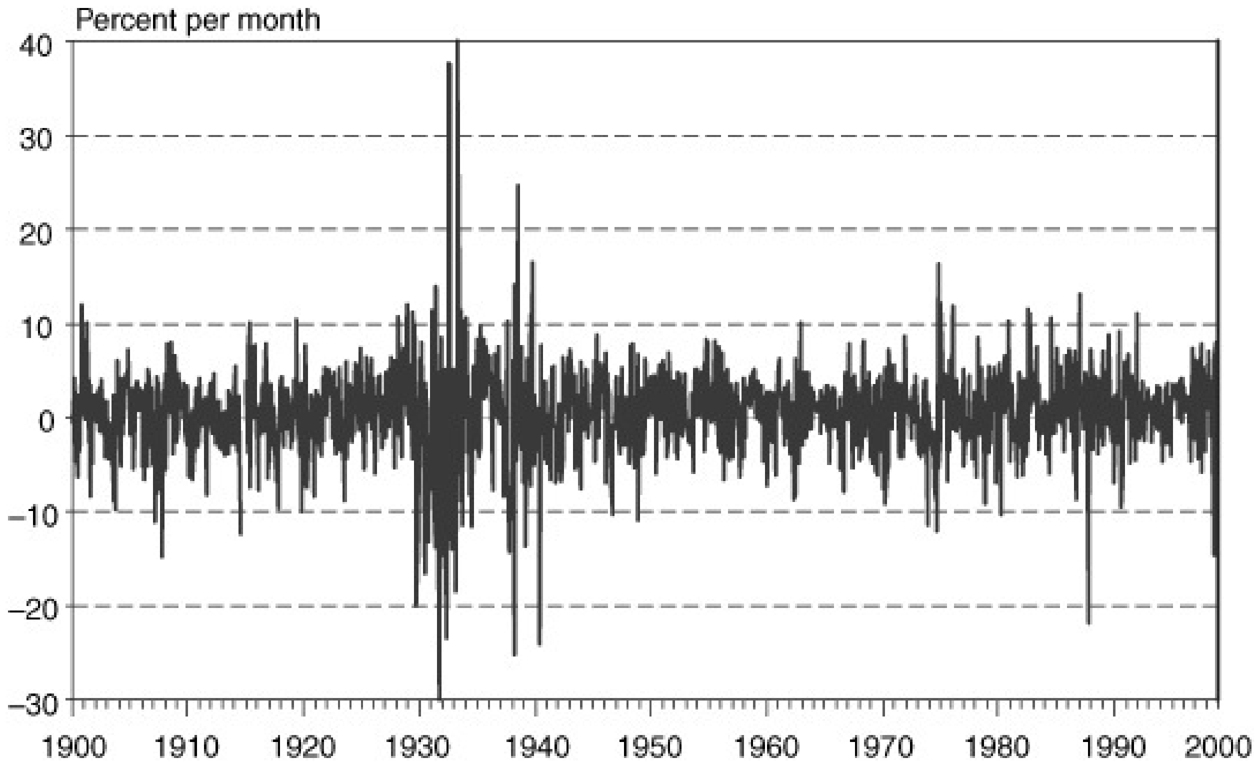
	Equities				Fixed Income			Currencies	
Date	U.S., S&P	Japan, Nikkei	U.K., FTSE	Germany, DAX	Fed Funds	3-Month T-Bill	30-year T-Bond	Yen/ $	DM/ $
Oct 19	−20.4%	−2.4%	−9.1%	−9.4%	−0.14 bp	−0.52 bp	0.01 bp	−0.2%	1.3%
Oct 20	5.3%	−14.9%	−11.4%	−1.4%	−0.77 bp	−0.62 bp	−0.76 bp	1.5%	1.7%

As an example, Table 14-4 details the vector of movements in risk factors for the stock market crash of October 1987. On Monday, October 19, the S&P Index lost more than 20 percent of its value. The sheer size of this movement had dramatic effects on other prices. The next day, the Nikkei fell by 15 percent. In order to forestall the failure of financial institutions, the Federal Reserve injected liquidity into the financial system, pushing short-term interest rates down by 91 basis points in 2 days. This chaos, however, apparently had little effect on currency markets. Overall, there was little diversification benefit across global stocks. As is often the case, however, long positions in bonds helped to cushion the equity losses.

Historical scenarios are quite useful to measure joint movements in financial variables. Their drawback, from the risk manager's perspective, is the limited number of extreme events. Whenever possible, very long histories should be considered. These give a long-term perspective that may be absent from recent data.

The October 1987 crash, for instance, is viewed widely as an extremely unusual event. Perhaps so if one looks only at recent data. On the other hand, a different perspective is offered by Figure 14-3, which reports monthly returns on U.S. stocks since the beginning of the century. The figure shows that if one goes back sufficiently in time, there have been many other instances of losses exceeding 20 percent over a month and that the recent past has not been unusually volatile.

FIGURE 14-3
Monthly U.S. stock returns, 1900–2000.



Likewise, events such as sovereign defaults are extremely rare. Recently, however, Russia defaulted on its domestic debt and Ecuador on its external debt. One would need to go back to the 1930s to encounter sovereign defaults on external debt. Defaults should be expected to occur, however. Otherwise, there would be no rationale for the wide yield spread observed on some sovereign debt.

As should be clear by now, the generation of relevant scenarios is a time-consuming process that requires quantitative skills as well as a good economic understanding of the factors driving financial shocks.

14.5 STRESS-TESTING MODELS AND PARAMETERS

Going back to the schematics of scenario analysis in Figure 14-2, the risk manager should critically examine all the steps in the generation of risk measures. A stress-testing process should consider not only movements in market variables but also the other components of the risk-management system, that is, the securities valuation models and the risk engine.

A distinction usually is made between *sensitivity analysis*, which examines the effect of changing the functional form of the model, and stress-testing *model parameters*, which are inputs into the model.

Let us consider sensitivity analysis first. Derivative securities can be priced using a variety of models. Interest-rate derivatives, for instance, can be valued using one-or multiple-factor models, with parameters typically estimated from historical data. Mortgage-backed securities (MBS) must in addition model prepayments. All these assumptions introduce insidious risks. Current model prices may fit the current market data but may not provide a good approximation under large movements in key variables. The Askin story (see Box 1-2) provides an example of an MBS portfolio that was thought to be hedged but led to large losses under severe interest-rate shocks. Pricing models may fail in changing environments.

Likewise, simplifications in the risk measurement system also may create hidden risks. For example, bond mapping replaces a continuous yield curve by a finite number of risk factors. If there is insufficient *granularity*, or detail, in the choice of the risk factors, the portfolio could be exposed to losses that are not measured by the risk management system.

Turning to *model parameters*, pricing and risk management systems rely on particular input data, such as a set of volatilities and correlations. Correlations, however, may deviate sharply from historical averages in times of stress. A key issue is whether a traditional variance-covariance-based VAR system provides adequate measures of risk when historical correlation patterns break down.

To some extent, this question can be answered directly by scenario analysis based on historical data. It is also informative, however, to check how sensitive a VAR number is to changes in the risk measures. As will be seen in the example of LTCM, this is especially important if the same period is used to measure risk and optimize the portfolio.

As an example, consider a covariance matrix measured with recent data that shows a high correlation between two series. The risk manager, however, cannot believe that this high correlation will remain in the future and could alter the covariance matrix toward values that are considered reasonable.[3] The stress test then compares the new VAR measure with the original one.

In all these cases, there is no simple rule to follow for stress testing. Rather, the risk manager must be aware of limitations, assumptions, and measurement errors in the system. Stress testing can be described as the art of checking whether the risk forecasts are robust to changes in the structure of the system.

14.6 POLICY RESPONSES

The main function of stress tests is to communicate and understand the risks that the institution is exposed to. The next question is what to do if the size of the stress loss appears unacceptably large. This is the crux of the issue with stress testing. Too often stress-testing results are ignored because they create large losses that are dismissed as irrelevant.

Indeed, institutions do not need to withstand every single state of the world. Central banks, in particular, are supposed to provide protection against systemic banking crises. Likewise, there is little point in trying to protect against a widespread nuclear war.

Relevant scenarios, however, require careful planning. One response is for the institution to set aside enough *economic capital* to absorb the worst losses revealed by stress tests. In many cases, however, this amount may be much too large, which will make it uneconomical. A number of other actions can be considered, though. The institution could

BOX 14-2

STRESS TESTING'S BENEFITS

A risk manager at a U.S. investment bank recalls that in December 1997 stress tests showed that the firm could be put in jeopardy should Russia default on its debt. The firm reduced its exposure to Russia in part through the purchase of credit derivatives.

The bank was able to ride the turmoil but still suffered losses owing to the fact that some counterparties defaulted on the credit protection. This illustrates that stress testing generally is useful but still is a subjective exercise that cannot possibly cover all contingencies.

- Purchase protection or insurance for the events in question (although this may transform market risk into counterparty risk)
- Modify the portfolio to decrease the impact of a particular event through exposure reduction or diversification across assets
- Restructure the business or product mix for better diversification
- Develop a plan for a corrective course of action should a particular scenario start to unfold
- Prepare sources of alternative funding should the portfolio liquidity suffer

Risk limits often are based on stress-test results, in addition to the usual VAR or notional-amount limits. This plan of action should help to ensure that the institution will survive this scenario, as shown in Box 14-2.

14.7 CONCLUSIONS

While VAR focuses on the dispersion of revenues, stress testing instead examines the tails. Stress testing is an essential component of a risk management system because it can help to identify crucial vulnerabilities in an institution's position. The methodology of stress testing also applies to credit and operational risks.

In some sense, stress testing can be viewed as an extension of the historical simulation method at increasingly higher confidence levels.

Stress testing, however, is a complement to standard VAR methods because it allows users to include scenarios that did not occur over the VAR window but nonetheless are likely. It also allows risk managers to assess "blind spots" in their pricing or risk management systems. Stress testing can help to ensure the survival of an institution in times of market turmoil.

The drawback of the method is that it is highly subjective. Bad or implausible scenarios will lead to irrelevant potential losses. Even worse, plausible scenarios may not be considered. The history of some firms has shown that people can be very bad at predicting extreme situations. Generally, stress-test results are presented without an attached probability, which makes them difficult to interpret. Unlike VAR, stress testing can lead to a large amount of unfiltered information. There may be a temptation for the risk manager to produce large numbers of scenarios just to be sure that any likely scenario is covered. The problem is that this makes it harder for top management to decide what to do.

Overall, stress testing should be considered an essential complement to rather than a replacement for traditional VAR measures. Stress testing is useful to evaluate the worst-case effect of large movements in key variables. This is akin to drawing a few points in the extreme tails: useful information, but only after the rest of the distribution has been specified. Still, stress testing provides a useful reminder that VAR is no guarantee of a worst-case loss.

QUESTIONS

1. Why should VAR measures be supplemented by portfolio stress testing?
2. How is stress testing different from backtesting when evaluating risk

models?

3. “Stress testing is not necessary because the same results are obtained by a VAR model with an increasing confidence level.” Comment.
4. Why is unidimensional scenario analysis not sufficient for stress-testing purposes?
5. If historical scenarios automatically take into account correlations, why not rely exclusively on them?
6. What is the major difficulty in generating hypothetical stress tests across large portfolios?
7. Describe the SPAN system for setting margins on a portfolio of futures and options on the same currency.
8. A pension fund has a portfolio with $1 billion invested in U.S. stocks and $1 billion invested in Japanese stocks. The 99 percent 1-week VAR analysis reveals a VAR of $112 million. The risk manager, however, is concerned about extreme moves not reflected in VAR. Compute the stress loss for the following situations:

   (a) A univariate scenario where U.S. stocks fall by 20 percent.

   (b) A univariate scenario where Japanese stocks fall by 25 percent.

   (c) A prospective scenario where U.S. stocks fall by 20 percent and Japanese stocks by 15 percent.

   (d) A prospective scenario where U.S. stocks fall by 5 percent and Japanese stocks by 25 percent. As a risk manager, which stress loss do you think would be most relevant?
9. Just to be sure, the risk manager runs a regression of the Japanese stocks on U.S. stocks and finds a slope coefficient of 0.9. Assuming that a shock of 20 percent originates from the United States, compute the predicted stress loss for the portfolio. What is the danger of this approach?
10. What is the scenario most used for stress tests of firm-wide positions, (a) The 1987 equity crash, (b) money tightening by central banks, (c) a country default, or (d) widening of credit spreads?
11. The crash of 1987 was a 20 standard deviation event. Based on the normal density function, a movement of this magnitude should never

happen. In addition, it was aggravated by a failure of the stock exchange to absorb the volume of trading, which has been fixed since then. Based on this information, a portfolio manager argues that this event should not be used in stress tests. Discuss.

12. If a scenario analysis reveals unacceptably large losses, what is a possible response?

PART IV
APPLICATIONS OF RISK MANAGEMENT SYSTEMS

CHAPTER 15
Using VAR to Measure and Control Risk

> At the close of each business day, tell me what the market risks are across all businesses and locations.
>
> —*Dennis Weatherstone, J.P. Morgan*

So far this book has discussed the motivation, building blocks, and various approaches to value-at-risk (VAR) systems. It is now time to turn to the implementation and applications of VAR.

By now, VAR has established itself as a key building block of financial risk management systems. VAR provides a top-level view of financial risk. This is ideally suited to institutions that engage in proprietary trading but also to asset managers and other financial institutions. It is also taking hold with nonfinancial corporations such as multinationals that have significant exposure to financial risks. Indeed the VAR methodology applies not only to balance-sheet values but also to cash flows. *Cash flow at risk* (CFAR) is a straightforward extension of VAR.

At the time VAR burst onto the scene, in 1994, it was devised as a method to measure and report market risk. Financial institutions have established global risk management committees that aggregate companywide risks into a single VAR measure that is easy to communicate to top management and shareholders. For most users, however, VAR was simply a *passive* application. They were content to use VAR to report “risk numbers” to stakeholders.

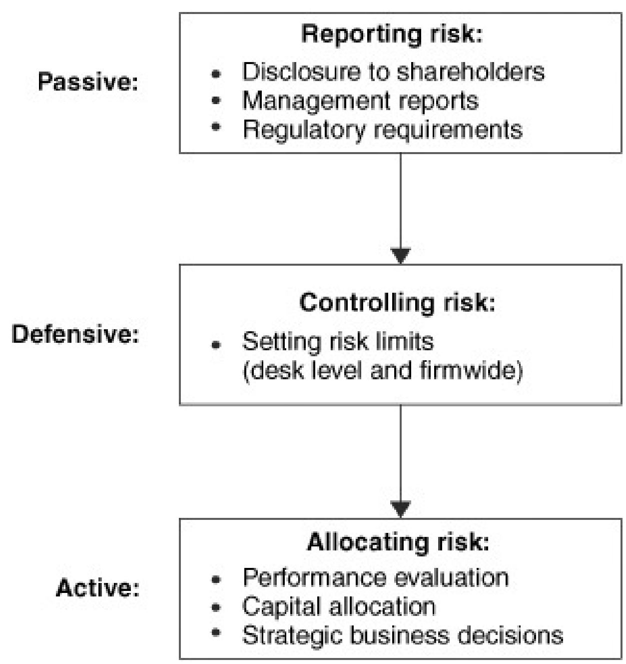
Since then, VAR has evolved into much more than a method to measure risk. Institutions have learned to apply VAR as a risk-control tool. Once a global risk management system is in place, it can be used to control risk more tightly than before. For instance, position limits for traders can be complemented by VAR limits that properly account for the leverage and risk of various instruments. At the firmwide level, VAR allows the institution to monitor its global risk exposure, taking into account diversification across business units. The firm can identify whether too many bets create unacceptable risks and, if so, reverse engineer the VAR process to identify where to cut risks. This second stage of applications represents a notable improvement over the passive reporting of risk. It is still *defensive* in nature, however.

Most recently, VAR has developed into an *active* risk management tool. With VAR tools on hand, institutions can decide how to trade off risk and return. Economic capital can be allocated as a function of business risks. Traders can be evaluated in terms of their risk-adjusted performance. Among the most advanced institutions, VAR systems are now used to identify areas of competitive advantage or sectors where they add risk-adjusted value. The evolution of VAR applications is described in Figure 15-1.

This chapter deals with the passive and defensive applications of VAR systems. Section 15.1 first reviews factors that create a need for global risk systems. It discusses situations where VAR systems are likely to be more valuable. Applications of VAR as an information-reporting tool and as a risk-control tool are analyzed in Sections 15.2 and 15.3.

FIGURE 15-1

Evolution of VAR applications.



The active risk management function, because it is so important, will be examined in Chapter 16. We also examine the application of VAR to asset managers in Chapter 17. VAR-type approaches are being extended beyond market risk to the credit risk, operational risk, and firmwide risks, which are explained in later chapters.

15.1 WHO CAN USE VAR?

15.1.1 The Trend to Global Risk Management

VAR methods represent the culmination of a trend toward *centralized* risk management. For a number of years, financial institutions have maintained local risk management units, especially around derivatives that need to be controlled carefully because of their leverage. But only recently have institutions started to measure risk on a global basis.

This trend to global risk management is motivated by two driving factors, exposure to *new sources of risk* and the *greater volatility* of new products. With the globalization of financial markets, investors are now exposed to new sources of risk such as foreign-currency risk. Greater volatility is induced by greater risk in some underlying variables, such as exchange rates, or by the design of "exotic" products that are more sensitive to financial variables.

This trend toward centralized risk management goes back to the creation of customized over-the-counter (OTC) derivatives, such as swaps, in the 1980s. Initially, these OTC derivative transactions were immediately offset with opposing transactions, that is, swaps with similar risks. Intermediaries were essentially acting as brokers. Later, derivatives were *warehoused*, that is, kept in inventory, with dealers temporarily hedging the transaction until an offsetting transaction could be found. This led to the need for a good inventory system, as well as a good accounting system to track transactions. The next step was the transition to a *portfolio approach*. Each transaction was disaggregated into component cash flows and aggregated with other instruments in the portfolio. This is what started the process of computing VAR.

For credit-risk management, centralization is also essential. The continued expansion of derivatives markets has created new entrants with lower credit ratings and greater exposure to counterparties. A financial institution may have myriad transactions with the same counterparty, coming from various desks, such as currencies, fixed income, commodities, and so on. Even though all the desks may have a reasonable exposure when considered on an individual basis, these exposures may add up to an unacceptable risk.

Moreover, with netting agreements, the total exposure depends on the net current value of contracts covered by the agreements. All of this becomes intractable unless a global measurement system for credit risk is in place.

Large commercial and investment banks were the first to monitor on a centralized basis counterparty exposure, country, and market risks across all products and geographic locations. Asset managers and nonfinancial corporations, however, also benefit from global risk management systems.

Implementing a global risk management system, however, is no small feat. It involves integrating systems, software, and database management, which can be very expensive. In addition, it requires substantial investment in intellectual and analytical expertise. As such, it may not be appropriate for all institutions (see, for instance, Box 15-1). This is why it is useful to delineate factors that favor the development of such systems.

Diversity of Risk

Institutions exposed to a diversity of financial risks, interest rates, exchange rates, and commodity prices certainly would benefit from a global risk management system. They need a system that consistently accounts for correlations, various exposures, and volatility across risk factors. This is especially so when the institution has a large number of independent risk-taking units whose risks need to be aggregated at the highest levels. In contrast, institutions that are exposed to one source of risk only may not require a sophisticated global risk management system. Savings and loans institutions, for instance, are exposed mainly to domestic interest-rate risk, in which case a simple duration measure may be sufficient.

BOX 15-1

MERRILL'S APPROACH

Merrill's approach to global risk management differs from that of other banks. A much smaller proportion of revenues is generated by position trading. Most of its profits come from customer orders, which generally are hedged immediately.

Given Merrill's large volume of trading, VAR reports produced at the close of the previous day quickly become outdated. Perhaps this explains why Merrill's risk managers do not rely much on computer models. In their view, their best risk management tool is "distribution."

Merrill also takes the view that it has natural "business" exposure to volatility that offsets the exposure of its financial portfolio. When volatility increases, more customer orders flow in, which generates additional profits. These profits offset potential falls in the value of its inventory. The firm also keeps a positive vega (long volatility) position on its options books, just to be sure. This overall approach to the risk of the institution is an example of informal *integrated risk management*.

Active Position Taking

Firms that take aggressive proprietary positions do require the discipline imposed by a global risk management system, especially if positions change quickly and if their leverage is high. On the other hand, firms that routinely match all trades have less of a need for such a system. One such example is foreign-exchange "brokers," who simply match buyers and sellers without ever taking positions. For them, a VAR system is not useful.

Complexity of Instruments

Firms that deal with complex instruments do require a centralized risk management system that allows consistent measures and controls of risk. Another benefit is that such a system requires a central repository for all trade processing, price quotes, and analytics. This provides some protection against operational risk, including fraud and model risk.

15.1.2 Proprietary Trading Desks

Proprietary trading desks are the prime example of institutions that satisfy all the criteria just listed. Their business has become exposed to global sources of risk. At the same time, the desks can take aggressive positions, can operate generally independently of each other, and can deal with complex products.

Consider, for instance, an investment bank where traders are awaiting U.S. unemployment numbers. Currency traders may short the dollar; they bet on unexpectedly high figures, leading to a fall in U.S. interest rates that should push the dollar down. Bond traders also may expect joblessness to rise and go long Treasury bonds. The fall in inflationary expectations may push commodity traders to short gold. Individually, these risks may be acceptable, but as whole, they sum to a sizable bet on just one number. Global risk management provides a uniform picture of the bank's risk. It fully accounts for correlations across locations and across asset classes. It allows firms to understand their risk better and therefore to control their risk better.

TABLE 15-1

J.P. Morgan's Trading Business

	Fixed Income	Currency	Commodities	Derivatives	Equities	Emerging Markets	Proprietary	Total
Number of active locations	14	12	5	11	8	7	11	14
Number of independent risk-taking units	30	21	8	16	14	11	19	120
Thousands of transactions per day	>5	>5	<1	<1	>5	<1	<1	>20
$ billions in daily trading volume	>10	>30	1	1	<1	1	8	>50

One of the earliest applications is the famous 4:15 P.M. report at J.P. Morgan. Table 15-1 shows the global trading business of the bank in 1994. Trading activities are grouped into seven business areas, each of them active in up to 14 locations. Altogether, the bank had 120 independent risk-taking units that handle over 20,000 transactions per day with a total volume exceeding $50 billion. Although decentralized trading appears very profitable, strong central risk controls are essential to understand the global risk exposure of the bank.

At the end of the day, all trading units report their estimated profit and loss for the day, their position in a standardized mapping format, and their estimated risk profile over the next 24 hours. Corporate risk management then aggregates the information with centrally administered volatilities and correlations. This leads to the global consolidated 4:15 P.M. report, which is discussed by business managers before being sent to the board's chair. Before such reports became commonplace, banks essentially were ignorant of their aggregate risk.

15.1.3 Nonfinancial Corporations

VAR is also taking hold in the corporate world, albeit more slowly than for financial institutions. Nonfinancials usually focus more on variability in cash flows than on market values of assets and liabilities. Thus the VAR methodology can be modified to measure what has been called *cash flow at risk* (CFAR), which is the worst loss in cash flows at some confidence level.[1]

The first step for measuring CFAR requires delineating *economic exposures*, which represent the sensitivity of cash flows to movements in the price of the financial variable. Consider first *contractual cash flows*, such as a contract to sell goods in a foreign currency, say, the euro. This contract can be "mapped" to a long position in the euro with an economic exposure equal to the notional amount. *Anticipated exposures* are similar except that they involve some uncertainty as to the actual payment. These items can be generalized to the entire cash-flow statement. As with the usual VAR methods, this is a bottom-up approach and may be called a *pro forma approach*.

Suppose, for instance, that a U.S. corporation exports to Europe and is planning to receive a series of four quarterly payments, as described in Table 15-2. The table also reports the budgeted exchange rate and the total cash flow in dollars, which is $8.51 million. The question is, what is the CFAR?

The next step consists of describing the risk distribution of key financial variables, commodity prices, exchange rates, and interest rates. This can be done via Monte Carlo simulations. The horizon usually is selected to match the business planning cycle. Note that with longer horizons, the modeling of expected returns is increasingly important, justifying the use of cointegration techniques described in Chapter 12. This is not such a problem with short-term VAR measures because volatility dominates expected returns over short horizons.

TABLE 15-2

Cash-Flow Exposure

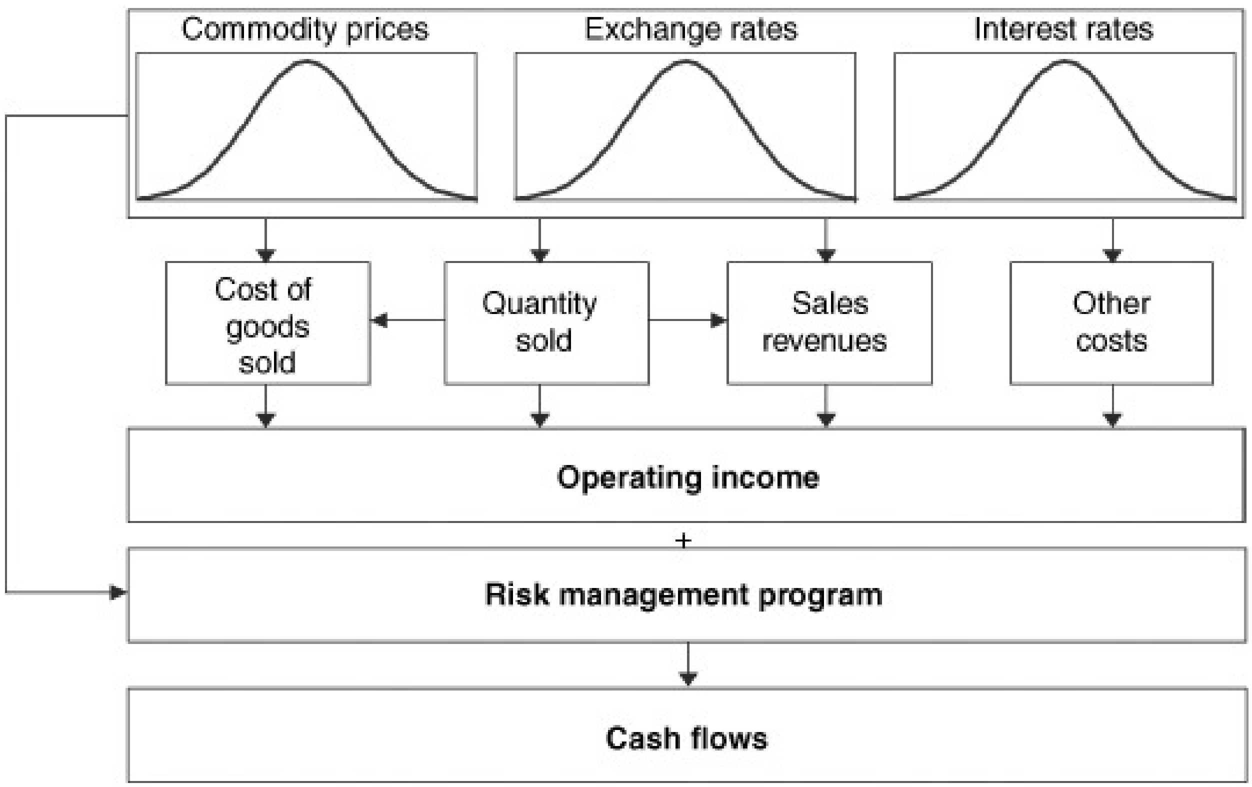
	Period				
	1	2	3	4	Total
Cash flow (€ million)	€2	€2.1	€1.5	€2.5	€8.1
Exchange rate ($/€)	1.02	1.04	1.06	1.08	
Cash flow ($ million)	$2.04	$2.18	$1.59	$2.70	$8.51

Finally, these financial variables need to be combined with economic exposures. This is akin to attaching a simulation engine to the business cash-flow model. In the preceding example, if we assume an annual volatility of 12 percent and a trend given by the budgeted rates, simulations yield an average cash flow of $8.52 million and a 95 percent lower value of $7.40 million. Hence the worst cash-flow shortfall, or CFAR, is $1.12 million. Stein *et al.* (2001) analyze the distribution of historical cash flows for comparable companies, which results in a large sample of data points. They report that for an average company with $100 in assets, the CFAR for a horizon of 1 year and at the 95 percent confidence level is about $10.

This approach can be generalized to all earnings, not just specific cash flows, in which case the risk measure is *earnings at risk* (EAR). Financial variables affect *operating cash flows* through quantities sold, sales revenues, the cost of goods sold, and other costs, as shown in Figure 15-2. For instance, costs may be affected by commodity prices or, if imported, exchange rates. Sales revenues may be affected by exchange rates, if exported. Quantities and sale prices will depend on competition, however, as shown in Appendix 15.A. This *quantity uncertainty* makes it more difficult to measure CFAR than traditional VAR, where the exposures are easier to assess.

FIGURE 15-2

Measuring cash flow at risk.



Also difficult to assess are the effects of *strategic options*, also known as *operational hedges*, whereby firms can alter their marketing strategy (product or pricing) or production strategy (such as product sourcing or plant location) over the horizon in response to movements in financial variables. These options, as in the case of stop-loss or other risk-mitigating techniques, in general reduce market risk.

Once this model is constructed, risk can be measured using the VAR of the operating cash flows. A risk management program then can be set up with derivatives to lower this risk, as shown in Figure 15-2. The effectiveness of the hedging program will be measured by the reduction in VAR.

Although CFAR may not be easy to measure, there is no question that the gathering of current companywide information provides useful information. Box 15-2 illustrates the benefits of VAR in a Treasury operation.[2]

BOX 15-2

TOYOTA'S VAR

Toyota Motor Credit Corporation (TMCC) is one of the largest corporate issuers in the global bond markets. Its goal is to facilitate the sale of Toyota cars to U.S. consumers. The company raises about $7 billion a year to provide funds for car leases, which typically involve level payments over a 3-year period.

TMCC simply could lock in fixed rates to cover its assets and liabilities. The treasury manager, Jerome Lienhard, however, takes the view that raising funds at floating rates is cheaper in a positively sloped yield-curve environment.

This, however, involves taking some interest-rate risk, which is measured using VAR. TMCC runs Monte Carlo simulations of its cash inflows and outflows and discounts them to the present. These simulations allow realistic interest-rate paths as well as the inclusion of caps, or call options, that provide protection if floating rates increase.

VAR is computed using a 95 percent confidence interval over a 30-day period. This horizon gives the treasurer enough time to react if rates increase unexpectedly. Since TMCC put its VAR system in place, the portfolio VAR has been reduced from $85 million to about $30 million. This represents 1.3 percent of its capital of about $2 billion. Furthermore, TMCC estimates that hedging expenses have been reduced by $10 million, or 20 percent. Says Lienhart, "There is no question that we have gained enormous understanding of risk through the process of creating an in-house system."

15.2 VAR AS AN INFORMATION-REPORTING TOOL

15.2.1 Trends in Disclosure

Traditionally, companies are reluctant to reveal any information about their positions for fear of losing a competitive advantage. One benefit of VAR, however, is that it provides an aggregate measure of risk that is nondirectional. This reveals no information about the sign of the positions and hence should alleviate some of these concerns.

Disclosure about trading activity usually appears in two places in annual reports:

- *Management discussion and analysis.* This section typically contains a narrative statement of the types of risks the firm is exposed to. More detailed information includes a qualitative description of risk management procedures, objectives, and strategies for using derivatives and quantitative information about market and credit risks.
- *Financial statements.* This section describes the financial position of the firm and, depending on national accounting rules, can include information about derivative positions in footnotes. Annual financial statements and footnotes are *audited* by independent accountants.

Table 15-3 provides a summary of disclosure of market risks by major banks, securities firms, and insurance companies. Of the 44 institutions surveyed in 2002, 98 percent provided quantitative information about their market risk. Ten years before, only 5 percent of this group disclosed such information. Thus nearly all large financial institutions now report quantitative information about their market risk with VAR.

This vast improvement came at the prodding of regulators. The Joint Forum, a group composed of central bankers, securities regulators, and insurance regulators, provided detailed templates for disclosure of various financial risks.[3]

TABLE 15-3

Disclosure of Market Risks by Financial Institutions

		Institutions Providing Quantitative Info	
Year	Total Examined	Total	Percent
1993	79	4	5%
1994	79	18	23%
1995	79	36	46%
1996	79	50	63%
1997	78	63	81%
1998	71	47	66%
1999	57	49	86%
2000	55	47	85%
2001	54	48	89%
2002	44	43	98%

Source: Basel Committee surveys. Definitions changed after 1998, which induces a break in the series.

The Joint Forum (2001) reached three main conclusions. First, to be meaningful, disclosures should include a balance of quantitative and qualitative information. Qualitative information provides context and perspective but is not sufficient. Second, disclosures should be consistent with firms' own risk management practices after due reporting of the parameters used. This lowers the cost of disclosures but unfortunately makes it more difficult to compare risks across institutions. Third, disclosures should cover more than the end of each period because excessive reliance on end-of-period information can create incentives for "window dressing," or the manipulation of positions before the disclosure dates. Ideally, institutions should provide summary VAR figures on a daily, weekly, or monthly basis, perhaps in a graph. They should also compare their daily profit and loss (P&L) information to VAR numbers to give some indication of the effectiveness of the risk measurement system. An example is shown in Box 15-3.

15.2.2 Why Risk Management Disclosures?

The Basel Committee (1995b) states that disclosure

> . . . can reinforce the efforts of supervisors to foster financial market stability in an environment of rapid innovation and growing complexity. If provided with meaningful information,

investors, depositors, creditors and

BOX 15-3

DEUTSCHE BANK'S VAR DISCLOSURES

Deutsche Bank (DB) is among the banks providing the most detailed analysis of its trading risk. DB breaks down its total trading VAR of 66.3 million euros as of December 2004 into business lines. This shows, for instance, that the largest VAR is for the interest-rate trading unit.

	VAR in Millions of Euros			
	Average	Maximum	Minimum	Year End
Interest-rate risk	61.7	91.1	39.7	41.1
Equity-price risk	30.8	45.1	19.9	42.6
Foreign-exchange risk	10.6	25.9	2.9	17.2
Commodity-price risk	7.0	10.8	3.8	5.1
Diversification effect	(38.4)	(61.5)	(28.1)	(39.8)
Total	71.6	97.9	54.5	66.3

In addition to year-end values, DB gives average, maximum, and minimum VARs. The annual report also plots the daily VAR. It provides a discussion of the performance of the model, which is compared with the bank's actual and hypothetical trading revenues.

counterparties can impose strong market discipline on financial institutions to manage their trading and derivatives activities in a prudent fashion and in line with their stated business objectives.

The view is that disclosure of quantitative information about market risk is an effective means of *market discipline*, or scrutiny by shareholders, debtors, and financial analysts. Firms that fail to reveal this information may be susceptible to market rumors, possibly resulting in loss of business or funding difficulties. Market discipline should manifest itself by requiring "a higher return from funds invested in, or placed with, a bank that is perceived to have more risk."[4]

Disclosure of market risks is also one of the three pillars of the new Basel II agreement (2005c).

Transparency also should lead to greater *financial market stability*. Indeed, the LTCM saga is a perfect illustration of how an institution can build up unreasonable amounts of leverage while disclosing very little information to the rest of the world, subsequently creating a near disaster in financial markets.[5]

Arguments for disclosure also apply to nonfinancial institutions. Lev (1988) develops a theory that rationalizes *mandated* disclosure requirements. The gist of the argument is that disclosure may be in the best interest of the company itself. The reason is that uninformed investors who feel they are not receiving enough information from a company can react by choosing to do less trading in that company's stock. Thus *asymmetric information* leads to lower trading volumes, higher trading costs, and perhaps lower equity values, which is not socially optimal.

Indeed, the industry often fails to make voluntary disclosures of information that would be relevant to investors. For instance, a *coordination problem* arises if each firm benefits from disclosure only if all other firms likewise disclose. If so, this market failure may require disclosure regulations. This is why disclosure regulations are present "across practically all free-market economies."

The issue then boils down to an evaluation of whether the benefits of disclosure of market risks are greater than the cost imposed on corporations. The Securities and Exchange Commission (SEC, 1998a) finds that its new market risk disclosures "provide investors and analysts with new and useful information." For example, analysts said that disclosures may allow investors to avoid investments in companies that are deemed too risky. A bank also said that it is now using the new market risk disclosure in its evaluation of loan applications. On the cost side, companies required to disclose found the rules "not terribly costly," with estimates ranging from $10,000 to $50,000. More generally, if disclosures are based on a firm's own risk management system, the marginal cost should be low. In addition, not only do these quantitative disclosures provide information on market risk otherwise difficult to assess, but they also bring some reassurance that a risk management system is in place.

15.2.3 Disclosure Examples

Table 15-4 compares the information provided in annual reports for a group of global financial institutions. The table shows trading VARs over 1 day at the 99 percent confidence level. It also displays total assets, the actual market risk charge (for commercial banks only), and risk capital.

As an example, J.P. Morgan Chase reports a trading VAR of $72 million. Following the Basel rules, multiplying this number by 3 X $\sqrt{10}$ translates into a general-market risk charge of $684 million. The next column shows the actual regulatory market-risk charge of $5,576 million, which includes other factors.[6] This is a small fraction of its total riskbased capital of $97,000 million. Of course, the bank's capital also must absorb other risks, in particular credit risks.

TABLE 15-4

VAR Reporting (Millions of Dollars)

Institution	Total Assets	Trading VAR 99%, 1 day	VAR Times $3\sqrt{10}$	Market-Risk Charge	Risk Capital
U.S. Banks					
Bank of America	$1,111,000	$48	$455	$5,184	$92,000
Citigroup	$1,481,000	$80	$759	$840	$87,000
JPM Chase	$1,157,000	$72	$684	$5,576	$97,000
U.S. Securities Firms					
Goldman Sachs	$531,000	$93	$886	NA	$23,000
Merrill Lynch	$648,000	$58	$550	NA	$28,000
Morgan Stanley	$771,000	$113	$1,073	NA	$29,000
Non-U.S. Banks					
Deutsche Bank	$1,142,000	$90	$855	$1,095	$39,000
HSBC	$1,267,000	$38	$357	$4,313	$91,000
RBS	$1,005,000	$28	$267	$2,647	$83,000
UBS	$1,522,000	$99	$942	$1,274	$31,000

Source: Financial reports as of December 2004. VAR numbers have been adjusted to a 99 percent confidence level and horizon of 1 day.

These numbers are useful, for instance, to assess claims that banks are now taking too much risk. Deutsche Bank, for instance, increased its trading VAR from $62 to $90 million from 1998 to 2004. Even though its risk capital has remained relatively constant over this period, around $39,000 million, the current market-risk charge of $1,100 million can be absorbed easily by its capital.

In addition, disclosures of VAR numbers should be useful to compare the trading-risk profiles of different institutions after harmonizing the VAR parameters. The table shows that Morgan Stanley takes on more trading risk than

Merrill Lynch, for instance.

Indeed, a burgeoning empirical literature reports that these VAR disclosures are informative. Jorion (2002b) finds that quarterly VAR numbers are useful forecasts of the variability in trading revenues, especially when comparing across banks. Banks with large VARs tend to have bigger swings in trading revenues, even after controlling for other factors such as the size of trading assets or derivatives notionals. Liu *et al.* (2004) find that the predictive power increases with bank technical sophistication and over time.[7]

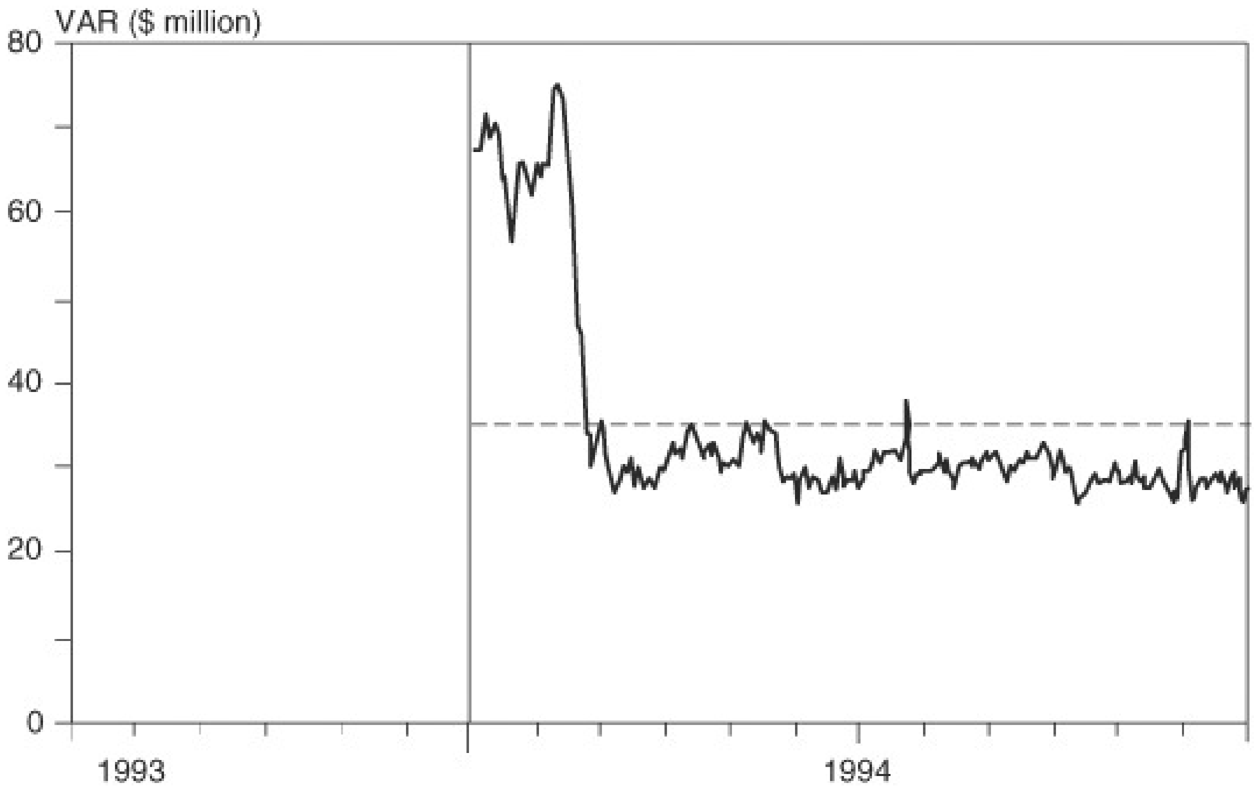
15.3 VAR AS A RISK-CONTROL TOOL

VAR is not only useful for reporting purposes but also as a risk-control tool. VAR limits can be used to control the risk of traders as a supplement to traditional limits on notional amounts. Such limits also can be used at the level of the overall institution. Often, the mere act of quantifying risk is sufficient to provoke the institution into risk reduction. Goldman Sachs, the U.S. investment bank, for instance, was caught by the U.S. interest-rate hikes of 1994. In response, it developed what is considered one of the best risk management groups on Wall Street. Interestingly, its VAR chart showed a much lower risk profile after risk management measures were in place. Yet the company made more money than ever.

15.3.1 Adjusting Firmwide VAR

These VAR limits can be adjusted as a function of the perceived return-to-risk tradeoff. In the face of an increasingly volatile environment, for instance, a sensible response is to scale down positions. An example of this reaction is presented in Figure 15-3, which plots the evolution of the daily VAR for Bankers Trust's combined portfolio during 1994.

FIGURE 15-3
Bankers Trust's VAR.



The figure shows that the bank's VAR started at about $70 million at the beginning of January 1994, then declined sharply during February to about $30 million, and, except for minor fluctuations, remained at that level for the rest of the year.

Bankers Trust explains this drastic change as follows:

> The year began with a sharp, global increase in interest rates. . . . The Corporation responded to this adverse and unsettled market environment through an orderly withdrawal in the first quarter of 1994 from substantial market positions. . . . The risk reduction that occurred during February 1994 reflected the Corporation's decision to reduce its exposure in its Trading and Positioning accounts due to fluctuations in interest rates. . . . Also, interest rate risk was the single largest source of market risk during the year with an average Daily Price Volatility of $29 million. In comparison, the Corporation's average Daily Price Volatility across all market risks was $35 million in 1994.

BOX 15-4

DEUTSCHE BANK'S VAR LIMITS

Deutsche Bank's annual report for 2004 states:

> Our value-at-risk disclosure is intended to ensure consistency of market risk reporting for internal risk management, for external disclosure and for regulatory purposes. The overall value-at-risk limit . . . was 80 million euros in the time period from January 1 to March 9, 2004 and 90 million euros from March 10 to December 31, 2004 . . . . Four temporary excesses to the Group limit were approved by our Board of Managing Directors in 2004.

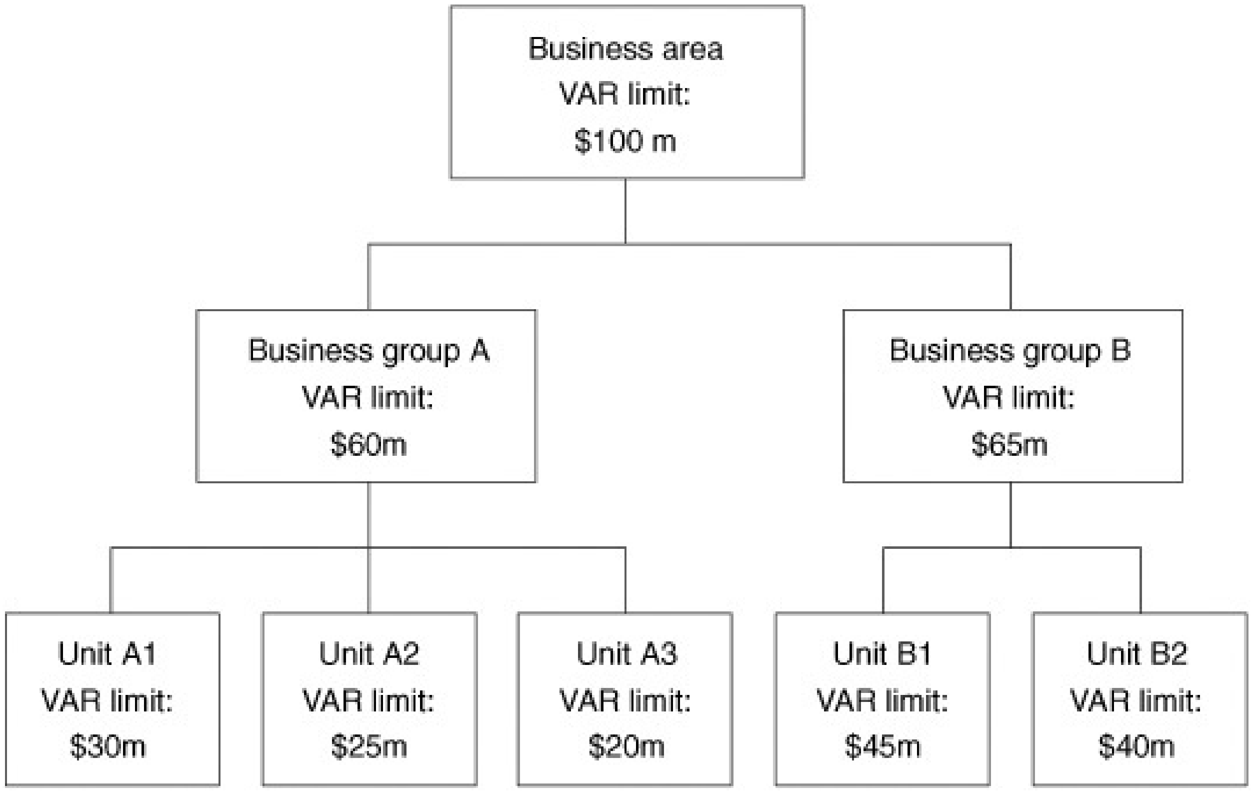
In other words, this withdrawal is rationalized by the increased volatility of the fixed-income market because the Federal Reserve started a series of interest rates hikes in February 2004. Bankers Trust, like most financial institutions, must have had long positions in long-term bonds financed by short-term debt. This positive duration could have exposed the bank to substantial losses. In response, Bankers Trust scaled down its positions substantially. Thus VAR can help as a guide to decide how much risk should be allowed. Another example is in Box 15-4.

15.3.2 Adjusting Unit-Level VAR

At the business area or unit level, VAR can be used to set position limits for traders and to decide where to allocate limited capital resources. A great advantage of VAR it that it creates a common denominator with which to compare various risky activities.

Traditionally, position limits are set in terms of notional exposure. A trader, for instance, may have a limit of $10 million on overnight positions in 5-year Treasuries. The same limit for 30-year Treasuries or in Treasury bond futures, however, is substantially riskier. Thus notional position limits are not directly comparable across units.[8] Instead, VAR provides a common denominator to compare various asset classes and can be used as a guide to set position limits for business units.

FIGURE 15-4
Setting VAR limits.



In addition, since VAR accounts for correlations, position limits can be set such that the risk limit at higher levels can be lower than the sum of risk limits for individual units. As Figure 15-4 shows, diversification allows the risk limit for business group A to be $60 million, which is less than the sum of $75 million for units A1, A2, and A3, owing to diversification benefits.

Figure 15-5 gives an example of a summary position report for a firm with exposure to foreign currencies and fixed-income markets. The two business units have differing positions in U.S. dollars, euros, and yen. All positions and risk measures are expressed in millions of dollars.

Figure 15-5 shows the FX position for each unit as well as the estimated VAR and the VAR limit. For instance, unit A has a 1-day VAR at the 95 percent confidence level of $1.28 million against a limit of $2 million. Unit B has a higher FX VAR of $2.73 million and limit of $3 million. Note that the total FX VAR for the two units, $1.94 million, is substantially lower than the sum of individual VARs because of diversification. The total FX VAR limit of $4 million also reflects diversification.

FIGURE 15-5

Example of VAR reports.

	Unit A	Unit B		Total
FX Position	–150	120	$	–30
(spot equiv. $MM)	100	80	Euro	180
	50	–200	Yen	–150
VAR	$1.28	$2.73		$1.94
Limit:	*$2.00*	*$3.00*		*$4.00*
Interest Rate Position	–300	0	$	–300
(2- year equiv. $MM)	90	150	Euro	240
	100	–500	Yen	–400
VAR	$0.68	$0.67		$0.81
Limit:	*$2.00*	*$3.00*		*$4.00*
Total Position				
VAR	$1.27	$2.74		$2.01
Limit:	*$3.00*	*$4.50*		*$5.00*

The next panel shows the bond positions, expressed in 2-year equivalents. In other words, the notionals have been adjusted by the ratio of the duration to that of a 2-year note. The total interest-rate VAR is $0.81 million, below its limit of $4 million. Finally, the last lines display the VAR for each unit, their limits, and the total portfolio VAR, which is only $2.01 million. Such a report gives a broad perspective on positions and risk profiles. Because it contains the essential information in a simple summary form, any exceedence of limit can be detected quickly.

VAR limits, however, cannot be the sole deciding factor for positions. If market volatility jumps up suddenly and is directly reflected in VAR, the risk manager may want to give some leeway in the enforcement of limits or even increase the VAR limit accordingly. Otherwise, it may prove too costly to liquidate positions under difficult market conditions. It is also useful to bear in mind that various VAR models have different responses to temporarily higher volatility. The RiskMetrics model, for instance, has a very fast and permanent response to changes in risk. In contrast, the Basel model is designed to have a slower reaction time. Thus implementation of limits should be subject to the educated judgment of the risk manager.

15.4 CONCLUSIONS

We identified three steps in the application of VAR: passive, for risk measurement; defensive, for risk control; and active, for risk management. By now, VAR has become a standard method for measuring and reporting market risk. VAR is ideally suited to large-scale portfolios and provides a risk measure built up in a comprehensive and consistent fashion. It has been endorsed most strongly by regulators and is now used widely.

VAR is also used to control risk as a supplement to traditional limits on notional amounts, limits on exposures, and stop-loss limits. Here again, VAR allows consistency and comparability across various units.

Finally, VAR is starting to be used as an active management tool for the allocation of capital. For the first time, institutions have the tools to evaluate the tradeoff between expected profits and risk. This will be the subject of Chapter 16.

APPENDIX 15.A
Modeling Economic Exposure

Setting up a cash-flow-at-risk (CFAR) model requires the modeling of the cash flows' economic exposures to financial risk factors. Exposures can be complex for nonfinancial corporations, though. They depend on notional amounts as well as on the competitive environment in which the firm operates.

Consider, for instance, a U.S. exporter to Europe. The question is, how are revenues affected by the exchange rate? If the company competes with other U.S. firms, a depreciation of the euro will affect all exporters equally, and they may be able to raise prices in euros to cover their costs if the demand for the product is inelastic. In this case, the company has low currency exposure. In contrast, if the U.S. exporter competes with foreign firms, it may not be able to raise its prices, leading to potentially large losses. In this case, it has no market power, and the exposure can be substantial.

To be more specific, let us write export revenues as a function of foreign currency prices P^*, of quantities sold Q, and of the exchange rate S expressed in dollars. Assume that the price P^* is set so as to maintain Q. Now define the *elasticity* of P^* with respect to S as the ratio of the percentage change of P^* for a given percentage change of S. This elasticity η is defined by

$$\frac{(P_1^* - P_0^*)}{P_0^*} = \eta \frac{(S_1 - S_0)}{S_0} = \eta \frac{\Delta S}{S} \tag{15.1}$$

If quantities are unchanged, we can write dollar revenues as

$$R = P_1^* Q S_1 = P_0^* \left(1 + \frac{\Delta S}{S}\eta\right) Q S_0 \left(1 + \frac{\Delta S}{S}\right) \tag{15.2}$$

If the U.S. exporter has no market power in the foreign market, the foreign price the company sets cannot be affected by the exchange rate and $P_1^* = P_0^*$, which implies that η = 0. In this case, Equation (15.2) shows that dollar revenues will fall by the full amount of the depreciation in *S*. This is the example of fixed exposures presented in Table 15-2.

In contrast, if prices are set in dollars all over the world, any fall in the value of the foreign currency can be offset by an increase in the local price *P**. With a perfect offset, the elasticity is η = –1. The two terms in Δ*S*/*S* will cancel, and dollar revenues will not be affected.

In the intermediate case, the U.S. exporter may be able only to raise prices partially to offset the fall in the exchange rate. For example, say that η = –0.5. In this case, the Monte Carlo simulations used to derive the distribution of cash flows can be easily modified to take into account this competitive effect.

Another interesting example is that of a commodity producer whose revenues are defined by

$$R = PQ \tag{15.3}$$

The variability in revenues in large part depends on the correlation between movements in prices and quantities, which reflects the nature of shocks to market prices.

Supply shocks are, for instance, due to weather. This implies that particularly good weather across the region increases the size of the crop, which will push prices down. Thus *Q* and *P* move in different directions. This negative correlation between *P* and *Q* makes revenues more stable, lessening the need to hedge price risk.

On the other hand, *demand shocks* are due to changes in consumption patterns. For instance, a drop in the demand for the commodity will decrease both prices and quantities sold. Thus *Q* and *P* move in the same direction. This will make revenues more volatile than otherwise, creating a sharper need for risk management.

QUESTIONS

1. What was the initial purpose of VAR?
2. What is the three-stage evolution of VAR applications in risk management?
3. What general factors motivate the trend toward centralized risk management?
4. Explain company-specific factors that would favor the development of VAR systems.
5. Would you expect JPM Chase or Fannie Mae to rely more on VAR? (JPM Chase is exposed to a great variety of financial risks; FNMA is exposed mainly to U.S. interest-rate risks.)
6. Would you expect JPM Chase or Merrill Lynch to rely more on VAR? (JPM Chase takes large proprietary trading positions; Merrill largely hedges its risks.)
7. Define CFAR. How is it measured?
8. How can CFAR be reduced with derivatives?
9. Why is CFAR more complex than traditional VAR models?
10. Toyota builds cars in Japan that are exported to the United States. Explain how this exposure could be hedged using forward contracts on the yen/dollar rate or by borrowing in yen or dollars.
11. A company exports a U.S.-built product to Germany, expecting revenues of $100 million. Its forecast profit margin is 10 percent. The German customers, however, only want to pay in euros. The dollar/euro rate is approximately normally distributed with volatility of 10 percent per annum. What is the probability of losing money on the exports?
12. Assume now that if the euro falls by 10 percent, the company will be able to increase prices in euros, thereby maintaining the same price in dollars. What is the probability of losing money on the exports?
13. “Companies have been known to smooth earnings with *earnings management*. This occurs when managers use judgment in financial

reporting to alter financial reports to mislead stakeholders about the company's economic performance." With the recent corporate scandals, earnings management is now more difficult. Should this increase or decrease the use of financial derivatives?

14. Why is it important to disclose quantitative information on market risk?
15. According to the Joint Forum (2001), why should disclosures cover more than the end of each reporting period?
16. If disclosure of risk is beneficial to companies, why do we need regulatory intervention? Wouldn't companies disclose the information voluntarily anyway?
17. Does the empirical evidence show that risk disclosures are useful forecasts of future risks?
18. Assume that we observe that the VAR of a bank has gone up by 20 percent from one year to another. Give two interpretations of this observation.
19. Explain why the sum of VAR limits at the level of individual units is greater than the VAR at the level of the business unit.

CHAPTER 16
Using VAR for Active Risk Management

> Returns these proprietary businesses produce look very different on a risk-adjusted basis.
>
> —*Analyst's comment at the announcement that Travelers was disbanding Salomon's bond arbitrage unit*

Investors have long recognized that financial management is about balancing return against risk. Before the widespread application of value at risk (VAR), however, institutions did not have the tools to evaluate the risk-return tradeoff for business lines. Banks relied on measures such as return on assets (ROA) or return on [book] equity (ROE) that totally ignored differences in the risk of various activities. Since then, risk-adjusted performance measures based on VAR have become a key building block for modern financial risk management.

This chapter discusses how VAR can be used to manage risk actively. Active risk management functions include performance evaluation, capital allocation, and strategic business decisions.

The chapter starts by showing how VAR can be viewed as a measure of "economic" risk capital necessary to support a position. This generalization of the equity investment concept is driven by the complexity and leverage of financial products and institutions. This has made it crucial to evaluate activities that require little up-front capital but create contingent liabilities. Section 16.1 shows how such activities can be charged risk-based capital.

Assigning economic risk capital to a trading activity allows *risk-adjusted performance measurement* (RAPM), as shown in Section 16.2. This gives incentives to traders to pay attention to the risks they are incurring. Economic capital is best measured using position data, as with VAR. Another approach is to use the volatility of historical earnings, which is discussed in Section 16.3.

The purpose of these methods is to provide a uniform yardstick to measure the performance of traders or business units but also to guide capital-allocation decisions. *Capital allocation* is the process by which a firm assigns economic capital to transactions, products, or business lines. In this context, Section 16.4 discusses whether risk adjustments should take into account correlations with other business units.

Of course, VAR is only one facet of capital decisions. It just quantifies the existing risk. It does not tell us whether we should or should not take a risk. To answer this question, we need to also account for expected profits. The tradeoff between profits and risks can be analyzed using the framework of shareholder value added (SVA). Section 16.5 explains how to use VAR as a strategic decision tool.

16.1 RISK CAPITAL

16.1.1 VAR as Risk Capital

VAR can be viewed as a measure of *risk capital*, or economic capital required to support a financial activity. This resolves the paradox of how to calculate rates of return on investments that require no up-front investment, such as futures. Consider, for instance, a bank with one investment only, a futures position with a notional of \$100, a margin of \$10, and a payoff (or dollar return) of \$5, perhaps in millions. This example is not unlike the situation of highly leveraged financial institutions such as commercial banks, investment banks, and hedge funds.

How do we compute the rate of return on this \$5 payoff so as to compare it with other investments? In other words, what is the relevant denominator? The notional is not relevant because it is never paid. Nor is the margin, which is a performance bond and may not provide the cushion desired by the bank.

Instead, we could consider the amount of *equity capital* that needs to be set aside to cover most of the potential losses at a predetermined confidence level. Taking into account only market risks, this equity capital is basically a market VAR measure. Therefore, VAR measures the *economic capital* (EC), defined as the aggregate capital required as a cushion against unexpected losses, that is,

$$\mathrm{EC} = \mathrm{VAR} \tag{16.1}$$

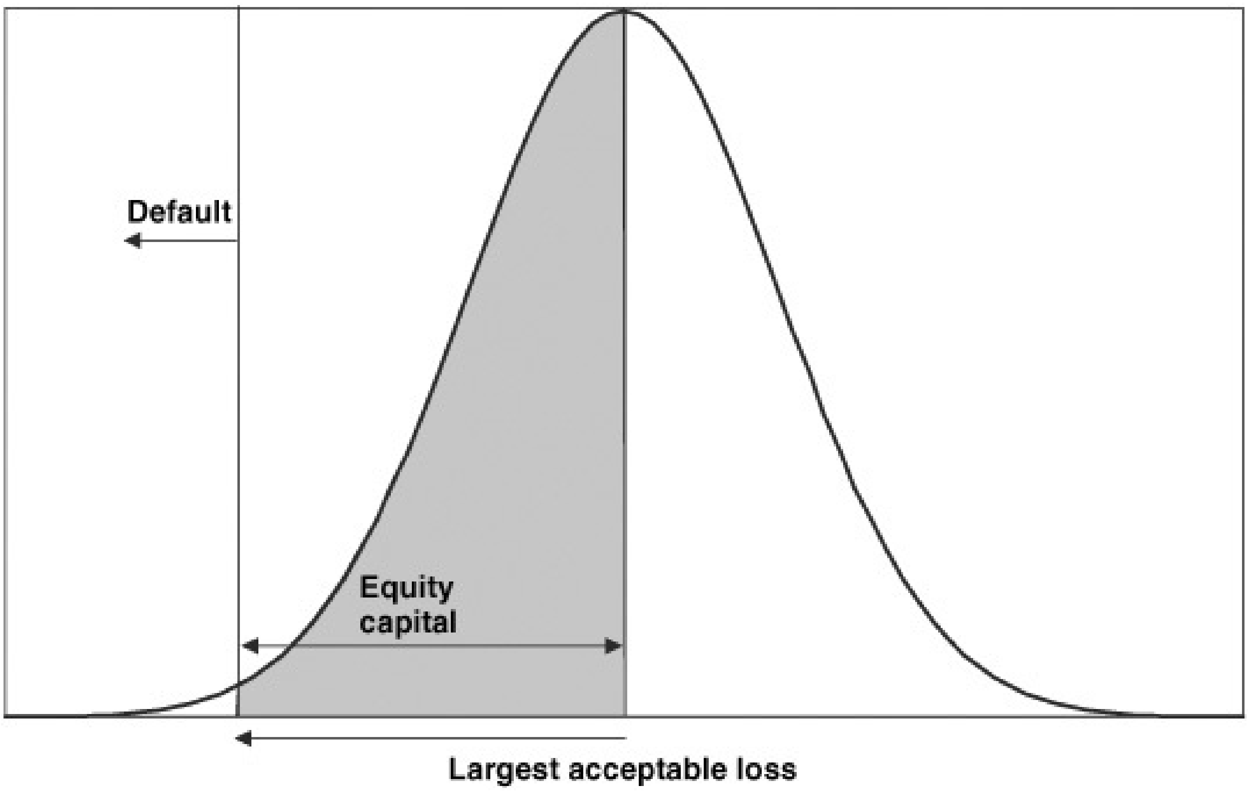
Note that this economic capital in general will differ from the bank's book equity, as well as from the regulatory risk capital.

Suppose, for instance, that VAR is estimated to be \$10 at the 99 percent level. As before, VAR is interpreted as the "largest" acceptable loss the bank is willing to suffer. To cover this loss, the bank must maintain adequate equity capital. In other words, VAR is the amount of capital a firm allocates to self-insurance. This is illustrated in Figure 16-1. Appendix 16.A shows how to translate the VAR measure on the target horizon to a current amount of economic capital.

Note that this capital is leveraged. That is, the equity is supported by borrowing. Ideally, the bank should select the amount of leverage to balance the marginal benefits of increased leverage with the marginal costs of increased default probability.[1] In this case, the rate of return on economic capital is $5/$10 = 50 percent. With a lower capital base, the rate of return would be even higher. The drawback, of course, is that the investment is more risky. The choice of this optimal capital structure should be reflected in the choice of the confidence level.

FIGURE 16-1

Equity capital as a VAR measure.



16.1.2 Choosing the Confidence Level

This interpretation of VAR also helps to determine the appropriate confidence level. Consider a hypothetical bank for which all the risks are captured by the VAR measure. A 1-year horizon is selected, assuming that it corresponds to the horizon needed to adjust the level of capital, that is, to raise additional equity. If the bank suffers a loss greater than the VAR within a year, its equity capital is wiped out, and the bank defaults.

The bank then can set the VAR confidence level in relation to the desired credit rating. This is illustrated in Table 16-1, which reports the relationship between the credit ratings supplied by a rating agency, Moody's, and their historical 1-year default rates. The last columns report the *equity coverage*, measured as the number of standard deviations necessary to achieve the desired credit rating.

TABLE 16-1

Choosing Equity Coverage from the Credit Rating (Multiples of Standard Deviation Assuming Various Distributions)

Desired Rating (Moody's)	1-Year Probability of Default	Equity Coverage (Multiple of SD)	
		Normal	Student *t*
Aaa	0.01%	3.72	9.08
Aa1	0.02%	3.54	8.02
Aa2	0.02%	3.54	8.02
Aa3	0.03%	3.43	7.46
A1	0.05%	3.29	6.79
A2	0.06%	3.24	6.56
A3	0.07%	3.19	6.37
Baa1	0.13%	3.01	5.67
Baa2	0.16%	2.95	5.44
Baa3	0.70%	2.46	4.01
Ba1	1.25%	2.24	3.52
Ba2	1.79%	2.10	3.23
Ba3	3.96%	1.76	2.62
B1	6.14%	1.54	2.30
B2	8.31%	1.38	2.08
B3	15.08%	1.03	1.65

Note: Equity coverage is defined as the number of standard deviations (SD) necessary to achieve desired default probability. Two distributions are used, the standard normal and student *t* with 6 degrees of freedom.

Choice of confidence level VAR can be set at a value such that the probability of losses exceeding VAR is equal to the probability of default for this risk.

As an example, assume that the market risk of the bank is such that the annual standard deviation of profits and losses (P&L) is $1 billion. How much capital should the bank set aside to achieve an Aa2 credit rating? The table shows that if the distribution of losses is normally distributed, the required amount of equity is 3.54 times $1 billion, or $3.54 billion. If, however, the bank estimates that the distribution of losses has fatter tails than the normal distribution, it could select the multiplier from a student *t* distribution, for example. In this case, the required amount of equity is higher, at $8.02 billion. Alternatively, EVT could be used to model the tails. Banks routinely provide their economic capital measure using a 99.98 percent confidence level, which indeed corresponds to a target credit rating of Aa. Box 16-1 shows how Bank of America linked its desired credit rating to a confidence level and to a multiple of standard deviations.[2]

Suppose then that the bank sets its equity capital at $8 billion. It has $4 billion in subordinated debt. With a loss equal to $8 billion, the value of equity is wiped out. With a loss equal to $12 billion, holders of the subordinated debt are also wiped out. With a greater loss, depositors and the government deposit insurance fund are also at risk.

16.2 RISK-ADJUSTED PERFORMANCE MEASUREMENT

Armed with this measure of risk capital, the VAR methodology allows us to compare traders, or business units, or investment portfolios, that generate large revenues with little apparent need for capital. In the preceding example, we computed a *risk-adjusted performance measurement* (RAPM) as the dollar profit over the dollar VAR, which is $5/$10, or 50 percent.

BOX 16-1

RISK MANAGEMENT AT BANK OF AMERICA

In 1993, Bank of America decided to implement risk-adjusted performance measurement throughout the organization. The bank identified major categories of risk for which unexpected losses required holding economic capital.

The amount of economic capital attributed to all business units was set so as to guaranteee the solvency of the bank at a 99.97 percent confidence level over 1 year. This 0.03 percent probability of default was determined to reduce the risk of the bank to the level of AA-rated companies.

The time horizon was set at 1 year as a compromise between the time frame for credit risk and market risk. While relatively arbitrary, the most important aspect of this choice was to ensure consistency across all business units and sources of risk.

For most market risks with normal distributions, the bank estimated a capital coverage of 3.4 times the standard deviation of unexpected losses. For asymmetric risks such as credit risks, the capital coverage was set at 6 standard deviations.

TABLE 16-2

Computing RAPM

	Profit	Notional	Volatility	VAR	RAPM
FX trader	$10 million	$100 million	12%	$28 million	36%
Bond trader	$10 million	$200 million	4%	$19 million	54%

RAPM allows institutions to compare units that have very different risk capital needs. Let us go back, for example, to the example of traders in Chapter 4 (Box 4-2). Two traders achieved profits of $10 million each over the last year. How do we compare their performance? How do we decide which unit should be given additional risk capital? Table 16-2 shows how RAPM helps to answer this question.

Assume a constant notional amount of $100 million and $200 million for the FX and bond traders, respectively, and an annual volatility in currency and bond markets of 12 and 4 percent. We can compute the risk capital at the 99 percent level over 1 year as $28 million and $19 million for the two traders.

The risk-adjusted performance then is measured as the profit divided by the economic capital risk charge, that is,

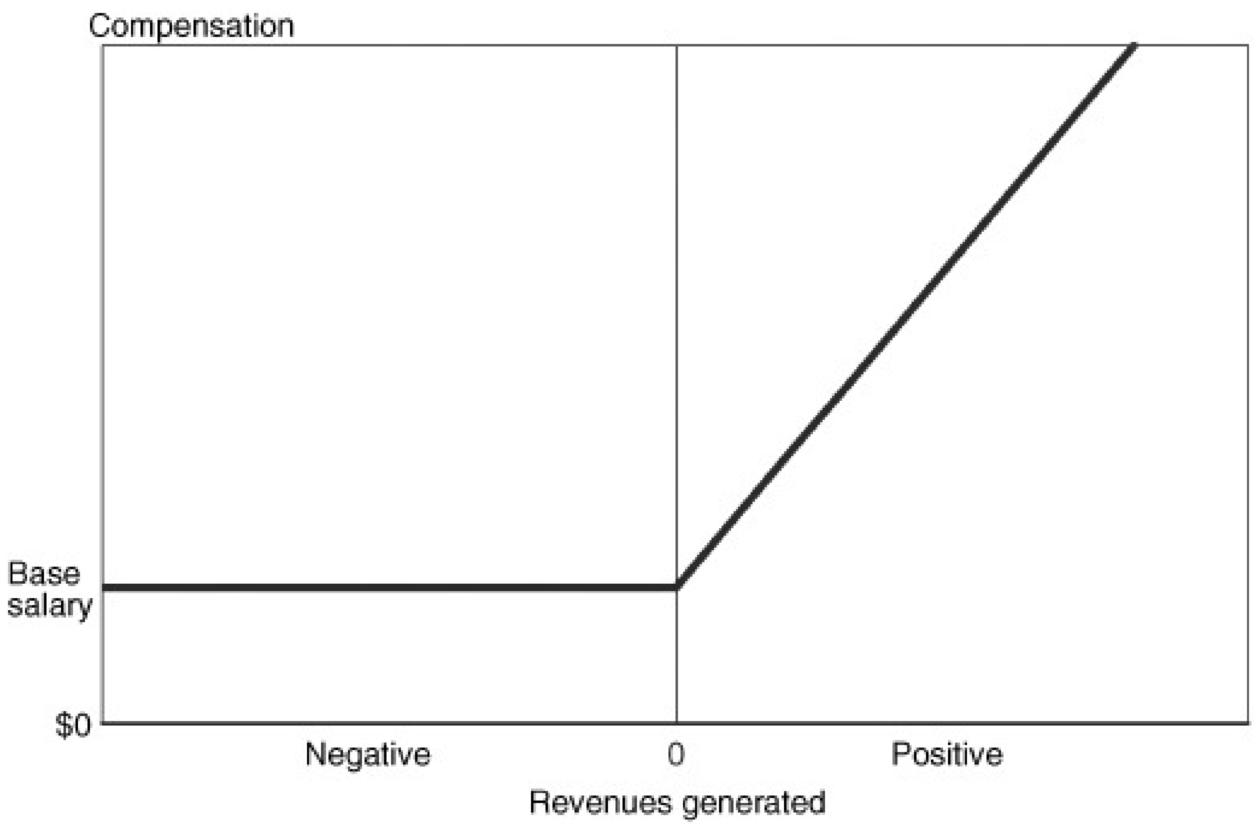
$$\text{RAPM} = \frac{\text{profit}}{\text{EC}} \tag{16.2}$$

Table 16-2 shows that the bond trader has a higher RAPM than the currency trader, at 54 versus 36 percent, owing to the lower risk capital requirement. Thus this trader makes better use of scarce equity capital.

The RAPM approach allows managers to adjust the profit performance of traders for the risk they are taking, using VAR as the measure of economic capital. This risk adjustment provides a solution to the *moral hazard problem* inherent in linking trader bonuses to profits. Without controlling for risks, traders may have an incentive to take more aggressive positions. This is due to the asymmetric payoff profile of a trader's compensation. Traders typically receive a percentage of (or bonus based on) profits, which can be quite large. In contrast, if they lose a large amount of money, they still draw their base salary. The worst penalty they can suffer is to be fired. This limited-liability profile is described in Figure 16-2.

FIGURE 16-2

The optionlike bonus profile.



BOX 16-2

ADJUSTING BONUS STRUCTURES

Typically, traders are given a bonus that depends on their profits, when positive. Attempts at changing this compensation structure have met with little success.

One such experiment was attempted by Warren Buffett as part of his efforts to reshape Salomon Brothers after he became chairman in 1991. Buffett pledged to curb the excesses of what he called an "irrational" compensation system. Salomon cut bonuses paid to managing directors by an average of 25 percent. At the same time, a bigger share of the bonus was paid in company stock rather than in cash.

These changes, however, were undermined by competitors willing to pay top dollar to attract the most talented individuals. Several Salomon managers walked out, and the new structure had to be scrapped.

Regulators have complained that this incentive system encourages excessive risk-taking. They have not taken action so far, owing to the global nature of trading. The Bank of England, for instance, has recognized that any regulatory action on this front simply will prompt a trading business to move to another jurisdiction.

More fundamentally, it is up to the shareholders to decide whether they want to be in the business of granting a license to speculate to their traders. No doubt this explains why the famous U.S. bond arbitrage unit at Salomon was disbanded in 1998. The Travelers Group, the parent company that had just acquired Salomon, was unwilling to accept the high profit volatility in the bond arbitrage business.

This payoff profile is akin to a long position in an option. Since option values increase with volatility, traders have an incentive to increase the risk of their positions. Such behavior may be optimal for them but not for the corporation (see Box 16-2). As with risk-based capital requirements for financial institutions, imposing an ex ante penalty for higher risk attempts to curb this behavior.

More generally, RAPM should take into account other financial risks. It can be measured as

$$\text{RAPM} = \frac{\text{revenues} - \text{costs} - \text{expected losses}}{\text{VAR}} \tag{16.3}$$

where

$$\text{VAR} = \text{market VAR} + \text{credit VAR} + \text{operational VAR} - \text{diversification} \tag{16.4}$$

This definition includes not only market but also credit and operational risks and possibly diversification benefits. Profits should take into account directly observable costs but also expected futures losses owing to credit and operational risks, called *reserves*. Here, VAR measures the capital necessary to cushion against unexpected market, credit, and operational risks and can be defined as *overall economic capital*.

16.3 EARNINGS-BASED RAPM METHODS

VAR measures of economic capital are based on a structural analysis of the current positions. Another, simpler approach is to use information in earnings. Suppose that we are running a trading operation for which we have a *history* of daily profits and losses (P&L) for all the units. The standard deviation of P&L can be used to compute a measure of *earnings at risk* (EAR). The earnings-based RAPM consists of average profits divided by the EAR measure.

Earnings-based methods are easy to implement because they only require historical data on trader or unit performance. They also give some indication of the risks undertaken.

This approach, however, has shortcomings. First, EAR measures historical risks rather than giving a forward-looking measure of risk. It reflects past decisions rather than current profiles. Second, the risks that are not realized over the sample period will be missed by the EAR approach. As Box 16-3 shows, this approach is also prone to model risk if it relies on the trader's valuation model or parameters to measure profits. Finally, focusing on earnings does not help in the understanding of the drivers of risk, nor does it provide tools to control volatility.

On the other hand, earnings-based measures are more useful at a broader level of aggregation, for example, when analyzing business units or when comparing disparate business activities. Because they focus on the actual earnings volatility, such measures cover all sources of risk, including business and operational risks, that cannot be measured easily with structural position-based models, which are best developed for measuring market risks.

Earnings-based measures also provide a link between required risk capital and the perception of risk by investors, earnings volatility. The methodology developed in the following sections, however, applies as well to position-based (VAR) and to earnings-based risk measures. Table 16-3 summarizes the pros and cons of position-and earnings-based VAR measures.

BOX 16-3

SHORTFALLS IN EARNINGS MEASURES

In 1987, Bankers Trust had a close encounter with disaster. Apparently, the firm had enjoyed profits of $600 million from foreign currency trading, half of which was attributed to Andy Krieger, its star trader, who specialized in selling long-dated currency options. He was so successful that he was given a trading limit of $700 million which represented a quarter of the bank's capital. The problem was that the positions were illiquid and difficult to value.

After he left, the firm unraveled his trades and then realized that his trading revenues had to be lowered by $80 million. In addition, the bank was shocked to learn that it sometimes had $2 billion at risk. "Its whole future was riding on the judgment and trades of one 32-year-old banker," an executive said.

The bank vowed this would never happen again. Top management put in place more robust risk-control systems and made sure that traders were evaluated on their risk-adjusted performance.

This episode illustrates the shortcoming of earnings-based volatility measures. Marking to market the positions is notoriously difficult for illiquid markets such as long-dated options in exotic currencies. Even if measured correctly, the profit pattern of short option positions gives a misleading picture of the true risk. These positions generate small stable profits as long as markets behave normally but can lose large amounts of money in times of increased volatility.

16.4 ALLOCATION OF RISK CAPITAL

RAPM methods represent a generalization of a well-known performance measure, the *Sharpe ratio*, developed by William Sharpe in 1966. This ratio measures the ratio of average return, in excess of the risk-free rate, to the total volatility of returns, that is,

$$S_i = \frac{\bar{R}_i - R_F}{\sigma(R_i)} \qquad (16.5)$$

TABLE 16-3

Comparing Position-and Earnings-Based VAR

	Position-Based	Earnings-Based
Approach	Structural model: bottom-up	Aggregate model: top-down
Horizon	Forward-looking: uses current profile	Backward-looking: uses historical data
Best application	Similar businesses such as trading	Disparate activities such as fee income
Usefulness for control	Very useful: details risk drivers	Not so useful: no info on risk drivers
Ease of implementation	Difficult to set up: process data	Easy to calculate: uses P&L time series
Coverage	Covers only modeled risks (e.g., market)	Captures all risks (in historical data only)

where $\bar{R}_i$ is the average return on asset *i*, and $\sigma(R_i)$ its volatility. RAPM generalizes this formula to dollar numbers using risk capital instead of an up-front investment.

The performance-measurement literature, however, has long recognized that this risk adjustment ignores diversification considerations. The Sharpe ratio is appropriate when total risk matters, that is, when all an investor's wealth is invested in this asset. When the asset, however, is considered only in relation to a large, diversified portfolio, measuring risk by total volatility is inappropriate.

To address this problem, Treynor (1965) proposed an alternative performance measure emphasizing systematic rather than total risk. The *Treynor ratio* is the ratio of average return, in excess of the risk-free rate, to the contribution of this asset to the portfolio's total risk, that is,

$$T_i = \frac{\bar{R}_i - R_F}{\beta_i} \tag{16.6}$$

where β_i is the systematic risk of asset *i* relative to the bank's total portfolio *p*.

Similarly, the conventional RAPM definition is not entirely appropriate owing to the fact that the trader's risk is only one component of the bank's total risk. Generally, trading profits may be related to the risks of other departmental units. Consider, for instance, an institution with two units, a bond-trading desk and a futures desk. If both departments take long positions in anticipation of a decrease in interest rates, the total risk to the corporation will be very large. In contrast, if the futures desk goes short, the combined positions may be nearly risk-free. This is the case, for instance, when the two desks engage in arbitrage between the cash and futures markets. Applying VAR separately to each unit overstates the combined risk of each.

The Treynor ratio, however, can be generalized to deal with this issue. In fact, this is equivalent to the concept of *marginal VAR*, which was developed in Chapter 7. We defined the marginal VAR for security i, ΔVAR_i, as the change in total VAR owing to an increase in the allocation to i. This can be generalized to the level of the business unit or trader. The marginal RAPM measure is now

$$\text{Marginal RAPM}_i = \frac{\text{profit}_i}{(\Delta\text{VAR}_i)} = \frac{\text{profit}_i}{(\text{VAR} \times \beta_i)} \qquad (16.7)$$

This measure can be used as a guide to make decisions to enter or exit a particular business line. Its drawback, however, is that it does not fully allocate all the bank's capital to each unit.

To do this, we can resort to the previously defined concept of *component VAR*, which provides a partition of the firm's VAR into various components. The component RAPM measure (sometimes called *diversified*) is now where all the CVAR_i measures are assured to sum to the firm's total VAR.

$$\text{Component RAPM}_i = \frac{\text{profit}_i}{(\text{CVAR}_i)} = \frac{\text{profit}_i}{(\text{VAR} \times w_i \beta_i)} = \frac{\text{profit}_i}{(\text{VAR}_i \times \rho_i)} \qquad (16.8)$$

Table 16-4 gives an example of a bank with four lines of trading activity. The data represent the actual quarterly earnings of a large bank. The first column reports the annual profit; the second, the quarterly volatility. This can be translated into an annual VAR at the 99 percent confidence level by multiplying by 2.33 times the square root of 4, assuming normal distributions.

The conventional RAPM measure is based on the individual (sometimes called *undiversified*) VAR. Here, the return-to-risk ratio for the interest-rate desk appears to be much higher than that of other units.

The bank's total VAR is only about half the sum of the units' VARs owing to diversification effects. The next column reports the correlation of each unit with the total. The interest-rate desk displays a high correlation, whereas the FX desk has a negative correlation.

TABLE 16-4

Profit and Risk Measures ($ Millions)

Unit	Annual Profit	Quarterly Volatility	Individual VAR	RAPM	Correlation	Component VAR	Percent VAR
Interest rate	$1,636	$152	$708	2.31	0.851	$603	80%
FX	$381	$92	$426	0.89	−0.131	($56)	−7%
Equity	$123	$71	$332	0.37	0.554	$184	24%
Commodity	$50	$16	$74	0.68	0.322	$24	3%
Sum			$1,541				
Diversification			($786)				
Total	$2,189	$162	$755	2.90		$755	100.0%

The component VAR analysis, scaled in percent VAR, indeed shows that 80 percent of the bank's VAR is due to the interest-rate desk. In addition, the FX desk serves as a hedge against other risks because its component VAR is negative.[3] Thus additional capital should be allocated in priority to this unit because this will decrease the total risk of the bank. This example illustrates that the application of VAR in performance measurement depends on its intended purposes.[4]

- *Internal performance measurement* aims at rewarding units that produce the best performance within their allowed parameters. It requires a risk measure which does not depend on what other units do. For instance, the composition of a bonus pool should only depend on actions that a unit has full control over. Here, the individual (undiversified) VAR seems the appropriate choice.
- *External performance measurement*, in contrast, aims at allocation of existing or new capital to existing or new business units. Such decisions should be made with the help of marginal and component (diversified) VAR measures. Without an allowance for diversification effects, the risk of each activity becomes a function of the organizational structure. Using stand-alone risk measures could lead to the puzzling result that a reorganization of business units could change the aggregate capital requirement even though the total risk of the firm has not changed.

In practice, for performance-measurement purposes, risk capital typically is allocated according to the unit's own risk. The bank then reaps the benefits of diversification, which are allocated to the central unit. Indeed, this central unit is directly responsible for setting position limits for each unit and for deciding how much to allocate across business units. This method has the advantage of being robust and easy to implement.

16.5 VAR AS A STRATEGIC TOOL

More generally, VAR can be used at the strategic level. Risk-adjusted performance measures can be used to identify where shareholder value is being added throughout the corporation. The objective is to help management make decisions about which business lines to expand, maintain, or reduce, as well as about the appropriate level of capital to hold.

16.5.1 RAROC and EVA

Risk-adjusted return on capital (RAROC), developed by Bankers Trust in the late 1970s, extends RAPM by charging profits for the cost of economic capital in the numerator. RAROC is formally defined as where profits are charged a cost for immobilizing capital of the form *k* X EC, where *k* is the appropriate discount rate.[5]

$$\text{RAROC} = \frac{(\text{profit} - \text{risk adjustment})}{\text{EC}} \tag{16.9}$$

RAROC is now used widely as a performance metric. Zaik *et al.* (1996), for example, explain that Bank of America compares each business unit's RAROC with the cost of the bank's equity, which is the minimum rate of return required by shareholders. If RAROC is greater, the unit is deemed to add value for shareholders.[6]

RAROC is also consistent with *economic value added* (EVA).[7] EVA focuses on the creation of value during a particular period, measuring "residual" economic profits as where profits are adjusted for the cost of economic capital. EVA is simply the numerator in the RAROC equation.[8]

$$\text{EVA} = \text{profit} - (\text{capital} \times k) \tag{16.10}$$

16.5.2 Shareholder Value Analysis

RAROC and EVA are linked directly to *shareholder value analysis* (SVA). To maximize the total value for shareholders, investments can be evaluated using a net present value (NPV) analysis, where expected free cash flows are discounted at the rate *k* that reflects the risks of the project.

SVA dictates that any project must be undertaken only when it generates a positive NPV. Alternatively, if the corporation cannot generate such profitable projects, it simply should return capital to its shareholders in the form of dividends or share buybacks. Or the capital could be used for acquisitions. Indeed, this focus on SVA explains why many banks have been repurchasing their shares recently.

EVA is a one-period measure that can be used to evaluate historical performance. SVA is a prospective multiperiod measure that can be used to compare prospective investments. The two methods, however, are consistent with each other and lead to similar decision rules provided the same inputs are used.[9]

16.5.3 Choosing the Discount Rate

The choice of the capital charge or of the discount rate *k*, however, is a complex issue. The preceding section has shown that the capital charge can be a function of the unit's undiversified VAR or of the component VAR. The rationale for the latter choice is that the component VAR measures the effect on the bank's total risk.

It is not obvious, however, that the bank should focus on its total risk. If the bank's shareholders are well diversified, they should penalize the bank for its marginal contribution to the "market" portfolio volatility instead of its total risk. This insight forms the basis for the capital asset pricing model (CAPM), which holds that shareholders only worry about systematic risk.

The CAPM states that the required return on a project, or hurdle rate k_i, should be constructed as the sum of the risk-free rate R_F and the market risk premium times the project's systematic risk β_i^m, that is,

$$k_i = R_F + [E(R_m) - R_F]\beta_i^m \qquad (16.11)$$

If so, the risk penalty should include the unit's systematic risk vis-à-vis the market, not the bank.

As with all financial theories, the usefulness of the CAPM is that it forces us to identify conditions under which its precepts may not hold. One could say, for instance, that empirical tests do not entirely support the CAPM and that earnings volatility does seem to matter to investors. Alternatively, we could look into the literature that tries to rationalize why firms do seem to hedge financial risks. The fact that firms do undertake hedging activities has presented somewhat of a puzzle to finance researchers. With "perfect" capital markets, there would be no need to hedge. Hedging produces profits and losses that should average out to zero over the long run. In the meantime, hedging simply lowers the variability of outcomes. Firms could decide not to hedge and simply raise external funds whenever they are subject to unfavorable financial shocks. Under these conditions, hedging does not provide any benefits.[10]

Thus, rationalizing hedging requires some "frictions" in capital markets. (In fact, most bankers would argue that their very *raison d'être* stems from inefficiencies in capital markets.) One such theory is that reducing the bank's volatility is important owing to the high costs of financial distress. Other theories focus on financial market imperfections. If raising external capital is much more costly than using internal cash, then the bank should try to smooth out its earnings so as to provide steady investments from internal funds. Froot and Stein (1998) show that with such capital constraints, the hurdle rate the bank should impose on investment projects takes the form of which now depends not only on the market beta but also on the beta of the project with respect to the bank's existing portfolio β_i.

$$k_i = R_F + c_1\beta_i^m + c_2\beta_i \tag{16.12}$$

Finally, some factors are specific to financial institutions. A unique aspect of commercial banks is that they face regulatory constraints that link their capital to their risk level. Assuming that these constraints are binding, it then makes sense for the bank to allocate its precious capital as a function of each unit's marginal contribution to the bank's risk. For investment banks, financial distress can prove very expensive. They are as leveraged as commercial banks but do not have access to low-cost deposits as a source of funds. As a result, they have to rely on shortterm funding. In times of distress, liquidity constraints can cause major disruptions because funding sources can dry up quickly.

In practice, banks seem to worry mainly about their total risk instead of their systematic risk, perhaps for all the reasons just listed.

16.5.4 Implementing SVA

Institutions have found that a strong capital-allocation process produces substantial benefits. Experience has shown that the main benefit of these methods is that the process itself nearly always leads to improvements. Financial executives are forced to examine prospects for revenues, costs, and risks in all their business activities. Invariably, managers learn things about their business they did not know.

One bank consultant study, for example, has found that typical banks have only one-third of their business with high risk-adjusted returns, one-third with returns around 12 percent, and fully one-third with returns below 8 percent. Without a risk-adjustment method, it is not easy to tell where a particular business line or product would fit.

Even though this approach is a first step toward better business decisions, other factors must be considered. Customer satisfaction and repeat business are important considerations that are better captured by a multiperiod model. Development cycles are also overlooked in single-period models. EVA systems also ignore business synergies. For instance, businesses with low returns may have cross-selling benefits that make them worth keeping. Still, RAPM methods represent a big improvement in the allocation of capital. In practice, these methods are proving successful.

16.6 CONCLUSIONS

Financial institutions are realizing that efficient allocation of capital must be supported by risk-adjusted performance measures. The starting point for these methods is VAR, which measures the amount of economic capital necessary to support a business activity. This leads the way for RAPM methods, such as RAROC, which deduct risk charges from profits.

This approach provides a powerful mechanism to counterbalance the structural incentive of traders to take on more risk than they should. Thus we should hope that “risk penalties” will become a common element of traders’ compensation structure.

RAPM-based methods are entirely in line with shareholder value analysis, which carefully balances profits against risks. For the first time, we now have the tools to apply risk management analysis to the whole firm and to make strategic decisions based on the best return-to-risk ratio. There is no doubt that these developments have been made possible by the widespread use of VAR.

APPENDIX 16.A
A Closer Look at Economic Capital

This appendix gives a more formal interpretation of VAR as economic capital. Originally, VAR was developed as a method to measure shortterm risks. For longer horizons, however, discounting or trends become an issue. VAR, which is a shortfall measure on a target horizon, needs to be converted into economic capital valued in *current* dollars.

This link is best explained in the Merton (1974) framework that views equity as a call option on the value of an indebted firm. The debt is risky because there is no guarantee that the value of the firm will be sufficient to repay the face value of the debt, which plays the role of the strike price. Define

S_t = value of equity

B_t = value of debt

$V_t = S_t + B_t$ = value of the firm

K = face value of debt

μ = expected return on V

r = risk-free rate

σ = volatility of V

T = length of horizon

To briefly summarize the essence of option pricing theory, we can price an option in a *risk-neutral world* by (1) assuming that all assets grow at the risk-free rate and (2) discounting at the same risk-free rate. This explains why the actual growth μ does not appear in the Black-Scholes model. This shortcut does not mean that investors are risk-neutral but instead happens to provide the correct solution.

Now, if the value of the firm follows a lognormal process, we can write

$$\ln(V_T) = \ln(V_t) + \left(\mu - \tfrac{1}{2}\sigma^2\right)T + \sigma\sqrt{T}\epsilon \qquad (16.13)$$

Note that the $\frac{1}{2}\sigma^2$ term is due to the transformation of arithmetic returns to logarithms (see, for instance, Hull, 2005). We can find the breakeven level such that the probability of falling below V^* is equal to the desired value of p, that is, where ∈ is a standard normal variable, and α is the deviate corresponding to the probability p. Note that μ is the actual (i.e., objective instead of risk-neutral) trend in V. This breakeven level defines the VAR on the target date as

$$p = P(V_T \le V^*) = P\left\{\epsilon \le \left[\ln(V^*/V_t) - \left(\mu - \tfrac{1}{2}\sigma^2\right)T / \sigma\sqrt{T}\right]\right\} = P(\epsilon \le -\alpha) \quad (16.14)$$

$$\text{VAR}_T = E(V_T) - V^* \tag{16.15}$$

If the value of the firm drops below V^*, the firm defaults.

Once the strike price is set, the amount of economic capital required to provide the desired confidence level can be calculated as the value of the stock or a call option on the firm with strike price $K = V^*$, that is,

$$\text{EC} = S_t = C(V_t, K, T, r, \sigma) = V_t N(d_1) - Ke^{-rT} N(d_2) \tag{16.16}$$

Example

Consider a firm with initial value V_t = \$100 and parameters μ = 8 percent, r = 5 percent, and σ = 30 percent. The firm is financed with a mix of equity and zero-coupon debt. We wish to find the economic capital such that the probability of default is 1 percent over T = 1 year.

Solving Equation (16.14) with the associated deviate α = 2.33, that is, we find V^* = \$51.48. We now can compute the value of equity in the capital structure using Equation (16.16), which gives S_t = \$51.09. Debt must account for the remainder, or B_t = \$100 – \$51.09 = \$48.91.

$$\left[\ln(V^*/V_t) - \left(\mu - \tfrac{1}{2}\sigma^2\right)T\right] / \sigma\sqrt{T} = -2.33$$

Hence the economic capital required to achieve this default probability should be \$51.09, expressed in current dollars. The VAR at the 1-year horizon is

$$\text{VAR}_T = E(V_T) - V^* = V^* = V_t e^{\mu T} - V^* = \$108.33 - \$51.48 = \$56.85$$

Note that this says nothing yet about how much economic capital should be held at the present time.

Suppose that, mistakenly, this VAR number is taken as the current equity capital. If so, the actual probability of default will be lower than 1 percent.[11] This bias increases for longer horizons and greater expected returns.

We can take a shortcut to translate the VAR number into economic capital. When the probability of default is very low, $N(d_1)$ and $N(d_2)$ are close to unity, and the Black-Scholes model collapses to where the expectation is taken with respect to the risk-neutral measure, that is, assuming that V_t will grow at the risk-free rate. Thus VAR_T^{RN} is the VAR on the target date that assumes that the value of the firm grows at the risk-free rate. To translate this VAR into a current number, we simply discount at the risk-free rate.

$$S_t \approx V_t - Ke^{-rT} = e^{-rT}(V_t e^{rT} - V^*) = \\ e^{-rT}\,[E^{RN}(V_T) - V^*] = e^{-rT}\text{VAR}_T^{RN} \tag{16.17}$$

In our example, this yields

$$\text{VAR}_T^{RN} = V_t e^{rT} - V^* = \$105.13 - \$51.48 = \$53.65$$

and discounting into the present, we get

$$S_t \approx e^{-0.05} \text{ X } \$53.65 = \$51.03$$

which is quite close to the true value of $51.09. This shows that equity capital should be viewed as a VAR measure discounted into the present.

QUESTIONS

1. Companies traditionally have compared different business lines using measures such as return on assets or return on book equity. What is the drawback of such measures?
2. What is risk capital? Why can VAR be viewed as a measure of risk capital?
3. When using VAR as a measure of risk capital, how should the confidence level be chosen? Discuss the cost and benefits of having a high confidence level.
4. How can the desired credit rating be used to set a confidence level?
5. Why is a confidence level of 99.98 percent used often for economic capital?
6. Consider a foreign-exchange trader with an annual profit of $1 million trading a notional amount of $10 million. The annual volatility is 12 percent. What is the VAR at the 99 percent level over 1 year, assuming a normal distribution? What is the RAPM?
7. Why is it important to deduct expected losses from revenues when evaluating a lending operation with RAPM?
8. Explain whether the compensation of traders and risk managers is equivalent to a long or short option position.
9. How is earnings at risk measured?

10. Does earnings-based VAR or position-based VAR give a forward-looking measure of risk?
11. Discuss whether earnings-based VAR or position-based VAR provides a better method to help in understanding and controling risk.
12. What is the Sharpe ratio? How is it related to RAPM?
13. Does the Treynor ratio use total risk or beta risk?
14. What is the drawback of the conventional RAPM approach?
15. How are marginal RAPM and component RAPM related to marginal VAR and component VAR, respectively?
16. Normally, the CAPM predicts that systematic risk, that is, relative to the stock market, is the only priced risk. List factors that lead banks to focus on their total risk instead of systematic risk.

CHAPTER 17
VAR and Risk Budgeting in Investment Management

> We have a standard deviation for our total plan, but it only tells you what happened in the past. By contrast, VAR looks forward.
>
> —*Director of Chrysler's pension fund*

By now, value at risk (VAR) has spread well beyond the Wall Street trading departments where it originated. The investment management industry is also discovering the benefits of VAR systems.

Many of the reasons that made VAR successful in the banking industry also apply to asset managers. VAR is a forward-looking measure of the risk profile of a fund based on current positions. The more traditional returns-based approach, in contrast, is purely historical; it does not offer timely measurement of risk.

As seen in Chapters 15 and 16, VAR can be used to measure, control, and manage risk. VAR is comprehensive because it accounts for leverage, volatility, and diversification. VAR is a simple measure of risk that can be explained easily to portfolio managers and investors. VAR systems also can be used to set consistent guidelines that improve over traditional guidelines using limits on notionals or sensitivity measures. As a bonus, comprehensive risk management systems provide some protection against rogue traders, thereby helping to avoid embarrassing financial losses.

This chapter shows how VAR can benefit the investment management industry, which includes mutual funds, pension funds, endowment funds, insurance companies, and hedge funds. This new risk management technique has led to the development of *risk budgeting*. Risk budgeting is the process of allocating and managing risk using a top-down approach to different aspects of the investment process. Risk budgeting builds on VAR measures that are applied to asset classes, asset managers, and even securities.

Risk budgeting is fast spreading as a best-practice method to manage risk. Like VAR, the concept is not new. Like VAR, its main advantage is to provide a top-down, comprehensive, and practical method to manage risk. It provides a dynamic comparison of current risk profiles with prespecified risk budgets.

Section 17.1 compares risk measures of proprietary bank trading to the investment management industry. The definition of risk, however, depends on the investment objectives, as discussed in Section 17.2. Section 17.3 shows how to use VAR to monitor and control risk. VAR can be used to manage risk, as described in Section 17.4. A particular example of this is the risk-budgeting process, which is described in Section 17.4.

17.1 VAR APPLICATIONS TO INVESTMENT MANAGEMENT

17.1.1 Sell Side versus Buy Side

The investment management industry usually is called the "buy side" of Wall Street, in contrast with banks, the "sell side" that developed VAR. Whereas VAR has been widely, and rather quickly, accepted by the banking industry, it has spread more slowly to the investment management industry. Perhaps this is so because investment management differs in many fundamental respects from the fast-paced trading environment of dealing banks. Table 17-1 compares the characteristics of the buy side with those of the sell side.

TABLE 17-1
Risk Management for the Sell and Buy Sides

Characteristic	Sell Side (e.g., Banks)	Buy Side (e.g., Investors)
Horizon	Short-term (1 day, intraday)	Long-term (month, quarter, years)
Turnover	Rapid	Slow
Leverage	High	Low
Risk measures	VAR Stress tests	Asset allocation Tracking error
Risk controls	Position limits VAR limits Stop-loss rules	Diversification Benchmarking Investment guidelines

Consider first bank trading portfolios, where the horizon is short, turnover rapid, and leverage high. VAR is particularly appropriate for such an environment. In this case, historical measures of risk basically are useless because yesterday's portfolio profile may have nothing to do with today's. In an investment environment, in contrast, the horizon, as measured by the portfolio evaluation period, is much longer, monthly or quarterly. Positions change more slowly.

Bank trading portfolios are also highly leveraged, which makes it particularly important to control their risk. A sequence of adverse events easily could bankrupt the institution, as shown by the Barings crisis. In contrast, pension funds, whose positions are guided by a “prudent investor” philosophy, do not allow much leverage. Thus there is a less crucial need to control the downside risk.

In sum, the daily application of VAR measures has become a requirement of bank trading portfolios owing to short horizons, rapid turnover, and high leverage. Risk is controlled through position limits, VAR limits, and stop-loss rules. Although the investment management industry operates with different risk parameters, the proper measurement of risk is also a critical function. This chapter will demonstrate the benefits of VAR methods for the investment management industry.

17.1.2 Investment Process

To understand the requirements of the investment management industry, it is useful to start by describing the investment process of large investors such as pension funds. Generally, this process consists of two steps. In the first step, a consultant provides a strategic, long-term *asset-allocation* study usually based on mean-variance portfolio optimization, that balances off expected return against risk. This study determines the amounts to be invested in various asset classes, for example, domestic stocks, domestic bonds, foreign stocks, foreign bonds, and perhaps additional classes such as emerging markets, real estate, venture capital, and total-return funds, also known as *hedge funds*. The asset allocation relies on *benchmarks*, or *passive indices*, that represent a feasible investment strategy.

In the second step, the fund may delegate the actual management of funds to a stable of *active* managers. These managers are reviewed periodically for performance relative to their benchmark, measured in terms of their *tracking error*. Risk typically is controlled through a list of *investment guidelines* defining the universe of assets they can invest in, with some additional restrictions such as duration, maximum deviations from equity-sector weights, or maximum amounts of foreign currency to hedge or cross-hedge. Generally, risk is measured ex post, that is, from historical data.

Chapter 15 indicated that institutions exposed to a diversity of risks, to complex financial instruments, and to changing positions should benefit from VAR risk management systems. Let us see how these criteria apply to the investment management industry.

First, investments are becoming more global in nature, creating a need for risk measures that take diversification into account. Before 1974, for example, few pension funds invested in foreign markets. By now, funds invest all over the world. They also invest in new asset classes, such as hedge funds.

Second, financial instruments are becoming more complex over time. This creates a need for stronger, centralized risk management systems.

In practice, however, institutional investors seem to have too many risk systems. Risk measures usually are based on historical tracking error. Risk controls include "prudent investor" rules based on diversification principles, benchmarking, and investment guidelines. Such systems can have serious flaws, as illustrated in Box 17-1. In contrast, VAR provides a simple, transparent, and consistent measure of overall risk.

BOX 17-1

LESSONS FROM WISCONSIN

In March 1995, the State of Wisconsin Investment Board, which controls over $34 billion in assets, revealed that it had lost $95 million on currency and interest-rate swaps. While the loss was small in relation to the asset pool, it led to great embarrassment.

Of the total loss, $35 million came from just one contract, an interest-rate swap that paid

$$\$10 \text{ million} \times (2.95 \text{ percent MexSpread})/2.95 \text{ percent}$$

where MexSpread was defined as the yield spread between Mexican and U.S. government bonds. Apparently, the staff had not done a proper sensitivity analysis of the value of the swap and thought that the amount at risk was only $10 million. In fact, it was much greater owing to the leverage effect induced by the denominator. This loss would have been avoided had the swap been marked to market or, even better, evaluated with a VAR method.

Third, most investment portfolios are dynamic, with changing positions. Because the assets of the fund typically are dispersed over a number of managers, it is difficult to create a current picture of the overall risk of the fund. During a quarter, for instance, many fund managers may have increased their exposure to one particular industry. Taken separately, these risks may be acceptable, but as a whole, they may amount to an unsuspected large bet on one source of risk. In addition, money managers sometimes change their investment strategy, either deliberately or inadvertently. If so, the fund should be able to detect and correct such changes quickly. This explains why VAR-based, forward-looking risk measurement systems are also essential to the investment management industry.

17.1.3 Hedge Funds

Hedge funds, however, pose special risk measurement problems. This group is very heterogeneous. Most hedge funds have leverage. Some groups have greater turnover than traditional investment managers. Long Term Capital Management is an extreme example of a hedge fund that went nearly bankrupt owing to its huge leverage. Such hedge funds are more akin to the trading desks of investment banks than to those of pension funds. As such, they should use similar risk management systems.

Another category of funds, however, invests in *illiquid* assets, such as convertible bonds, which are traded infrequently, even within a month. When this is the case, risk measures based on monthly returns give a misleading picture of risk because the closing *net asset value* (NAV) does not reflect recent transaction prices. This creates two types of biases.

First, correlations with other asset classes will be artificially lowered, giving the appearance of low systematic risk. This can be corrected using enlarged regressions with additional lags of the market factors and summing the coefficients across lags.[1]

Second, volatility will be artificially lowered, giving the appearance of low total risk. Such illiquidity, however, will show up in positive serial autocorrelation in returns. Biases in volatility measures can be corrected by taking this autocorrelation into account when extrapolating risk to longer horizons.[2]

Finally, hedge funds can pose special problems owing to their *lack of transparency*. Many hedge funds refuse to reveal information about their positions for fear of others taking advantage of this information. For clients, however, this makes it difficult to measure the risk of their investment both at the hedge-fund level and in the context of their broader portfolio.

17.2 WHAT ARE THE RISKS?

First, we have to define the risks in investment management. Risk can be clearly defined for a bank trader. It is the risk of loss on the marked-to-market position. Investment asset managers, however, can have different perceptions of risk.

17.2.1 Absolute and Relative Risks

Risk can be defined as the possibility of losses measured in the base currency, dollar or other. This is the most common definition of risk. For managers who have a mandate to beat a benchmark, however, risk must be measured in relative terms. We can distinguish between two definitions:

- *Absolute risk*, which is the risk of a dollar loss over the horizon. This is the usual definition of risk in a trading environment. Sometimes this is called *asset risk*. The relevant rate of return is R_{asset}.
- *Relative risk*, which is the risk of a dollar loss in a fund relative to its benchmark. This shortfall is measured as the dollar difference between the fund return and that of a like amount invested in the benchmark. The relevant return is the *tracking error* $E = R_{\text{asset}} - R^b$, which is the excess return of the asset over the benchmark. If this is normally distributed, VAR can be measured from the standard deviation of the tracking error σ_E as $\text{VAR} = \alpha W_0 \sigma_E$.

17.2.2 Policy Mix and Active Management Risk

Consider next a fund that allocates its investment to a pool of active managers in various asset classes. The absolute performance of the fund can be broken down into two components, one owing to policy (or benchmark) choice and the other to active management. Hence total asset risk can be attributed to two sources, the risk of the total policy mix and the risk of active manager deviations from the policy mix:

- *Policy-mix risk*, which is the risk of a dollar loss owing to the policy mix selected by the fund. Since the policy mix generally can be implemented by investing in passive funds, this risk represents that of a passive strategy.
- *Active-management risk*, which is the risk of a dollar loss owing to the total deviations from the policy mix. This represents the summation of profits or losses across all managers relative to their benchmark. Thus there may be diversification effects across managers, depending on whether they have similar styles or not. In addition, the current asset-allocation mix may deviate temporarily from the policy mix.

The absolute risk can be measured from fund returns and can be defined as

$$R_{\text{asset}} = \sum_i w_i R_i \tag{17.1}$$

Where w_i is the weight on fund i with return R_i. This return can be decomposed into

$$R_{\text{asset}} = R_{\text{policy mix}} + R_{\text{active mgt.}} = \sum_i w_i^b R_i^b + \sum_i (w_i R_i - w_i^b R_i^b) \tag{17.2}$$

where R_i^b represents the return on the benchmark for fund i, and w_i^b is its policy weight. If the pension plan deviates from its policy mix $(w_i \neq w_i^b)$, the active-management portion can be decomposed further into a term that represents policy decisions and manager performance.

The funds' total VAR can be obtained from the policy-mix VAR, the active-management VAR, and a cross-product term. As an example, the Ontario Teachers' Pension Plan Board (OTPPB) estimates that its annual VAR at the 99 percent level of confidence can be decomposed as follows (in percent of the initial fund value):

Source of Risk	VAR
Policy-mix VAR	19.6%
Active-mgt. VAR	1.6%
Asset VAR	19.3%

This table points to a number of interesting observations. First, most of the risk is due to the policy mix. This is a general result owing to Brinson *et al.* (1986), who demonstrated that most of the variation in portfolio performance can be attributed to the choice of asset classes. In other words, the choice of mix of stocks and bonds will have more effect on the portfolio performance than the choice of a particular equity or bond manager.

The second interesting result is that the active-management VAR is rather small. Apparently, the fund diversifies away much of the risk of managers deviating from their benchmarks through a careful choice of various styles or many managers. Another explanation is that most of the assets are invested in indexed or closely indexed funds.

Finally, the table shows that the policy-mix VAR and active-management

VAR do not add up to the total-asset VAR. In fact, there is a slightly negative correlation between the two, leading to a lower overall asset VAR. If this occurs, active managers could take greater deviations from their benchmark without affecting the plan's total VAR.

This analysis is a good example of insights created by a VAR analysis. Such decomposition can help sponsors to make more informed decisions.

17.2.3 Funding Risk

Focusing on the volatility of assets alone, however, may not be appropriate if the assets are supposed to cover fixed liabilities. Notably, a pension fund with *defined benefits* promises a stream of fixed payments to retirees. If the assets are not sufficient to cover these liabilities, the shortfall will have to be made up by the fund's owner. On the other hand, *defined-contribution* plans put the risk on the employees. In other words, risk should be viewed in an *asset/liability management* (ALM) *framework*. We can define *funding risk* as the risk that the value of assets will not be sufficient to cover the liabilities of the fund.

The relevant variable is the *surplus S*, defined as the difference between the value of assets A and liabilities L. The change then is $\Delta S = \Delta A - \Delta L$. Normalizing by the initial value of assets, we have

$$R_S = \frac{\Delta S}{A} = \frac{\Delta A}{A} - \frac{\Delta L}{L}\frac{L}{A} = R_{\text{asset}} - R_{\text{liabilities}}\frac{L}{A} \qquad (17.3)$$

where $R_{\text{liabilities}}$ is the rate of return on liabilities.

While the value of assets can be measured by marking to market, liabilities are more difficult to evaluate. For pension funds, this represents *accumulated-benefit obligations*, which measure the present value of pension benefits owed to employees discounted at an appropriate interest rate. When liabilities consist mainly of nominal payments, their value in general will behave like a short position in a long-term bond. Thus decreases in interest rates, while beneficial for equities on the asset side, can increase even more the value of liabilities, thereby negatively affecting the surplus. If liabilities are indexed to inflation, they behave like inflation-protected bonds.

The minimum-risk position then corresponds to an *immunized* portfolio, where the duration of the assets matches that of the liabilities. In practice, it may not be possible to immunize the liabilities completely if the existing pool of long-term bonds is insufficient. More generally, immunization carries an opportunity cost if other asset classes generate greater returns over time.

This funding risk represents the true long-term risk to the owner of the fund. If the surplus turns negative, it will have to provide additional contributions to the fund. Sometimes this is called *surplus at risk* (SAR). An example is given in Box 17-2.

As an example, consider a hypothetical pension plan. Call it Public Employee Retirement Fund (PERF). PERF has $1000 in assets and $900 in liabilities, for a surplus *S* of $100 million. The duration of liabilities is 15 years; this high number is typical for pension funds. Assume that the expected return on the surplus, scaled by assets, is 5 percent. Using Equation (17.3), this translates into an expected growth of $50 million over 1 year, creating an expected surplus of $150 million. For Canadian pension funds, the typical volatility of the surplus is 9.4 percent, leading to an annual VAR of 22 percent, or $220 million at the 99 percent confidence level.[3] Taking the deviation between the expected surplus and VAR, we find that there is a 1 percent probability that over the next year the surplus will turn into a deficit of $70 million or more. The tradeoff between this number and an expected surplus growth of $50 million defines the risk profile of the fund. If acceptable, risk budgeting then allocates the SAR of $220 million to different aspects of the investment process.

BOX 17-2

SURPLUS AT RISK AT OTPPB

The Ontario Teachers' Pension Plan Board (OTPPB) has been at the forefront of applying VAR techniques among institutional investors. OTPPB is the biggest pension fund in Canada, with about C$90 billion (US$78 billion) in assets in 2005.

The plan is required to deliver *defined benefits* to Ontario's teachers during their retirement years. Its stated objective is to earn a high rate of return, at least as great as the rate of inflation plus 5 percent per annum, while minimizing the risk of a contribution increase. Until 1990, OTPPB could invest in Ontario bonds only. Starting in 1990, the plan embarked on an ambitious drive to expand into broader asset classes, with the guidance of a risk management system.

The OTPPB has decided on a policy mix of 45 percent equities, 23 percent fixed-income and absolute strategies, and 32 percent inflation-sensitive investments (i.e., commodities, real estate, and real-return bonds). The goal of this mix is to achieve a long-term surplus growth of 1.3 percent per annum. This translates into a surplus VAR of 22 percent, which is also C$20 billion when applied to assets.

In 1996, OTPPB purchased a firmwide risk management system sold by Sailfish that cost about $500,000. Management has access to daily risk reports, and the board receives monthly risk reports. VAR is measured as the worst loss at the 99 percent confidence level over 1 year. Thus we would expect on average a loss worse than C$20 billion in 1 year out of a hundred. Of course, this invites quips about risk managers not likely to be around for a century. The risk managers then patiently explain that these parameters are equivalent to a confidence level of 90 percent over 4 years.* Thus this loss would be expected in one of ten periods of 4 years.

* See De Bever *et al.* (2000). This transformation assumes normal and i.i.d. returns, in which case the ratio of 90 and 99 percent normal deviates is 1.28/2.33, which multiplied by $\sqrt{4}$ indeed gives a number close to 1.

17.2.4 Sponsor Risk

This notion of surplus risk can be extended to the risk to the owner of the fund, the plan sponsor, who ultimately bears responsibility for the pension fund. One can distinguish between the following risk measures:

- *Cash-flow risk*, which is the risk of year-to-year fluctuations in contributions to the pension fund. Plan sponsors that can absorb greater variations in funding costs, for instance, can adopt a more volatile risk profile.
- *Economic risk*, which is the risk of variation in total economic earnings of the plan sponsor. The surplus risk may be less of a concern, for instance, if falls in the surplus occur in an environment where the firm enjoys greater operating profits.[4]

From the viewpoint of the plan sponsor, risk is measured not only by movements in the assets, or even the surplus, but also by the ultimate effect on the economic value of the firm. Thus pension-plan management should be integrated with the overall financial goals of the plan sponsor. This is in line with the trend toward enterprisewide risk management, which will be analyzed in Chapter 20.

17.3 USING VAR TO MONITOR AND CONTROL RISKS

Chapter 15 demonstrated that VAR systems can be used to measure and control market risks. This also applies to the investment management industry. VAR systems allow investors to check that their managers comply with guidelines and to monitor their market risks. *Credit risk* usually is controlled through limits on exposures on a name-by-name basis. VAR systems provide some protection against *operational risk*, which is also controlled by policies and procedures. Such applications are still *passive* or *defensive* in nature.

17.3.1 Using VAR to Check Compliance

The impetus for centralized risk management in the investment management industry came from the realization that the industry is not immune to the “rogue trader” syndrome that has plagued the banking industry. Indeed Box 17-3 explains how the Common Fund lost $138 million from unauthorized trading. This has led to a reorganization of the fund with a centralized risk management function.

BOX 17-3

NONCOMPLIANCE AT THE COMMON FUND

In 1995, the Common Fund, a nonprofit organization that manages about $20 billion on behalf of U.S. schools and universities, announced that it had lost $138 million from unauthorized trading by one of its managers. Apparently, Kent Ahrens, a trader at First Capital Strategies, had deviated from what should have been a safe index-arbitrage strategy between stock index futures and underlying stocks. One day he failed to complete the hedge and lost $250,000. He then tried to trade his way out of this loss but with little success. The growing loss was concealed for 3 years until Ahrens confessed in June 1995.

This loss was all the more disturbing because after the Barings affair, the Common Fund had specifically asked First Capital to demonstrate that a rogue trader could not do the same thing at First Capital. The firm answered that market neutrality was being verified daily. In this case, it seems that proper checks and balances were not in place.

Although the dollar loss was not large compared with the size of the asset pool, it severely damaged the reputation of the Common Fund. Several fund investors left, taking $1 billion with them. The fallout also forced the president of the Common Fund the resign.

In retrospect, the Common Fund realized that running an operation with a large number of fund managers requires strong centralized controls. To prevent such mishaps, the fund created the new position of “independent risk oversight officer.” The fund also set up new risk management committees, one of which is at the board level. Its custodian, Mellon Trust, developed online software that checks for violations of investment policies. To reduce operational risk, the fund also cut the number of active managers and custodial agreements.

The lessons from this loss are applicable to any “manager of managers,” that is, a manager who delegates the actual investment decisions to a stable of managers. While rogue traders, fortunately, are rare, minor violations of investment guidelines occur routinely. Some securities may be prohibited because of their risks or for other reasons (e.g., political or religious). Bank custodians, however, indicate that fund managers sometimes trade in and out of unauthorized investments before the client realizes what happened. With monthly reporting, it is hard to catch such movements. Centralized risk management systems, in contrast, can monitor investments in real time.

Such occurrences have moved the pension-fund industry toward centralized risk management. VAR systems provide a central repository for all positions. Independent reconciliation against manager positions makes fraud a lot more difficult. VAR systems also allow users to catch deviations from stated policies quickly.

17.3.2 Using VAR to Monitor Risk

With a VAR system in place, investors can monitor their market risk better. This applies to both passive and active allocations.

Passive allocation, or benchmarking, does not keep risk constant because the composition of the indices can change substantially. The late 1990s, for example, witnessed a high-tech bubble that increased the market capitalization of firms in high-tech industries. As a result, market-capitalization benchmarks such as the Standard & Poor’s (S&P) 500 became increasingly exposed to the high-tech industry, which sharply increased the volatility of the indices. Such trends would be picked up by a forward-looking risk measurement system.

Active portfolio management can change the risk profile of the fund. Suppose, for instance, that the investor notices a sudden jump increase in the reported VAR of the fund. The key is to identify the reason for the jump. Several explanations are possible, each requiring different actions.

- *A manager taking more risk*. VAR allows dynamic risk monitoring of managers, who are given a VAR limit or risk budget. Any exceedence of the VAR limit will be flagged and should be examined closely. For instance, if this is an unauthorized trade, the infraction should be corrected at once. Otherwise, the exceedence requires a discussion with the manager. There may be good reasons to increase the risk profile. Perhaps the risk increase is temporary or justified by current conditions. In any event, it is important to understand the reason behind the change.
- *Different managers taking similar bets*. This can happen, for instance, when managers increase their allocation to a particular sector, which is perhaps becoming more attractive or has performed well in the recent past. Because active managers operate in isolation, such a problem can be caught only at the portfolio level. To decrease the portfolio risk, managers can be given appropriate instructions.
- *More volatile markets*. VAR can increase if the current environment becomes more volatile, assuming that time variation in risk is explicitly modeled, such as with GARCH models. The plan sponsor then will have to decide whether it is worth accepting greater volatility. If the risks are deemed to be too large, positions can be cut. Increased volatility, however, often is associated with falls in asset prices leading to correspondingly higher expected returns. Thus the rebalancing decision involves a delicate tradeoff between risk and return. As seen in Box 17-4, however, some investors prefer to set up their system so as to smooth out spikes in risk.

More generally, VAR can be reverse engineered to understand where risk is coming from using the VAR tools explained in Chapter 7. Measures of marginal and component VAR can be used to identify where position changes will have the greatest effect on the total portfolio risk.

This assumes, however, that all the relevant risks are captured by the risk management system. As explained earlier, risk cannot be measured easily for some important asset classes such as real estate, venture capital, and some categories of hedge funds owing to illiquidity. Other series may have very short histories, such as emerging markets, or none at all, such as initial public offerings. In some cases, the missing series can be replaced by a *proxy*, using a mapping approach. The risk manager should be aware of the limitations of the system.

BOX 17-4

SMOOTHING OUT RISK AT OTPPB

The OTPPB risk measurement system is based on historical simulation because of its ability to represent nonnormal market movements. The system loads more than 10,000 positions on a daily basis, which are combined with historical data going to January 1987. As of 2005, this represents an expanding window spanning 19 years.

This long history was selected for two reasons. First, the risk managers wanted to include the crash of October 1987 in the sample period so as to model the possibility of a future crash. Second, this long window decreases the weight on recent observations, which smoothes out the volatility process. Shorter windows lead to greater fluctuations in risk measures, which create problems when investment managers are subject to strict risk limits. OTPPB's risk managers indicated that "using a lot of history avoids a potential conflict between risk control and investment strategy."

17.3.3 The Role of the Global Custodian

The philosophy behind VAR is centralized risk management. The easiest path to centralization is to use one global custodian only.

This explains why many investors now are aggregating their portfolio holdings with a single custodian. With one global custodian, position reports directly give a consolidated picture of the total exposure of the fund. Custodians become the natural focal point for this analysis because they already maintain position information and have market data. The next level of service is to combine the current position with forward-looking risk measures.

Not all agree, however, that the risk measurement function can be delegated to the custodian. Some larger plans have decided to develop their own internal risk management system. Their rationale is that they have tighter control over risk measures and can better incorporate VAR systems into operations. Larger plans also benefit from economies of scale, spreading the cost of risk management systems over a large asset base, and also require tighter control when their assets are partly managed internally.

These clients are the exception, however. Most investors may be content with risk management reports developed by custodians. Such systems, however, are not cheap to develop. As a result, the trend will be toward fewer custodians that can provide more services. Already, large custodian banks such as Deutsche Bank, JPM Chase, Citibank, and State Street are providing risk management products. State Street, for instance, is already providing a Web-based system, called *VAR Calculator*, that allows users to perform VAR calculations on demand.

17.3.4 The Role of the Money Manager

On the money management side, managers are now under pressure from clients to demonstrate that they have in place a sound risk management system. More and more clients are explicitly asking for risk analysis because they are no longer satisfied with quarterly performance reports only.

Increasingly, clients are asking their managers, “What is your risk management system?” Leading-edge investment managers already have adapted VAR systems into their investment management process. Managers who do not have comprehensive risk management systems put themselves at a serious competitive disadvantage. Indeed, the “Risk Standards” developed in 1996 for institutional investors recommend measuring the risk of the overall portfolio, as well at that of each instrument. The report also notes that manager differentiation increasingly is created by providing risk management services to clients.

17.4 USING VAR TO MANAGE RISKS

VAR systems can be used to manage risk, which is an *active* application. VAR can be used to improve investment guidelines for active managers and to help with the investment process. In theory, VAR also could be used to compute the risk-adjusted performance of investment managers, as is done for bank traders.

17.4.1 Using VAR to Design Guidelines

VAR systems can be used to design better investment guidelines. Managers’ guidelines generally are set up in an ad hoc fashion to restrict the universe of assets in which the managers can invest and, to some extent, to control risk. Typically, guidelines include limits on *notionals*, for example, maximum sector weight deviations for equities and maximum currency positions, or limits on *sensitivities*, such as duration gaps between fixed-income portfolios and their benchmarks.

Banking institutions, however, have learned the hard way that limits on notionals and sensitivities are insufficient. Limits on notionals work best with simple portfolios with no derivatives and leverage. They do not account for variations in risk nor correlations. Limits on sensitivities are an improvement but still have blind spots, such as for hedged portfolios. In contrast, VAR limits are comparable across assets and account for risk, diversification, leverage, and derivatives (provided the system is well designed).

Says Leo de Bever, risk manager at Ontario Teachers, “Typically, you control positions by saying, ‘Thou shalt not have more than *X* million of this.’ When you do that, you end up with a whole bunch of rules on what you can and cannot do, but not a handle on how much you might lose on any given day.”

Another problem is that the spirit of these limits can be skirted with new financial instruments. For example, a manager may not be allowed to trade in futures that may be viewed as too “risky,” such as futures contracts.

Instead, investments may be allowed in high-grade medium-term notes, often viewed as safe because they have no credit risk. The problem is that these notes can be designed as *structured* notes with as much market risk as futures contracts. Hence detailed guidelines, like government regulations, are one step behind continuously changing financial markets. Traditional guidelines cannot cope well with new instruments or leverage. They also totally ignore correlations.

This is precisely what VAR attempts to measure. Instead of detailed guidelines, plan sponsors could specify that the anticipated volatility of tracking error cannot be more than 3 percent, for instance. Position limits can be set consistently across markets.

17.4.2 Using VAR for the Investment Process

A good risk management system can be used to improve the investment process, starting with the top-level asset-allocation process all the way down to trading decisions for individual stocks.

As explained earlier, the strategic asset-allocation decision is the first and most important step in the investment process for pension funds. It is usually based on a mean-variance optimization that attempts to identify the portfolio with the best risk-return tradeoff using a set of long-term forecasts for various asset classes.

In practice, the optimization usually is constrained in an effort to obtain solutions that look “reasonable.” This adjustment, however, partly defeats the purpose of portfolio optimization and fails to recognize the effects of marginal adjustments from the selected portfolio.

Since VAR is, after all, perfectly consistent with a mean-variance framework, VAR tools can be used to allocate funds across asset classes. Box 17-5 shows how incremental VAR can yield useful insights into the risk drivers of a fund.

Risk management systems are also useful at the trading level. Portfolio managers are paid to take bets. Presumably, they’ve developed skills in one dimension of the risk-return space. They are expected to identify expected returns on various investments. While expected returns can be estimated on an individual basis, assessing the contribution of a particular stock to the total portfolio risk is much less intuitive. Even if analysts could measure the individual risk of the particular stock they are considering, they cannot possibly be aware of the relationships between all existing positions of the fund. This is where VAR systems help.

BOX 17-5

VAR AND CURRENCY HEDGING

Bankers Trust (now Deutsche Bank) recently provided its RAROC 2020 risk management system to the Chrysler pension fund. The system provides measures of total and incremental VAR for the various asset classes in which the fund is invested. It can be used, among other things, to evaluate the effectiveness of hedging strategies. In particular, the fund was considering adding a currency hedge to protect the currency position of its foreign stock and bond investments.

The RAROC system showed that the “individual” risk of a $250 million currency position was $44 million at the 99 percent level over 1 year, which appears substantial. However, the pension fund realized that the “incremental” contribution to total risk of a passive currency hedge program was only $3 million.* This means that currency risk already was largely diversified in the existing portfolio. The fund decided not to hedge its currency exposure, thereby saving hefty management fees. These savings more than offset the modest cost of the RAROC system, which was priced at around $50,000 per year.

This result parallels the discussion in the academic literature, where currency hedging initially was advocated as a "free lunch," that is, lower risk at no cost.† Indeed, currency hedging reduces the volatility of individual asset returns, but this is not the relevant issue. Absent currency views, what matters is total portfolio risk. Empirically, total risk generally is not much affected by currency hedging if the proportion of assets invested abroad is small. Thus there is not much benefit from currency hedging.

* As seen in Chapter 7, incremental risk is the change in risk when the position is dropped.

† As in Pérold and Schulman (1988). Jorion (1989), however, argues that the benefit of hedging must be viewed in the context of total portfolio risk.

For each asset to be added to the portfolio, analysts should be given a measure of its marginal VAR. If two assets have similar projected returns, the analyst should pick the one with the lowest marginal VAR, which will lead to the lowest portfolio risk. Assume, for instance, that the analyst estimates that two stocks, a utility and an Internet stock, will generate an expected return of 20 percent over the next year. If the current portfolio is already heavily invested in high-tech stocks, the two stocks will have a very different marginal contribution to the portfolio risk. Say that the utility stock has a portfolio beta of 0.5 against 2.0 for the other stock, leading to a lower marginal VAR for the first stock. With equal return forecasts, the utility stock is clearly the preferred choice. Such analysis is only feasible within the context of a portfoliowide VAR system.

17.5 RISK BUDGETING

Advances in VAR have led to *risk budgeting*, which is spreading rapidly in investment management. This concept is equivalent to a top-down allocation of economic risk capital starting from the asset classes down to the choice of the active manager and even to the level of individual securities.

17.5.1 Budgeting across Asset Classes

Consider again our pension fund, PERF, that needs to allocate $1000 million to four asset classes, U.S. and foreign stocks and bonds. Table 17-2 displays the average returns, volatilities, and correlations estimated from 1978–2005 data. These are taken as inputs into the portfolio-allocation process.

Assume now that PERF's board of trustees has decided on a total volatility profile for the fund of 10 percent. This translates into an annual VAR of 23.3 percent, or $233 million, at the 99 percent confidence level. In what follows, we assume normal distributions and compute VAR as $\alpha\sigma W$ = 2.33 X 0.10 X $1,000 = $233 million.

Given this risk appetite, the optimal asset allocation is listed in Table 17-3. These optimal weights can be converted into risk budgets, or individual VARs. For instance, for U.S. stocks, this is 2.33 X 0.151 X $525 = $184 million. Note that the risk budgets sum to $308 million, which is the undiversified VAR. The actual fund VAR is lower, at $233 million, owing to diversification effects.

TABLE 17-2

Asset Classes: Risk and Expected Returns

				Correlations			
Asset		Expected Return	Volatility	1	2	3	4
U.S. stocks	1	13.4%	15.1%	1.00			
Non-U.S. stocks	2	12.8%	16.7%	0.54	1.00		
U.S. bonds	3	8.0%	7.3%	0.19	0.11	1.00	
Non-U.S. bonds	4	9.0%	10.9%	0.05	0.52	0.40	1.00

TABLE 17-3

Risk Budgeting across Asset Classes (Millions)

Asset	Weight	Volatility	Principal	Risk Budget
U.S. stocks	52.5%	15.1%	$525	$184
Non-U.S. stocks	10.4%	16.7%	$104	$41
U.S. bonds	12.2%	7.3%	$122	$21
Non-U.S. bonds	24.9%	10.9%	$249	$63
Portfolio	100.0%	10.0%	$1000	$233

The same process can be continued at the next level. Assume that PERF allocates the $525 million principal in U.S. stocks to two active managers. The two managers are equally good and receive the same amount, $262.5 million each. They run portfolios with volatility of 16 percent and correlation of 0.78 with each other. This gives a risk budget of $\alpha\sigma W$ = 2.33 X 0.16 X $262.5 = $98 million each. Again, the risk budgets sum to an amount that is greater than the risk budget for this asset class owing to diversification effects. The total risk budget is

$$\sqrt{\$98^2+\$98^2+2\times 0.78\times \$98\times \$98} = \$184$$

= $184 million.

Each manager then is charged to earn the highest return on these risk units. Another advantage of this approach is that it avoids micromanaging the investment process. As long as managers stay within their risk guidelines, they can execute new transactions without requiring approval of senior management.

17.5.2 Budgeting across Active Managers

This approach can be refined further if we are willing to make assumptions about the expected performance of active managers. For better or for worse, active managers usually are evaluated in terms of their *tracking error* (TE), defined as the active return minus that of the benchmark.[5] Define μ as the expected TE and ω as its volatility (TEV). The *information ratio* then is defined as

$$\mathrm{IR} = \mu/\omega \tag{17.4}$$

Managers are commonly evaluated on the basis of their IR. Grinold and Kahn (1995), for example, assert that an IR of 0.50 is "good," meaning in the top quartile of the active managers. Logically, a greater risk budget should be allocated to managers with better performance, as measured by the IR criterion.

The optimization problem for active manager allocation attempts to maximize the IR for the total portfolio subject to a TEV constraint. Define x_i as the fraction invested in manager i, who has a tracking error of ω_i and excess return of μ_i. The value added for the total portfolio p is

$$\mu_p = \sum_i x_i \mu_i = \sum_i x_i (\mathrm{IR}_i \times \omega_i) \tag{17.5}$$

Now assume that the deviations for each manager are independent of each other.[6] The portfolio TEV is fixed at

$$\omega_p = \sqrt{\sum_i x_i^2 \omega_i^2} \tag{17.6}$$

Maximizing the portfolio information ratio subject to a fixed TEV gives the following solution:

$$x_i \omega_i = \mathrm{IR}_i \left(\frac{1}{\mathrm{IR}_p} \omega_p \right) \tag{17.7}$$

Thus the relative risk budgets should be proportional to the information ratios.

Table 17-4 gives an example. PERF wants to allocate \$525 million to a pool of active managers so as to maximize the information ratio of the fund subject to an overall TEV of 4 percent. This is equivalent to a risk budget of \$48.9 million. Each manager has a TEV of 6 percent. To achieve an exact TEV of 4 percent, we also need some residual investment in the benchmark, which has a TEV of zero. The fund managers have different capabilities; their IRs are 0.60 and 0.40, respectively. Equation (17.7) gives a solution of 55 percent weight for manager 1, 37 percent for manager 2, and the residual of 8 percent in the index.

TABLE 17-4

Risk Budgeting across Active Managers (Millions)

	Inputs		Outputs			
	TEV ω_i	Information Ratio IR_i	Weight x_i	Allocated Principal x_iW	Excess Return $x_i\mu_i$	Relative Risk Budget
Manager 1	6.0%	0.60	55%	\$291	2.0%	\$40.6
Manager 2	6.0%	0.40	37%	\$194	0.9%	\$27.1
Index	0.0%	0.00	8%	\$40	0.0%	\$0.0
Portfolio	4.0%	0.72	100%	\$525	2.9%	\$48.9

This portfolio is expected to return 2.9 percent. With 4 percent TEV, this translates into an information ratio of 2.9/4.0 = 0.72. This is higher than the IR of both managers and is due to the fact that we assumed that the active returns were independent of each other, leading to substantial diversification benefits.[7]

Such structured, top-down risk allocation adds tremendous rigor to the investment management process. It has been of great value to early adopters, such as the OTPPB. The VAR system has reduced the number of investment rules and has facilitated the closer supervision of risks. OTPPB's focus on risk and diversification has led to greater investment in alternative asset classes. Since 1990, the plan has beaten its benchmark by an average of 4 percent per annum and has provided $16 billion in value added.

17.6 CONCLUSIONS

Centralized risk management systems, by now widely adopted on Wall Street, are also taking hold in the investment management industry. Even though institutional investors have a longer-term horizon than bank trading departments, they also greatly benefit from the discipline provided by VAR systems.

Traditionally, risk has been measured using historical returns or as the occurrence of a big loss. While useful for some purposes, these risk measures have severe shortcomings because they are backward-looking. In contrast, VAR provides forward-looking measures of risk, using a combination of current positions with risk forecasts.

When implemented at the level of the total plan, VAR allows improved control of portfolio risk and of managers. It cuts through the maze of diversification rules, benchmark portfolios, and investment guidelines. VAR systems allow analysts to make better risk-return tradeoffs. VAR, of course, will not tell you where to invest. The goal is not to eliminate risk but rather to get the just reward for risk that managers elect to take.

Such risk management systems are spreading quickly among institutional investors, changing the face of the industry. They are affecting the custody business, forcing custodians to offer risk management reporting capabilities. Managers are affected, too. Those who do not have a risk management system put themselves at a serious competitive disadvantage.

It is somewhat ironic that the investment management industry, which has long relied on modern portfolio theory, is only now turning to fund-wide risk measurement systems. These systems have been developed by "quants" on Wall Street who were originally trying to get a grip on their short-term derivatives risk. What we are learning now is that these methods can be extended usefully from the short-term trading environment to the longer-term framework of patient investors.

This turn of events was inevitable. Since advances in technology and communications create almost instantaneous flows of information across the globe, plan sponsors cannot continue to rely on monthly or quarterly hard-copy reports on their investments.

QUESTIONS

1. Does the investment management industry have greater leverage than bank trading portfolios?
2. Discuss factors that should favor the use of forward-looking VAR measures as opposed to historical risks.
3. Traditionally, pension funds control risk using investment guidelines, for instance, limits on notionals in each market. What is the drawback of this approach?
4. What is relative risk? Why is it important for investment managers?
5. A portfolio manager of U.S. stocks has a negative view of the market and allocates 40 percent of his portfolio to cash. The benchmark is the S&P 500 Index. Does the portfolio have more or less risk relative to the index?
6. In the context of investment management, is the sum of the policy-mix VAR and active-management VAR greater than the total-asset VAR?
7. Typically, is most of the downside risk of a fund due to policy-mix VAR or active-management VAR?
8. What is funding risk? Why is it important to examine funding risk for a pension fund?
9. Will a drop in interest rates reduce funding risk?
10. A pension fund has $100 billion in assets and $90 billion in liabilities. Assets are fully invested in stocks, and liabilities have a modified duration of 15 years. In 2002, stocks fall by 22.1 percent, and yields drop by 1.24 percent. Compute the surplus at the end of 2002 and the return (scaling by initial asset value).
11. Assume now that the volatility of stock is 15 percent annually. The volatility of changes in yields is 1 percent annually. Assuming zero correlation between stocks and bonds and normal distributions, compute the 99 percent VAR for the surplus (where dollar returns are scaled by

the initial asset value). Next, compare with the loss in the preceding question.

12. In the preceding question, would a negative correlation between stocks and bonds (not yields) increase or decrease the surplus risk?
13. “Relative VAR is no different from forward-looking tracking error risk. So this is old wine in a new bottle?” Discuss.
14. Assume that a pension fund has many active stock managers who can take bets on different countries, industries, and stocks. List reasons for a sudden jump in the reported VAR.
15. Consider an investment in an unhedged portfolio of global stocks and corporate bonds. What risks is this portfolio exposed to?
16. Suppose that a pension fund imposes strict VAR limits on the active managers. What are the pros and cons of measuring VAR over a short window?
17. A fund manager is considering two types of assets to add to the existing portfolio. The two assets are predicted to have the same expected return. How can the manager use VAR analysis to decide on the best investment?
18. Which of the following are risk budgets conceptually closest to: marginal VAR, component VAR, individual VAR, or diversified VAR?
19. Two active managers are expected to perform as follows. Their index returned 10 percent.

Manager	Expected Return	Total Risk	Tracking Risk
1	13%	17%	6%
2	14%	17%	5.7%

Which manager has the best information ratio?

20. An investment fund allocates $50 million each to two active stock managers. Both managers have volatility of 17 percent and correlation of 0.7. Assuming normal distributions, compute the 99 percent VAR of the combined allocation.
21. In fact, the $100 million stock allocation had a total risk budget of $35

million, which is less than the actual risk. Explain how the allocations could be changed to maintain the target risk budget.

22. After further investigation, the first manager should have an information ratio of 0.5 and the second of 0.7. If $50 million is invested in the first, how much should be invested optimally in the second?

PART V
EXTENSIONS OF RISK MANAGEMENT SYSTEMS

CHAPTER 18
Credit Risk Management

> Don't focus on derivatives. One of the most dangerous activities of banking is lending.
>
> —*Ernest Patakis, Federal Reserve Bank of New York*

Credit risk can be defined broadly as the risk of financial loss owing to counterparty failure to perform its contractual obligations. As Chapter 2 has shown, the historical record of financial institutions indicates that credit risk is far more important than market risk. Time and again, lack of diversification of credit risk has been the primary culprit for bank failures. The dilemma is that banks have a comparative advantage in making loans to entities with whom they have an ongoing relationship, thereby creating excessive concentrations in geographic or industrial sectors.

It is only recently that the banking industry has learned to measure credit risk in the context of a portfolio. These newer models truly started to blossom as a result of the risk management revolution started by value at risk (VAR). After all, the main idea behind VAR is the aggregation of risks across an institution, taking into account portfolio effects. Once measured, credit risk can be managed and better diversified, like any financial risk. This is why the banking sector is busily developing sophisticated *internal models* for portfolio credit risk. These models are broader than loan portfolios, however. They are applicable to any portfolio of *credit-sensitive assets*, such as corporate bond funds.

Credit risk, unfortunately, is much more difficult to quantify than market risk. There are more types of risk factors, including the risk of default or downgrade, recovery risk, and credit exposures. Default probabilities are difficult to assess because of the infrequency of defaults. Comovements must be modeled within each type of risk and across types of risks. This requires the use of causative models that take observable financial variables as input data and generate default probabilities and correlations.

Nevertheless, the industry has made immense strides in the direction of greater diversification of credit risk across geographic and industrial sectors, which ultimately should lead to a safer financial environment. In addition, portfolio credit-risk models have led to the rise of the *credit derivatives* market.

This chapter provides an introduction to credit risk. The quantification of credit risk has by now become a large subject area, and an entire book could be devoted to this topic alone.[1] Instead, the emphasis of this chapter will be on the extension of traditional VAR methods to credit risks. Many of the tools developed for market risk also apply to credit risk. Factor models, for instance, can be used to simplify the correlation structure. As with VAR, we will show how to construct the loss distribution for the portfolio, calculate economic capital, and use VAR tools such as marginal VAR to identify the positions that are causing more risk to the portfolio.

Section 18.1 discusses the broad characteristics of credit risk. Default risk and recovery risk are introduced in Section 18.2. Section 18.3 discusses credit exposure, including the effect of netting arrangements. Next, Section 18.4 shows how to combine this information to measure portfolio credit risk. The management of credit risk and recent portfolio credit-risk models are discussed in Section 18.5. This section also discusses the new Basel Accord, dubbed *Basel II*, that was finalized in June 2004.

18.1 THE NATURE OF CREDIT RISK

18.1.1 Sources of Risk

Portfolio credit-risk models originated from commercial banks whose main assets, loans, are exposed to credit risk. A loss occurs when an obligor defaults on its loans. The actual loss depends on the amount at risk and the fraction of this recovered. Thus credit risk includes three risk factors:

1. *Default risk*, which is the risk of default by the counterparty and is measured by the *probability of default* (PD).
2. *Credit exposure risk*, which is the risk of fluctuations in the market value of the claim on the counterparty; at default, this is also known as *exposure at default* (EAD).
3. *Recovery risk*, which is the uncertainty in the fraction of the claim recovered after default; this is also 1 minus the *loss given default* (LGD).

Overall, the risk management function for credit risk focuses on a set of issues that are quite different from those facing market risk managers, as shown in Table 18-1. First, credit risk deals with the combined effect of market risk, default, and recovery risk. Market risk appears through movements in credit exposures. Second, risk limits apply to different units. For market risk, limits apply to levels of the trading organization (such as business units, trading desks, or portfolios); for credit risk, limits apply to the total exposure to each counterparty, a legally defined entity. Third, the time horizon generally is quite different, usually very short (days) in the case of market risk but much longer (years) for credit risk. This longer horizon makes it important to consider changes in the portfolio, as well as any mean reversion in the risk factors. Fourth, legal issues are very important for evaluating credit risk, whereas they are not applicable to market risk. Recovery from credit losses depends on national laws and on the application of bankruptcy rules. Thus credit risk is much more complex than market risk.

TABLE 18-1

Comparison of VAR with Credit Risk

Item	Traditional VAR	Credit Risk
Source of risk	Market risk	Market risk, default and recovery
Unit to which risk limits apply	Some level of trading organization	Legal entity of counterparty
Time horizon	Short term (days)	Long term (years)
Portfolio	Static portfolio	Dynamic portfolio
Mean reversion	Not important	Essential
Legal issues	Not applicable	Very important

Source: Adapted from Evan Picoult, Citibank.

18.1.2 Credit Risk as a Short Option

Credit risk creates an asymmetric risk profile. At best, the bond or loan is paid in full. When this happens, the upside is small. At worst, default wipes out the value of the asset. Thus the downside is large.

When the exposure is stochastic, credit risk is more complex. To create losses, two conditions must be satisfied. First, there must be a net claim against the counterparty (or credit exposure), and second, that counterparty must default.

Traditionally, credit risk only applied to bonds and loans, for which the exposure is simply the face value of the investment. Derivatives, in contrast, can have either positive value (a net asset to the solvent party) or negative value (a liability of the solvent party). Credit exposure only exists when the contract has positive value, or is *in the money*.

Consider, for instance, the credit risk of a forward contract on a foreign currency contracted with a client. The credit exposure is the positive value of the contract, whose value depends on movements in exchange rates.

In effect, the loss owing to default is much like that of an option. Define V_0 as the current, or replacement, value of the asset to the solvent party. Assuming no recovery in case of default, the loss is the *current exposure* V_0, if positive, that is,

$$\text{Loss}_t = \max(V_0, 0) \tag{18.1}$$

This asymmetric treatment stems from the fact that if the counterparty defaults while the contract has negative value, the solvent party is typically not free to “walk away” from the contract, as shown in Box 18-1. In contrast, a loss may occur if the defaulting party goes bankrupt, in which case payment will be only a fraction of the funds owed. Therefore, the current exposure from default has an asymmetric pattern, like a short position in an option. This asymmetry will reappear in the shape of the credit loss distribution.

18.1.3 Portfolio Effects

Traditionally, credit risk was viewed on a *transaction-by-transaction basis*, which essentially ignores portfolio effects. Consider, for instance, a portfolio consisting of a long yen forward position and a short yen forward with two different counterparties. The portfolio is hedged as to market risk. The transaction-by-transaction approach would consider the effect of default on each position separately. A loss on the long position occurs if the yen appreciates and the first counterparty defaults. Conversely, a loss on the short position occurs if the yen depreciates and the second counterparty defaults. In the traditional approach, the potential credit losses from the two positions are added up.

BOX 18-1

WALKAWAY FEATURES IN DREXEL’S COLLAPSE

The collapse of the Drexel Burnham Lambert Group (DBL Group) in 1990 provides an illustration of the asymmetry in payoffs when default occurs.

DBL Group’s bankruptcy placed its swap subsidiary, DBL Products, in default. Most of DBL’s swap agreements contained a walkaway clause that permitted the solvent party to cease payment even if it owed money to the defaulting party (the standard documentation for swaps has been changed since).

Even so, nearly all counterparties paid DBL Products the money they owed, for a number of reasons. DBL Products threatened to challenge the right to walk away through litigation. Counterparties settled to avoid expensive litigation because there were unresolved legal issues as to the enforceability of these contracts. A number of counterparties also feared that other institutions would be less likely to do business with them if they took advantage of the walkaway clause with Drexel. As a result, DBL was paid 100 percent of what it was owed but negotiated to pay only about 70 percent of the value of the contracts that were in the money for the solvent parties.

Since appreciation and depreciation of the yen are two mutually exclusive events, however, this method overstates the true potential loss from credit risk. Instead, a *portfolio approach* would take into account interactions between market movements and then determine the potential loss. In this case, assuming equal probability of appreciation/depreciation and of default by the two counterparties, the potential loss is only half the previous measure.

Similarly, adding up separate capital charges for each exposure ignores correlations across defaults. Accounting for portfolio effects, however, is no simple matter. This is the purpose of internal portfolio credit-risk models. To understand how these are constructed, we now turn to the various components of credit risk.

18.2 DEFAULT AND RECOVERY RISK

18.2.1 Default Risk

This is the risk that the counterparty will default on its obligation, which is a discrete state represented by the variable b_i, where default is defined as $b_i = 1$ and no default as $b_i = 0$. The expectation of this variable is the *probability of default* (PD).

Perhaps the most delicate part of credit-risk modeling consists of assessing default probabilities. These can be based on actuarial models or market prices.

Actuarial models forecast objective default probabilities by analyzing factors associated with historical default rates. One such model is the z-score, developed by Altman (1968), which predicts bankruptcy from a combination of accounting variables. Another, more sophisticated approach is that of *credit-rating agencies*, which classify issuers by *credit ratings*.

To be useful, these classifications should be related to actual default rates. Table 18-2 shows historical default rates reported by Standard & Poor's for various credit ratings, ranging from AAA for the best credits to C for the riskiest obligors. A borrower with an initial rating of BBB, for example, had an average 0.29 percent default rate over the next year and 6.95 percent over the next 10 years. The table shows that lower-rated borrowers have indeed higher default rates. Thus we could use this information as estimates of default probabilities for an initial rating class.

TABLE 18-2

Standard & Poor's Cumulative Default Rates (Percent)

	Year									
Rating	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
AAA	0.00	0.00	0.04	0.07	0.12	0.21	0.31	0.48	0.54	0.62
AA	0.01	0.03	0.08	0.16	0.26	0.40	0.56	0.71	0.83	0.97
A	0.04	0.13	0.26	0.43	0.66	0.90	1.16	1.41	1.71	2.01
BBB	0.29	0.86	1.48	2.37	3.25	4.15	4.88	5.60	6.21	6.95
BB	1.28	3.96	7.32	10.51	13.36	16.32	18.84	21.11	23.22	24.84
B	6.24	14.33	21.57	27.47	31.87	35.47	38.71	41.69	43.92	46.27
CCC/C	32.35	42.35	48.66	53.65	59.49	62.19	63.37	64.10	67.78	70.80

Note: Static pool average cumulative default rates, 1981–2004 (adjusted for "not rated" borrowers).

Table 18-2 reports *cumulative* default rates c_n, which represent the total probability of defaulting at any time between now and year n, starting from a fixed initial credit rating. This information can be used to recover *marginal* or annual default rates d_i during year i. This represents the proportion of firms that default in year i from the set that is still alive at the beginning of the year.

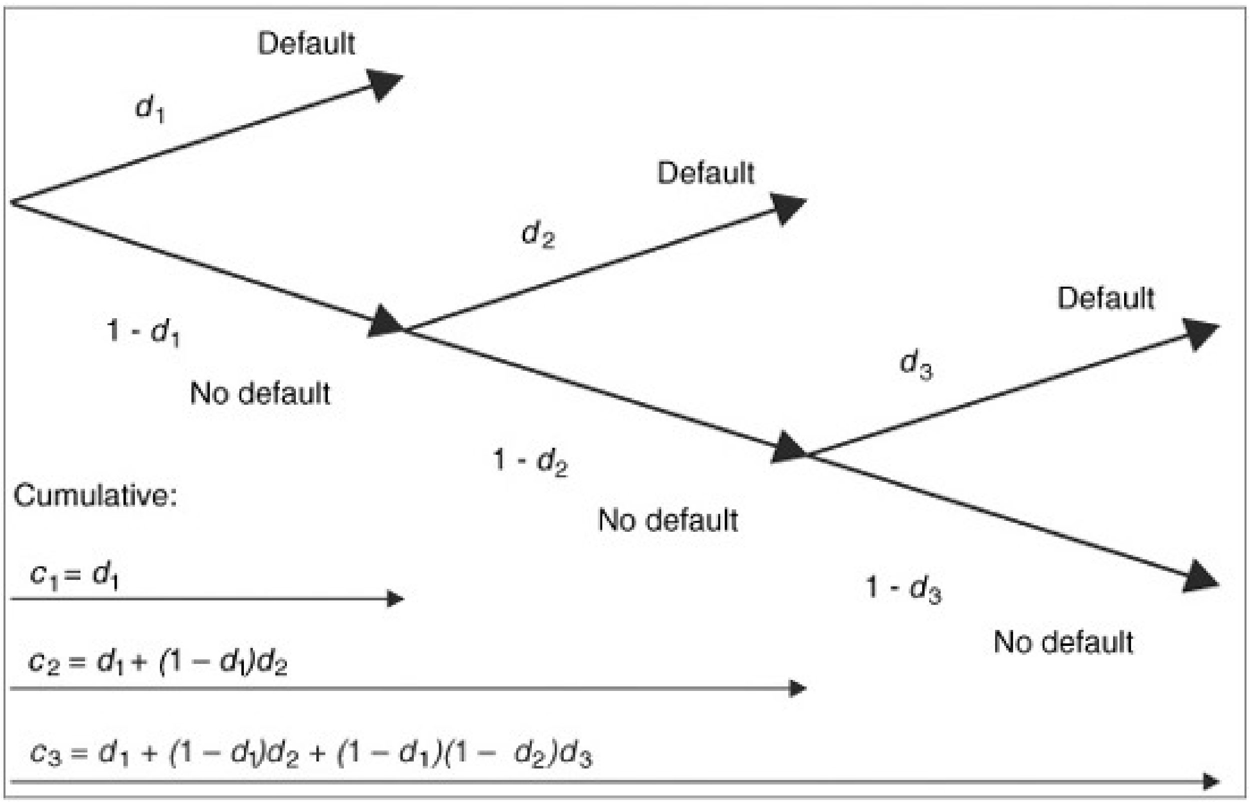
Figure 18-1 describes the multiperiod default process. For a firm to survive year n, it needs to have survived up to year $n – 1$ and not defaulted in year n. Hence we can write the *survival rate* up to year n as which can be used to solve recursively for d_i. For example, for year 1 of our BBB-rated credit, $c_1 = d_1 =$ 0.29 percent. For year 2, $c_2 = 0.86$ percent. Solving for year 2, we set (1 –

0.0086) = (1 – 0.0029) (1 – d_2), which yields d_2 = 0.57 percent, and so on.

$$(1-c_n) = (1-c_{n-1})(1-d_n) = \prod_{i=1}^{n}(1-d_i) \tag{18.2}$$

Next, we can compute the total probability of defaulting exactly in year *i*, starting from now, as

FIGURE 18-1 Sequential default process.



$$k_i = (1 - c_{i-1})d_i \tag{18.3}$$

The cumulative probability is the sum of these probabilities, that is,

$$c_n = d_1 + k_2 + \cdots + k_n \tag{18.4}$$

18.2.2 Recovery Risk

Another component of credit risk is the loss given default (LGD). This represents the fraction of the exposure lost on default, or 1 minus the *fractional recovery rate*, defined as *f*. Because *f* varies across defaults, this is a stochastic variable taking values between 0 and 1.

LGD depends on a number of factors. Most important is the priority of the debt. Debt that is secured by assets should have a higher recovery rate than debts not secured by assets because the lender can seize and sell the assets on default. Likewise, senior debt has higher priority during bankruptcy proceedings than subordinated debt and should have a higher recovery rate. Another factor is the industry. Some industries, such as utilities, have tangible assets that can be auctioned off on default, leading to higher recovery rates. Others, such as Internet firms, have no tangible assets. The legal environment is another important factor affecting recovery rates, which vary systematically across countries.

Table 18-3 displays typical recovery rates for U.S. debt. The table confirms that as the seniority of a bond decreases (going down the column), so does the average recovery rate. The average recovery rate for senior unsecured debt, for instance, is estimated at $f = 37$ percent. Typically, derivatives have the same status as senior unsecured debt, and the same recovery rate can be used. Bank loans have higher recovery rates than do bonds.

TABLE 18-3

Historical Recovery Rates for U.S. Corporate Debt

		Recovery Rates	
Ranking	**Number**	**Average**	**SD**
All bank loans	310	62%	23%
Senior secured bonds	238	53%	27%
Senior unsecured bonds	1095	37%	27%
Senior subordinated bonds	450	32%	24%
Subordinated bonds	477	30%	21%
All bonds	2368	37%	26%

Source: Moody's (2003), from 1982–2002 defaulted bond prices.

This allows us to infer the expected credit loss for a bond or loan. As an example, take a BBB-rated bond with a maturity of 5 years. From Table 18-2, we have $c_5 = 0.0325$. Thus the expected credit loss on a \$100 million bond is \$100 × 0.0325 × (1 – 0.37) = \$2.05 million.

The recovery rate, however, varies widely, creating another source of uncertainty. Typically, this is modeled by a probability density function such as the beta, $b(f)$, where f can vary between 0 and 1. This distribution is fitted to the average value and standard deviation reported in Table 18-3.

18.2.3 Market-Based Models

Credit risk also can be assessed from the price of traded assets whose value is affected by default. This includes bonds, credit default swaps, and equities for the reference entity. Because the prices are set in freely functioning financial markets, they incorporate the expectations of traders about potential losses owing to default.

For instance, consider the price of a *credit-sensitive* bond whose ultimate payoff depends on the state of default. Figure 18-2 describes a simplified default process for this bond over one period. At maturity, the bond can be either in default or not. Its value is $f \times \$100$ if default occurs and \$100 otherwise. Define π as the cumulative default rate from now to maturity T.

If bond prices carry no risk premium, the current price must be the mathematical expectation of the discounted values in the two states. Define y^* and y as the yields on the credit-risky bond and on an otherwise identical risk-free bond. Hence

$$P^* = \frac{\$100}{(1+y^*)^T} = \left[\frac{\$100}{(1+y)^T}\right] \times (1-\pi) + \left[\frac{f \times \$100}{(1+y)^T}\right] \times \pi \qquad (18.5)$$

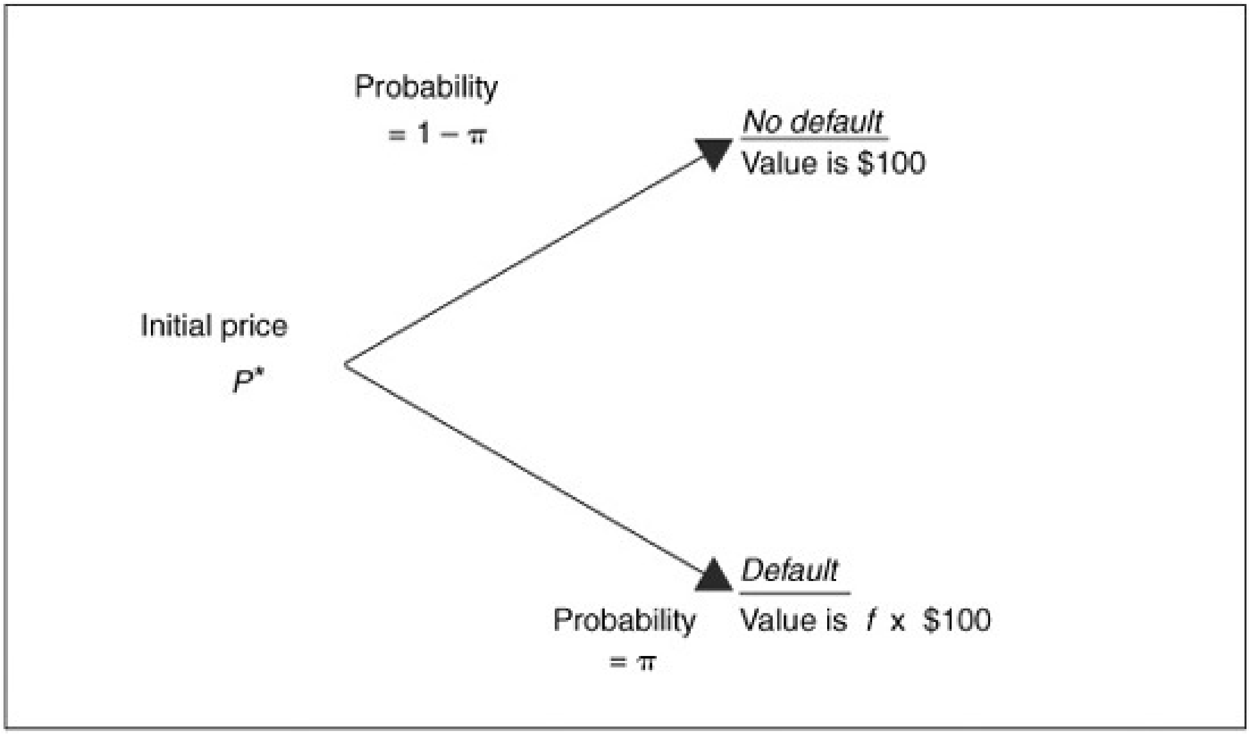
Note that we discounted at the risk-free rate y because we assumed that there was no risk premium. After rearranging terms,

$$(1+y)^T = (1+y^*)^T\,[1 - \pi\,(1-f)] \qquad (18.6)$$

To simplify, let us drop second-order terms and assume that $T = 1$. We find

$$y^* \approx y + \pi\,(1-f) \qquad (18.7)$$

FIGURE 18-2 A simplified bond default process.



Hence the credit spread $y^* - y$ measures the cumulative probability of default π times the loss given default $(1 - f)$. Thus we could use the information in yields to measure the expected credit loss. Making some assumption about the recovery rate f, we can back out the default probability π from y and y^*.

For example, assume that a 1-year bond issued by Ford Motor trades at a 7.2 percent yield versus 4.5 percent for an otherwise identical Treasury bond. We assume a recovery rate of 37 percent. What is the implied default probability? Using Equation (18.7), this gives $\pi = (y^* - y)/(1 - f) = (7.4 - 4.5)/(1 - 0.37) =$ 4.29 percent.

This probability, however, is a *risk-neutral* probability because, like all option pricing models, it assumes that investors are risk-neutral. In practice, the relationship is made more complicated by the presence of a *risk premium*. Defaults happen more often when the economy is doing badly. As a result, credit-sensitive debt has some systematic risk, for which investors require an additional compensation. *Liquidity effects* may be another factor. Less liquid bonds, trading at a wider bid-ask spread, are less attractive to investors. This translates into lower prices or higher yields than otherwise. Finally, *tax effects* also can cause distortions owing to the fact that the coupon payments on corporate and Treasury bonds are taxed differentially at the level of U.S. states.[2] To measure the *actual*, or *physical*, default probability, we would have to model these other factors. This is a problem common to all methods that use market prices. For instance, we note that Ford is rated BB, which from Table 18-2 is associated with a historical 1-year default rate of 1.28 percent, which is less than the risk-neutral number $\pi = 4.29$ percent found before.

Even so, this approach is useful only when the counterparty has issued publicly traded bonds that have meaningful transaction prices. An alternative is to turn to default-risk models based on stock prices, which are available for a larger number of companies and in addition are more actively traded than are corporate bonds.

We can decompose the value of the firm into debt and equity. Debt is an obligation that has to be repaid at a fixed price in the future. If the value of the firm is insufficient to repay this debt, the firm is in default. In theory, the stock price then goes to zero. Merton (1974) has shown that the firm's equity can be viewed as a call option on the assets of the firm, with an exercise price given by the face value of debt. The current stock price therefore embodies a forecast of default probability in the same way that an option embodies a forecast of being exercised. These kinds of models, which are based on the capital structure of the company, are called *structural models*. In particular, *KMV Moody's* uses this approach to sell *estimated default frequencies* (EDFs) for a large number of firms. Developing this approach, however, is well beyond the scope of this book. In contrast, the approach based on bond-price yields is called a *reduced-form model* because it models the default probability directly.

18.2.4 Default Comovements

In addition to measuring individual default risk, we also need to measure the comovements of defaults. As we have seen in the case of market risk, correlations are crucial drivers of portfolio risk. Chapter 7 shows that in large portfolios the portfolio risk converges to the average volatility times the square root of the correlation coefficient.

Default correlations cannot be measured directly. Obviously, we cannot observe occurrences of historical defaults on existing companies (which, by definition, cannot be bankrupt.) Defaults are also relatively scarce, especially for investment-grade firms, which makes it even more difficult to measure joint defaults. In practice, the industry has adopted *causative models* that take observable financial variables as input data and generate default probabilities and correlations.

These models start with a *latent*, or unobserved, variable, which is, for instance, the market value V of the company's assets and modeled via simulations. The default process can be modeled by choosing a cutoff point below which the company goes into default. Mapping this variable onto a set of common factors then creates comovements in asset values, which generates comovements in defaults.

An example is presented in Table 18-4. We consider two firms, both rated B. We assume that the latent variables have a joint standard normal distribution with correlation of $\rho_v = 0.20$. From Table 18-2, the probability of default over one year is $d_1 = 6.24$ percent. Thus we choose the cutoff point of $z = -1.535$, which is such that the area to its left is equal to d_1. In the first simulation in the table, both V_1 and V_2 are above this point, so there is no default, $b_1 = b_2 = 0$. In the second simulation, $V_2 < z$, and we have one default, $b_2 = 1$. In the third simulation, we have one default, $b_1 = 1$. In the fourth simulation, we have two defaults, $b_1 = b_2 = 1$. And so on. We can compute a correlation across the default variables that is close to $\rho_d = 0.10$. Typically, default correlations are much lower than asset correlations owing to the infrequency of defaults. The latter can be calibrated to give appropriate default correlations.

Thus this approach uses the panoply of tools developed for market risk. The same problems arise, consequently. As explained in Chapter 8 on multivariate models, modeling default correlations involves estimating a large number of parameters, which increases with the square of the number of obligors. Chapter 8 has shown how factor models can help to reduce the dimensionality of the problem. Likewise, the joint movements in asset values can be based on a normal copula, which is used most commonly because of its simplicity but may not represent dependencies in the tails adequately.

TABLE 18-4

Example of Default Simulations

	Latent Variables			Default Variables	
Step *i*	V_1	V_2	**Cutoff *z***	b_1	b_2
1	1.79	−0.21	−1.54	0	0
2	−1.09	−1.62	−1.54	0	1
3	−1.58	−0.93	−1.54	1	0
4	−1.60	−2.29	−1.54	1	1
⋮					
Correlation	$\rho_v = 0.20$			$\rho_d = 0.10$	

In fact, the Basel II rules are calibrated to a one-factor model with average asset correlation of $\rho = 0.20$ and a normal copula. Higher correlations or greater tail dependencies will increase the tails of the credit-risk distribution, however. Grundke (2005), for instance, performs sensitivity analyses of the tail of the credit distribution. For a portfolio of 500 Baa-rated assets, the 99.9 percent quantile is 3.1 percent of the notional amount when $\rho_v = 0.20$ but increases to 7.4 percent when $\rho_v = 0.50$. Similarly, using a student copula instead of the normal copula increases the quantile from 3.1 to 8.6 percent. As seen in Box 8-2, the industry is still learning, sometimes the hard way, how to model these default comovements.

18.3 CREDIT EXPOSURE

Credit exposure (CE) is defined as the replacement value of the asset, if positive, on the target date. This is derived from the *exposure profile,* which is a function of time until the maturity of the instrument. It includes not only the current replacement value V_0 but also the *potential*, or future, value, that is where ΔV_t represents the increase in value to time *t*. This is also the market price and so is a random variable. Its distribution can be usefully characterized by an expected value and worst value at some confidence level.

$$\mathrm{CE}_t = \max\,(V_0 + \Delta V_t, 0) \tag{18.8}$$

18.3.1 Bonds versus Derivatives

In the case of risky debt, credit exposure at expiration is the principal. Before expiration, the exposure can vary if the market value of the bond fluctuates, but overall, this will be close to the principal, or notional.

With derivatives, in contrast, credit exposure is much more complex. The exposure represents the positive value of the contract, which is much less than the notional amount. Consider, for instance, a fixed-to-floating interest-rate swap. There is no exchange of principal at initiation or at expiration. Each period, payments are netted and represent small proportions of the principal. The exposure stems from the fact that the rate on the fixed payments may differ from prevailing market rates. At maturity, this risk is zero because there are no remaining coupon payments.

To illustrate, Figure 18-3 presents the exposure profile of a 5-year interest-rate swap.[3] Initially, the exposure is 0 if the contract is fairly priced. After 1 year, the average exposure rises to about 2 percent of the notional. Eventually, the value of the swap converges to 0 at maturity because there is no exchange of principal. This profile is a combination of two factors, the *amortization effect*, which decreases risk as the maturity nears, and the *diffusion effect*, which increases the dispersion of interest rates as time goes by. Over long horizons, it is important to factor in mean reversion effects, which can affect the extent of this dispersion dramatically. The graph also shows that the worst exposure, measured at the 95 percent confidence level, peaks at about 8 percent of the notional after 1 year.

FIGURE 18-3 Exposure profile for a 5-year interest-rate swap.

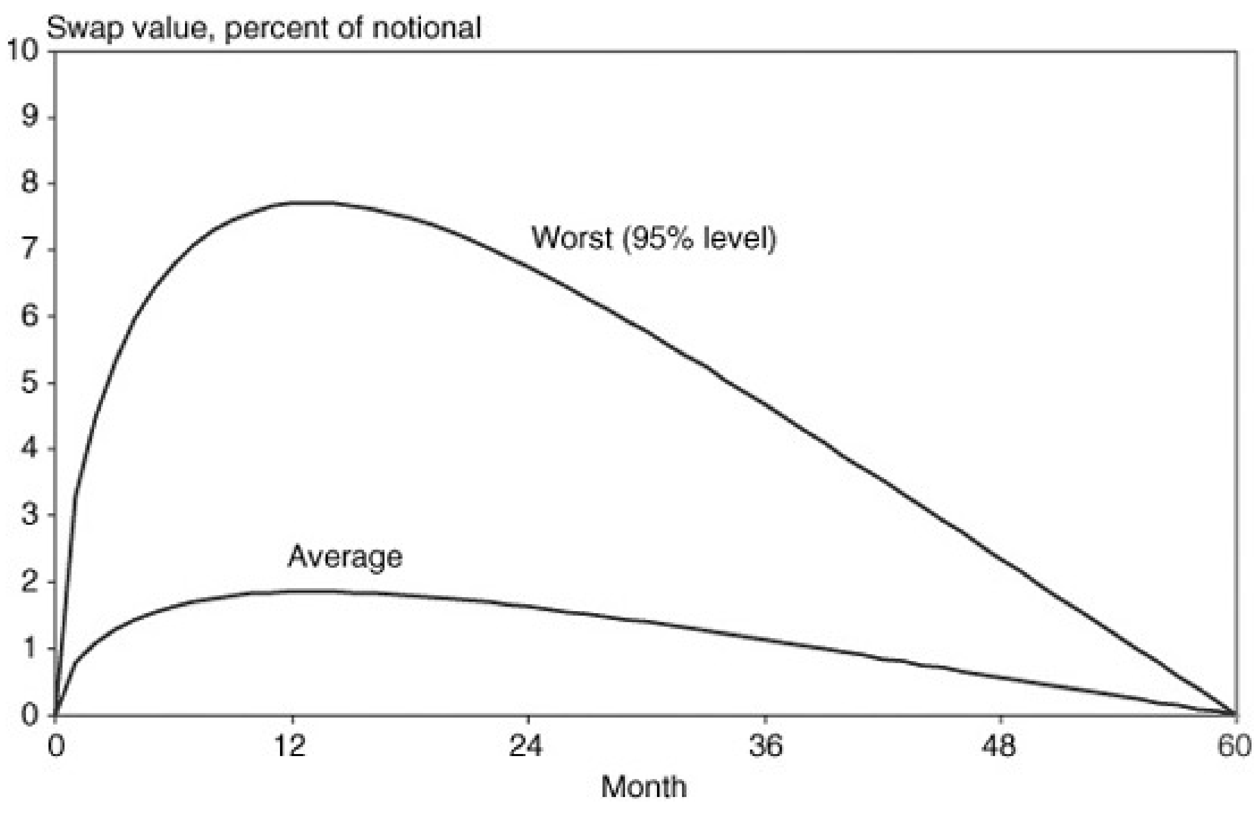
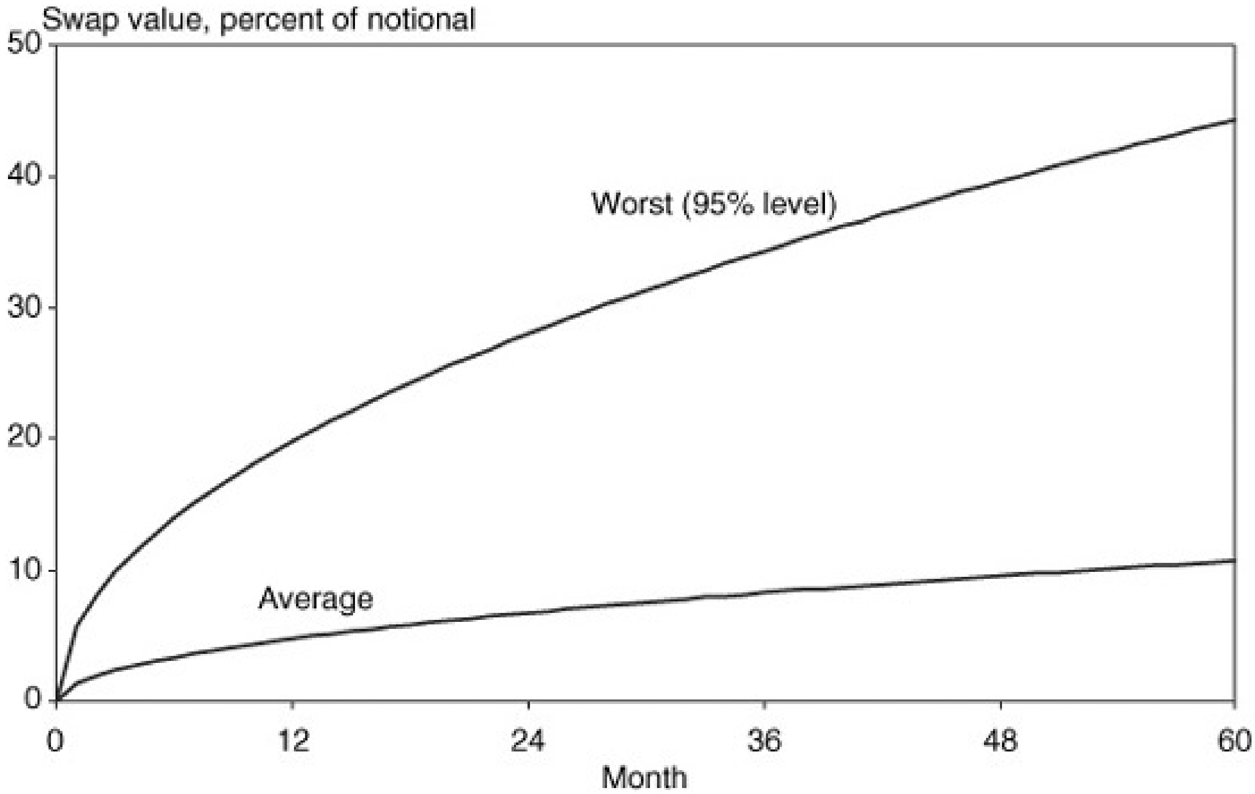


FIGURE 18-4 Exposure profile for a 5-year currency swap.



In contrast, the exposure on a currency swap increases steadily with the passage of time. This is so because exchange rate risk applies to all coupon payments and also to the principal, which is exchanged in two different currencies. There is no amortization effect and, in addition, very little mean reversion in exchange rates. Figure 18-4 shows an average exposure of about 10 percent at maturity.[4] The maximum exposure can be quite large.

18.3.2 Expected and Worst Exposure

Expected credit exposure (ECE) is the expected value of the asset replacement value x, if positive, on a target date, that is, where $f(x)$ is the distribution function of x. Note that the credit exposure is intertwined with market risk. This formula is also akin to an option.

$$\text{ECE} = \int_{-\infty}^{+\infty} \max(x, 0) f(x)\, dx \qquad (18.9)$$

Worst credit exposure (WCE) is the largest (worst) credit exposure at some level of confidence c. This is also sometimes defined as *credit at risk* (CAR). Like VAR, it is implicitly defined as the largest value such that

$$1-c=\int_{\mathrm{CAR}}^{\infty} f(x)\,dx \tag{18.10}$$

As an example, suppose that the payoff is normally distributed. The expected credit exposure then is $\mathrm{ECE}=\frac{1}{2}\,E(x|x>0)=\sigma/\sqrt{2\pi}$.[5] The worst credit exposure at the 95 percent level is given by WCE = 1.65σ.

Consider, for instance, an outstanding forward or swap contract. If the current in-the-money value of the contract is x_0, we have

$$\mathrm{ECE}=\text{notional}\times(x_0+\sigma/\sqrt{2\pi}) \tag{18.11}$$

For a bond or a loan, we could assume that changes in the market value are small relative to the principal, that is,

$$\mathrm{ECE}=\text{principal} \tag{18.12}$$

This also applies to *receivables*, *trade credits* (where default applies to the face amount at maturity), and *financial letters of credit* (which are guarantees against default and are fully drawn when default occurs). For *short* positions in options, for which the premium has been paid, the option contract can expire either worthless or as a liability. Hence there is no credit exposure to the counterparty, that is,

$$\mathrm{ECE}=0 \tag{18.13}$$

For *long* positions in options, the current exposure is the value of the option.[6] All the instruments with the same counterparty should be analyzed in this fashion.

More generally, exposure should take into account *exposure modifiers*, which attempt to decrease the exposure to the counterparty. These include *exposure limits*, which, when reached, require a payment from the counterparty. Also, *collateral* can be held as a means to reduce exposure.

The ultimate form of credit exposure reduction is *daily marking to market*, in which case changes in the value of the derivative are settled daily, reducing the exposure to intraday volatility. This, of course, creates other types of risk, namely, liquidity and operational risk, because the cash flows must be managed daily.

18.3.3 Netting Arrangements

An important method to control credit-risk exposures is *netting agreements*. Netting allows the offsetting of obligations under the same agreement, resulting in one single net claim against the counterparty. Closeout netting is by now a standard provision in the legal documentation of over-the-counter (OTC) derivative contracts that have been helped by a standardized agreement established in 1992 by the *International Swaps and Derivatives Association* (ISDA).

For instance, assume that Bank A has two derivative contracts with Bank B that fall under the same master netting agreement. Say that the first contract has a positive value of $100 million and that the second has a negative value of $80 million. Without netting, the exposure if Bank B defaults is $100 million. In contrast, with netting, the exposure is the difference, or $20 million, which is considerably less.

More generally, with a set of N derivatives contracts between two parties and no netting, the potential loss is the sum of all positively valued contracts, that is,

$$\text{Gross loss} = \sum_{i=1}^{N} \max(V_i, 0) \tag{18.14}$$

In contrast, with a netting agreement, the exposure is reduced to the positive sum of the market value of *all* contracts in the agreement, that is,

$$\text{Net loss} = \max(V, 0) = \max\left(\sum_{i=1}^{N} V_i, 0\right) \tag{18.15}$$

This can be further reduced if collateral is held against the exposure. This is always less than the gross exposure.[7]

Netting can be applied to the *current exposure*. Without netting agreements or collateral, the *gross replacement value* (GRV) is the sum of the worst-case loss over all counterparties K, that is,

$$\text{GRV} = \sum_{k=1}^{K} \text{gross loss}_k = \sum_{k=1}^{K} \left[\sum_{i=1}^{N_k} \max(V_i, 0)\right] \tag{18.16}$$

With netting agreements, the exposure is defined as the *net replacement value* (NRV), that is,

$$\text{NRV} = \sum_{k=1}^{K} \text{net loss}_k = \sum_{k=1}^{K} \left[\max\left(\sum_{i=1}^{N_k} V_i, 0 \right) \right] \quad (18.17)$$

In the preceding example, we had GRV = \$100, and NRV = \$20. The netting benefit in this case can be defined as 1 – NRV/GRV = 1 – \$20/\$100 = 80%.

To illustrate the importance of netting, Table 18-5 compares derivatives information provided in annual reports for a group of major commercial banks and securities firms. The notional amounts in the first column are huge, ranging from \$6 trillion to \$45 trillion. This is many times greater than these institutions' capital, which ranges from \$20 to \$100 billion. For JPM Chase, for example, the total derivatives notional amount is close to 500 times its capital base.

TABLE 18-5

Derivatives Credit Risk: 2004 (Billions of Dollars)

	Derivatives					
Institution	Notional Amount	Gross Replacement Value	Net Replacement Value	NRV after Collateral	Risk-Weighted Assets	Capital
U.S. banks						
JPM Chase	45,010	753	66	52	72	97
Bank of America	17,896	254	40	30	25	92
Citicorp	16,950	245	49	44	41	87
U.S. securities firms						
Goldman Sachs			57	48		23
Merrill Lynch			36	25		28
Morgan Stanley			67	39		29
Non-U.S. banks						
Deutsche Bank	29,288	442	92			39
UBS	16,884	250	69			31
RBS	15,748	174	35			83
HSBC	6,436	71	30			91

Of course, the notionals give no indication of the potential loss. A more appropriate measure of exposure is the gross replacement value. For JPM Chase, for example, GRV only accounts for 1.7 percent of the notional. The NRV is substantially lower, at 0.15 percent of notional, and is further reduced after the collateral is applied. This leads to a current exposure of $52 billion, which looks more manageable.

Ideally, we would like to transform this number into a bond-equivalent exposure. For this, we would need to account for potential increases in the exposure over the horizon, including diversification effects. Figure 18-5 illustrates the evolution of this exposure for JPM Chase's derivatives portfolio. Assuming that no new trades are added to the portfolio, the average exposure declines from the current $52 billion to about $10 billion after 10 years. The worst credit exposure is computed at the 97.5 confidence level, taking into account portfolio diversification effects. For example, take an offsetting swap trade with two clients. Even if the two clients were to default simultaneously, only one of the two trades could generate a potential credit exposure. Thus this peak exposure must be computed using simulations of market-risk factors. The graph shows that the peak exposure increases to about $65 billion after 1 year and then declines to close to $20 billion. Considering that the bank had $402 billion in outstanding loans, the credit exposure owing to derivatives appears reasonable.

FIGURE 18-5 Credit exposure profile: JPM Chase.

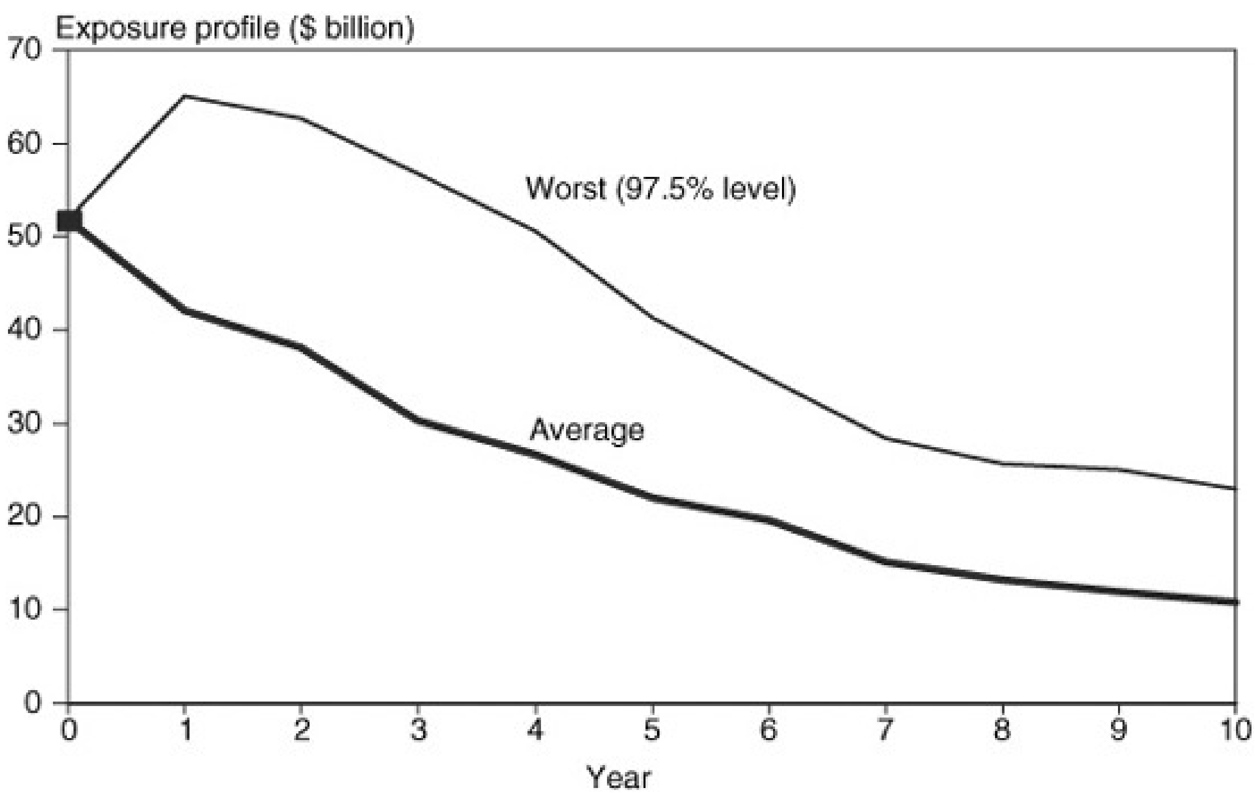


Figure 18-5 shows the output of an exposure simulation model. The Basel risk charges for derivatives, which are detailed in Appendix 18.A, provide a simplistic approximation that replaces derivatives positions by bond-equivalent exposures The penultimate column in Table 18-5 shows this *risk-weighted asset* equivalent for derivatives, which is $72 billion for JPM Chase. This is still a worst-case scenario, however, which assumes that all derivatives counterparties will default at the same time, thereby ignoring default probabilities and diversification effects.

18.4 MEASURING CREDIT RISK

18.4.1 Pricing Credit Risk

Risk measurement starts with an assessment of *expected credit losses*. In other words, the price of the asset should be low enough to cover average credit losses. Alternatively, the yield or coupon should be high enough to absorb expected losses. This assessment is relatively straightforward because it involves expectations, which are additive.

For pricing purposes, we need to compute the distribution of expected credit losses over the entire life of the asset. The first step consists of chopping up the maturity T into time intervals, say, 1 year. The expected credit loss (ECL) at each point in time t is defined as where the probability is that of defaulting during the year ending at t, measured as $k_t = (1 - c_{t-1})d_t$.

$$\mathrm{ECL}_t = \mathrm{ECE}_t \times (1 - f) \times \mathrm{prob(default)}_t \tag{18.18}$$

As an example, consider a 5-year bond rated BBB with a notional of $100 million. The exposure can be taken as constant and equal to the notional. We assume a recovery rate of 37 percent. The cumulative 5-year default probability is 0.0325, from Table 18-2. The total credit loss over the life of the bond then is ECL = $100 × (1 – 0.37) × 0.0325 = $2.047 million. This is also 205 basis points.

For more precision, this calculation should involve the time variation in the exposure and in the probability of default and a discounting factor. Define PV_t as the present value of a dollar paid at t. The *present value of expected credit losses* (PVECL) can be obtained as the sum of the discounted expected credit losses, that is,

$$\mathrm{PVECL} = \sum_t \mathrm{ECL}_t \times \mathrm{PV}_t = \sum_t [\mathrm{ECL}_t \times (1 - f) \times k_t] \times \mathrm{PV}_t \tag{18.19}$$

A shortcut consists of taking the average default probability and the average exposure, that is,

$$\text{PVECL}_2 = \left[(1/T)\sum_t \text{ECE}_t\right](1-f)\left(\sum_t k_t\right) \times \left[(1/T)\sum_t \text{PV}_t\right] \quad (18.20)$$

Ignoring the discounting factor PV, this is also the average exposure times the loss given default times the cumulative default probability. This, however, may be oversimplifying if the default probabilities or exposure profiles change over time in a related fashion.

Consider a 5-year interest-rate swap with a counterparty initially rated BBB and a notional of $100 million. This extends our previous example of a bond to a swap. Table 18-6 illustrates the computation of PVECL using a discount factor of 6 percent.

The first column reports the cumulative default probability at different points in time. The second column shows the marginal probability, and the third column shows the probability of defaulting in each year, conditional on not having defaulted before. The next column reports the annual ECE for this swap. Combining with (1 – 0.37) = 0.63 and PV leads to the factor in the last column. The total is $0.01465 million on a swap with notional of $100 million, which is about 1.5 basis point. The shortcut gives PVECL_2 = $1.0384 × (1 – 0.37) × 0.0325 × [(1/5) 4.2124] = $0.0179 million, which is close. Thus the expected credit loss is quite low for this swap, a hundred times less than for our bond because the exposure is less.

TABLE 18-6

Computation of Expected Credit Loss

Year	c_t	d_t	k_t	ECE_t	$(1-f)$	PV_t	Factor
1	0.0029	0.00290	0.0029	1.862	0.63	0.9434	0.00321
2	0.0086	0.00572	0.0057	1.631	0.63	0.8900	0.00521
3	0.0148	0.00625	0.0062	1.130	0.63	0.8396	0.00371
4	0.0237	0.00903	0.0089	0.569	0.63	0.7921	0.00253
5	0.0325	0.00901	0.0088	0.000	0.63	0.7473	0.00000
Total			0.0325			4.2124	0.01465

PVECL provides essential information for pricing purposes. It can be used as the basis for computing a minimum bid-ask spread and a credit provision. It should be deducted from revenues when computing risk-adjusted return on capital measures.

18.4.2 Portfolio Credit Risk

Once information has been gathered on the exposures, default probabilities, and recovery rates for all the assets in the portfolio, the distribution of losses owing to credit risk at the selected horizon can be described as

$$L = \sum_{i=1}^{N} \mathrm{CE}_i \times (1 - f_i) \times b_i \qquad (18.21)$$

where CE_i is the credit exposure, f_i is the recovery rate, and b_i is a random variable that takes the value of 1 if default occurs and 0 otherwise, with probability p_i.

Table 18-7 gives an example of a $100 million portfolio with three issuers, rated BB, B, and C, respectively. For simplicity, assume that the exposures are constant, that there is no recovery, and that default events are independent across issuers. The top of the table displays exposures and default probabilities over the next year. The bottom part lists all possible states. In the first state, there in no default, which happens with probability given by $(1 - p_1)(1 - p_2)(1 - p_3) = (1 - 0.01)(1 - 0.06)\ (1 - 0.32) = 0.6328$. In the second state, only the first issuer defaults, with probability $p_1(1 - p_2)\ (1 - p_3) = 0.0064$, and so on.

We then can tabulate the frequency of credit losses. From the distribution, we can compute the expected loss, which is $16.45 million, and the standard deviation, which is $22.31 million. The worst loss at the 95 percent confidence level is $45 million.[8] More generally, numerical simulation methods are needed to take into account many more assets, varying exposures, uncertain default probabilities and recovery rates, and correlated defaults.

TABLE 18-7
Portfolio Exposures, Default Risk, and Credit Losses

Issuer	Exposure	Probability
A	$25	0.01
B	$30	0.06
C	$45	0.32

Default i	Loss L_i	Probability $p(L_i)$	Cumulative Probability	Expected $L_i p(L_i)$	Variance $(L_i - EL_i)^2 p(L_i)$
None	$0	0.6328	0.6328	0.0000	171.24
A	$25	0.0064	0.6392	0.1598	0.47
B	$30	0.0404	0.6796	1.2118	7.42
C	$45	0.2978	0.9774	13.4006	242.73
A,B	$55	0.0004	0.9778	0.0224	0.61
A,C	$70	0.0030	0.9808	0.2106	8.63
B,C	$75	0.0190	0.9998	1.4256	65.16
A,B,C	$100	0.0002	1.0000	0.0192	1.34
Sum				$16.45	497.59

18.4.3 Horizon and Confidence Levels

So far we have not discussed the choice of the horizon or confidence level. Increasing either will increase VAR or the unexpected loss. Chapter 5 explained that the choice of these parameters should reflect the purpose of VAR. If VAR is to be used as a measure of economic capital, we need a long horizon and high confidence level. Credit-risk models commonly choose a 1-year horizon and a confidence level of at least 99.9 percent.

The long horizon is also justified by the fact that commercial loans are illiquid assets, unlike trading portfolios, and that a 1-year horizon is the typical reporting period for evaluating loans and measuring default rates. This choice is also conservative. If risk-mitigating actions can be undertaken as a problem develops, the worst risk rarely will be attained. In fact, the advent of a liquid credit derivatives market shortens the period required for corrective action.

The Basel Committee requires banks to hold sufficient tier 1 and tier 2 capital to cover unexpected losses over a 1-year horizon at the 99.9 percent confidence level. This seems very conservative. It means that a bank would fail, on average, once in a thousand years. In practice, however, tier 2 capital does not offer the same protection as tier 1 capital. Also, “the high confidence level was also chosen to protect against estimation errors” in the parameters and other model uncertainties.

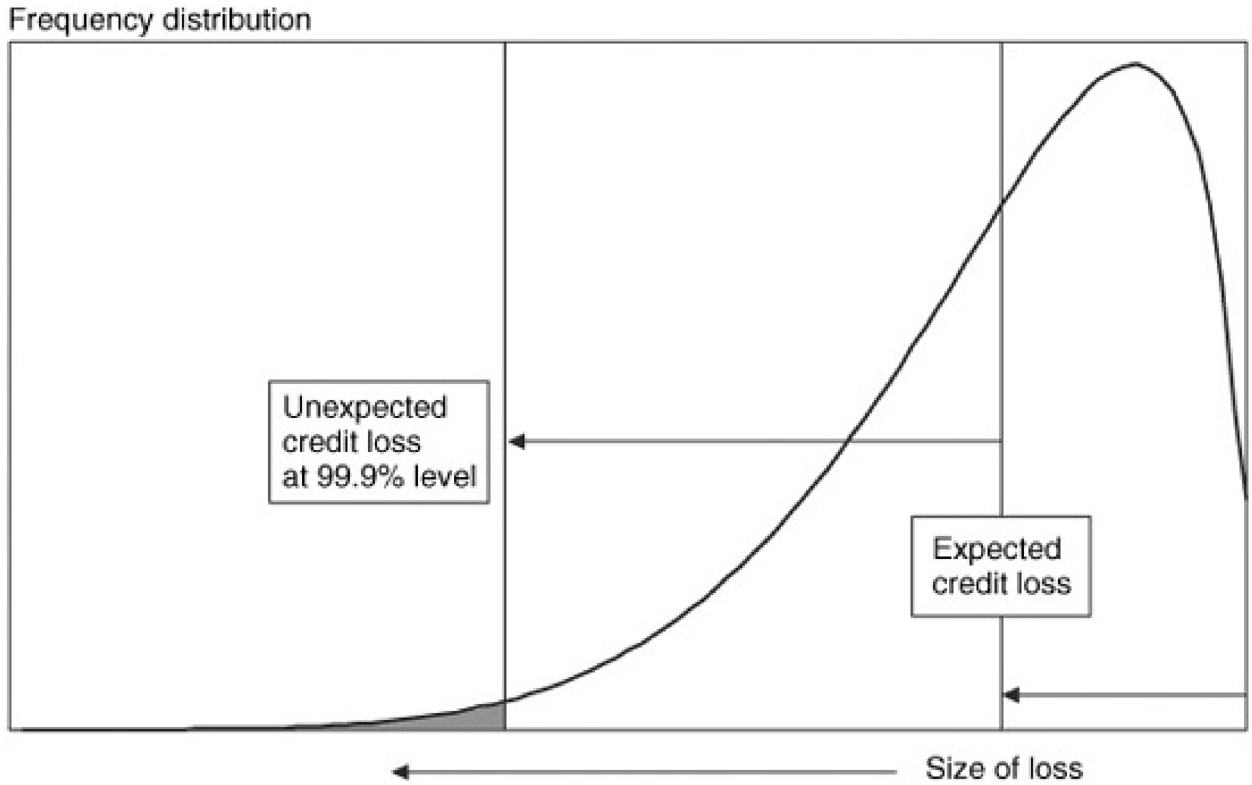
18.4.4 Credit-Risk Distribution

Figure 18-6 displays a typical distribution of credit losses, which is heavily skewed to the left. Note that this pattern is akin to a short position in an option, as discussed in Section 18.1.2. The distribution can be described by its main characteristics.

The *expected credit loss* is the average of the distribution. It represents the ongoing cost of the lending business over the next year. This should be covered by the *credit reserve*, which is the amount to set aside in anticipation of expected credit losses. At the instrument level, the pricing should provide a sufficient buffer for expected credit losses over the life of the instrument, as we have seen in Section 18.4.1.

The distribution also can be described by its volatility. More important is the *unexpected credit loss* at some confidence level. This is the deviation between the quantile and the expected loss. The *equity reserve* is the amount to set aside as a buffer to cover unanticipated credit losses, as in the case of market VAR. It is also called *economic capital*.

FIGURE 18-6 Credit-risk distribution.



For the more advanced *internal ratings-based* (IRB) approaches, the Basel II Accord distinguishes between expected loss (EL) and unexpected loss (UL). Capital is supposed to absorb unexpected losses, which means that it cannot support expected losses as well. Banks typically fund accounts called *general provisions* or *loan-loss reserves* to absorb expected credit losses. Hence Basel II withdraws general provisions from tier 2 capital.[9]

18.5 MANAGING CREDIT RISK

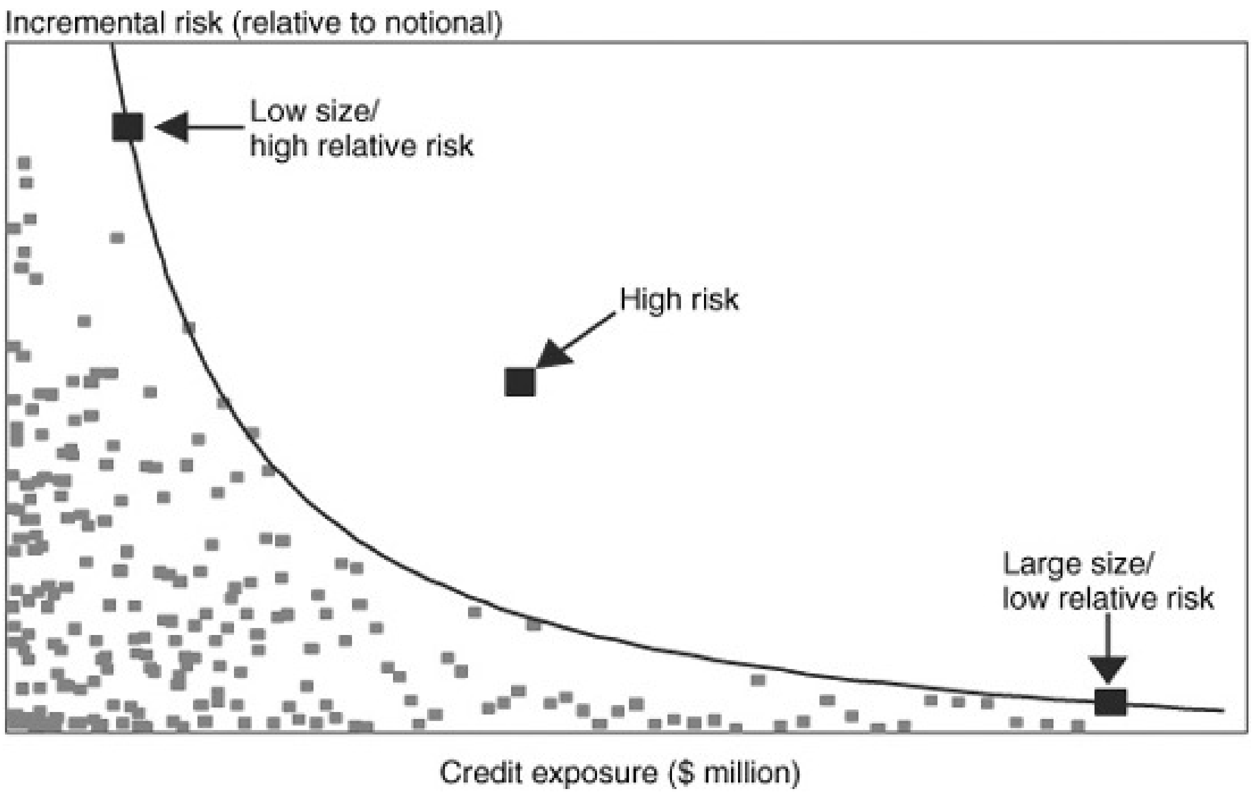
Risk that is measured can be managed. Portfolio credit-risk models can help banks to decide whether to extend credit based on the incremental credit VAR, as opposed to decisions made on a stand-alone basis. They allow banks to manage their credit risk better by identifying concentrations of credit risk by name, industry, country, or product. By assessing credit risk more finely, they make it possible to carry a smaller level of economic capital, increasing returns to the existing capital.

Portfolio management involves trading off expected profit against risk. The first step consists of measuring the expected net profit on each position, taking into account the pricing of the instrument and expected credit losses. The second step compares these expected net profits with the marginal or incremental contribution to risk. Ultimately, this should help to move the portfolio toward the best risk-return profile.

18.5.1 Risk Concentrations

Figure 18-7 illustrates the identification of risk concentrations. The horizontal axis measures the credit exposure in millions of dollars, for example. The vertical axis measures the incremental VAR in percent of exposure, or relative risk. The dollar incremental VAR is the product of the two coordinates. As explained in Chapter 7, *incremental risk* is (minus) the change in VAR when the position is deleted from the portfolio. The solid line represents the locus of points with the same dollar incremental VAR.

FIGURE 18-7 Identifying risk concentrations.



Consider two points on this line. On the right, a position has large size but low relative risk. On the left, another position has high relative risk, which means that it is too similar to the rest of the portfolio, but low size. These two points have identical dollar incremental risk.

The point in the center, however, has even higher risk. This position should attract immediate scrutiny from the risk manager because eliminating it will have the biggest impact on the portfolio. The asset could be sold off or hedged using a credit default swap, which provides protection in case the obligor defaults. In practice, such optimization techniques typically reduce economic capital by at least 30 percent.

18.5.2 Portfolio Credit Risk Models

The need to take into account portfolio effects led to the development of portfolio credit-risk models. Such models can account for the time profile of credit exposure, for realistic default rates, and for correlations.

This is generally achieved through simulation methods, which explicitly model movements in asset prices, defaults, changes in exposure, and losses across the whole portfolio. Such methods can account for the legal structure of transactions by assigning individual transactions to master netting agreements, master netting agreements to counterparties, and even counterparties to countries. These methods account for portfolio diversification effects across transactions and counterparties.

Table 18-8 compares leading credit-risk models that are available commercially. The models differ in a number of key dimensions. In terms of risk definitions, models can consider either losses owing to the occurrence of default only or, more generally, as any change of the market value in the debt. Portfolio Manager and Credit Manager are based on a Merton-type structural model where defaults are driven by a causal model using latent variables. The models are similar in philosophy but differ in implementation. Portfolio Manager uses more firm-level information. Both models assume random recovery rates and derive the credit distribution from simulations.

Credit Risk+, in contrast, is a reduced-form model based on a top-down, actuarial approach where defaults are drawn from a fixed distribution. No economic causality is assumed. The model assumes fixed recovery rates and provides an analytical solution to the distribution. A recent survey of 52 institutions revealed that 50 percent use Portfolio Manager, 15 percent Credit Manager, 5 percent Credit Risk+, and 40 percent an internal model, creating some overlap between responses.

TABLE 18-8
Comparison of Credit-Risk Models

	Portfolio Manager	Credit Manager	Credit Risk+
Originator	KMV Moody's	RiskMetrics	Credit Suisse
Risk definition	Market value	Market value	Default
Philosophy	Structural, bottom-up, causal	Structural, bottom-up, causal	Reduced-form, top-down, no causality
Default probability	Merton model	Exogenous	Exogenous
Risk drivers	Asset values	Asset values	Poisson distribution
Correlation	Factor model	From equities	Across groups
Recovery rates	Random	Random	Constant
Solution	Simulation	Simulation	Analytical

TABLE 18-9

Comparison of Credit-Risk Models (Percentage of Total Exposure)

Model	Expected Loss	Volatility	Economic Capital
Portfolio Manager	0.61%	1.60%	7.96%
Credit Manager	0.79%	1.04%	6.25%
Credit Risk+	0.73%	1.12%	7.42%

Source: Adapted from Rutter Associates (2003).

While these models seem to have taken disparate approaches to credit risk, in fact, they have a very similar underlying mathematical structure.[10] Table 18-9 compares the outputs of the three major risk models applied to the same portfolio. The test portfolio is representative of a large bank portfolio and is made up of 2903 obligors with a total exposure of $61 billion. The portfolio has average credit rating of BBB and is fairly well diversified across industries.

The table shows similar numbers for expected losses, around 0.7 percent of notional. Economic capital, measured at the 99.85 percent confidence level, is also similar across models, around 6 to 8 percent of notional.

18.5.3 Regulatory Capital

In Table 18-9, the sum of expected loss and economic capital is close to 8 percent of notional. This explains the simplistic approach of Basel I, which imposes a capital charge of 8 percent to the typical bank portfolio. Of course, economic capital is not a fixed number but depends on default probabilities and correlations. More (less) risky assets should carry greater (less) economic capital. This explains why *Basel II* moves to more credit-sensitive regulatory capital charges.

Under the new rules, banks have a choice between a standardized approach, which is a simple extension of the Basel I rules, and a more complex *internal ratings-based* (IRB) *approach*. Under the standardized approach, risk weights are assigned to external credit ratings, as indicated in Chapter 3 (Table 3-2).

Under the *foundation IRB approach*, banks estimate the *probability of default* (PD), and supervisors supply other inputs. Under the *advanced IRB approach*, banks can supply other inputs as well. These include *loss given default* (LGD) and *exposure at default* (EAD). The combination of PDs and LGDs for all applicable exposures then are mapped into regulatory risk weights. The credit-risk charge is obtained by multiplying the risk weight RW by EAD by 8%, that is,

$$\mathrm{CRC} = 8\%\left(\sum_i \mathrm{RW}_i \times \mathrm{EAD}_i\right) = \sum_i K_i \times \mathrm{EAD}_i \qquad (18.22)$$

There is something strange about Equation (18.22), though. The Basel capital charges add up the risk charges for each individual credit. VAR is fundamentally nonlinear in the positions, however. Even in the VAR decomposition explained in Chapter 7, component VAR depends on the portfolio. Thus this formula can't be exact.

More precisely, a VAR decomposition cannot be generally *portfolio invariant*. Portfolio invariance implies that the capital for a given loan depends on the risk of that loan only and does not depend on the portfolio to which it is added. Gordy (2003) shows that portfolio invariance only obtains for *asymptotic single-risk-factor models*. These models assume a large number of positions, asymptotically tending to infinity, where the latent variables are driven by one common factor. Owing to perfect diversification of idiosyncratic risk, the capital charge depends on the risk of each loan only.

Thus the Basel capital charges are only invariant under a restricted set of conditions. Appendix 18.B gives more detail on the construction of the risk weights. Capital has been calibrated to typical portfolio correlations. As a result, the new rules penalize banks with greater than average diversification. Compromises and simplifications were necessary in the absence of the next-best alternative, which was to allow banks to use their own, internal portfolio models. With the increased level of knowledge, use, and comfort about portfolio credit-risk models, however, it is possible to envision a next set of rules, predictably called *Basel III*, that would allow banks to use their internal credit models, as has happened with market risk.

18.6 CONCLUSIONS

In recent years, financial institutions have developed formal models for quantifying the credit risk of their portfolios. Such models are based on the traditional VAR framework first developed from market risk.

We have seen that these top-down models are driven by various risk factors: the state of default, exposures at default, and loss given default. Integrating these risk factors leads to a distribution of credit losses, which is considerably more complex than market-risk models.

Some fundamental insights, however, are common to traditional VAR models. For large portfolios, correlations are important. In particular, the extent of the tails is heavily influenced by correlations between defaults. With zero default correlations, increasing the number of credits leads to diminishing tails, through the central limit theorem. Higher default correlations increase risk. In practice, we observe positive default correlations. Proper assessment of this parameter is a crucial component of these models.

These portfolio credit-risk models have revolutionized the industry. For the first time, banks and portfolio managers have been able to assess the amount of economic capital required to sustain credit portfolios. Better measurement leads to more efficient use of capital. It also allows banks to price credit at a transaction level, thereby gaining an advantage over the competition. In this process, we should gain a safer banking system.

Indeed, U.S. banks weathered the 2001 recession remarkably well in large part because of risk management.[11] Chairman Greenspan (2003) stated the ". . . application of more sophisticated methods for measuring and managing risk are key factors underpinning the enhanced resilience of our largest financial intermediaries."

These models are still evolving, however. In particular, credit-risk models suffer from a verification problem. Unlike market risk, for which backtesting can be performed on a daily basis, the longer horizon of credit-risk models makes it difficult to compare risk forecasts with their realization. Measures of economic capital depend crucially on default correlations, which are difficult to assess. The true test of these new models will come during a major economic downturn.

APPENDIX 18.A
The Basel Risk Charges for Derivatives

The Basel Accord requires commercial banks to hold capital in excess of 8 percent of their total risk-weighted assets (see BCBS, 1995c, 2005c). For on-balance-sheet (ONBS) items, the notional is the credit exposure. For off-balance-sheet (OFFBS) items, however, notionals do not represent the worst exposure. Under the standardized approach, items are converted into credit-exposure equivalents through the use of *credit conversion factors* (CCFs). Commitments to provide funding, for example, are OFFBS items. Such commitments with maturity over 1 year receive a CCF of 50 percent because they are less risky than outright loans.

Derivatives are another class of OFFBS items. For these, the Basel I rules and the *current exposure method* under the new Basel II rules are as follows: A bond-equivalent *credit exposure* is computed as the sum of the current replacement value plus an *add-on* that is supposed to capture the potential exposure, that is, where the add-on factor depends on the tenor and type of contract, as listed in Table 18-10. NGR is the *net-to-gross ratio*, or ratio of current net market value to gross market value, which is always between 0 and 1.

$$\text{Credit exposure} = \text{NRV} + \text{add-on} \tag{18.23}$$

$$\text{Add-on} = \Sigma\,[\text{notional} \times \text{add-on factor} \times (0.4 + 0.6 \times \text{NGR})]$$

The add-on factor roughly accounts for the maximum credit exposure, such as described in Figures 18-2 and 18-3, which depends on the volatility of the risk factor and the maturity. This explains why the add-on factor is greater for currency swaps than for interest-rate instruments. Finally, the purpose of the NGR formula is to reduce the capital requirement for contracts that fall under a netting agreement by, at most, 60 percent.

TABLE 18-10

Add-on Factors for Potential Credit Exposure (Percent of Notional)

Residual Maturity (Tenor)	Contract: Interest Rate	Exchange Rate, Gold	Equity	Precious Metals, Except Gold	Other Commodities
< 1 year	0.0	1.0	6.0	7.0	10.0
1–5 years	0.5	5.0	8.0	7.0	12.0
> 5 years	1.5	7.5	10.0	8.0	15.0

Risk-weighted assets then are obtained by applying counterparty risk weights to the credit exposure. For instance, take a $100 million interest-rate swap with a domestic corporation and a residual maturity of 4 years. Say that the current market value of the swap is $1 million. Using the 0.5 percent add-on factor, the total credit exposure is CE = $1 + $100 × 0.005 = $1.5 million. This number must be multiplied by the counterparty-specific risk weight to compute its risk-weighted asset value.

Note that in this Basel framework, credit risk is evaluated on a *transaction-by-transaction basis*. Default risk is taken into account indirectly through the risk weights, which vary across types of counterparties. The main drawback of this approach, of course, is that it completely ignores the potential for diversification across time and counterparties.

APPENDIX 18.B
The Basel IRB Risk Weights

Table 18-11 illustrates the link between PD and the risk weights for various asset classes under the IRB approach. For instance, a corporate loan with a 1.00 percent probability of default would be assigned a risk weight of 92.32 percent, which is close to the standard risk weight of 100 percent from Basel I. Note that retail loans have much lower risk weights than the other categories, which makes sense because retail credits are more diversified than other types of loans.

TABLE 18-11
IRB Risk Weights

Probability of Default	Corporate	Residential Mortgage	Other Retail
0.03%	14.44%	4.15%	4.45%
0.10%	29.65%	10.69%	11.16%
0.25%	49.47%	21.30%	21.15%
0.50%	69.61%	35.08%	32.36%
0.75%	82.78%	46.46%	40.10%
1.00%	92.32%	56.40%	45.77%
2.00%	114.86%	87.94%	57.99%
3.00%	128.44%	111.99%	62.79%
4.00%	139.58%	131.63%	65.01%
5.00%	149.86%	148.22%	66.42%
10.00%	193.09%	204.41%	75.54%
20.00%	238.23%	253.12%	100.28%

Note: Illustrative weights for LGD = 45 percent, maturity of 2.5 years, and large corporate exposures (firms with turnover greater than 50 million euros).

The IRB capital charge is computed as follows[12]: where N is the cumulative normal distribution function, N^{-1} its inverse, T is the effective maturity, and ρ(PD) and b(PD) are predefined functions.

$$K = \left\{ \text{LGD}\, N\left[\frac{N^{-1}(\text{PD}) + \sqrt{\rho(\text{PD})}N^{-1}(0.999)}{\sqrt{1-\rho(\text{PD})}}\right] - \text{PD LGD}\right\}\left[\frac{1+(T-2.5)b(\text{PD})}{1-1.5b(\text{PD})}\right] \quad (18.24)$$

While this formula looks complicated, it can be parsed to a few components. First, the PD LGD term in the middle is simply the expected loss. Banks have to carry capital separately for expected loss, so this is subtracted from the total loss.

The last term, between brackets, is an adjustment for the maturity of the asset. Longer-term credits are more risky than short-term credits. The formula is calibrated to a standard maturity of 2.5 years. The adjustment is a function of PD because safer credits, with low PD, have more potential for downgrades over a long maturity.

Finally, the first term $N(\ldots)$ represents the component VAR under an asymptotic single-risk-factor (ASRF) model. Using a decomposition similar to Equation (7.34), Gordy (2003) shows that component VAR under this ASRF model is which is proportional to the probability that the obligor asset return R_i is less than a cutoff value Z_i, conditional on a common factor value m.

$$\text{VAR} = \sum_i \text{EAD}_i \times \text{LGD}_i \times P(R_i \leq Z_i \mid m) \tag{18.25}$$

In the one-factor model, we decompose this return into a market return and an idiosyncratic return, that is, assuming that M and ϵ_i have standard, normal, independent distributions.

$$R_i = \sqrt{\rho_i}\; M + \sqrt{1-\rho_i}\, \epsilon_i \tag{18.26}$$

We note that Z_i can be chosen to fit the default probability for the obligor, that is, which implies $Z_i = N^{-1}(\text{PD}_i)$. Inserting this and Equation (18.26) into the conditional probability gives

$$P(R_i \leq Z_i) = \text{PD}_i \tag{18.27}$$

$$P(R_i \leq Z_i \mid m) = P\left[\sqrt{\rho_i}\, M + \sqrt{1-\rho_i}\, \epsilon_i \leq N^{-1}(\text{PD}_i) \mid m\right]$$

$$P(R_i \leq Z_i \mid m) = P\left[\epsilon_i \leq \frac{N^{-1}(\text{PD}_i) + \sqrt{\rho_i}(-M)}{\sqrt{1-\rho_i}} \mid m\right]$$

$$P(R_i \leq Z_i \mid m) = P\left[\epsilon_i \leq \frac{N^{-1}(\text{PD}_i) + \sqrt{\rho_i}\, N^{-1}(0.999)}{\sqrt{1-\rho_i}}\right]$$

where we set the market factor M to correspond to a 99.9 percent confidence level, $P(-M \leq m) = 0.999$, or $M = N^{-1}(0.999)$. This gives

$$P(R_i \leq Z_i \mid m) = N\left[\frac{N^{-1}(\text{PD}_i) + \sqrt{\rho_i}\, N^{-1}(0.999)}{\sqrt{1-\rho_i}}\right] \tag{18.28}$$

which is exactly the N (…) term in Equation (18.24). The model is calibrated to an average asset correlation of 0.20 but also includes further adjustments to correlations. The ρ(PD) function accounts for the empirical observation that defaults on low-rated obligors are less dependent than others.

QUESTIONS

1. What is credit risk?
2. Banks usually have a comparative advantage in making loans to a group of companies that they know well, which can be in the same industry or geographic location. What is the drawback of this advantage?
3. What makes credit risk more difficult for quantification and backtesting than market risk?
4. Explain why the distribution of credit-risk gains and losses is similar to a short position in an option. Discuss loans and derivatives.
5. For B-rated borrower, S&P gives cumulative default probabilities of 6.24 and 14.33 percent for a 1-and 2-year horizon, respectively. Compute the (marginal) probability that this borrower will default during year 2.
6. Can the cumulative default probability go down as the horizon extends? How about the marginal default probability?
7. According to Moody's, what type of debt carries the highest recovery rates?
8. A B-rated bond trades at a yield of 11 percent, versus 5 percent for a Treasury debt, with the same maturity of 1 year. Compute the implied default probability assuming a recovery of 37 percent.
9. Compare the number with the default rate in Table 18-2. Discuss what could explain the difference.
10. Explain the rationale for using stock prices to infer default risk.
11. As the portfolio gets larger, how is the portfolio volatility related to the correlation of defaults?
12. Typical credit-risk models infer default correlations from the movement of asset values that is assumed to have a joint normal distribution with average correlation of 0.20. This creates a measure of economic capital (EC) at some confidence level. Give two reasons the EC could be understated.
13. What is credit exposure in the case of a defaultable bond?
14. Why is there credit exposure for the fixed-to-floating interest-rate swap?
15. Which derivative has a zero credit exposure at expiration, a 5-year interest-rate swap or a 5-year currency swap?

16. What is the credit exposure for a derivative that has negative market value?
17. Why is the net loss in case of default with the netting agreement always less than without the agreement?
18. What is worst credit exposure? How is it different from expected credit exposure?
19. Is it correct to say that the credit reserve should be based on the present value of unexpected credit losses?
20. What is the main drawback of evaluating credit risk on a transaction-by-transaction basis under the Basel framework?
21. A B-rated bond has a cumulative default probability of 32 percent over 5 years. Assuming a recovery rate of 45 percent, compute the expected credit loss for a notional of $100 million.
22. Roughly, what would you expect for the expected credit loss of a swap with the same features as in the preceding question?
23. Explain why the confidence level is high and horizon long for computing unexpected credit losses.
24. A risk manager evaluates three credits and finds Compute the dollar incremental risk. Which credit should the risk manager examine first?

Credit	Exposure ($ Million)	Relative Incremental Risk (Percent)
A	100	1
B	50	2
C	70	2

25. Explain why the Basel I capital charge for typical bank portfolios should be about 8 percent of notional.
26. The Basel II capital charges are additive. Generally, economic capital cannot be decomposed into a simple sum. Explain the meaning of portfolio invariance and how this is applied to the Basel II credit-risk charges.
27. What is the main benefit of portfolio credit-risk models?

CHAPTER 19
Operational Risk Management

> [A]n informal survey . . . highlights the growing realization of the significance of risks other than credit and market risks, such as operational risk, which have been at the heart of some important banking problems in recent years.
>
> —*Basel Committee on Banking Supervision (June 1999)*

Operational risk is perhaps the most pernicious form of risk because it is indirectly responsible for numerous failures in financial institutions. Obviously, it is not a new risk. What is new, however, is the idea that operational risk management is a discipline with its own management structure, tools, and processes.

For a long time, institutions had narrowly focused on a subset of operational risk that involves transaction processing and settlements, ignoring other aspects of operational risk. This is now changing. The industry is embracing a broader and more systematic approach to operational risks. We are learning to quantify *oprisk* using tools borrowed from the insurance industry and from value-at-risk (VAR) techniques.

Once quantified, operational risk can be subject to capital charges. There is no better way to force the lines of business to think hard about this type of risk. This increased scrutiny of operational risk has been, as in the case of credit and market risks, also spurred by bank regulators who will impose new capital charges as a further incentive for managers to pay attention to operational risks.

Section 19.1 motivates the recent interest in operational risk. While our knowledge of market and credit risk has matured, operational risk management is still in its infancy. Identification and assessment of oprisk is covered in Section 19.2. Section 19.3 then discusses the measurement of this risk. It describes the loss-distribution approach as well as data-collection issues, which are particularly delicate for this type of risk. Section 19.4 discusses the management of operational risk. Finally, the Basel regulatory charges are described in Section 19.5.

19.1 THE IMPORTANCE OF OPERATIONAL RISK

Many of the great financial fiascos can be traced to a combination of market or credit risk and failure of controls—in other words, they involve some form of operational risk. The biggest such risk is unauthorized trading. Indeed, the biggest fear of banks is to have their name end up in the hall of infamy, along with the likes of Barings, Daiwa, and AIB, as described in Chapter 2. While shareholders understand that the very function of trading is to take financial risk, thus leading once in a while to trading losses, few are willing to forgive losses owing to lack of supervision that are entirely avoidable.

Arguably, the spectacular failures of Enron, WorldCom, and others involving fraud can be attributed to operational risk. Internal fraud arose at the highest level of the organization and was due to poor corporate *governance*, which is the process of high-level control of an organization.

This recognition, along with the pace of change in the industry, leading to bigger and more complex business operations, is bringing greater interest in operational risk. At the same time, we are witnessing advances in the quantitative measurement of operational risk that would have been hard to imagine just a few years ago. These advances, built on VAR methods, lead to measures of *economic capital* (EC) against operational risk.

Figure 19-1 illustrates the distribution of operational losses, taken as positive numbers by convention, over the selected horizon. On average, the institution should incur the *expected loss* (EL). On the other hand, the *unexpected loss* (UL), measured at a given confidence level, represents the economic capital. Typical parameters are a 1-year horizon and a confidence level of at least 99.9 percent.

Note that the distribution has a long right tail.[1] It has much higher skewness and kurtosis than distributions for market and credit risk. This represents the occurrence of rare but sometimes deadly losses that could bankrupt the institution. These fat tails generate high values of economic capital, which in addition will be imprecisely measured owing to the paucity of data. Typically, these distributions are modeled using extreme-value theory (EVT), developed in Chapter 5.

FIGURE 19-1 Distribution of operational losses.

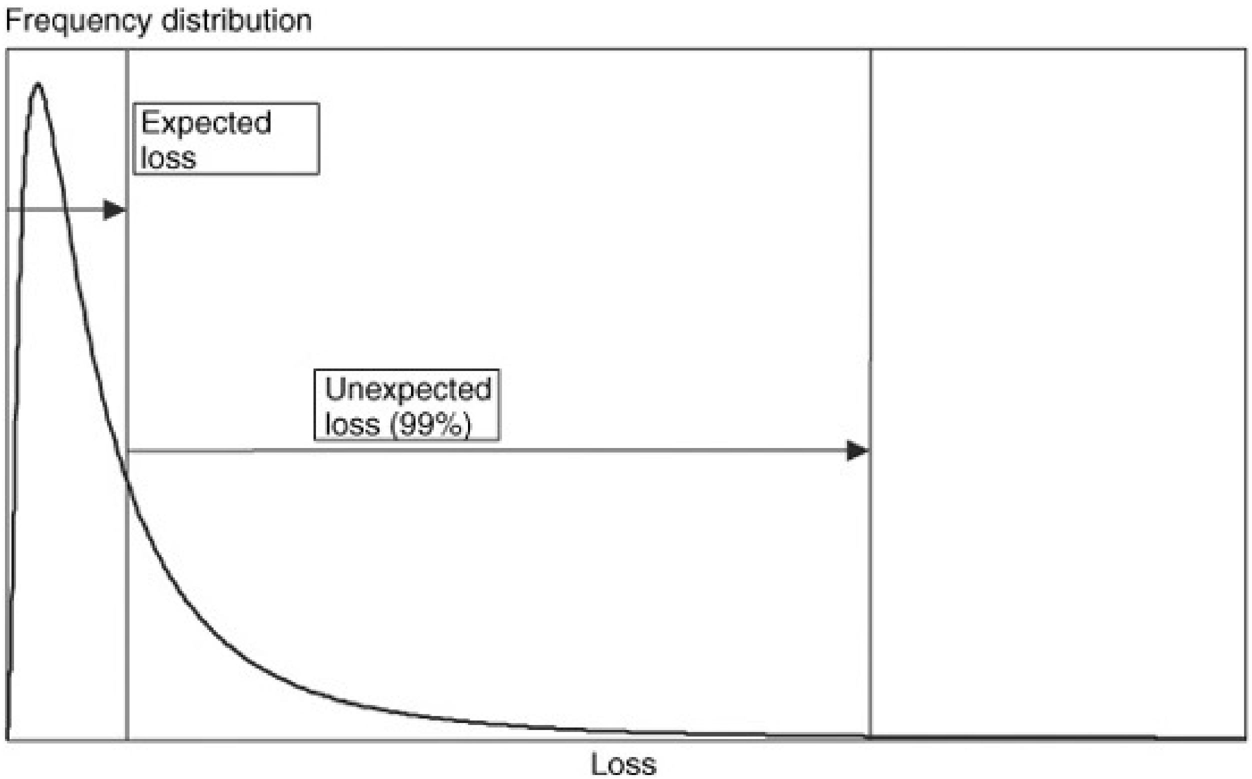


TABLE 19-1
Distribution of Economic Risk Capital (Billions) at 1-Year Horizon and 99.97 Percent Confidence Level

Risk Type	Citigroup	JPM Chase	Deutsche Bank
Credit risk	$33.2	$16.5	6.0
Market risk	$16.0	$7.5	5.5
Operational risk	$8.1	$4.5	2.2
Sum	$57.3	$28.5	13.7
Total EC	$52.2	$34.9	13.2
Actual capital	$74.4	$68.6	18.7

Source: 2004 annual reports. Actual capital is tier 1 capital. (Total EC includes diversification effects and other sources of risk.)

Table 19-1 compares measures of economic capital against credit, market, and operational risk for a group of commercial banks, all with assets in excess of $1000 billion. The table also lists the total economic capital reported by each institution, including other risk sources and diversification effects. This can be compared with *actual capital*, which is the amount of equity that can serve as a buffer against these risks.

Credit risk is the largest component of EC. Second comes market risk, from trading as well as from the interest-rate risk of banking assets. Operational risk comes in third place. On average, it accounts for 15 percent of the sum of these three categories, which is fairly high. For the average bank, this proportion would be even higher because the banks described in this table are very active in proprietary trading, which increases their market risk.

Operational risk arises in the course of all financial activities. It is less visible than traditional financial risks such as market and credit risk. This may give a false impression of safety, however. Consider the asset management business. The owners are not directly affected by market risk, which is borne by the investors. This is an attractive business because it generates a steady flow of revenues and apparently has no risk. Box 19-1, however, gives a case where a compliance failure led to a huge operational loss.

BOX 19-1

DEUTSCHE MORGAN GRENFELL'S RISK

In September 1996, the investment bank Deutsche Morgan Grenfell (DMG) announced that it had suspended a star fund manager, Peter Young, in its asset management unit. DMG also halted trading on its three main European equity funds, worth some $2.2 billion.

Apparently, Peter Young had breached the limit of 10 percent that such funds can invest in unlisted securities. This limit is imposed because of the difficulty of confirming market values for these securities. While the funds he managed had a stellar performance in 1995, they ranked dead last in their category in the first half of 1996.

Deutsche Bank, the German owner of DMG, agreed to compensate the shareholders in the funds. It later announced that it had set aside some $720 million to cover the total losses. The total cost must have been even higher as a result of the lost business from the bank's tarnished reputation.

19.2 IDENTIFICATION AND ASSESSMENT OF OPRISK

19.2.1 Identification

The first step in the measurement of any risk is its precise definition. Proper definitions are essential to assign responsibilities and risk capital. In recent years, an increasing number of banks have appointed specialist operational risk managers, with the rank equivalent of heads of credit risk and market risk, who report directly to committees in charge of overall risk. To avoid overlaps or holes in the coverage of various risks, each of these functions needs to be well defined. Further, one cannot measure operational risk without first defining it.

Previously, the industry took a narrow view of operational risk, which was restricted to the *risk arising from operations*. This involves transactions processing and systems failures, for example. This definition, however, misses significant risks such as fraud that should be of major concern to financial institutions.

At the other extreme, operational risk has been defined as any financial risk *other than market and credit risk.* This definition, however, is too broad to be useful. It also includes business risk owing to movements in prices, quantities, and costs that the firm assumes to create shareholder value.

After much discussion and with the guidance of the Basel regulators, the industry has now settled on the following definition:

> *Operational risk* is the risk of loss resulting from inadequate or failed processes, people and systems or from external events.

This definition includes *legal risk*, owing to fines and penalties resulting from supervisory actions, as well as private settlements. However, it excludes *business risk*, *strategic risk*, and *reputational risk*, where losses would be difficult to ascertain. Reputational risk is the risk of losses beyond the direct operational loss owing to the firm's damaged reputation. In extreme cases, the loss of reputation may spell the end of the institution.

This definition can be broken down further into specific categories, which are listed in Table 19-2. For detailed examples of these definitions, see BCBS (2003). The table also gives an example of loss within each category. Each loss is above $100 million, reinforcing the importance of controlling operational risk.

TABLE 19-2
Event Types for Operational Risk

Event Type	Examples, Specific Loss
Internal fraud (IF)	Employee theft, misreporting of positions: AIB, $691 million, rogue trader (2002)
External fraud (EF)	Robbery, computer hacking: Republic NY, $606 million, aided fraud by client (2001)
Employment practices and workplace safety (EPWS)	Discrimination, violation of rules: Merrill Lynch, $100 million, gender discrimination (1999)
Clients, products, and business practices (CPBP)	Improper selling, money laundering: Household Intl, $484 million, improper lending (2002)
Damage to physical assets (DPA)	Terrorism, earthquakes, fires, floods: Bank of New York, $242 million, Sept. 11 attack (2001)
Business disruption and system failures (BDSF)	Hardware or software failures: Salomon, $303 million, unreconciled accounts (1994)
Execution, delivery, and process management (EDPM)	Settlement failures, failed implementation: London Stock Exchange, $630 million, abandonment of Taurus system (1999)

19.2.2 Assessment

Once identified, operational risk can be *assessed* using the following tools, listed in order of increasing sophistication[2]:

- *Critical self-assessment*, where each department submits a *subjective* evaluation of sources of operational risk, as well as their expected frequency and costs
- *Key risk indicators*, where a centralized unit develops *subjective* risk forecasts through risk indicators, such as trading volume, number of mishandled transactions, staff turnover, and so on
- *Formal quantification*, where operational risk managers measure an *objective* distribution of operational risk losses from an event database

19.3 MEASURING OPERATIONAL RISK

19.3.1 Modeling Issues

Formal quantification of operational risk is based on VAR techniques. The essence of VAR is to aggregate all the losses an institution could suffer owing to one or several joint categories of financial risks. This applies to operational risk as well. Table 19-3 compares methods to measure market, credit, and operational risks.

Admittedly, operational risk is very different from market risk and credit risk. There is no exogenous source of data, such as movements in financial prices or defaults. Whatever data are observed are inextricably linked to the quality of the internal control environment.

Another major difference is that financial institutions seek exposure to market and credit risk to add value. Operational risk can only generate losses and is intertwined with the business process.

Otherwise, all the arguments we used for VAR in Chapter 16 are valid here, including the comparison of *bottom-up* and *top-down* approaches. Top-down approaches have the advantage of simplicity and low data requirements. For example, movements in earnings owing to operational risk can be constructed from subtracting from total earnings the effects of market risk and credit risk. This approach, however is backward-looking and includes other risk such as business risks. It provides little insight into the drivers of operational losses.

TABLE 19-3

Measurement of Financial Risks

Step	Market Risk	Credit Risk	Operational Risk
Define risk categories	Interest rate Equity Currency Commodity	Default Downgrade	Processes People Systems External events
Measure risk factors	Volatility Correlations	Default and recovery distributions	Loss frequency
Measure exposure	Duration Delta Mapping	Current and potential exposure	Loss severity
Calculate risk	Market VAR	Credit VAR Expected loss	Operational VAR Expected loss

On the other hand, bottom-up approaches are more informative, like VAR. These can be grouped into *process approaches* and *actuarial approaches*. These approaches are more detailed but provide tools for process improvement. The process approach starts with a step-by-step analysis of the procedures used for all activities. These can be linked by *causal networks* that explain dependencies between these various steps. For example, a trade-settlement failure can be attributed to a confirmation problem, a staff error, or a telecom failure, with associated probabilities. The confirmation problem itself has several causes.[3] Thus the process approach leads to a structural model of the probability of failure. Actuarial approaches are explained next.

19.3.2 Loss Distributions

The actuarial approach starts by examining separately two types of variables, the loss frequency and the loss severity when it occurs. The *loss frequency* is a measure of the number of loss events over a fixed interval of time. The *loss severity* is a measure of the size of the loss once it occurs. The *loss-distribution approach* (LDA) then combines these two variables into a distribution of total losses over the period considered.

Loss severities can be tabulated from a combination of internal and relevant external data. Say that the risk manager measures the loss severity y_k from historical observation k and adjusts it for inflation and some measure of current business activity. This gives a *scaled* loss x.

The second random variable of interest is n, the number of occurrences of losses over the horizon, taken as 1 year. Define the pdf for this random variable as

$$\text{pdf of loss frequency} = f(n) \qquad n = 0, 1, 2, \ldots \tag{19.1}$$

Next, the pdf of the loss severity x is

$$\text{pdf of loss severity} = g(x \mid n = 1) \qquad x \geq 0 \tag{19.2}$$

The total loss then is the summation of these random losses over a random number of occurrences, that is,

$$S_n = \sum_{i=1}^{n} X_i \tag{19.3}$$

TABLE 19-4
Sample Loss Frequency and Severity Distributions

Frequency Distribution			Severity Distribution	
Probability	Frequency		Probability	Severity
0.5	0		0.6	$1,000
0.3	1		0.3	$10,000
0.2	2		0.1	$100,000
Expectation	0.7		Expectation	$13,600

Table 19-4 provides a simple example of two such distributions.

Assuming that X and N are independent considerably simplifies the analysis, although in practice this assumption would have to be examined closely. If we just need the expected total loss, we can find it simply as the product of the expected frequency and severity, which here is $E(S) = E(N) \times E(X) = 0.7 \times \$13,600 = \$9,520$. The computation of the variance is more complex, however. One can show that, assuming independence of N and S, the variance is $V(S) = E(N) \times V(X) + V(N) \times E(X)^2$. To find the quantile, however, we need to recover the full distribution.

Assuming that the frequency and severity are independent, the two distributions can be combined into a probability distribution of aggregate loss through a process known as *convolution*. Convolution can be implemented through a variety of methods. We illustrate here the process through tabulation. *Tabulation* consists of systematically tabulating all possible combinations with their probability. This is only feasible with a small number of combinations, however.

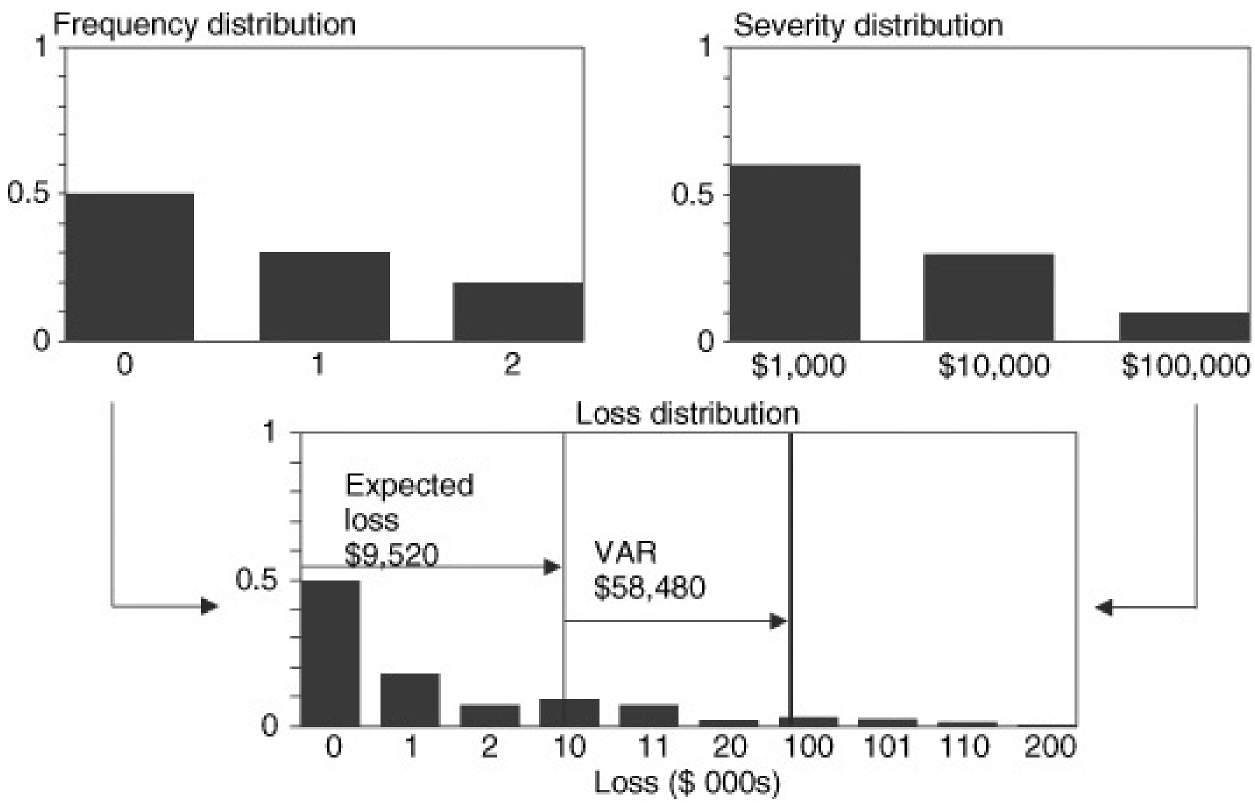
Table 19-5 illustrates this method. We can have at most two occurrences of a loss. Thus we start with a situation with no loss, which has probability 0.5. Next, we go through all occurrences of one loss. A loss of $1000 can occur with probability of $P(n = 1) \times P(x = \$1000) = 0.3 \times 0.6 = 0.18$. After that, we compute the probability of a one-time loss of $10,000 and $100,000. Next, we go through all occurrences of two losses. A loss of $1000 can occur twice with a probability of $0.2 \times 0.6 \times 0.6 = 0.072$. And so on. Finally, we collect all total-loss occurrences and their associated probabilities.

The distribution is constructed in Figure 19-2. Note that even with only three possible values for N and X each, the loss distribution is already quite rich. We can verify that the expected loss is $9520, as reported previously. To compute the 95 percent quantile, we assume that we can interpolate between the data points or that the underlying distribution is continuous. The 95 percent quantile is $68,000. Hence the VAR measure, or *unexpected loss*, is $68,000 – $9,520 = $58,480.

TABLE 19-5
Tabulation of Loss Distribution

Number of Losses	First Loss	Second Loss	Total Loss	Probability
0	0	0	0	0.500
1	1,000	0	1,000	0.180
1	10,000	0	10,000	0.090
1	100,000	0	100,000	0.030
2	1,000	1,000	2,000	0.072
2	1,000	10,000	11,000	0.036
2	1,000	100,000	101,000	0.012
2	10,000	1,000	11,000	0.036
2	10,000	10,000	20,000	0.018
2	10,000	100,000	110,000	0.006
2	100,000	1,000	101,000	0.012
2	100,000	10,000	110,000	0.006
2	100,000	100,000	200,000	0.002

FIGURE 19-2 Constructing the loss distribution.



This operation is equivalent to finding the loss pdf as the integral of

$$\text{pdf of loss} = h(s) = \int g_s\,(s \mid n)\,f\,(n)\ \ dn \tag{19.4}$$

where g_s is the pdf of a sum of severity variables. For some combinations of the functions f and g we can find a closed-form analytical solution for $h(s)$, as illustrated in Appendix 19.A. If not, we have to resort to simulation methods.

19.3.3 Collecting Data

Perhaps the greatest challenge to the measurement of operational risk is the collection of relevant data. Unlike market and credit risk, the sources of operational risk lie within each firm. In practice, the database of operational losses must be built from both *internal data*, specific to the institution, and *external data*, from the experience of other firms.

Internal data provide a history of occurrences of losses from operational failures, which include loss estimates and frequency indicators. Internal data take into account the quality of current and past controls within the institution. The institution must establish a systematic process for collecting internal data.[4]

Operational losses include many small errors with low costs, such as transaction fails, called *high-frequency/low-value losses.* These are amenable to statistical analysis and can be related to risk indicators, which can be used to forecast changes in operational losses or to estimate the cost of improvements. If all losses were of this type, internal data would be sufficient.

At the other end of the spectrum are *low-frequency/high-value losses*. For these, one needs to cull *external data*, ideally from comparable institutions. Internal losses do not include, for instance, large losses that could bankrupt the institution (otherwise the data would not be collected, which is a form of survivorship bias).

This raises several issues. The first one is that not all losses are publicly disclosed. Institutions may be understandably reluctant to reveal failures in their internal systems. Thus the database may be biased toward public information such as legal settlements and ignore technological failures, which are rarely acknowledged. At the lower end, losses below $1 million usually are not included in public databases. The lower threshold for internal databases typically is much lower, around $10,000.

Another problem is that external losses could correspond with different business profiles and internal controls and may not be directly applicable to another institution. In the words of some observers, external data should allow some "marking to operations," that is, adapting the loss distribution for different internal controls. Scaling to the size of assets and to the quality of controls is no simple operation, however. Even so, under the not-so-gentle prodding of regulators, the industry is slowly moving toward centralized databases for operational risk.

19.3.4 Public Databases

Databases such as SAS OpRisk and Fitch Risk/OpVantage provide information on the distribution of operational losses. The databases contain thousands of publicly reported operational losses exceeding $1 million. The distribution of loss severity is summarized in an interesting paper by De Fontnouvelle *et al.* (2006) and is excerpted in Table 19-6. The top panel breaks down losses by event type, using the Basel definitions of Table 19-2. The bottom panel breaks down losses by business type.

For all event types, the 50th percentile or median loss was $6 million. The 95th percentile involved a loss of $88 million. The middle column gives the fraction of losses by type. The highest fraction (55.5 percent) is due to clients, products, and business practices (CPBP); the next highest category is internal fraud, at 23 percent of the total occurrences, followed by external fraud. These three categories lead to large losses, all above $40 million for the 95th percentile.

They help to explain the distribution of losses across business lines. Retail banking has the largest fraction of losses, but the size of losses is limited because this activity deals with retail client accounts, which, on average, involve small amounts. In contrast, *trading and sales* has a 95th percentile loss of $334 million, which is the highest entry. Thus trading and sales is intrinsically a dangerous activity because the amounts involved can be very large, because of the complexity of products, and because of the very high financial stakes for the traders, which create incentives to cheat.

TABLE 19-6
Loss Severity Distribution

Classified by Event Type				
		Percentiles (Millions)		
Business Line	Percent of Losses	50%	75%	95%
All	100.0%	$6	$17	$88
Internal fraud	23.0%	$4	$10	$42
External fraud	16.5%	$5	$17	$93
EPWS	3.0%	$4	$14	—
CPBP	55.5%	$7	$20	$95
Damage physical assets	0.4%	$18	—	—
BDSF	0.2%	$36	—	—
EDPM	1.3%	$9	$27	—

Classified by Business Type				
		Percentiles (Millions)		
Business Line	Percent of Losses	50%	75%	95%
All	100%	$6	$17	$88
Corporate finance	6%	$6	$23	—
Trading and sales	9%	$10	$44	$334
Retail banking	38%	$5	$11	$52
Commercial banking	21%	$7	$24	$104
Settlement	1%	$4	$11	—
Agency services	2%	$22	$110	—
Asset management	5%	$8	$20	—
Retail brokerage	17%	$4	$12	$57

Source: De Fontnouvelle et al. (2006). Loss data from SAS OpRisk expressed in 2002 dollars for losses that occurred in the United States.

In practice, the distributions are smoothed using extreme-value theory (EVT). A simple approach is to compute the logarithm of losses minus the log of the $1 million threshold. Define this transformed variable as *y*. The chosen EVT density function then is where the coefficient *b* is a scale parameter that determines the width of the tail.[5]

$$f(y) = (1/b)e^{-y/b} \tag{19.5}$$

De Fontnouvelle *et al.* (2006) report fitted values at around 0.65 for b, on average. We can use this information to compute the worst-severity event at the 99.9 percent confidence level, for example. The cumulative distribution function for the exponential pdf in Equation (19.5) is

$$F(y) = 1 - e^{-y/b} \tag{19.6}$$

Setting this to 0.999 with $b = 0.65$ and solving gives $y = -0.65 \ln (1 - 0.999) = 4.49$. Taking the exponential gives $x = \$89$ million. For trading, which is more risky, $b = 0.75$ and $x = \$178$ million. For retail operations, $b = 0.55$ and $x = \$45$ million.

The final step consists of combining this fitted distribution with the number of losses in a year. The Risk Management Group (2003) reports the results of a recent loss data-collection exercise that reveals that a typical large bank experiences on average 70 losses above $1 million per year. Using simulations for the convolution of the two distributions gives an unexpected loss, or economic capital, of $1600 and $3200 million at the 99.9 and 99.97 percent confidence levels, respectively. The latter number is broadly consistent with the economic capital estimates reported by the commercial banks in Table 19-1.

19.4 MANAGING OPERATIONAL RISK

19.4.1 Monitoring Operational Risk

The first step in any risk management process is the monitoring of losses. The simple act of recording losses may reveal unanticipated vulnerabilities. Armed with this information, the institution can evaluate the cost and benefit of investments in process improvements. At a broader, firmwide level, these charges should lead to more informed strategic decisions. Top management then may discover that a particular business line that looks attractive without considering operational risk is actually barely profitable once operational losses are factored in. Once operational risk has been measured, it can be better controlled, financed, and managed.

19.4.2 Controlling and Mitigating Operational Risk

Operational risk can be better controlled with measures of the costs and benefits of *alternative actions*. Once vulnerabilities are identified, corrective actions can be framed in the following terms:

- *Loss reduction*, or reduction in the severity of the losses when they occur
- *Loss prevention*, or reduction in the frequency of occurrences, for sources of risk internal to the firm
- *Exposure avoidance*, which is an extreme form of loss prevention, where the activity is completely avoided, for example, by exiting the business line

Loss reduction can be achieved by strategies that mitigate the cost of operational errors. One sample is *contingency planning*. While insurance can be purchased as protection against natural disasters such as fire, floods, and earthquakes, it may only cover physical structures. Loss of business activity may be substantial if disaster strikes without an institution having adequate backup facilities. Contingency planning can offer protection against unexpected sources of risk, as seen in Box 19-2.

Loss prevention can be achieved by purchasing better equipment that will decrease failure rates or restructuring processes to make them less prone to errors. Reducing the frequency of occurrences, however, is only achievable for sources of risk internal to the firm. Like "total quality management" or "six-sigma quality control systems" in manufacturing firms, measuring operational risk in itself should pave the road for process improvements. *Redundant and automated control systems* are other loss-prevention measures. "Straight through" processing, for instance, interfaces the front-and back-office systems so that deals entering the front-office system are automatically sent to the back office, which eliminates manual intervention and the potential for human errors. Some systems now require double validations for trades above certain thresholds.

More generally, the key to controlling operational risk lies in *control systems* and *competent managers*. BCBS (2003) provides common-sense advice. Operational risk is greater in some situations, such as new products, unfamiliar markets, and geographically distant locations that are more difficult to control. Institutions should have independent risk management functions with authority to set and monitor risk limits. *Independence of functions* is a basic principle for protection against risk. It explains the rise of new functions, such as *chief risk officer, chief compliance officer*, for asset-management firms, and even *chief governance officer*. In itself, the implementation of market risk management systems should provide some protection against operational control risks such as rogue trading or fraud.

BOX 19-2

CONTINGENCY PLANNING BENEFITS

Contingency planning sometimes creates protection against sources of risk that are totally unexpected. Consider, for example, the *2000 problem*, also known as the *Y2K problem* or *millennium bug*. The common practice of abbreviating the year to the last two digits would have caused many date-related programs to operate incorrectly as of January 1, 2000. The industry spent enormous resources fixing the problem, checking millions of lines of programming code. Financial institutions, realizing the critical importance of information technology infrastructure to their business, established backup centers to restore capabilities in case of disaster.

In the end, perhaps because of these preparations, Y2K was a nonevent. There was no major computer failure. Planes did not fall from the sky.

The backup facilities established by the financial industry, however, proved critical in the recovery after the September 11 attacks on the World Trade Center that devastated Wall Street. Many New York banks had installed computer backup systems in New Jersey. As a result, markets reopened very quickly, as soon as Monday September 17, which was astonishing considering the scale of the disaster. In large part, this was due to risk management.

19.4.3 Funding Operational Risk

Once the appropriate control structure is in place, the next aspect of operational risk management is the *financing of unexpected losses.* The decision can be viewed in terms of a choice of *preloss financing* or *postloss financing*.

Postloss financing simply uses the available capital to absorb a loss after it occurred. Preloss financing builds up a reserve in anticipation of risk of losses. This risk can be *retained* or *transferred*.

Institutions can decide to guard against unexpected losses by *self-insurance*, that is, putting aside capital in an internal reserve fund, or a captive insurance company, against such losses (retaining risk). Alternatively, they can purchase external insurance (transferring risk), as shown in Box 19-3. When considering external insurance, the obvious issue is whether the insurance premium is reasonably priced. One could argue that self-insurance should be cheaper. Insurance premiums need to cover administrative fees and risk-adjusted return on capital for the insurance company.

BOX 19-3

OPERATIONAL RISK INSURANCE

Swiss Re, one of the largest companies in the global reinsurance industry, introduced in 1999 an innovative product called *financial institution operational risk insurance* (FIORI). This provides coverage against a range of losses owing to operational risk, including rogue traders and so on. The policy pays out after losses exceed a high threshold, typically $50 to $100 million.

On a claim, FIORI advances the payment immediately. Whether the claim is ultimately allowable is examined later. In fact, insurance payouts often involve litigation, thus creating legal risk.

Also, buying insurance decreases some of the incentives to control risk, creating *moral hazard* that increases the cost of insurance. Indeed, prevention and control are a form of self-insurance against some risks. In practice, traditional insurance coverage is most effective for two categories of risk, damages to physical assets (DPA) and business disruption and system failure (BDSF). These categories are least likely to be affected by the moral-hazard problem because these events generally originate from outside the firm.

Purchasing insurance does provide protection against extreme losses. Even large losses can be diversified by the insurance company. In addition, the insurer has access to industrywide data on losses and may be able to assess expected losses more precisely.

19.5 REGULATORY CAPITAL

As in the case of market and credit risk, the industry is also being prodded into action by bank regulators. Indeed, the new Basel II rules will impose a new *operational risk charge* (ORC) in exchange for lowering capital charges for credit risk. The 8 percent charge of the 1988 Capital Accord implicitly accounts for operational risk. The view of regulators is that the current level of global bank capital is adequate, having been “stress tested” through some tumultuous recent times. The new ORC is expected to account for approximately 12 percent of the total minimum regulatory capital.

In addition, these new rules will have an effect beyond the commercial banks that are under the purview of the Basel regulators. In the United States, for instance, the Securities and Exchange Commission imposed similar rules for certain broker-dealers.[6]

The new rules give three alternatives methods. The simplest is called the *basic indicator approach*. This is a top-down approach based on an aggregate measure of business activity. The capital charge equals a fixed percentage, called the *alpha factor*, of the exposure indicator defined as gross income (GI),[7] that is, where α has been set at 15 percent, based on studies that relate the target ORC to gross income for the typical bank. The advantage of this method is that it is simple and transparent and uses readily available data. The problem is that it does not account for the quality of controls or the business lines, like all top-down methods. As a result, this approach is expected to be used mainly by nonsophisticated banks.

$$\mathrm{ORC}^{\mathrm{BIA}} = \alpha \times \mathrm{GI} \tag{19.7}$$

The second method is the *standardized approach*. Here, banks’ activities are divided into eight *business lines*, as we indicated in Table 19-6. Within each business line, gross income is taken as an indicator of the scale of activity. The capital charge is then obtained by multiplying gross income by a fixed percentage, called the *beta factor*, and summing across business lines, that is,

$$\mathrm{ORC}^{\mathrm{SA}} = \sum_{i=1}^{8} \beta_i \times \mathrm{GI}_i \tag{19.8}$$

The β factors are described in Table 19-7. This approach is still simple but better reflects varying risks across business lines.[8] Trading, for instance, is assigned a high factor. This standardized approach can be used only if the bank demonstrates effective management and control of operational risk.

TABLE 19-7
Beta Factors

Business Line	Beta Factor
Corporate finance	18%
Trading and sales	18%
Retail banking	12%
Commercial banking	15%
Payment, settlement	18%
Agency services	15%
Asset management	12%
Retail brokerage	12%

The third class of method is the *advanced measurement approach* (AMA). This allows banks to use their own internal models in the estimation of required capital using quantitative and qualitative criteria set by the Basel Accord. The qualitative criteria are similar to those for the use of internal-market VAR systems.[9] Once these are satisfied, the risk charge is obtained from the unexpected loss (UL) or VAR at the 99.9 percent confidence level over a 1-year horizon, that is, provided that the expected loss is accounted for. Banks are also subject to other quantitative criteria.[10] Finally, insurance can be used to offset up to 20 percent of the operational risk charge. This approach offers the most refined measurement of operational risk and is expected to be used by more sophisticated institutions.

$$\text{ORC}^{\text{AMA}} = \text{UL}(1\text{-year, } 99.9 \text{ percent confidence}) \qquad (19.9)$$

19.6 CONCLUSIONS

Operational risk has only very recently come under close scrutiny by the financial industry. Indeed, institutions and regulators now realize that many financial disasters can be traced to a fatal combination of operational risk with some other form of financial risk.

In response, institutions anxious to avoid the fates of Barings, Daiwa, and AIB have begun recently to develop a framework for measuring and monitoring operational risks. This quantification should allow them to understand their risks better and to control and manage their risks more efficiently. To accelerate this trend, the Basel Committee will impose a capital charge against operational risk.

Operational risk management is still an evolving art form, however. In particular, the collection of relevant data is a major issue. Unlike market and credit risk, operational risk is internal to the firm. Since firms are understandably not eager to reveal their failings, public data on losses caused by operational risk are nowhere as rich as for other forms of risks. As a result, some observers argue that the measurement of operational risk is unreliable and too subjective. Nevertheless, the industry is now busily collecting internal and external data on operational losses. These have revealed, for example, that the average annual loss for a typical U.S. bank is 0.06 percent of assets. For a large bank with $1000 billion in assets, this translates into an annual loss of $600 million. Thus, even in ordinary situations, operational losses create very large ongoing costs.

The most advanced institutions already have put in place structures to measure and manage operational risk. Their annual reports disclose economic capital charges for oprisk close to 15 percent of the total, which is significant. For a bank with $10 billion in economic capital, this represents $1500 million.

Even more important, efforts to get a grip on operational risk seem to be paying off. As Susan Bies (2005) said, “Value is added to the firm when operational risk measurement is integrated with the business-unit management processes. . . . business-line staff can add significant value to this effort through their understanding of inherent risks and controls in their areas.” Recent surveys indicate that the *management* of operational risk could yield reductions of 10 percent of economic capital. In our example, this results in savings of $150 million. This warrants investing substantial resources into the new discipline of operational risk management.

APPENDIX 19.A
Constructing Loss Distributions

The purpose of this appendix is to illustrate analytical methods for modeling loss distributions. The risk manager could tabulate a distribution of relevant losses from historical data, but this is unlikely to be smooth, especially with limited sample sizes. Instead, we can fit the pdf for the loss frequency from a parametric distribution.

Take, for instance, the geometric distribution for the loss frequency n, that is, where the parameter p must be $0 < p \leq 1$. For instance, with $p = 0.5$, we have $f(1) = 0.5$, $f(2) = 0.25$, and so on. Thus the probability decreases geometrically. The expected loss frequency then is $E(N) = 1/p$, and its variance is $V(N) = (1\ p)/p^2$. Other frequency distributions include the Poisson and negative binomial, of which the geometric is a special case.

$$f(n) = p\,(1 - p)^{n-1} \qquad n = 1, 2, \ldots \tag{19.10}$$

Next, suppose that the loss severity x comes from an exponential distribution, that is, which is characterized by the parameter $\lambda > 0$. This implies a probability of a loss that decreases exponentially with the size of the loss. The expected value and standard deviation of the loss are given by $E(X) = SD(X) = 1/\lambda$. Other severity distributions include the lognormal, Weibull, and gamma distribution, of which the exponential is a special case.

$$g(x) = \lambda e^{-\lambda x} \qquad x \geq 0 \tag{19.11}$$

We seek to find the distribution of the total losses over the period, which is $S_n = \Sigma_{i=1}^{n} X_i$. This is the sum of a *random number* of random variables. We assume that n is independent of the realizations X.

The total probability of observing a sum less than s then is

$$P(S \leq s) = \sum_{n=1}^{\infty} P(S_n \leq s \mid n) f(n) \tag{19.12}$$

Next, we use the fact that a sum of i.i.d. exponential random variables has a gamma distribution

$$P(S_n \leq s \mid n) = \int_0^s \frac{1}{(n-1)!} \lambda^n u^{n-1} e^{-\lambda u}\, du \tag{19.13}$$

After integration, we find that

$$P(S \leq s) = 1 - e^{-\lambda p s} \tag{19.14}$$

or that the loss has itself an exponential distribution with parameter λp, that is,

$$h(s) = (\lambda p) e^{-(\lambda p)s} \tag{19.15}$$

From this we can compute the expected loss as well as the worst deviation at some confidence level. The expected loss is $E(S) = (1/\lambda p)$, which is indeed the product of the two expected distribution values $E(N) \times E(X)$. The VAR at the c level of confidence then is $s^* - E(S) = (1/\lambda p)\,[\ln(1/c) - 1]$.

Other distributions are feasible but may not combine analytically as easily as these do. If so, we could approximate the loss distribution by running simulations from processes sampled from $f(n)$ and $g(x)$.

QUESTIONS

1. Define the four risk drivers of operational risk.
2. List the three types of financial risks facing commercial banks and asset management firms in decreasing order of importance.
3. What tools can be used to assess operational risk?
4. What are the pros and cons of bottom-up models versus top-down models?
5. Describe the loss-distribution approach. What are the two types of risk factors used to determine the distribution of annual losses?
6. Assume that the expected annual loss owing to oprisk is 0.06 percent of $500 billion in assets for our bank. If the institution expects to incur 60 instances of losses per year, what is the expected loss severity?
7. Describe the shape of the loss distribution. Does it have light or heavy tails, and why?
8. Why is it hard to measure operational risk?
9. In practice, what business line creates the largest potential losses?
10. Assume that the tail loss distribution for losses in millions of dollars, in excess of $1 million, is modeled by an exponential distribution with scale parameter $b = 0.75$. Compute the worst loss at the 99th percent confidence level.
11. What are the pros and cons of mitigating operational risk by purchasing insurance?
12. Describe the AMA approach to regulatory capital, and explain why it should be superior to the basic indicator approach and the standardized approach.

CHAPTER 20
Integrated Risk Management

> The most important function of a Treasurer has to be managing risk. . . . Up to now, risk management has been event-and transaction-driven. Now we are trying to comprehend the total picture.
>
> —*Susan Stalnecker, Treasurer of DuPont (a global chemical company)*

The revolution in risk management that started with value-at-risk (VAR)–based measures of financial market risk is now spreading to firmwide risk management. After all, the common denominator for any risk management activity is efficient use of capital. This is nothing more than an extension of the VAR approach, whose essence is centralization, to firmwide risks. Thus the ideas behind the VAR revolution are quickly spreading to enterprisewide risk management (ERM).

Like VAR, ERM considers aggregate risks, including market risk, credit risk, operational risk, and business risk. This integrated view brings powerful economies of scale. In the past, risks were considered separately from each other and hedged one at a time. In the old *silo* approach, risks were managed independently in separate compartments. Most of these risks, however, are uncorrelated, which means that such a piecewise approach provides unnecessary coverage. Considerable cost savings can be achieved by hedging only *net* risks. In addition, a firmwide approach can reveal natural hedges and guide the firm's strategy toward activities that are less risky when taken as whole. Thus ERM should be an essential strategic tool for corporations. By providing an aggregate measure of risk, ERM allows chief financial officers (CFOs) to decide on how much equity the firm should hold. For the first time, we now have the tools to evaluate fundamental questions such as the amount of leverage a firm should carry.

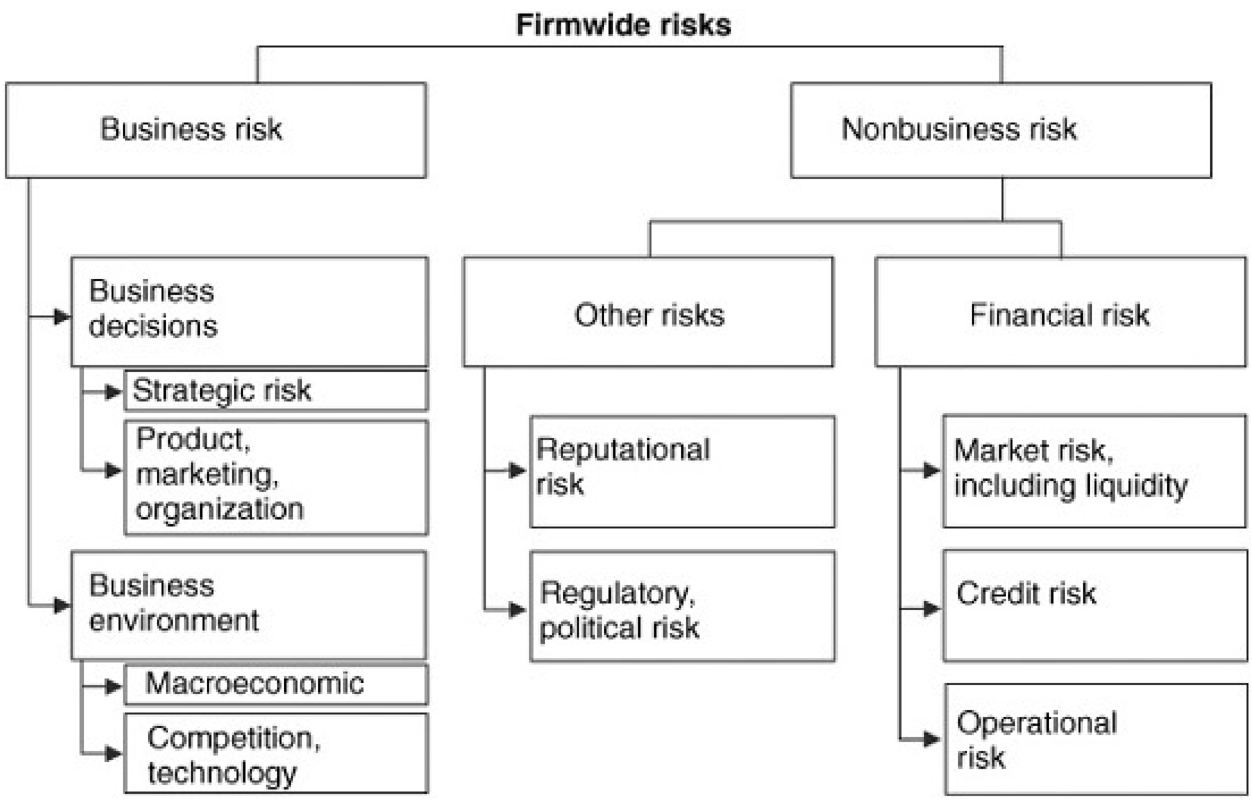
Integrated risk management has other advantages. It could help to stabilize earnings, whose volatility appears to worry CFOs, by careful neutering of undesirable risks. In addition, the quantification of some risks such as market and credit could push institutions to take other types of risk that are less visible and may be more dangerous. The only solution is to take a comprehensive view of firmwide risks.

Section 20.1 presents a classification of firmwide risks. This book has already covered market risk, credit risk, and operational risk. The first section presents remaining risks, including business, strategic, and reputational risks. Section 20.2 introduces firmwide risk management. Section 20.3 then discusses how integrated risk management can add value to corporations.

20.1 THE GALAXY OF RISKS

We first provide a broad description of the risks facing financial and other institutions. Figure 20-1 classifies the risks facing institutions into business and nonbusiness risks, the latter being further classified into financial and other risks. Admittedly, these classifications are somewhat arbitrary because some of these risks overlap categories.

FIGURE 20-1 Firmwide risks.



20.1.1 Business Risks

Business risks refer to the risks the corporation willingly assumes to create a competitive advantage and add value for shareholders. These risks include the *business decisions* companies make and the *business environment* in which they operate. Business decisions risk include product-development choices, marketing strategies, and the choice of the company's organizational structure. Broader in nature is *strategic risk*, which reflects decisions made at the level of the company's board or top executives. Business environment risk includes *macroeconomic risks*, which result from economic cycles or fluctuations in incomes and monetary policies, as well as *competition risk* and *technological innovations risk*. Business risk is symmetric in that it can create both gains and losses. In some sense, corporations are "paid" to take business risk.

Business risk can affect earnings if the institution does not react quickly to changing conditions. A key component of this flexibility is the cost structure. High fixed costs make is more difficult to adapt. Some institutions now give estimates of their business risk based on *earnings at risk* (EAR), defined as the worst fall in earnings over a 1 year horizon at a high confidence level.

20.1.2 Nonbusiness Risks

Financial risks relate to possible losses owing to financial market activities. These include market, credit, and operational risks. Market and credit risks are symmetric because they can create both gains and losses. On the other hand, operational risk mainly creates losses. *Operational risk* is the risk of loss resulting from inadequate or failed processes, people, and systems or from external events. As explained in Chapter 19, this definition includes *legal risk* as well as some *event risk*, such as natural disasters.

Industrial firms can manage financial risk so that they can concentrate on what they do best, that is, manage business risks. In contrast, the primary function of *financial institutions* is to assume, intermediate, or advise on financial risk. *Insurance companies* also face market and credit risk through their investments but take exposure to disasters and other risks in exchange for receiving an insurance premium. Such *insurance risks* usually are classified separately from the risks shown here.

Other risks include reputational risk and regulatory or political risk. *Reputational risk* is the risk of losses beyond the direct operational losses from the firm's damaged reputation. *Regulatory* or *political risk* is the risk of losses resulting from changes in the regulatory or political environment. These risks are very difficult to assess and are not amenable to formal measurement.

20.1.3 Legal Risk

Legal risk is the risk of losses owing to fines and penalties resulting from supervisory actions as well as private settlements. This risk can be limited through policies developed by the institution's legal counsel, as approved by senior management. Prior to engaging in trades, institutions should ensure that their counterparties have the legal authority to do so and that the terms of the contracts have a sound basis. Even so, contracts that lead to large losses for counterparties often end up in a lawsuit. Such contracts are invariably claimed *unsuitable* to the client's needs or level of expertise. Losing parties often claim a form of *financial insanity*; that is, they were temporarily unable to judge financial contracts. Here, VAR can provide additional protection, as Box 20-1 shows.

The financial industry is also working to reduce legal risks through the use of standardized contracts such as master netting agreements. The language in such contracts has been formulated carefully so as to reduce mistakes and misunderstanding. Even when there are differences of interpretation, the use of such standardized contracts makes it more costly for a financial institution to renege on them unilaterally. Such behavior would be badly received by the rest of the community and actually increase the cost of entering future contracts.

BOX 20-1

USING VAR TO CONTROL LEGAL RISKS

VAR is now used to control legal risks. *Suitability* now can be defined in terms of VAR limits.

Some banks now require their traders to obtain signatures from counterparties based on VAR limits. Above a VAR level of $1 million, for instance, the deal must be approved by the finance director of the client institution. Above some other level, say, $5 million, the deal must be signed off on by a senior manager as well. This makes it more difficult for clients to claim financial insanity later.

20.1.4 Reputational Risk

Reputational risk can be viewed as the damage, in addition to immediate monetary losses, caused to the ongoing business of an institution from a damaged reputation. It is particularly important for banks because the nature of their business requires maintaining the confidence of the marketplace.

One such example is the story of Bankers Trust (Box 20-2), which before 1994 was widely admired as a leader in risk management but at some point became a victim of the backlash against derivatives.

BOX 20-2

BANKERS TRUST'S STRATEGIC RISKS

Charles Sanford transformed Bankers Trust from a sleepy commercial bank into a financial powerhouse using risk management as a competitive tool. In 1994, however, the bank became embroiled with a high-profile lawsuit with Procter & Gamble that badly damaged the bank's name. Many customers shied way from the bank after the bad publicity.

In an attempt to restore its reputation, the bank brought in a new chief executive, Frank Newman, in 1996. Mr. Newman, a well-respected former deputy secretary of the U.S. Treasury, quickly reached an out-of-court settlement with P&G and attempted to deemphasize the bank's trading activities.

The bank also recognized that its profit-driven culture often placed the bank's profit before the client's interest. Focusing on financial risks alone can become harmful if it detracts from the client relationship, which is still an important part of the banking business. Bankers Trust later implemented changes in its compensation schemes to reward salespeople for improving relationships with customers.

The new plan was to create a full-service investment bank serving growth companies in the U.S. market. The strategic transformation of Bankers Trust failed to take hold, however. By October 1998, the bank's stock price was back at its level of early 1996, underperforming its peer group. In November 1998, Bankers Trust announced it had agreed to an acquisition by the German behemoth Deutsche Bank. The price was right, at $9.2 billion, or 2.1 times Bankers Trust's book value.

20.1.5 Regulatory and Political Risk

Regulatory risks are the result of changes in regulations or interpretation of existing regulations that can negatively affect a firm. For instance, as a result of the Bankers Trust case, the Commodities and Futures Trading Commission (CFTC) and the Securities and Exchange Commission (SEC) have extended their jurisdiction over market participants by declaring swaps to be “futures contracts” and “securities.” As a result, Bankers Trust agreed to pay $10 million to settle charges brought by regulators.

Political risks arise from actions taken by policymakers that significantly affect the way an organization runs its business. It seems that big financial losses, either attributed to derivatives or to hedge funds, regularly lead to threats of legislative intervention. The private sector then has to demonstrate that self-policing is preferable to new laws. Political risk is also common in emerging markets, taking the form of nationalization or the imposition of capital controls. Often these actions are totally unexpected, as illustrated in Box 20-3.

20.2 INTEGRATED RISK MANAGEMENT

Integrated risk management, *enterprisewide risk management* (ERM), and *firmwide risk management* all have the same meaning:

BOX 20-3

POLITICAL RISK IN ARGENTINA

In December 2001, Argentina announced it would stop paying interest on its $135 billion foreign debt. This was the largest sovereign default ever. In January, the fixed-exchange-rate system was abandoned for the Argentinian peso, which promptly devalued from 1 peso per U.S. dollar to more than 3 pesos. The default and devaluation represent credit and market risk, which are integral to foreign investments.

What was totally unexpected, however, was the government’s announcement that it would treat differentially bank loans and deposits. Dollar-denominated bank deposits were converted into devalued pesos, but dollar-denominated bank loans were converted into pesos at a one-to-one rate. This mismatch rendered much of the banking system technically insolvent because loans (bank assets) overnight became less valuable than deposits (bank liabilities). This type of political risk is not amenable to formal measurement.

> ERM is a coordinated process for managing risk on a firmwide basis, across types of risk, locations, and business lines.

Like market risk management, ERM aims at measuring, controlling, and managing the overall risk of the institution. This is slowly made possible by a convergence in methods used to quantify financial risk, based on VAR methods.

20.2.1 Measuring Firmwide Risk

The measurement of overall risk starts with the analysis of different types of risks, including market, credit, operational risk, and perhaps others. *Economic capital* (EC) can be defined as the worst loss for each category of risk over a 1-year horizon at a high confidence level. Table 19-1 shows examples of disclosures of EC by risk category.

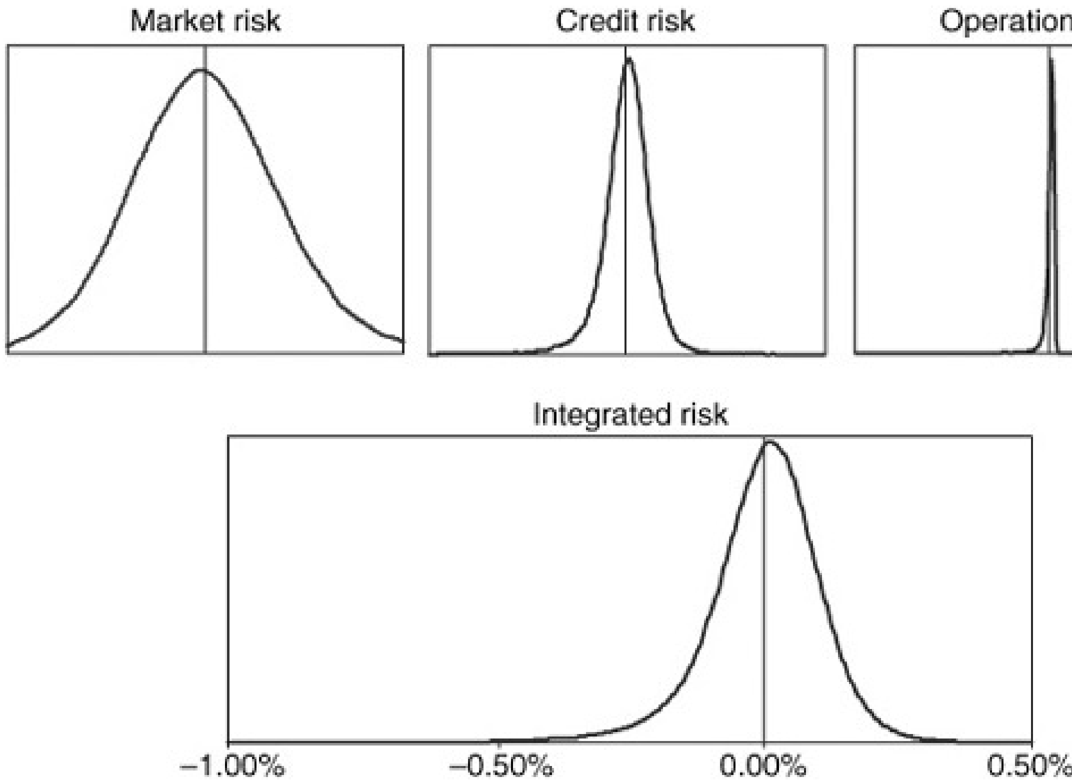
The next step then usually adds up the EC measures across risk classes, as is done under the Basel II approach. This is very conservative, however, because it would be very unlikely to suffer extreme losses in all the categories at the same point in time. Diversification effects should create a lower amount of economic capital.

A better approach would combine the marginal distributions for the risk classes into a total distribution. Rosenberg and Schuermann (2006) show how this can be done using data for the typical U.S. bank. They combine the marginal distributions into a total distribution using *copulas*, a statistical tool developed in Chapter 8, which take into account typical correlations. Table 20-1 compares the characteristics of the distributions, which are also described in Figure 20-2.

TABLE 20-1
Risk Distributions

	Market	Credit	Operational	Total
Volatility	0.58%	0.19%	0.04%	0.11%
Skewness	0.2	−1.3	−4.5	−1.1
Kurtosis	3.7	16.1	35.3	9.6
99.9% quantile	−1.81%	−1.20%	−0.37%	−0.43%
99.9% VAR	−0.06%	−0.35%	−0.25%	−0.43%

FIGURE 20-2 Integrating market, credit, and operational risk.



Market risk has high volatility, zero skewness, and low excess kurtosis. This means that the distribution is broadly symmetric and does not have fat tails. Credit risk has negative skewness, reflecting losses from defaults, and higher kurtosis. Operational risk, in contrast, has low volatility but very high kurtosis, reflecting a very long left tail. The last column describes the total distribution for a bank with the typical mix of market, credit, and operational risk, taking into account correlation effects. The last line reports VAR using the same denominator for the three risk classes. The total VAR is 0.43 percent of this common measure, taken as the sum of trading assets, lending assets, and total assets.

Rosenberg and Schuermann (2006) also analyze the effect of approximations to the total VAR. Simply adding up individual VARs for each risk category overestimates the total VAR by 52 percent. Summing the last line in Table 20-1 gives −0.06 − 0.35 − 0.25 = −0.66 percent, which is indeed 52 percent higher than the total VAR of 0.43 percent. This demonstrates that the practice of adding up VAR is too conservative. They also compute a total VAR measure that assumes, incorrectly, given the data in the table, that distributions are jointly normal. This underestimates the true VAR by 46 percent. Finally, they compute a *hybrid VAR*. For two risk factors, this is where VAR_1 and VAR_2 are the marginal measures for the two sources of risk, and ρ is their correlation. This simple measure gets close to the true VAR, with a slight overestimation of 13 percent. Thus this is a useful shortcut that avoids the need to model the entire joint distribution.

$$\mathrm{VAR}_H = \sqrt{\mathrm{VAR}_1^2 + \mathrm{VAR}_2^2 + 2\rho \mathrm{VAR}_1 \mathrm{VAR}_2} \qquad (20.1)$$

This analysis allows the CFO to investigate the impact of different business mixes. A trading bank, for example, has greater weight on market risk. The model shows that this increases the total VAR of the bank. At the other extreme, a bank could have a large asset management business, which primarily has operational risk. With increasing weight on oprisk, the total VAR first decreases, reflecting diversification effects across the three types of risk, and then increases slightly.

Of course, the choice of the risk mix also should reflect the bank's competitive advantage, in other words, the *expected returns* for exposure to each type of risk.[1] A bank may decide to engage in trading even if it increases its overall risk profile because it is a profitable activity. Without the measure of aggregate risk generated by ERM, however, it is very difficult to decide on the proper risk-return positioning.

20.2.2 Integrated Risk Management and Consolidation

A powerful implication of integrated risk management is that various types of financial risks diversify each other, which saves on economic capital. Diversification arises within market risks, within credit risk, and within operational risk. It also arises across these three categories of risk. Insurance risks also generally are uncorrelated with financial risks.[2] This explains much of the consolidation we observe nowadays in the financial industry, within banks, securities firms, and insurance companies, and across these sectors.[3]

Two types of risks, however, do not diminish with scale. The first is *reputational risk*. Problems in one part of a diversified firm may affect confidence in other parts of the firm, for instance, when the market fears that losses in a troubled unit may extend to other units.[4] This can lead to funding problems that may spell the end of the institution. The second is *liquidity risk*. Because of market impact, doubling the size of a portfolio will incur more than twice the transactions costs during a liquidation. Thus large portfolios are costly to maneuver. Size was indeed one of the causes of the downfall of Long-Term Capital Management (LTCM), whose saga is explained in Chapter 21. In addition, larger firms make it more difficult to provide proper incentives for managers. Creativity is usually better rewarded in a smaller, boutique-type environment. Finally larger firms may be plagued by conflicts of interests owing to the fact that the same firm performs multiple functions in capital markets. During the dot-com bubble, for example, a number of Wall Street analysts recommended buying stocks in poorly performing companies because these firms were bringing in investment-banking revenues. As a result, a number of Wall Street firms paid large amounts to settle conflict-of-interest allegations with respect to tainted stock research.

20.2.3 Controlling Firmwide Risk

Once risk is measured, it can be controlled better. This is becoming essential because businesses are becoming more complex, with more products that reach across various risk categories. Financial institutions in particular are discovering complex and unanticipated interactions between their risks. Most disturbingly, it seems that risk has a way of moving toward areas where it is not well measured.

Attempts at controlling one type of risk often end up creating another one. The syndicated *eurodollar loan*, for example, provides an interesting illustration of how market risk can be transformed into credit risk. In the late 1970s, a number of U.S. banks made loans to Latin American countries structured so as to minimize their market risk. The loans were denominated in dollars (no currency risk), were payable on a floating-rate basis (no interest risk), and were made to governments (which were supposed to be safe). After U.S. interest rates skyrocketed in the early 1980s, countries such as Mexico and Brazil went into default: They were unable to make the (floating) interest payments on their loans. In short, market risk had created credit risk.

Wrong-way trades are those where credit risk and market risk amplify each other. Box 20-4 shows another interesting example. A key insight of this analysis is that the counterparty's motivation for the trade plays an important role in the correlation between market risk and credit risk. When the counterparty uses the trade as a hedge, a loss on the trade should be offset by an operating gain. Thus the market gain for the bank or loss for the counterparty does not increase the probability of default. In contrast, when the counterparty uses the trade to speculate, a default is more likely when the counterparty has suffered a large loss. Generally, it is safer for a financial institution to enter trades with counterparties that hedge instead of speculate. This is also why grain futures exchanges require lower margins for farmers hedging their production rather than for the general public.

BOX 20-4

WRONG-WAY TRADES

An example of *wrong-way trades* is cross-currency swaps with Asian counterparties during the Asian currency crisis of 1997. A number of Asian institutions had borrowed in U.S. dollars to take advantage of low U.S. interest rates and to invest in local currencies such as the Thai baht or Korean won at higher interest rates.

At initiation, the contracts had little market and credit risk. As the local economies deteriorated, however, these Asian currencies devalued sharply, creating large losses on the contracts. At the same time, most Asian institutions suffered large operating losses owing to the contraction in local business activity. This combination led to numerous defaults. Basically, this was due to the fact that the counterparties were using the swaps to speculate instead of for hedging.

Similarly, the increasing practice of marking to market over-the-counter (OTC) swaps decreases credit risk, but at the expense of more frequent collateral payments that increase operational risk owing to the need for additional operations. Contracts need to be valued on a regular basis and cash exchanged to the counterparty. This cash requirement also creates liquidity risk.

Finally, an increasing number of instruments now mix different types of risks. Credit derivatives, for instance, involve both market and credit risk. So do tradable loans. These *risk interactions* create a need for integrated risk management systems.

20.2.4 Managing Firmwide Risk: The Final Frontier

Integrated risk management systems should allow institutions to manage their risk much better. Even if some risks are difficult to quantify, the process itself creates insights into a company's overall risk.

An immediate benefit of ERM is the discovery of *natural hedging*. Some firms have discovered that some risks offset each other. For instance, a California telecommunications company suffered damage to many of its phone poles during a 1994 earthquake. This loss, however, was offset by the increased phone traffic as worried families called relatives. This is an example of negative correlation between the value of assets and business revenues.

Another example is Countrywide Financial, a financial services company engaged primarily in residential mortgage banking. It has two lines of business. *Loan origination* generates revenues from new mortgage loans. This business performs well when interest rates are falling, causing many homeowners to refinance their loans. Conversely, this business suffers when rates increase. The other business is *loan servicing*, which generates revenues from a small fraction of the monthly loan payments. This is akin to an annuity with a maturity equal to the life of the loan. When interest rates increase, the life of the loans is extended because homeowners refinance less, thus increasing the present value of existing servicing agreements. Thus the two business lines have negative correlation. The company states that "when properly balanced and managed as countercyclical businesses, they form a natural economic hedge, which is designed to produce a stable, growing income stream."

BOX 20-5

MULTIPLE-RISK INSURANCE

In line with the ERM trend, the insurance industry is now providing contracts covering a broad range of risks. Honeywell, the U.S. controls-technology company, entered in June 1997 a 3-year insurance program that covers currency risk along with workers' compensation coverage. Losses beyond $30 million a year are underwritten by AIG, the U.S. insurance giant. By blending together these risks, Honeywell estimates that it saved at least 25 percent on annual premiums.

These programs, which are coming in vogue with CFOs anxious to stabilize corporate earnings, are also known as *holistic risk* and *enterprise risk*. Also, since these insurance contracts embed some derivatives, they have been dubbed, less gracefully, *derivatives in drag*.

Recently, some of these contracts have come under close regulatory attention because they did not transfer sufficient risk to the insurer, in which case they could not be treated as insurance.

Another tangible benefit of ERM is *cost reduction* for insurance against firmwide risks. By treating their risks as part of a single portfolio, institutions do not need to buy separate insurance against each type of risk, thereby taking advantage of diversification benefits. Insurance companies are now developing ERM-based insurance products, as seen in Box 20-5.

Some companies have taken the cost saving one step further by cutting down on the purchase of external insurance. British Petroleum, for instance, after carefully reviewing its portfolio of risks, has decided to discontinue all purchase of external insurance, except when required by law. This saves a bundle in insurance premiums.

Even more tangibly, centralized risk management can help to save transactions costs. Up to the mid-1990s, hedging systems consisted of focusing on sources of risk one at a time and perhaps covering risks individually. For instance, multinationals would evaluate their transactions risks in various currencies and hedge them individually. The problem with this approach is that it is inefficient because it ignores correlations between financial variables. Transactions costs can be saved if the hedging problem is viewed on a companywide basis.

Finally, credit-rating agencies have a favorable opinion of enterprise-level approaches to risk.[5] In some cases, the existence of an ERM program has led to a slightly higher credit rating, which can translate into sizable savings in annual capital costs.

20.3 WHY RISK MANAGEMENT?

Through better control of their risks, firmwide risk management can help corporations to stabilize their cash flows or earnings. The question is, why bother?

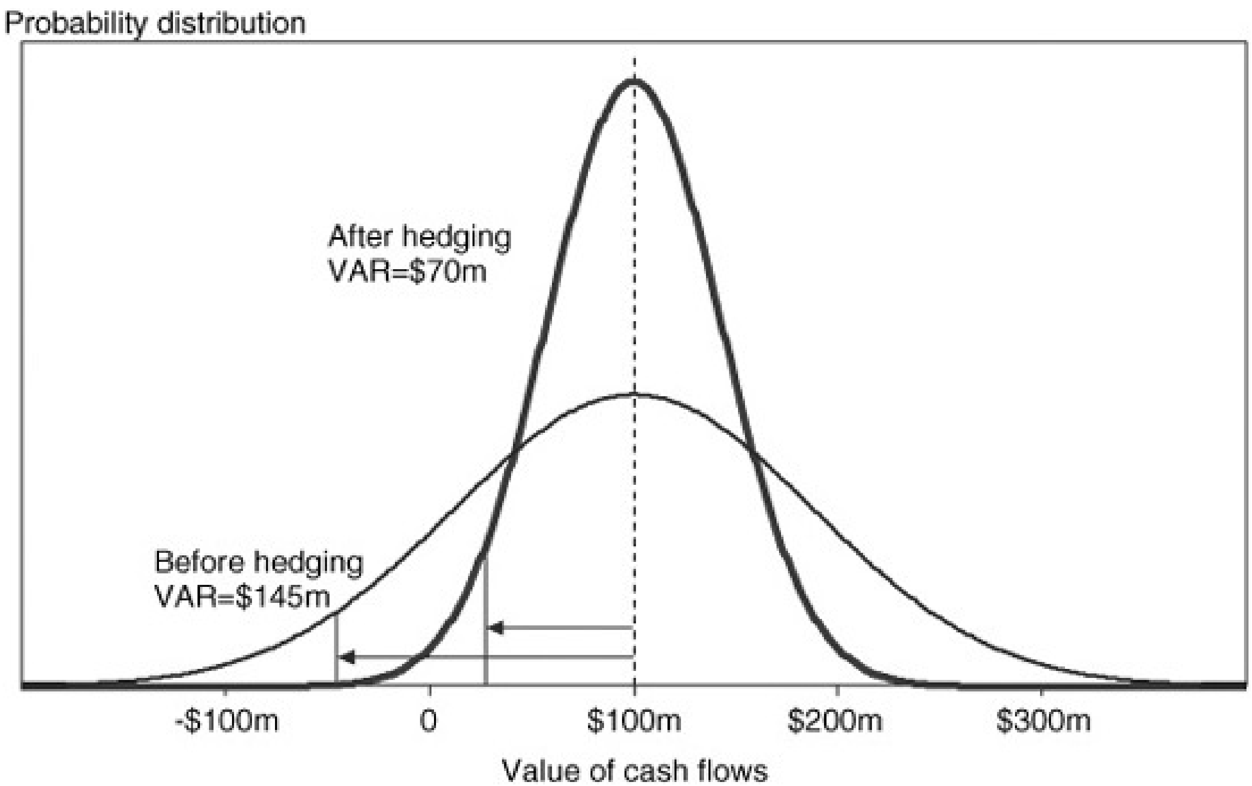
20.3.1 Why Bother?

This is not as obvious as it seems, for in the absence of market "frictions," investors in corporations should be able to replicate whatever risk management action the firm is taking. Hence it is not clear that risk management should add value. Indeed, the Modigliani-Miller (M-M) theorem (1958) states that under these conditions the value of a firm should be unaffected by its financial policies. Thus risk management is supremely irrelevant.

Take, for example, ExxonMobil, a large oil producer that is heavily exposed to oil-price risk. The company could hedge by a number of means, such as financial derivatives indexed to oil prices. Investors, however, are perfectly aware of this. Some may even buy shares in this company to acquire exposure to oil prices. The exposure is transparent because the company's production and reserves are disclosed in financial reports. It would be very easy for investors to cover the risk of their investment in ExxonMobil by selling oil futures. As a result, it is not clear why oil hedging by ExxonMobil should add value to the firm.

To understand the effect of hedging with VAR, Figure 20-3 gives an example of cash-flow distributions. Without hedging, the 95 percent VAR is $145 million. If the firm decides to hedge with derivatives, VAR provides a consistent measure of the effect of hedging on total risk, including correlations. This is a significant improvement over traditional hedging programs, which typically focus on individual transactions only.

FIGURE 20-3 VAR and corporate hedging.



Assume, for instance, that the firm has decided to hedge with linear contracts, such as forwards or swaps contracts. As shown in the figure, hedging narrows the distribution of cash flows. Say that, after hedging, the VAR number is reduced to $70 million. Hedging, however, does not change the mean of the distribution when the derivatives contracts are fairly priced. Thus, without market imperfections, hedging does not add value.

20.3.2 Why Hedge?

The real usefulness of the M-M theorem is not its conclusion of irrelevance but rather the focus it brings on market imperfections. Since the theorem was established, finance researchers have identified conditions under which hedging, meaning activities that lower the volatility of cash flows or firm value, should add value.[6]

- *Hedging can lower the cost of financial distress*. As Figure 20-3 shows, hedging reduces the probability of unfavorable left-tail outcomes. This is valuable if financial distress has *deadweight costs,* such as legal fees and costs incurred because the firm cannot be managed efficiently when undergoing bankruptcy proceedings.[7] Some of these costs may take place earlier, as soon as a firm's situation becomes unhealthy. For example, potential customers may become reluctant to deal with an ailing firm, leading to lost business.
- *Hedging can lower taxes*. Greater earnings stability can reduce average taxes paid when the firm's tax function is convex. Tax rates start at zero for negative incomes and then grow positive and higher for increasing levels of income. With such a convex tax function, taxes paid when income is really high are not offset fully by tax refunds when losses are incurred. The schedule of the tax authority is akin to a perpetual call option on profits. By lowering volatility, the firm can lower the value of this option, thereby enhancing firm value.
- *Hedging can lower agency costs*. Corporations can be viewed as delicate collections of contracts between stakeholders, including shareholders, bondholders, and managers. Shareholders necessarily delegate decisions to managers. This, however, creates *agency costs* owing to the fact that the agents' interests (management) are not aligned with those of the shareholders. Some managers may be incompetent, wasting firm value. Shareholders, of course, are perfectly aware of this situation and are continuously trying to assess the performance of managers by watching earnings, for example. The problem is that earnings can fluctuate owing to factors outside the control of the firm. By making earnings less volatile through hedging, risk management makes earnings more informative, which should lead to better evaluation of managers.
- *Hedging can facilitate optimal investment*. Some companies need steady

cash flows to invest in research and development (R&D) programs. It would be impractical to cut down R&D programs whenever the firm incurred a temporary financial loss only to restart them later. In addition, firms may need cash to take advantage of new projects. In all these cases, companies could go to external markets, for example, borrow funds from banks or bondholders to raise cash when needed. If, however, external financing proves more costly than internal sources of funds, hedging may add value to the firm.

These theories predict that hedging financial risks should add value to the firm. A body of recent academic research is now exploring this issue, with intriguing findings. Allayannis and Weston (2001) claim that market valuations are higher for firms that make use of foreign currency derivatives to hedge. The value added is significant: Hedging firms have market values that are 4.9 percent higher than others, on average. With a median market capitalization of $4 billion in their sample, this translates into an average value added of $200 million for each firm with a risk management program. Similarly, Carter, Rogers, and Simkins (2006) examine the case of fuel hedging for U.S. airlines and report an even higher hedging premium of about 14 percent, albeit with a very large confidence interval. They argue that hedging allows airlines to expand operations when times are bad for the industry, that is, buying gates or planes at low prices, thereby alleviating the underinvestment problem.

The interpretation of these results is debated, however.[8] These increases in market values are very large and are perhaps correlated with other unobserved variables that affect market values. Suppose, for instance, that competent managers, for example, with an MBA degree, do increase the value of the firm. If hedging operations must be run by managers with the training provided by an MBA, then hedging programs will be associated with higher market values owing to this confounding effect. In this case, the observed relation does not reflect a causal link. In addition, if it was so easy to increase market value, we should observe all companies actively engaging in hedging. Thus the question of whether risk management adds value or not does not have a simple answer. Undoubtedly, this important question will be subject to further empirical research.

20.4 CONCLUSIONS

In the early 1990s, VAR methods revolutionized the industry by providing a centralized measure for the total market risks facing an institution.

Enterprisewide risk management extends VAR to the agglomeration of all financial risks. This extension is rationalized by the existence of a common buffer against all risks, which is capital. Another reason is that a silo view can miss significant risks or, worse, push risks into places less visible, thereby creating unexpected losses.

In response, institutions have embarked on ambitious programs to quantify all their financial risks. The bright side of this vast effort is a better understanding of risks facing institutions, leading to improved control and risk management. Companies are discovering that hedging costs can be lowered through pruning unnecessary transactions and taking advantage of natural hedging. Diversification allows lower levels of economic capital for combined market, credit, and operational risks.

Finally, ERM potentially can transform the way in which corporations make strategic decisions. For the first time, it provides top management with the tools to trade off expected profits against risks for different types of business activity. This completes the transformation of VAR methods from measuring market risk at the level of a trader's desk toward a tool for strategic decisions at the highest level of the institution.

QUESTIONS

1. Define the drivers of financial risk and business risk.
2. Define legal risk. How can the use of standardized contracts help to reduce legal risk?
3. Suppose that an institution loses $100 million in unauthorized trading. After the announcement, it loses customer business worth $200 million. How should the latter loss be classified?
4. Suppose that a bank has an investment initially worth $200 million in Argentinean government debt, of which half is in dollars and the other half in the local currency. The currency then depreciates by 30 percent. The government then defaults on the dollar debt and repays 20 percent. Compute the losses attributable to market, credit, operational, and other risks.
5. Among market, credit, and operational risk, which one has highest kurtosis?
6. Among market, credit, and operational risk, which one(s) have large

skewness?

7. Which is the best approximation of integrated VAR, (a) adding up VAR for each risk, (b) assuming normal distributions, or (c) using a hybrid combination of VAR for each risk source with correlations?
8. Explain whether larger banks benefit from diversification effects, reputational effects, and liquidity effects.
9. A bank enters a swap with a gas-producing company where the company pays a fixed price for gas. Discuss whether this type of trade is a wrong-way trade or not.
10. Brokerage firms benefit from increased volatility because this leads to increased business volume. Discuss whether these firms should be long or short options to create a natural hedge.
11. A company considers different combinations of insurance contracts, each with maximum coverage of $10 million. The first choice is a group of contracts covering separately (1) property and casualty and (2) workers' compensation. The second choice is one contract covering both risks. Which approach is cheaper?
12. What are the advantages of integrated risk management systems?
13. What is the effect of a firm's hedging with forward contracts on the shape of distribution of its cash flows? Discuss both the mean and volatility.
14. What are benefits of hedging?

PART VI
THE RISK MANAGEMENT PROFESSION

CHAPTER 21
Risk Management Guidelines and Pitfalls

> Risk management is asking what might happen the other 1 percent of the time.
>
> —*Richard Felix, chief credit officer at Morgan Stanley*

The impetus for today's risk management industry can be traced to the financial disasters of the 1990s. While unfortunate, these derivatives disasters have led to useful lessons. If one document must be singled out as having shaped the risk management profession, it must be the landmark review published by the Group of Thirty (G-30) in July 1993. The G-30 laid out a series of "best practices" that included measuring value at risk (VAR). These recommendations, however, have wider applicability than just derivatives and have become a benchmark for prudent management of all financial risks.

This chapter shows how the industry has responded to episodes of financial distress by periodically improving risk management techniques and, sometimes belatedly, realizing their limitations. Some limitations are or should be obvious. Other side effects may be more subtle. Owing to the complexity of the process leading to VAR, some users have a mistaken impression of absolute precision in the VAR number. This is not the case. VAR gives a first-order magnitude of financial risks and has, like all approximations, limitations. Users must be aware of these limitations when interpreting the data. The saga of the hedge fund Long-Term Capital Management (LTCM) is a good illustration of flaws in risk management systems.

Section 21.1 summarizes milestone documents in risk management. Section 21.2 then discusses VAR limitations. It reviews standard drawbacks of VAR, which should be well recognized. These include the risk of exceptions, the risk of changing positions, event and stability risks, and model risks. Section 21.3 turns to more fundamental dangers of VAR. These are illustrated in the context of the Long-Term Capital Management story in Section 21.4.

21.1 MILESTONE DOCUMENTS IN RISK MANAGEMENT

We now review the defining documents that shaped the risk management profession. These include the 1993 Group of Thirty (G-30) recommendations for managing derivatives, the Bank of England report on the Barings failure, and the Counterparty Risk Management Policy Group reports.

21.1.1 “Best Practices” Recommendations from G-30

The G-30 best practices report has been hailed as a milestone document for risk management. Initially developed to deal with derivatives, the G-30 recommendations, however, are much more general and truly apply to any investment portfolio.

The report provides a set of 24 sound management practices, the most important of which are summarized as follows (using the original G-30 numbering method):

1. *Role of senior management*. Policies governing derivatives should be clearly defined at the highest level.[1] Senior management should approve procedures and controls to implement these policies, which should be enforced at all levels. In other words, derivatives activities merit the attention of senior management because they can generate large profits or losses. Senior management, the board of directors, or the board of trustees is the first point of responsibility.

2. *Marking to market*. Derivative positions should be valued at market prices, at least on a daily basis. This is the only valuation technique that correctly measures the current value of assets and liabilities. Marking to market should be implemented regardless of the accounting method used.

5. *Measuring market risk*. Dealers should use a consistent measure to calculate daily the market risk of their position, which is best measured with a VAR approach. Once a method of risk measurement is in place, market-risk limits must be set based on factors such as tolerance for losses and capital resources.

6. *Stress simulations*. Users should quantify market risk under adverse market conditions. VAR systems usually are based on normal market conditions, which may not reflect potential losses under extreme market environments. Stress simulations should reflect both historical events and estimates of future adverse moves.

8. *Independent market-risk management*. Dealers should establish market-risk management functions to assist senior management in the

formulation and implementation of risk-control systems. These risk management units should be set up with clear independence from trading and should have enforcement authority. They should establish risk-limit policies, measure VAR, perform stress scenarios, and monitor whether actual portfolio volatility is in line with predictions.

10. *Measuring credit exposure*. Users should assess the credit risk arising from derivatives activities based on frequent measures of current and potential exposure. Current exposure is the market value, or replacement cost, of existing positions. Potential exposure measures probable future losses owing to default over the remaining term of the transaction.
11. *Aggregating credit exposure*. Credit exposure to each counterparty should be aggregated taking into account netting arrangements.
12. *Independent credit-risk management*. Users should establish oversight functions for credit risks with clear authority, independent of the dealing function. These units should set credit limits and monitor their uses.
16. *Professional expertise*. Users should authorize only professionals with the requisite skills and experience to transact. These professionals include traders, supervisors, and those responsible for processing and controlling activities.

All these recommendations are still applicable. Nowadays, however, firms tend to integrate their market, credit, and operational risk functions due to the relationship between these risks.

21.1.2 The Bank of England Report on Barings

The Barings failure served as a powerful object lesson in risk management. By one estimate, Barings had ignored half the G-30 recommendations. But new lessons were learned from this fiasco. The Bank of England's report mentioned for the first time *reputational risk*. This relates to the risk to earnings arising from negative public opinion. Reputational risk can expose an institution to litigation or financial loss from the disruption of relationships with clients.

The report also identified several lessons from this disaster.

- *Management teams have a duty to understand fully the businesses they manage*. Top management at Barings did not have (or claimed so) a good understanding of Leeson's business despite the fact that it was apparently creating huge profits for the bank.

- *Responsibility for each business activity must be clearly established*. Barings was using a “matrix” reporting system (by region and product) that left ambiguities in the reporting lines for the trader, Nick Leeson.
- *Clear segregation of duties is fundamental to any effective risk-control system*. Indeed, the failure has been ascribed to the fact that Leeson had control over both the front and back offices. The Barings affair demonstrated once and for all the need for independent risk management.

21.1.3 The CRMPG Report on LTCM

The Counterparty Risk Management Policy Group (CRMPG) was formed in the wake of the LTCM near failure to strengthen risk management practices in the industry.[2] As with the G-30, this private-sector initiative also aimed at forestalling heavy-handed regulation of financial markets.

Indeed, the brokerage industry had come under fire for allowing LTCM to build up so much leverage. Chase Manhattan, for example, had a $3.2 billion exposure to LTCM, equivalent to 13 percent of the bank’s equity. Apparently, much of this *current* exposure was collateralized because the loans were marked to market. There was no margin, or haircut, though, to provide further protection. Had LTCM been forced into default, the *potential* credit exposure could have been quite large, with total losses to brokers estimated at up to $6 billion. Brokers had underestimated the interactions between market, credit, and liquidity risks.

In its defense, the brokerage industry argued that it did not have a complete picture of LTCM’s positions. The hedge fund maintained a religious secrecy about its positions, even to its own investors. Yet a report by the President’s Working Group on Financial Markets (1999) found “serious weaknesses in how firms used what information they did have.”

In response, the CRMPG (1999) report provides a set of recommendations, summarized as follows:

- *Information sharing*. Financial institutions that engage in dealings likely to entail significant credit exposures should assess capital conditions and the market and liquidity risk of their counterparty. Since some of this information is considered confidential, institutions should have in place policies governing the use of proprietary information.
- *Integrated view of risk*. Financial institutions should apply an integrated framework to evaluate market, credit, and liquidity risk, especially for

highly leveraged counterparties.

- *Liquidation-based estimates of exposure*. Institutions should measure their credit exposure not only using current exposure but also using potential exposure assuming liquidation of positions. This is especially important when exposures are large or illiquid.
- *Stress testing*. Institutions should stress test their market and credit exposure, taking into account concentration risk to groups of counterparties and the risk that liquidating positions could move the markets.
- *Harmonization of documentation*. The report identified areas for improvements in standard industry documents, which should help to ensure that netting arrangements are carried out in a timely fashion.

The appendices to the report provide an analytical framework for evaluating the effects of leverage on market liquidity and credit risk. Several measures are also proposed to evaluate funding liquidity risk, some of which were explained in Chapter 13.

The LTCM affair has forced financial institutions to recognize that credit risk and market risk are related. Indeed, a survey by Capital Markets Risk Advisors revealed that the proportion of institutions having integrated the two functions rose from 9 percent before 1998 to 64 percent after the crisis. Similarly, the number of firms making adjustments for large or illiquid positions rose from 25 percent of respondents to 58 percent. Finally, many more institutions now perform systematic stress tests. Thus the industry is belatedly learning from this episode and moving toward better risk management practices.

CRMPG (2005) evaluated recent progress in a later report, called *CRMPG II*. The report provides a useful summary of recent developments in financial markets. It notes that the risk of systemic financial shocks had fallen since 1998 owing to a number of factors: (1) the strength of the key financial institutions at the core of the financial system, (2) improved risk management techniques, (3) improved supervision, (4) more effective disclosure, (5) strengthened financial infrastructure, and (6) more effective techniques to hedge and widely distribute financial risks. Yet CRMPG II points out a number of potential weaknesses. It recommends further improvement in risk management practices, including paying particular attention to assumptions underlying risk models. This is the subject of the next sections.

21.2 LIMITATIONS OF VAR

Although VAR provides a first line of defense against financial risks, it is no panacea. Users must understand the limitations of VAR measures. These drawbacks can be classified into limitations of the system that are (or should be) generally recognized. More fundamental criticisms are explored in the next section.

21.2.1 Risk of Exceedences

The most obvious limitation of VAR is that it does not provide a measure of the absolute worst loss. VAR only provides an estimate of losses at some confidence level. Hence there will be instances where VAR will be exceeded. The lower the confidence level, the lower is the VAR measure, but the more frequently we should observe exceptions. This is why backtesting is an essential component of VAR systems. It serves as a reminder that exceptions are expected to occur, hopefully at a rate that corresponds to the selected confidence level.

21.2.2 Changing-Positions Risks

VAR also assumes that the position is fixed over the horizon. This also explains why the typical adjustment from 1-day to multiple-day horizons uses the square-root-of-time factor. This adjustment, however, ignores the possibility that trading positions may change over time in response to changing market conditions.

There is no simple way to assess the effect of changing positions on the portfolio VAR, but it is likely that *prudent* risk management practices create less risk than is suggested by VAR. For instance, the enforcement of loss limits gradually should decrease the exposure as losses accumulate (assuming liquid markets). This is similar to a long position in an option, which can be replicated dynamically by buying more of the asset as its price moves up or selling as its price decreases. This dynamic trading pattern thus is similar to purchasing an option, which creates a skewed distribution with limited downside potential. It is also possible, however, as Barings has demonstrated, that traders who lose money increase their bets in the hope of recouping their losses.

21.2.3 Event and Stability Risks

Another drawback of VAR models based on historical data is that they assume that the recent past is a good projection of future randomness. As always, there is no guarantee that the future will not hide nasty surprises that did not occur in the past. Surprises can take two forms, either one-time events (such as a devaluation or default) or structural changes (such as going from fixed to floating exchange rates). Situations where historical patterns change abruptly will cause havoc with models based on historical data.

In particular, changing correlation coefficients can lead to drastically different measures of portfolio risk. Recent work now extends portfolio optimization to scenarios where markets can be quiet, with normal correlation patterns, and more “hectic” periods, with correlations breaking down.[3]

Stability risk can be evaluated by *stress testing*, which aims at addressing the effect of drastic changes on portfolio risk. To some extent, structural changes can be captured by models that allow risk to change through time or by volatility forecasts contained in options. An example of structural change is the 1994 devaluation of the Mexican peso, which is further detailed as follows.

VAR and the Peso’s Collapse

In December 1994, the emerging-market play turned sour as Mexico devalued the peso by 40 percent. The devaluation was viewed widely as bungled by the government and led to a collapsing Mexican stock market. Investors who had poured money into the developing economies of Latin America and Asia faced large losses as the Mexican devaluation led to widespread drops in emerging markets.

Figure 21-1 plots the peso/dollar exchange rate, which was fixed at around 3.45 peso for most of 1994 and then jumped to 5.64 by mid-December.

FIGURE 21-1 Peso/dollar exchange rate.

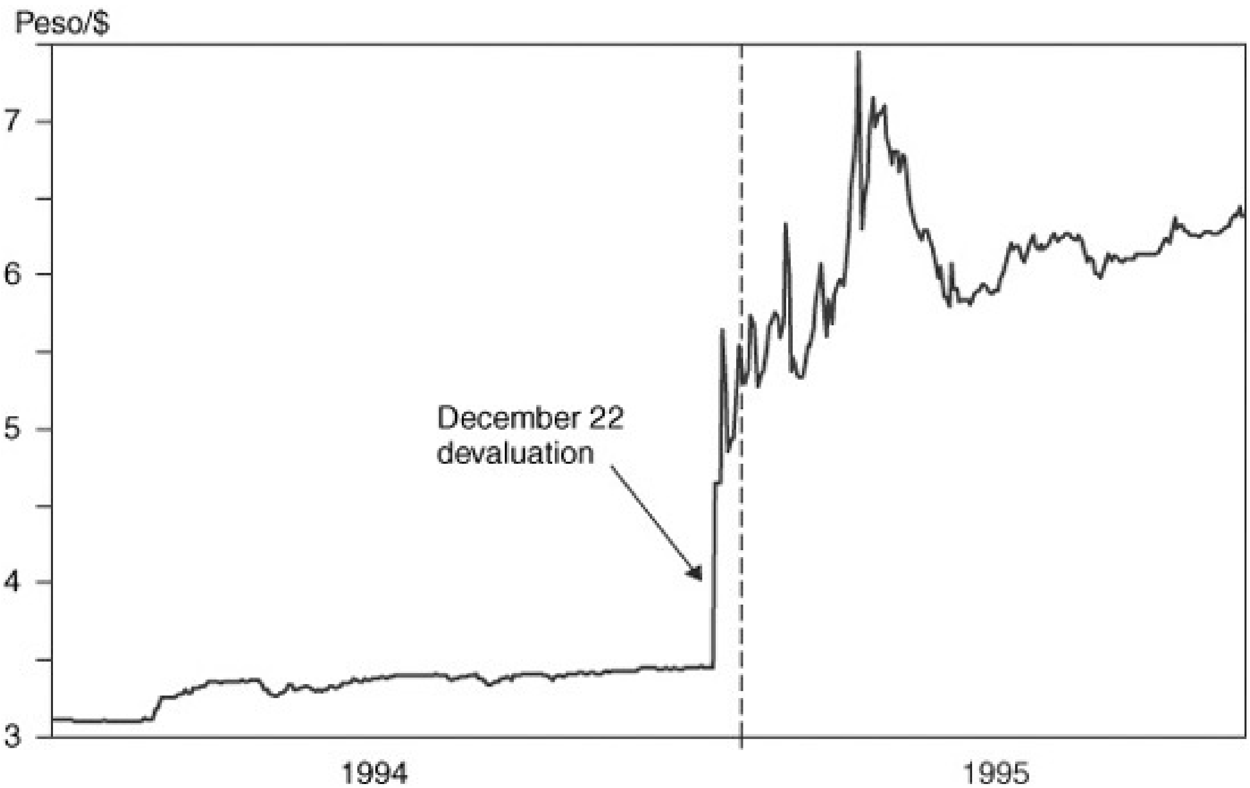
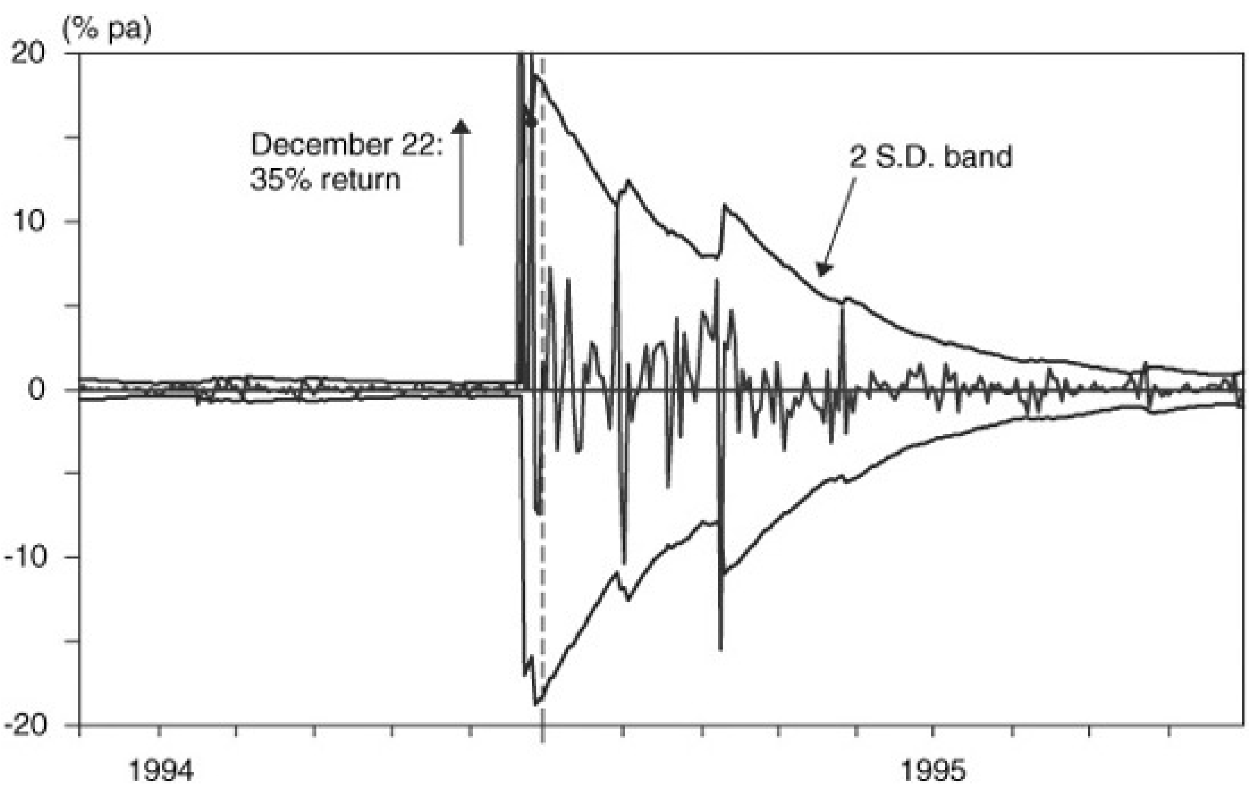


FIGURE 21-2 Peso/dollar volatility.



Apparently, the devaluation was widely unanticipated. This was despite a ballooning current-account deficit running at 10 percent of Mexico's gross domestic product (GDP) and a currency widely overvalued according to purchasing-power parity. A conventional VAR system would not have anticipated the magnitude of the devaluation. Based on an exponential volatility forecast, Figure 21-2 shows that the 35 percent devaluation was way outside the 95 percent confidence band. After December, the forecasts seem to capture reasonably well the turmoil that followed the devaluation. This was poor solace for investors caught short by the devaluation.

This episode indicates that especially when price controls are left in place for long periods, VAR models based on historical data cannot capture potential losses. These models must be augmented by an analysis of economic fundamentals and stress testing. Interestingly, shortly after the devaluation, the Mexican government authorized the creation of currency futures on the peso. It was argued that the existence of forward-looking prices for the peso would have provided market participants, as well as the central bank, an indication of market pressures. In any event, this disaster was not blamed on derivatives.

21.2.4 Transition Risk

Whenever there is a major change, a potential exists for errors. This applies, for instance, to organizational changes, expansion into new markets or products, implementation of a new system, or new regulations. Since existing controls deal with existing risks, they may be less effective in the transition. This creates *operational risk*.

Transition risk is difficult to deal with because it cannot be modeled explicitly. The only safeguard is increased vigilance in times of transition.

21.2.5 Data-Inadequacy Risks

Problem positions are in a category similar to transition. All the analytical methods underlying VAR assume that some data are available to measure risks. For some securities, such as infrequently traded emerging-market stocks, private placements, initial public offerings, and exotic currencies, meaningful market-clearing prices may not exist, however. Without adequate price information, risk cannot be assessed from historical data.

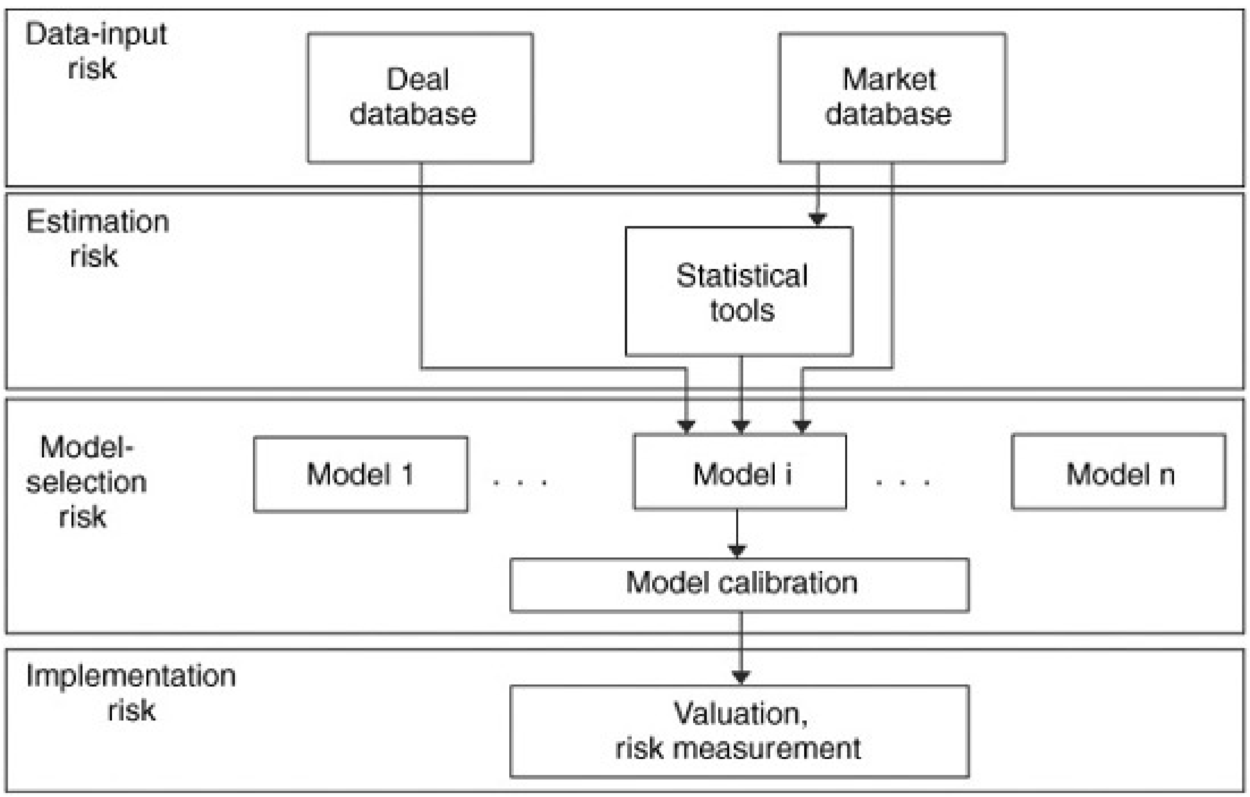
One solution is to use *proxies*, or similar series, instead. Alternatively, the mapping approach explained in Chapter 11 can be used to replace the missing series by major risk factors. In the absence of better information, risk can be controlled by diversification and exposure limits.

21.2.6 Model Risks

Model risk can be defined as the risk of loss occurring from the use of inappropriate models for valuing securities. This can result in misvaluation of the portfolio and hence of its risks. Model risk usually falls under the umbrella of operational risks.

Models are just abstractions of reality. As Emanuel Derman, head of quantitative strategies at Goldman Sachs, states, “A model is just a toy, though occasionally a very good one, in which case people call it a theory.” Models can fail for a number of reasons: (1) The input data can be wrong, (2) the parameters of the model can be incorrectly estimated, (3) the model can be incorrect, and (4) the model can be incorrectly implemented. This taxonomy is presented in Figure 21-3, adapted from Crouhy *et al.* (1998). Each of these is now examined in turn.

FIGURE 21-3 Model risk.



Data-Input Risk

Models rely on input data such as deal data and market data. At the most basic level, models can go wrong if the input data are flawed, owing to bad luck or by design. For instance, the Black-Scholes model requires inputting the implied volatility. Using wrong values can lead to major losses, as shown in Box 21-1.

Estimation Risk

Also known as *parameter risk*, estimation risk stems from the imprecision in the measurement of parameters estimated by statistical tools. We never observe the true values of volatilities and correlations. The estimates we compute depend on the sample selected, creating random errors in the process. This estimation-risk issue often is ignored in VAR analyses.

Chapter 5 has presented methods to assess the imprecision in VAR. More generally, the risk manager could assess the effect of estimation risk by resampling using Monte Carlo simulations. A simpler method consists of measuring VAR over a different historical period. Morgan Stanley is one of the few banks that provides such information. It reports VAR at the 99 percent confidence level using historical simulation. In 2004, the bank disclosed a VAR estimate of $72 million using a 1-year history and $73 million using a 4-year history. The fact that these two numbers are close to each other provides some confidence in the estimate. Note that it would be meaningless to report VAR with five significant digits.

BOX 21-1

NATWEST’S MODEL RISK

On February 28 1997, National Westminster Bank announced a loss of £77 million ($127 million) owing to mispricing derivatives. This revelation was shocking because NatWest, the biggest bank in the United Kingdom, was assumed to have a sophisticated risk management system. The size of its interest-rate option book alone was enormous, £267 billion in 1996.

NatWest said that a junior interest-rate trader, Kyriacos Papouis, “covered up losses and created false profits over a period of 2 years.” The trader had been dealing in long-dated OTC interest-rate options. Such options are used by companies, for instance, that borrow at a floating rate and purchase a “cap” on interest payments. Mr. Papouis would sell the cap and charge a premium presumably high enough to cover the bank’s risk.

The difficulty lies in valuing these options when they are relatively illiquid. Valuing options essentially reduces to estimating their implied volatility, which is easy to do for liquid at-the-money options that have a ready market. For out-of-the-money options, however, estimating volatility involves extrapolation.

Mr. Papouis calculated the price of these options by feeding his own estimates of volatility. Apparently, he overestimated the volatility and thus the model value of the options he had been selling, creating fictitious profits that built up over time. This conveniently translated into large bonuses for him.

To get an idea of the effect of mispricing, consider a typical £1 million 3-month cap starting in 1 year with a cap rate 1 percent higher than the current rate. Using Black’s model, the value of this cap is £520 with a 20 percent volatility. With a 25 percent volatility, the value grows to £805. Assuming that these prices are representative, extrapolating the difference to a £267 billion portfolio creates an error of £76 million, on the order of the reported loss.

It is not clear why NatWest’s risk managers accepted the trader’s volatility estimates. Observers have speculated that the bank’s culture made it difficult for financial controllers to question what traders were doing. One even said, “Little squirts in the middle office earning £20,000 a year don’t stand in the way of a producer earning £200,000 or £2 million.” This incident reflected badly on controls at the bank. Six managers resigned. Martin Owen, the chief executive of NatWest Markets, initially announced that he would give up £200,000 from his £500,000 bonus. He resigned soon thereafter. The bank’s reputation was badly damaged. Some analysts claim that this incident paved the way for its eventual takeover by the Royal Bank of Scotland in 2000.

A fundamental tradeoff always arises between using more data, which leads to more precise estimates, and focusing on more recent data, which may be more appropriate if risk changes over time. Unfortunately, the data may not be available for very long periods. Even worse, the available histories may give a distorted picture of risk merely owing to the *survival* of the series. Survivorship is an issue when an investment process only considers existing assets. The problem is that assets that have fared badly are not observed. Analyses based on current data therefore tend to be overly optimistic. Survivorship bias arises with the evaluation of *hedge funds*, which disappear at much higher rates than other investment funds. This phenomenon biases risk measures if the analysis only considers *live funds*. *Dead funds* have systematically higher risk.

More generally, unusual events with a low probability of occurrence but severe effects on prices, such as defaults, wars, or nationalizations, are not likely to be well represented in samples and may be totally omitted from survived series. Unfortunately, these unusual events are very difficult to capture with conventional risk models. This is why stress tests are essential complements to VAR.

Model-Selection Risk

This is the risk of losses owing to inappropriate models. It arises during the selection of a *pricing model* or *empirical relationships*.

Valuation errors can arise if the particular functional form chosen for valuing a security is incorrect. The Black-Scholes model, for instance, relies on a rather restrictive set of assumptions (i.e., geometric brownian motion, constant interest rates, and volatility). For conventional stock options, departures from these assumptions may not be consequential. However, there are situations where the model is totally inappropriate, such as options on short-term interest rate instruments. For these, one needs to model an interest-rate process, perhaps with a one-factor model, and then run simulations. But again, what may be satisfactory for these instruments may not be appropriate for another class of options, such as options on the slope of the term structure, which require more richness in the dynamics of interest rates.

Model risk also grows more dangerous as the instrument becomes more complicated. Pricing collateralized mortgage obligations (CMOs) requires heavy investments in the development of models, which may prove inaccurate under some market conditions.

Another most insidious form of risk is due to *data mining*. This occurs when analysts search over various empirical models and only report the one that gives good results. This is particularly a problem with nonlinear models (such as neural-network models), which involve searching not only over parameter values but also over different functional forms.

Data mining also consists of analyzing the data until some significant relationship is found. Take, for instance, an investment manager who tries to find “calendar anomalies” in stock returns. The manager wants to check whether stock returns systematically differ across months, weeks, days, and so on. So many different comparisons can be tried that in 1 case out of 20 one would expect to find “significant” results at the usual 5 percent level. Of course, the results are only significant because of the search process that discards nonsignificant models. Data-mining risk manifests itself in overly optimistic results, which often break down outside the sample period.

Data-mining risks can be best addressed by running *paper portfolios*, where an objective observer records the decisions and checks how the investment process performs on actual data.

Implementation Risk

Even if the model and its parameters are correct, implementation may be fraught with problems. With numerical methods, for instance, the solution may be badly approximated. Bugs can creep into the software and hardware. Even with the same program, *user risk* arises when different users obtain different solutions.

To get a sense of the magnitude of these problems, Marshall and Siegel (1997) surveyed 10 vendors of VAR software. The vendors were asked to report VAR numbers for standardized portfolios. Table 21-1 summarizes the distribution of vendors' estimates using the median and standard deviation (SD).

For the portfolio of forwards, for example, the median VAR across vendors is $425,800, with a standard deviation of $4800. The narrow range reflects the fact that forward contracts are easy to map and are linear in the risk factors.

TABLE 21-1
Implementation Risk

	Portfolio						
	Forwards	Money Market	FRAs	Global Bonds	Interest-Rate Swaps	FX Options	Interest-Rate Options
Notional (gross)	$130m	$46m	$375m	$350m	$311m	$374m	$327m
1-day 95% VAR							
Median	$425,800	$671,300	$79,000	$3,809,100	$311,100	$804,200	$416,700
SD	$4,800	$60,700	$7,500	$652,800	$66,600	$198,800	$115,200
Ratio (SD/median)	1%	9%	10%	17%	21%	25%	28%

As the products become increasingly more complicated, however, there is less agreement in VAR numbers. The most complicated products are interest-rate options (caps and floors), which are nonlinear and require modeling the term structure and its dynamics. Here, the standard deviation of VAR is 28 percent of the median. Model risk is greatest for the most complicated instruments.

There is no easy solution to model risk. As the industry moves toward more complex models, there is also a greater risk of mistakes. This is why modeling has become multidisciplinary and requires a good understanding of the process, from model development to coding and user interface. Modelers, programmers, and users need to work together to minimize model risk.

21.3 DANGERS OF VAR

21.3.1 False Sense of Precision

It has been argued that the widespread use of VAR is not only useless but even harmful because it gives a false impression of accuracy. This criticism of VAR is developed in Box 21-2.

The gist of this argument is that VAR is useless because it is not perfect, unlike measures in the physical sciences. As discussed in the preceding section, VAR is indeed not perfect. In defense of VAR, one might argue that our world is constructed by engineers, not physicists. This is why engineering has been described as the "art of the approximation." The same definition applies to VAR.

BOX 21-2

NASSIM TALEB'S ASSAULT ON VAR

Derivatives Strategy:*What do think of value at risk?**

VAR has made us replace about 2500 years of market experience with a covariance matrix that is still in its infancy. We made *tabula rasa* of years of market lore that was picked up from trader to trader and crammed everything into a covariance matrix. Why? So that a management consultant or an unemployed electrical engineer can understand financial-market risks.

To me, VAR is charlatanism because it tries to estimate something that is not scientifically possible to estimate, namely, the risks of rare events. It gives people misleading precision that could lead to the buildup of positions by hedgers. It lulls people to sleep.

Derivatives Strategy:*Proponents of VAR will argue that it has its shortcomings, but it's better than what you had before.*

That's completely wrong. It's not better than what you had because you are relying on something with false confidence and running larger positions than you would have otherwise. You’re worse off relying on misleading information than not having any information at all. If you give a pilot an altimeter that is sometimes defective, he will crash the plane. Give him nothing, and he will look out the window. Technology is only safe if it is flawless.

*©*Derivatives Strategy* (January 1997), reprinted with kind permission. The magazine followed up this interview with the Jorion-Taleb (April 1997) debate on the pros and cons of VAR.

Admittedly, risk managers must be aware of the limitations of VAR. It also behooves them to avoid creating an impression of undue precision when discussing VAR. Observers who claim risk management to be a “science” do a disservice to the profession. Risk management is as much an art as a science.

21.3.2 Traders Gaming the System

A potentially more serious danger is that traders could try to “game” the VAR system or evade risk limits when they are subject to VAR-based limits. This is called *VAR arbitrage.*

VAR arbitrage The deliberate creation of risky trades that appear to be low risk in a VAR framework.

For instance, traders could move into markets or securities that appear to have low risk for the wrong reasons. Currency traders could take large positions in currencies fixed against the dollar that have low historical volatility but high devaluation risk.

Historical simulation methods also create problems. For instance, if the window is very short, a dropoff day can create a VAR measure of risk that is predictably lower than the true or implied risk. Traders then could arbitrage by going long the asset and short options. If the window is too long, it will be slow to respond to increases in the true risk. Traders then could arbitrage in the same fashion.

Simple analytical methods may invite option trades. Traders exposed to a delta-normal VAR could take positions in short straddles with zero delta (like

Barings' Leeson). Such positions appear profitable, but only at the expense of future possible losses that may not be captured by VAR. More generally, a trader may be aware of measurement errors in the covariance matrix used to judge him or her. If so, he or she may overweight assets that have low estimated risk, knowing full well that this will result in a downward-biased risk measure.

Ju and Pearson (1999) provide estimates of this potential bias. Suppose that a trader is subject to a constraint on estimated VAR. If the trader knows the true covariance matrix, he or she will try to maximize the expected return on the position subject to this constraint. The true VAR, however, will be higher than the estimated VAR. Table 21-2 shows the ratio of true VAR to the estimated VAR for a various number of observations T and number of assets N. For instance, with 100 days and 50 assets, the true VAR is 201 percent, or twice the estimated VAR. The bias increases as the number of assets increases relative to the number of observations, reflecting increased measurement error in the covariance matrix.

With an exponential model for forecasting risk, the number of effective observations is small, leading to serious biases. With a decay factor of 0.94 and 50 assets, the true VAR is nearly 5 times the estimated VAR.

In the context of portfolio management, gaming by traders can be compared with the general problem of in-sample portfolio optimization, which is well known to create biased views of risk. This danger lies in relying on the same covariance matrix (i.e., in-sample) to perform the portfolio optimization and to measure risk.[4]

TABLE 21-2
Bias in VAR: Ratio of True to Estimated VAR

Model: Observations T	Number of Assets N 10	20	50	100
Moving average				
50	123%	164%		
100	110%	124%	201%	
200	105%	111%	133%	199%
1000	101%	102%	105%	111%
Exponential	135%	185%	485%	2174%

Note: The table reports the mean ratio of true VAR to estimated VAR assuming that the trader knows the true covariance matrix and maximizes expected return subject to a constraint on VAR. Because the trader may not know the true matrix, these numbers represent worst-case estimates of the bias. The exponential model uses 100 data points and a decay factor of 0.94.
Source: Adapted from Ju and Pearson (1999).

This behavior can be even more dangerous in the presence of *options*or asymmetric return distributions. Basak and Shapiro (2001) analyze the optimal behavior of managers subject to VAR limits and find that these managers incur large losses when losses occur. A manager with a 95 percent VAR limit of $10 million, for instance, may choose positions with losses that exceed VAR only 5 percent of time, but by a very large amount. In essence, the problem is that this simple VAR limit does not distinguish between expected losses beyond VAR of $20 million or $100 million. This simply reflects the fact that quantile-based measures are not “coherent,” as we saw in Chapter 5. The solution is to look not only at one quantile but also at the entire distribution or to compute the expected tail loss.

This is indeed a serious issue with VAR systems. This is why risk management is not simply a black box but a dynamic process where competent risk managers must be aware of the human trait for adaptation.

21.3.3 Creating Systemic Risk

Some take a more extreme view of the effects of risk management systems. The argument is that “attempts to measure risk in financial markets actually may be making them riskier.”[5]

Take, for instance, another high-tech computer-driven portfolio management technique, portfolio insurance. *Portfolio insurance* was developed in the mid-1980s as an application of the Black-Scholes model, which showed that a position in an option is equivalent to a dynamically adjusted position in the underlying asset. Hence the idea of replicating a long protective put strategy by buying “delta” of the asset. Variations in delta are such that as the asset falls in price, one would need to sell more of the asset to provide the protection.

In practice, these abstract models totally ignore liquidity. The problem is that as one starts selling large amounts in a falling market, the strategy can exacerbate price swings, thus amplifying volatility. Indeed, portfolio insurance has been blamed widely for having triggered the stock market crash of October 1987.[6] Note that the problem does not arise from automatic trading per se but rather from the practice of placing large sell orders in a falling market. More primitive trend-following systems have the same effect. The argument is that the widespread use of portfolio insurance could be destabilizing or create systemic risk.

A similar argument has been made with VAR systems. The Basel VAR-based risk charge came into effect in January 1998. This was shortly before the Russian default of August 1998, which led to falling prices that pushed up volatility. In turn, this could increase VAR and capital requirements.[7] The argument is that faced with binding VAR-based capital requirement, a “bank is then faced with two choices: put in extra capital or reduce its positions, whatever and wherever they may be. This is what happened last autumn.”[8] In turn, these forced sales depress prices, causing increased volatility, which further feeds into VAR. This is the vicious-circle hypothesis advanced by Persaud (2000).[9]

This line of argument should be a serious source of concern given the generalized trend toward risk-sensitive capital adequacy requirements. Such arguments are based primarily on anecdotal evidence, however. To be valid, this explanation requires most VAR-constrained institutions to start from similar positions. Otherwise, they could simply cross their trades with little effect on prices. These similar positions should translate into high correlations between mark-to-market returns. In practice, we do not observe such high correlations between the trading profits of U.S. commercial banks.[10] Thus there is no empirical evidence to support this story.

More generally, the question is whether volatility in financial markets is caused by risk management techniques or instead reflects changing fundamentals. We do know that, after the Russian default of 1998, there was a broad-scale reassessment of credit risk across global markets. And once the extent of LTCM’s problems became public, it was not clear that financial intermediaries would be able to weather an outright default by LTCM. This uncertainty was one of the causes of the widening of credit spreads that ensued. Thus the volatility of the market could have been due to fundamentals.

Despite the turmoil of 1998, the market risk management systems of financial institutions seemed to have worked better than expected. As Howard Davies (1999), chairman of the United Kingdom's Financial Services Authority, put it,

> It is fair to say . . . that—overall—financial institutions in the developed world survived the turmoil of '97 and '98 remarkably well. So their risk-management systems cannot have been as bad as all that.

Indeed, the Basel Committee (1999b) surveyed the performance of 40 banks during the second half of 1998. The report showed that although some banks experienced "yellow zone" exceptions, that is, 5 to 10 exceptions in a 250-day period, none suffered a trading loss that exceeded its capital requirement. A few banks had large trading losses, but none was seriously threatened. Considering the turbulence of 1998, this is a remarkable achievement. In other times, such events could have wiped out a few banks.

The spread of risk management techniques seems to have elicited two types of reaction. On one side, doomsayers (the "pessimists") argue that such systems create systemic risks. On the other side, observers (the "optimists") state that risk management techniques are beneficial because they force participants to pay more attention to their risks and help spread risks throughout the financial system. Alan Greenspan (2004), for instance, said that, "In my judgement, better risk measurement and risk management were noticeably important in moderating overall credit losses during the most recent recession." As time goes by, the optimists' view of risk management seems increasingly more appropriate.

21.4 RISK MANAGEMENT LESSONS FROM LTCM

21.4.1 LTCM's Risk Controls

The story of the hedge fund LTCM provides useful risk management lessons.[11] As described in Chapter 13, on liquidity risk, the core strategy of LTCM consisted of convergence-arbitrage trades, that is, trying to take advantage of small differences in prices among near-identical assets. Compare, for instance, a corporate bond yielding 7.5 percent and an otherwise identical Treasury bond with a yield of 6 percent. The yield spread of 1.5 percent includes compensation for the expected cost of default plus a premium for risk or liquidity. In the absence of default, a trade that is long the corporate bond and short the Treasury bond would be expected to return 1.5 percent annually. In the short term, the position will be even more profitable if the yield spread narrows further. The key is that eventually the two bonds should converge to the same value, which is the repayment of the same principal at the same time. Most of the time, this will happen, barring a default or market disruption.

The problem with such a strategy is that it only provides tiny returns. Thus the portfolio had to be *leveraged* to create the 30 to 40 percent returns investors were hoping for. Without some constraint on risk, however, leverage could become immense.

LTCM chose to limit its risk by targeting a level of volatility similar to that of an unleveraged position in U.S. equities, that is, about 15 percent per annum. At the end of 1997, the fund had $4.7 billion in equity, that is, investor funds. Combining this volatility with the equity, we find a monthly dollar volatility of \$4700 × 15 percent/$\sqrt{12}$ = \$204 million. The portfolio was positioned so as to maximize expected returns subject to the constraint that the fund's perceived risk was no greater than that of the stock market. At least, in theory.

21.4.2 Portfolio Optimization

To understand how portfolio optimization was used to set up LTCM's positions, this section presents a stylized example based on mean-variance optimization with two highly correlated assets. We use a government and a corporate bond portfolio series and assume that the investor maximizes expected returns subject to the constraint that the annual volatility is 15 percent.

Table 21-3 presents a worked-out example of a portfolio optimization with two risky assets, a BAA-rated corporate bond and a Treasury bond. The data are taken from the 5-year period 1993–1997. Note that the two series have very high correlation, at 0.9654. Expected returns are simply approximated by the yields to maturity as of December 1997. With a credit-yield spread of 1.53 percent and no allowance for defaults, the optimization should identify an "arbitrage" strategy that is highly leveraged.

Indeed, the portfolio takes a very large position in the corporate bond, $10.5, offset by a large short position of $8.3 in the Treasury bond, for every $1 of equity. Thus the leverage ratio is very high, above 10. The sum of the bond positions is $2.2, which implies a loan of $1.2 at the risk-free rate. The expected return on the portfolio is 22 percent per annum. This simple exercise captures the essence of LTCM's so-called arbitrage strategy, which is illustrated in Figure 21-

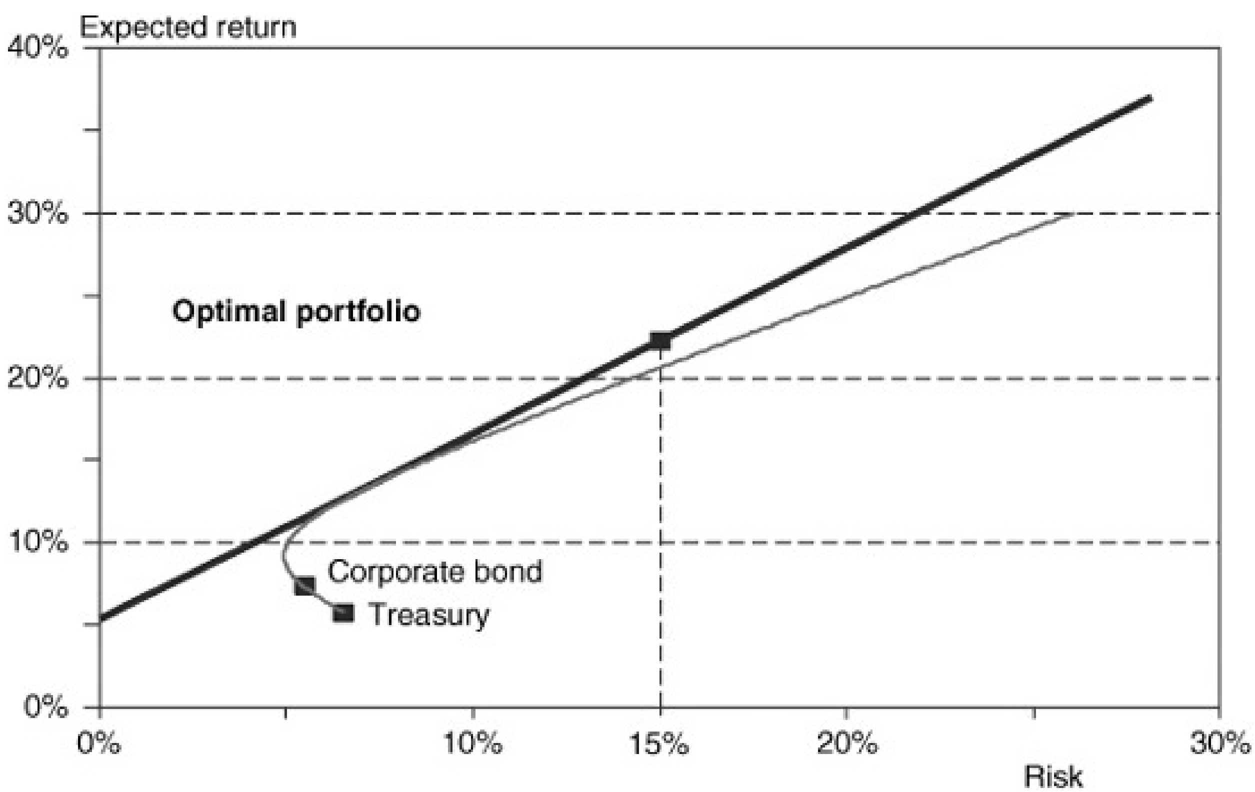
4. What is the problem with this strategy?

TABLE 21-3
Portfolio Optimization with Two Assets

Input Data	Corporate Bond	Treasury Bond	Risk-Free Asset
Expected return yield (%pa)	7.28%	5.75%	5.36%
Volatility of return (%pa)	5.47%	6.58%	
Correlation	0.9654		
Output data			
Position (for $1 equity)	$10.5	−$8.3	−$1.2

Optimal Portfolio	Monthly	Annual
Expected return	1.9%	22.2%
Volatility of return	4.3%	15.0%

FIGURE 21-4 Portfolio optimization.



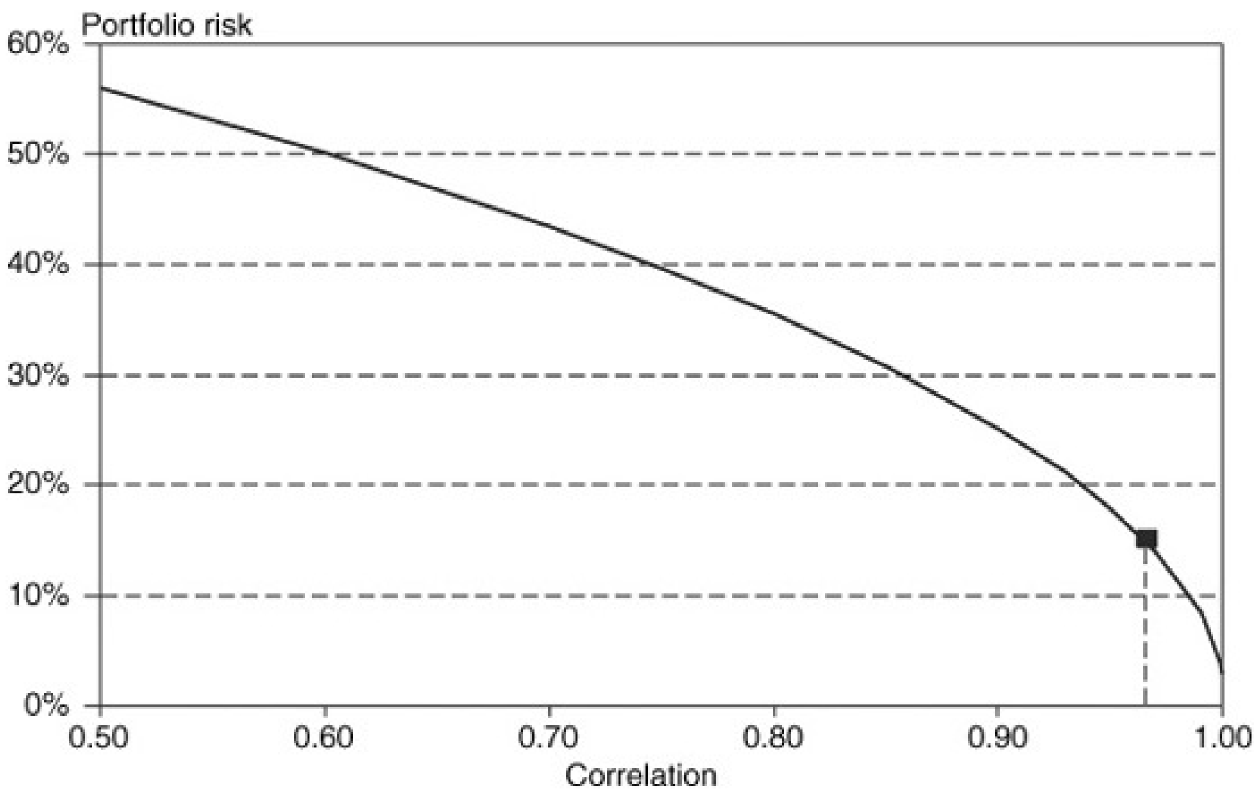
For every dollar invested, the annual volatility is 15 percent, assuming the correlation stays at $\rho = 0.9654$. The problem is that this number is likely to have been measured with some error and probably is too high. In addition, the correlation may change over time, more likely down because it is already so high.

The issue is how sensitive is the estimate of the portfolio risk to changes in this crucial parameter, the correlation coefficient. This stress-testing exercise is performed in Figure 21-5, which examines changes in the portfolio volatility as the correlation coefficient decreases.

We start with a volatility of 15 percent using the estimated correlation of 0.965. If the correlation drops to 0.80, however, the portfolio volatility rises sharply, to 36 percent.

In fact, this correlation, which had been high in the recent past, dropped sharply to 0.80 in 1998, explaining why the convergence strategy suddenly went bad. The portfolio returns were –$325 million in May, –$440 million in June, and –$1850 million in August. It is unlikely that these numbers could have resulted from a distribution with a volatility of $204 million only.

FIGURE 21-5 Effect of changing correlation on portfolio risk.



More likely, the portfolio risk had been underestimated owing to biases from portfolio optimization. As this exercise demonstrates, risk measures derived from an optimization exercise can be extremely sensitive to errors in input parameters.

21.4.3 LTCM's Short Option Position

In addition, the payoff profile of LTCM was strongly asymmetric. The fund took large positions in interest-rate swaps. Up to 1998, swap spreads had narrowed sharply. As a result, the distribution of these spreads had to be strongly asymmetric because swap spreads cannot go below zero. Thus LTCM was exposing itself to large losses in case of spreads widening.

LTCM also had taken positions in Russian bonds and other emerging-market debt. These provide high yields, but at the expense of a possibility of a large loss in case of default. Again, this type of risk exposure generates an asymmetric distribution.

LTCM also had short positions in option-implied volatilities. Volatilities are asymmetrically distributed because they cannot go below zero but can increase greatly. Here again, the distribution is asymmetric.

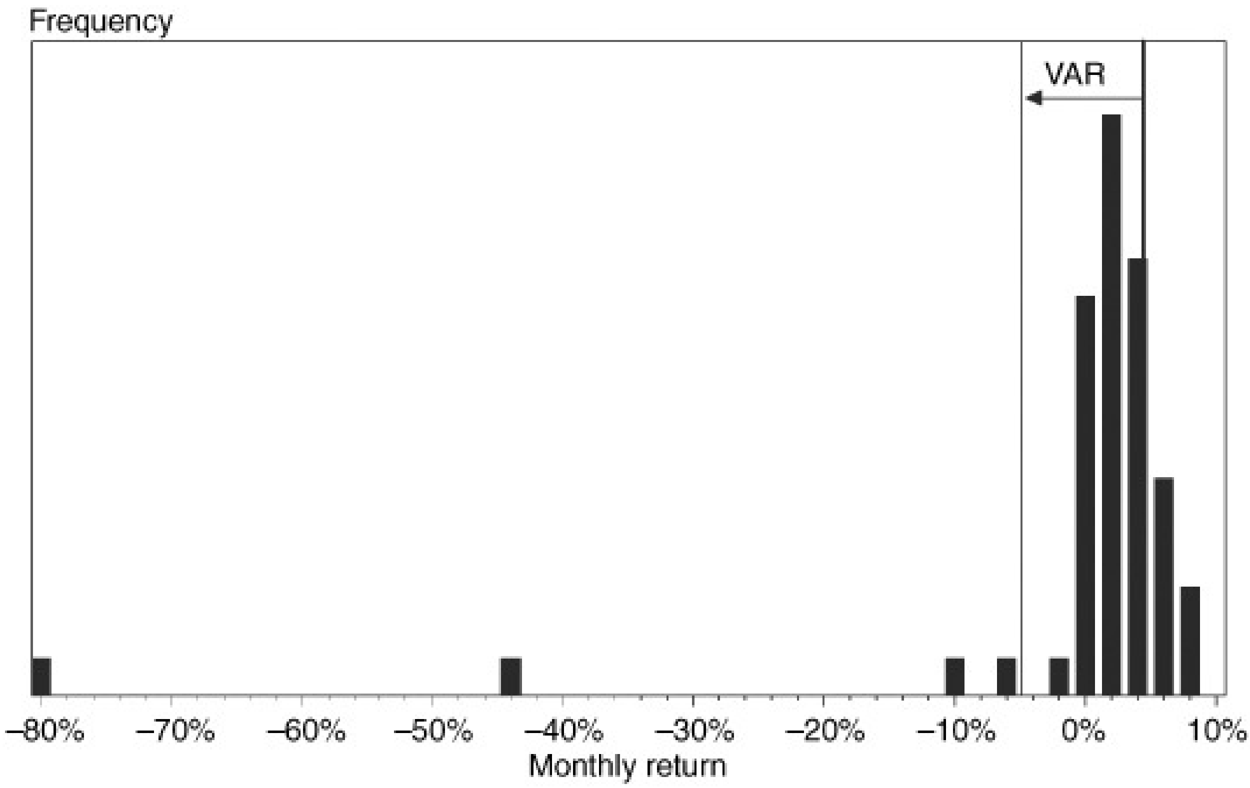
To make it even worse, the LTCM portfolio was exposed to liquidity risk owing to its huge size. As we learned from Chapter 13, liquidity is positively correlated with volatility.

All these positions added up to a rather undiversified portfolio with a distribution strongly skewed to the left, as shown in Figure 21-6. Based on history up to 1997, the 95 percent portfolio VAR would have been about minus 5 percent. Instead, the portfolio delivered several very large negative returns, culminating in an 80 percent loss in September of 1998.

Overall, the near failure of LTCM can be ascribed to inappropriate use of risk management tools. LTCM was not diversified in terms of strategy. Its payoff profile was strongly asymmetric, like a short position in a gigantic option, invalidating the use of volatility as a measure of risk. When positions involve infrequent events, risk cannot be measured from recent data only. These mistakes explain why LTCM's performance turned from stellar to disastrous in short order.

Overall, however, the LTCM episode was an aberration. It was the largest hedge fund in the industry by far, with an unprecedented leverage.

FIGURE 21-6 Distribution of LTCM's monthly returns.



It was so big that like a wounded tanker heading for a reef, it was unable to alter its positions after having lost half its capital. Hopefully, financial institutions and regulators will have learned their lesson and not allow this disaster to happen again.

21.5 CONCLUSIONS

While current risk management practices represent a huge step forward from unbridled risk-taking, the experiences of the last few years have shown that we still have much to learn in the application of risk management models.

In particular, users should be aware of limitations of VAR measures. VAR does not attempt to pinpoint the worst loss. Instead, one should expect regular exceedences. VAR also typically assumes some stability in the portfolio composition and, if based on historical data, in the risk measures. Finally, VAR is subject to model risk, which involves the choice of models, parameters, and their implementation.

VAR systems also may have more subtle, and perhaps dangerous, side effects. The technique may give users a false sense of accuracy, lulling portfolio managers into taking bigger positions than they otherwise might. Traders may attempt to game the system, exploiting flaws in VAR risk measures. Sometimes, as we have seen in the extreme case of LTCM, this involves positions equivalent to large sales of options, providing regular profits at the expense of being exposed to rare but huge losses.

This is why risk management will never be a pure science. Instead, it should be viewed as an evolving art form. VAR taught us that risk management requires a comprehensive approach to risk. Otherwise, a piecemeal approach can miss significant risks or, worse, create a misleading sense of safety. As we have seen, ideally, we should measure market, credit, operational, and hopefully other risks in a comprehensive fashion. We also learned that formal risk management models cannot be substituted for judgment and experience.

As James Leach (1998), chairman of the House Banking and Financial Services Committee that conducted hearings on the LTCM affair, put it,

> The fact that [modern financial engineering] failed does not mean that the science of risk management is wrong-headed; just that it is still an imperfect art in a world where the past holds lessons but provides few reliable precedents.

QUESTIONS

1. Discuss why marking to market is the preferred method to value derivatives positions as opposed to methods based on historical costs.
2. The G-30 report recommended centralized risk measurement for market risk and for credit risk, considered separately. How could this approach be improved?
3. What is the new risk manifest in the Barings failure mentioned in the Bank of England report?
4. Discuss whether estimates of credit exposure should be based on current exposure.
5. Despite advances in risk management systems, some weaknesses remain. Discuss.
6. Why are there instances in which actual losses exceed the VAR measure?
7. VAR assumes that the position is fixed over the horizon. Why could this assumption lead to inaccuracy in assessing risk?

8. List two situations when the VAR model based on historical data will not provide good risk measures.
9. A bank discloses a VAR of $50 million, but in reality, this is an estimate from a distribution with a mean of $70 million and a standard error of $15 million. What type of risk is this?
10. An option desk uses the Black-Scholes model to price eurodollar options. What type of risk could this create?
11. Which of these two instruments would you expect to have greater implementation risk, a futures contract on a foreign currency or a knockout option on a foreign currency?
12. Discuss the risk that NatWest was exposed to.
13. What is the danger of a false sense of precision in VAR?
14. Define VAR arbitrage and give an example.
15. Describe how the widespread use of VAR could create systemic risk and whether there is any empirical evidence supporting this hypothesis.
16. Which of the following was not a key factor in the failure of LTCM:

    (a) sharp drops in the correlations between fixed-income securities,

    (b) a strong asymmetric payoff profile owing to short option positions,

    (c) unauthorized position taken by a single trader, or (d) liquidity risk owing to the huge transaction size.

17. A corporate fixed-income portfolio has been duration-hedged using Treasury bond futures. As a result, what risk factors is the portfolio exposed to?
18. A risk manager computes VAR using historical data. Which one of the following positions is most likely to have underreported risk: (a) a long position in a floating exchange rate, (b) a short position in a floating exchange rate, (c) a long position in a fixed currency that is likely to devalue soon, or (d) a short position in a fixed currency that is likely to devalue soon.

CHAPTER 22
Conclusions

> [Risk managers] . . . The New Emperors of Wall Street.
>
> —*Risk Magazine*, *March 1999*

The risk management industry has truly experienced a revolution since the early 1990s. Once the domain of a few exclusive pioneers, risk management is now wholly embraced by the financial industry. It is also spreading fast in the corporate world. As a result, the financial risk manager function is now acquiring strategic importance within the corporate structure. Risk managers must be proficient in an amazing variety of topics, ranging from the practical knowledge of financial markets to derivatives pricing, probability, and even actuarial insurance modeling.

An essential part of their tool kit is value at risk (VAR). VAR provides a forward-looking view of a portfolio's overall risk. The spread of VAR owes to its elegant simplicity. It summarizes financial risk in one number, a potential dollar loss. The use of VAR has spread from simple quantification of risk to a control and management tool. Some firms also use it as the basis for the allocation of capital. The VAR methodology has been extended to other forms of risk besides market risk, that is, credit risk, operational risk, and liquidity risk. Thus the implementation of VAR is a large-scale aggregation problem that requires tools different from those used for derivatives pricing.

Without a doubt, the impetus behind the growth of the risk management industry was the series of derivatives debacles of the 1990s. At that time, it seemed that the technology behind the creation of ever-more complex financial instruments had advanced faster than our ability to control it. This has been corrected in large part by risk management techniques that give users a better understanding of financial instruments.

Another major factor behind the advent of risk management is the perverse incentive structure or moral-hazard problem. Profit-based compensation, as well as government insurance schemes, create incentives to take on extra risk, in some cases unwarranted risks. Risk management has grown as a counterbalancing factor to such problems.

But the explosive growth in risk management could not have evolved from these essentially defensive applications. The industry now realizes that risk management is a centerpiece of all financial market activity. Risk management, which grew up on a desk-by-desk basis, is now applied at the level of the whole corporation.

This chapter provides some concluding thoughts on VAR and risk management. Section 22.1 reflects on the evolution of risk management, which evolved from accrual methods to marking to market to VAR measures to optimization. Section 22.2 then discusses the newly established position of *risk manager*. Finally, Section 22.3 provides concluding comments on VAR.

22.1 THE EVOLUTION OF RISK MANAGEMENT

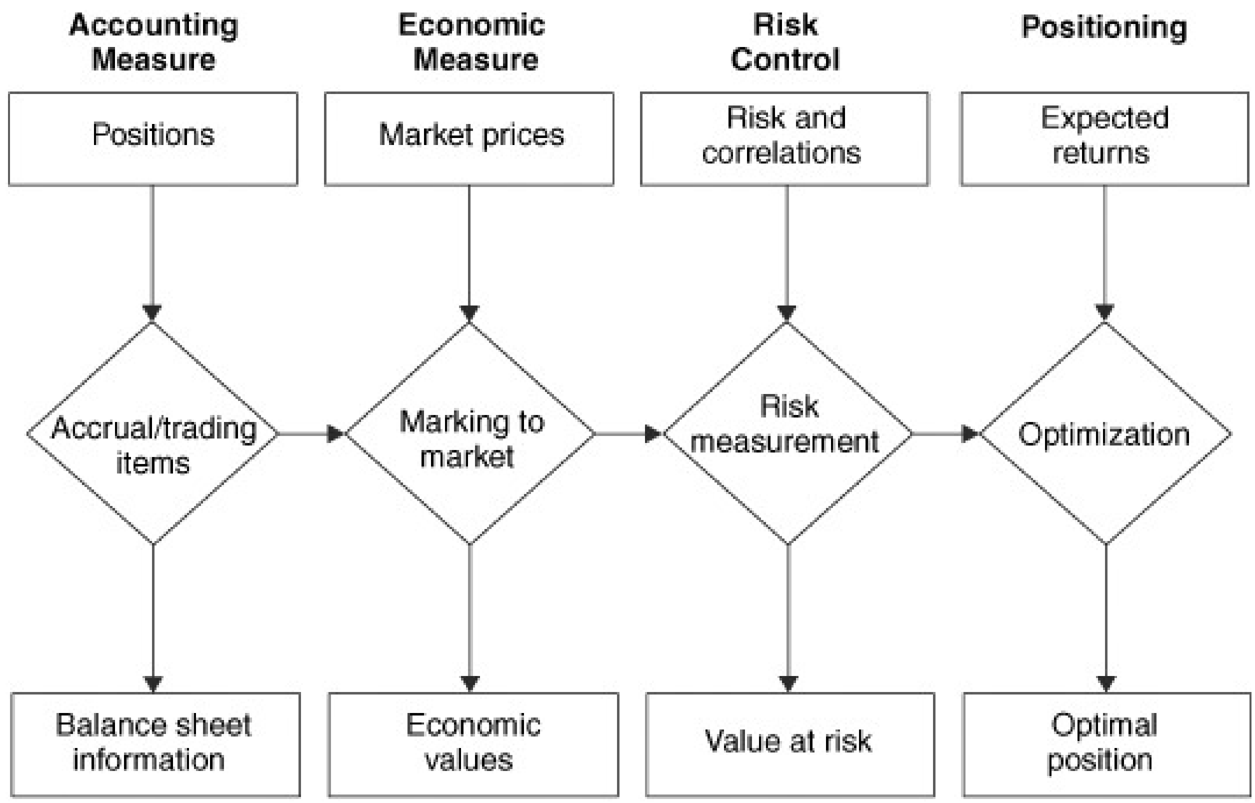
The steps leading us to VAR provide an interesting reflection on the evolution of modern financial management, as described in Figure 22-1.

VAR's antecedents can be traced to asset/liability management systems in place in the 1980s. At the time, banking institutions carried most of their financial assets and liabilities on the balance sheet using *accrual methods*; that is, transactions were booked at historical costs with adjustments for accruals. Some items, such as those held for the purpose of trading, were carried at market values.

The problem was that these accounting methods insulated the value of balance sheet items from their economic reality. Sometimes forecasts of future rates were used to project income over long periods, in a manner somewhat similar to what we describe as *scenario analysis*. Such accounting methods contributed to the great savings and loan debacle because it allowed institutions to present balance sheets that were in accordance with accounting rules but were hiding large losses.

Later, with the trend toward marking to market, some balance sheets started to be reported at market values. Once market values are available, the next logical step is to assess risk. A simple method for computing

FIGURE 22-1
Modern financial management.



VAR, for instance, consists of keeping track of the market value of all securities over a selected time interval, which gives an idea of the possible range of values for a trading portfolio. Thus the combination of positions, marking to market, and fluctuations in market values naturally leads to the concept of value at risk (VAR).

Once VAR is quantified, the next step consists of using the risk-control system as a feedback mechanism to evaluate business units. VAR provides a framework to compare the profitability of various operations on a risk-adjusted basis. Firms then can make informed decisions about maintaining or expanding lines of business or whether to hedge financial risks at the firm level. Thus risk measurement is the first step toward risk management and eventually *portfolio management*.

More generally, optimization makes the best use of return forecasts, combined with risk and correlations, to find the set of portfolios or businesses that provides the best tradeoff between risk and return. Chapter 7 has shown how to achieve this optimal combination. Thus risk management is now poised to take full advantage of Markowitz's portfolio theory.

22.2 THE ROLE OF THE RISK MANAGER

22.2.1 Controlling Risk

The impetus behind the rise of the risk manager was the realization of the perverse effects of compensation incentives. On the one hand, profit-based compensation pushes traders to do their very best and weeds out inefficient ones. On the other hand, the design of contracts with risk-taking agents gives them incentives to engage in activities that may not be in the best interest of their firm.

One of the lessons of these many financial disasters is the fundamental asymmetry, or convex profit pattern, in payoffs to traders (see Figure 16-2.) Absent fraud, a losing bet just means a lost job and some reputational damage. In contrast, a winning bet can lead to lifelong wealth.

Senior management in financial institutions is coming to realize that the negative effects of these convex patterns should be offset by some concave functions of profits. This can take the form of risk-based adjustment, such as RAROC, or by hiring risk managers. Obviously, the decision to implement risk management systems must come from the top of the institution.

22.2.2 Organizational Guidelines

Senior managers bear a particular responsibility because they define objectives, procedures, and controls. They can foster safe or unsafe environments through the choice of organizational structure.

Risk management practices vary widely. Less advanced companies may operate a credit-risk committee only and generally aggregate their risks at the business level only. More advanced institutions have global risk committees for market, credit, and operational risk and use quantitative measures of risk.

Figure 22-2 describes one implementation of a control model. The key to this flowchart is that the risk management unit is independent of the trading unit. Risk managers should not report to anybody whose compensation is linked to the success of a trading unit but rather should report directly to top management. Also, the compensation of risk managers and auditors cannot be associated with how well traders perform. In this structure, each unit has segregated duties and no overlapping management at lower levels. This provides for a system of checks and balances.

FIGURE 22-2
Organizational structure for risk management.

The implementation of risk management systems is spreading slower than one would wish, however, particularly outside industrialized markets. Besides the cost and intellectual development needed to develop the technological support for risk management, there is often a conflict of culture between the trading area and traditional bankers. Whereas traders typically are well versed in derivatives pricing and attendant risk measures, traditional loan officers are often less familiar with these concepts. The challenge is to convince the whole organization of the benefits from better control and pricing of risks. Faster-moving financial markets, regulatory pressure, and lessons from recent financial disasters all should prod global banking into better management of financial risks. Clearly, the impetus for change must come from the top of the institution.

22.2.3 Risk Managers

Unfortunately, there is a great temptation to cut corners on risk management and controls. Unlike traders, these units do not contribute directly to the bottom line of the firm. Jobs in back and middle offices are unglamorous jobs. In particular, risk managers serve a function similar to *selling an option*: At best, nothing happens; at worst, they fail to detect a problem, and they may be out of a job. This is the opposite of traders, for whom the performance-bonus link is similar to *buying an option*.

Risk managers are a special breed. They must be thoroughly familiar with financial markets, with the intricacies of the trading process, and with financial and statistical modeling. Box 22-1 describes the profile of a well-known risk manager.

Risk managers must be attentive to details because they continuously put their reputations on the line. To be effective, they need to develop positive relationships with traders, convincing them of the usefulness of active risk management. As we have seen in Chapter 7, the VAR methodology can assist portfolio managers in making strategic decisions by measuring the impact of a trade on portfolio risk. While traders are paid to take bets in various markets, it is certainly less intuitive to estimate the marginal contribution to risk. The challenge for risk managers is to convince traders that risk management can help them too.

Yet risk managers cannot receive huge bonuses as traders do. Their compensation is a delicate issue. Institutions that try to skimp on the remuneration of back-and middle-office personnel will fail to attract qualified staff. A recent G-30 survey, for instance, finds that there is "some concern that the development of staff in support areas lags behind."

BOX 22-1

PROFILE OF A RISK MANAGER

Jacques Longerstaey's career perfectly illustrates the development of the risk management profession.

Jacques Longerstaey received a bachelor's degree in economics from the University of Louvain in Belgium in 1985. He then began his career at J.P. Morgan in Brussels as a market strategist. This gave him the opportunity to develop the bank's first implementation of value at risk.

While in his next position as head of J.P. Morgan's Bond Index Group, he was asked to make the method available to the external community. In 9 months, his team developed *RiskMetrics*, which was unveiled in October 1994. This system added to the global impetus toward risk management.

Jacques Longerstaey then moved to Goldman Sachs in 1998. There he helped create the first risk management system for Goldman Sachs Asset Management. In 2003, he joined Putnam Investments, a global money management firm, as head of risk management. In 2004, he received the prestigious GARP Risk Manager of the Year Award, which is presented to individuals "who have achieved a level of excellence unique in financial services” and “whose career left a positive impact, both in terms of personal results and in the broader sense of advancing the risk management profession.”

BOX 22-2

RISK MANAGEMENT AT CHASE

Chase Manhattan was one of the few U.S. banks that emerged relatively unscathed from the summer of 1998. This came from the lessons from the year before. Chase had lost $78 million in the fourth quarter of 1997 owing to the Asian crisis. While not life-threatening, this loss prodded the bank into augmenting its risk management system with stress testing.

In particular, the stress tests allowed the bank to evaluate a scenario where credit spreads—then at historical lows—would widen again. The bank positioned its portfolio accordingly and suffered minimal losses despite a hectic summer. Marc Shapiro, Chase’s vice chairman and head of risk management, is convinced that the market risk team “saved the bank from disaster.”

This is where senior management again plays an important role. Strong internal controls are in the best interests of the institution because perceptions of a counterparty’s integrity are vital to the continuous flow of business. Effective oversight also reduces the likelihood that an institution will be exposed to litigation, financial loss, or reputational damage. The spectacular failures of institutions that lacked internal controls should serve as a powerful object lesson in the need for risk management.

Risk managers now have increasing responsibilities among the best institutions. No longer perceived as ex-traders tired of their job, they are given the power to allocate capital and, in some cases, even trader bonuses. As Box 22-2 shows, they can be crucial to the survival of the institution.

22.3 VAR REVISITED

The history of finance is littered with financial disasters. These expensive lessons have led the industry to adopt VAR as a universal benchmark for managing financial risk. VAR integrates market risk across all assets, derivatives, stocks, bonds, or commodities. VAR can be adapted to account for credit risk, liquidity risk, and operational risk.

Admittedly, VAR is no panacea. As we have seen, VAR makes no attempt to measure the losses beyond the specified limit. Even with a 99 percent confidence interval, unusual events happen, and they sometimes do so with a vengeance. Historical-based methods also have shortcomings. This is why VAR must be augmented by stress testing, which aims at assessing the effect of unusual market conditions. While VAR techniques are firmly grounded on a scientific basis, their interpretation remains more of an art than a science.

Thus VAR should be considered only as a first-order approximation. The fact that the value is generated from a statistical method should not hide the fact that it is only an estimate. Users should not be lulled into a state of complacency but rather recognize the limitations of VAR, which have been amply documented in this book. As Steven Thieke, chairman of J.P. Morgan's risk management committee says, "There has to be a point where this stops being a risk measurement methodology and becomes a management issue—what is the level of experience of the people in this business, and the firm's tolerance for risk."

Appropriate use of VAR, however, may have avoided some of the spectacular debacles of recent years, where investors had or claimed to have had no idea of their exposure to financial risks. In addition, the implementation of VAR forces integration of the front office (trading desk), of the back office, and of a newly created middle office, which performs a risk management function. This integration, although not necessarily easy in terms of logistics, has the side benefit in that it provides some protection against operational risk and is the only consistent approach to credit risk measurement.

This explains why VAR has become the new benchmark for managing financial risks. But clearly, the process of getting to VAR is as important as the number itself.