PEARSON　　ALWAYS LEARNING

2017 Financial Risk Manager (FRM®)

Exam Part II

Market Risk Measurement and Management

Sixth Custom Edition for the Global Association of Risk Professionals

GARP | Global Association of Risk Professionals

Excerpts taken from:
Options, Futures, and Other Derivatives, Ninth Edition
by John C. Hull

Excerpts taken from:

Options, Futures, and Other Derivatives, Ninth Edition
by John C. Hull

Grateful acknowledgment is made to the following sources for permission to reprint material copyrighted or controlled by them:

"Estimating Market Risk Measures," and "Non Parametric Approaches," by Kevin Dowd, reprinted from *Measuring Market Risk,* Second Edition (2005), by permission of John Wiley & Sons, Inc.

"Back Testing VaR" and "VaR Mapping," by Philippe Jorion, reprinted from *Value at Risk: The New Benchmark for Managing Financial Risk,* Third Edition (2007), by permission of The McGraw Hill Companies.

"Messages from the Academic Literature on Risk Measurement for the Trading Book," reprinted by permission of Basel Committee on Banking Supervision, January 2011.

"Some Correlation Basics: Properties, Motivation, Terminology," "Empirical Properties of Correlation: How Do Correlations Behave in the Real World?," "Statistical Correlation Models—Can We Apply Them to Finance?," and "Financial Correlation Modeling—Bottom Up Approaches," by Gunter Meissner, reprinted from *Correlation Risk Modeling and Management* (2014), by permission of John Wiley & Sons, Inc.

"Empirical Approaches to Risk Metrics and Hedges," "The Science of Term Structure Models," "The Evolution of Short Rates and the Shape of the Term Structure," "The Art of Term Structure Models: Drift," and "The Art of Term Structure Models: Volatility and Distribution," by Bruce Tuckman and Angel Serrat, reprinted from *Fixed Income Securities: Tools for Today's Markets,* Third Edition (2012), by permission of John Wiley & Sons, Inc.

Learning Objectives provided by the Global Association of Risk Professionals.

Pearson Education, Inc., 330 Hudson Street, New York, New York 10013
A Pearson Education Company
www.pearsoned.com

Printed in the United States of America

1 2 3 4 5 6 7 8 9 10 XXXX 19 18 17 16

000200010272074298

RM/KS

PEARSON

ISBN 10: 1-323-56912-X
ISBN 13: 978-1-323-56912-2

Contents

2017 FRM Committee Members

Dr. René Stulz*, Everett D. Reese Chair of Banking and Monetary Economics
The Ohio State University

Richard Apostolik, President and CEO
Global Association of Risk Professionals

Michelle McCarthy Beck, MD, Risk Management
Nuveen Investments

Richard Brandt, MD, Operational Risk Management
Citibank

Dr. Christopher Donohue, MD
Global Association of Risk Professionals

Hervé Geny, Group Head of Internal Audit
London Stock Exchange

Keith Isaac, FRM, VP, Operational Risk Management
TD Bank

William May, SVP
Global Association of Risk Professionals

Dr. Attilio Meucci, CFA
CRO, KKR

Dr. Victor Ng, CFA, MD, Chief Risk Architect, Market Risk Management and Analysis
Goldman Sachs

Dr. Matthew Pritsker, Senior Financial Economist
Federal Reserve Bank of Boston

Dr. Samantha Roberts, FRM, SVP, Retail Credit Modeling
PNC

Liu Ruixia, Head of Risk Management
Industrial and Commercial Bank of China

Dr. Til Schuermann, Partner
Oliver Wyman

Nick Strange, FCA, Head of Risk Infrastructure
Bank of England, Prudential Regulation Authority

Sverrir Thorvaldsson, FRM, CRO
Islandsbanki

* Chairman

Estimating Market Risk Measures

An Introduction and Overview

1

Learning Objectives

After completing this reading you should be able to:

- Estimate VaR using a historical simulation approach.
- Estimate VaR using a parametric approach for both normal and lognormal return distributions.
- Estimate the expected shortfall given P/L or return data.
- Define coherent risk measures.
- Estimate risk measures by estimating quantiles.
- Evaluate estimators of risk measures by estimating their standard errors.
- Interpret QQ plots to identify the characteristics of a distribution.

Excerpt is Chapter 3 of Measuring Market Risk, *Second Edition, by Kevin Dowd.*

This chapter provides a brief introduction and overview of the main issues in market risk measurement. Our main concerns are:

- *Preliminary data issues*: How to deal with data in profit/loss form, rate-of-return form, and so on.
- *Basic methods of VaR estimation*: How to estimate simple VaRs, and how VaR estimation depends on assumptions about data distributions.
- How to estimate coherent risk measures.
- How to gauge the precision of our risk measure estimators by estimating their standard errors.
- *Overview*: An overview of the different approaches to market risk measurement, and of how they fit together.

We begin with the data issues.

DATA

Profit/Loss Data

Our data can come in various forms. Perhaps the simplest is in terms of profit/loss (or P/L). The P/L generated by an asset (or portfolio) over the period t, P/L_t, can be defined as the value of the asset (or portfolio) at the end of t plus any interim payments D_t minus the asset value at the end of $t-1$:

$$P/L_t = P_t + D_t - P_{t-1} \tag{1.1}$$

If data are in P/L form, positive values indicate profits and negative values indicate losses.

If we wish to be strictly correct, we should evaluate all payments from the same point of time (i.e., we should take account of the time value of money). We can do so in one of two ways. The first way is to take the present value of P/L_t evaluated at the end of the previous period, $t-1$:

$$\text{present value}(P/L_t) = \frac{(P_t + D_t)}{(1+d)} - P_{t-1} \tag{1.2}$$

where d is the discount rate and we assume for convenience that D_t is paid at the end of t. The alternative is to take the forward value of P/L_t evaluated at the end of period t:

$$\text{forward value}(P/L_t) = P_t + D_t - (1+d)P_{t-1} \tag{1.3}$$

which involves compounding P_{t-1} by d. The differences between these values depend on the discount rate d, and will be small if the periods themselves are short. We will ignore these differences to simplify the discussion, but they can make a difference in practice when dealing with longer periods.

Loss/Profit Data

When estimating VaR and ES, it is sometimes more convenient to deal with data in loss/profit (L/P) form. L/P data are a simple transformation of P/L data:

$$L/P_t = -P/L_t \tag{1.4}$$

L/P observations assign a positive value to losses and a negative value to profits, and we will call these L/P data 'losses' for short. Dealing with losses is sometimes a little more convenient for risk measurement purposes because the risk measures are themselves denominated in loss terms.

Arithmetic Return Data

Data can also come in the form of arithmetic (or simple) returns. The arithmetic return r_t is defined as:

$$r_t = \frac{P_t + D_t - P_{t-1}}{P_{t-1}} = \frac{P_t + D_t}{P_{t-1}} - 1 \tag{1.5}$$

which is the same as the P/L over period t divided by the value of the asset at the end of $t-1$.

In using arithmetic returns, we implicitly assume that the interim payment D_t does not earn any return of its own. However, this assumption will seldom be appropriate over long periods because interim income is usually reinvested. Hence, arithmetic returns should not be used when we are concerned with long horizons.

Geometric Return Data

Returns can also be expressed in geometric (or compound) form. The geometric return R_t is

$$R_t = \ln\left(\frac{P_t + D_t}{P_{t-1}}\right) \tag{1.6}$$

The geometric return implicitly assumes that interim payments are continuously reinvested. The geometric return is often more economically meaningful than the arithmetic return, because it ensures that the asset price (or portfolio value) can never become negative regardless of how negative the returns might be. With arithmetic returns, on the other hand, a very low realized return—or a high

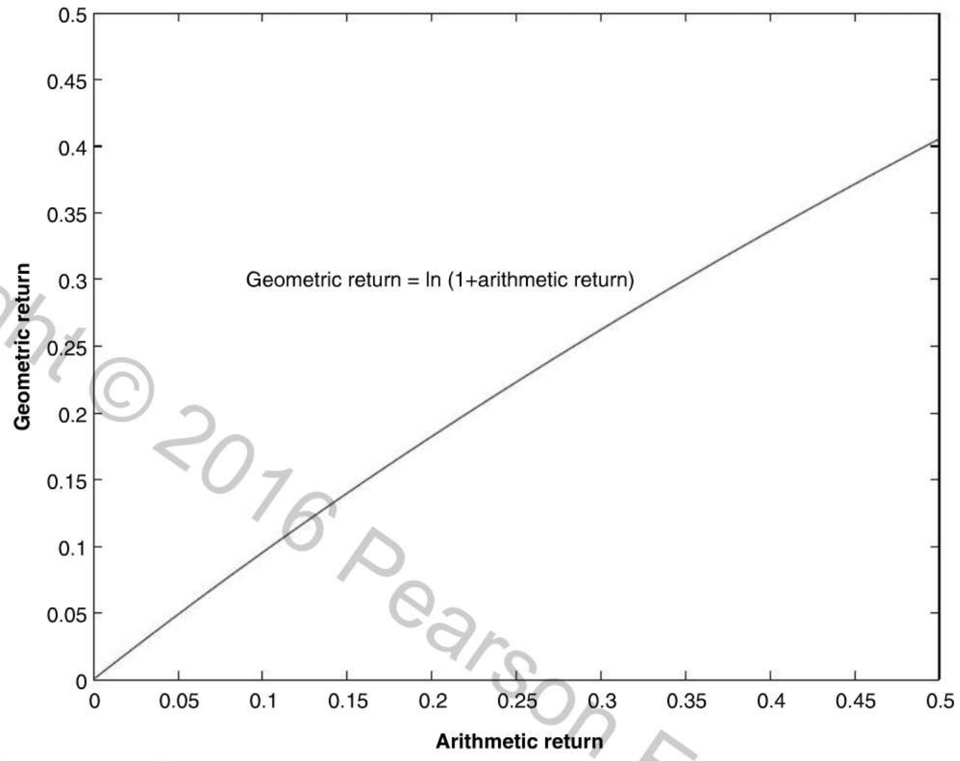

FIGURE 1-1 Geometric and arithmetic returns.

loss—implies that the asset value P_t can become negative, and a negative asset price seldom makes economic sense.[1]

The geometric return is also more convenient. For example, if we are dealing with foreign currency positions, geometric returns will give us results that are independent of the reference currency. Similarly, if we are dealing with multiple periods, the geometric return over those periods is the sum of the one-period geometric returns. Arithmetic returns have neither of these convenient properties.

The relationship of the two types of return can be seen by rewriting Equation (1.6) (using a Taylor's series expansion for the natural log) as:

$$R_t = \ln\left(\frac{P_t + D_t}{P_{t-1}}\right) = \ln(1 + r_t) = r_t - \frac{1}{2}r_t^2 + \frac{1}{3}r_t^3 - \cdots \quad \textbf{(1.7)}$$

from which we can see that $R_t \approx r_t$ provided that returns are 'small'. This conclusion is illustrated by Figure 1-1, which plots the geometric return R_t against its arithmetic counterpart r_t. The difference between the two returns is negligible when both returns are small, but the difference grows as the returns get bigger—which is to be expected, as the geometric return is a log function of the arithmetic return. Since we would expect returns to be low over short periods and higher over longer periods, the difference between the two types of return is negligible over short periods but potentially substantial over longer ones. And since the geometric return takes account of earnings on interim income, and the arithmetic return does not, we should always use the geometric return if we are dealing with returns over longer periods.

[1] This is mainly a point of principle rather than practice. In practice, any distribution we fit to returns is only likely to be an approximation, and many distributions are ill-suited to extreme returns anyway.

Example 1.1 Arithmetic and Geometric Returns

If arithmetic returns r_t over some period are 0.05, Equation (1.7) tells us that the corresponding geometric returns are $R_t = \ln(1 + r_t) = \ln(1.05) = 0.0488$. Similarly, if geometric returns R_t are 0.05, Equation (1.7) implies that arithmetic returns are $1 + r_t = \exp(R_t) \Rightarrow r_t = \exp(R_t) - 1 = \exp(0.05) - 1 = 0.0513$. In both cases the arithmetic return is close to, but a little higher than, the geometric return—and this makes intuitive sense when one considers that the geometric return compounds at a faster rate.

ESTIMATING HISTORICAL SIMULATION VaR

The simplest way to estimate VaR is by means of historical simulation (HS). The HS approach estimates VaR by means of ordered loss observations.

Suppose we have 1000 loss observations and are interested in the VaR at the 95% confidence level. Since the confidence level implies a 5% tail, we know that there are 50 observations in the tail, and we can take the VaR to be the 51st highest loss observation.[2]

We can estimate the VaR on a spreadsheet by ordering our data and reading off the 51st largest observation from the spreadsheet. We can also estimate it more directly by using the 'Large' command in Excel, which gives us the *k*th largest value in an array. Thus, if our data are an array called 'Loss_data', then our VaR is given by the Excel command 'Large(Loss_data,51)'. If we are using MATLAB, we first order the loss/profit data using the 'Sort()' command (i.e., by typing 'Loss_data = Sort(Loss_data)'); and then derive the VaR by typing in 'Loss_data(51)' at the command line.

More generally, if we have *n* observations, and our confidence level is α, we would want the $(1 - \alpha)$. $n + 1$th highest observation, and we would use the commands 'Large(Loss_data,(1 − *alpha*)**n* + 1)' using Excel, or 'Loss_data((1 − *alpha*)**n* + 1)' using MATLAB, provided in the latter case that our 'Loss_data' array is already sorted into ordered observations.[3]

[2] In theory, the VaR is the quantile that demarcates the tail region from the non-tail region, where the size of the tail is determined by the confidence level, but with finite samples there is a certain level of arbitrariness in how the ordered observations relate to the VaR itself—that is, do we take the VaR to be the 50th observation, the 51st observation, or some combination of them? However, this is just an issue of approximation, and taking the VaR to be the 51st highest observation is not unreasonable.

[3] We can also estimate HS VaR using percentile functions such as the 'Percentile' function in Excel or the 'prctile' function in MATLAB. However, such functions are less transparent (i.e., it is not obvious to the reader how the percentiles are calculated), and the Excel percentile function can be unreliable.

An example of an HS VaR is given in Figure 1-2. This figure shows the histogram of 1000 hypothetical loss observations and the 95%VaR. The figure is generated using the 'hsvarfigure' command in the MMR Toolbox. The VaR is 1.704 and separates the top 5% from the bottom 95% of loss observations.

In practice, it is often helpful to obtain HS VaR estimates from a cumulative histogram, or empirical cumulative frequency function. This is a plot of the ordered loss observations against their empirical cumulative frequency (e.g., so if there are *n* observations in total, the empirical cumulative frequency of the *i*th such ordered observation is *i/n*). The empirical cumulative frequency function of our earlier data set is shown in Figure 1-3. The empirical frequency function makes it very easy to obtain the VaR: we simply move up the cumulative frequency axis to where the cumulative frequency equals our confidence level, draw a horizontal line along to the curve, and then draw a vertical line down to the *x*-axis, which gives us our VaR.

ESTIMATING PARAMETRIC VaR

We can also estimate VaR using parametric approaches, the distinguishing feature of which is that they require us to explicitly specify the statistical distribution from which our data observations are drawn. We can also think of parametric approaches as fitting curves through the data and then reading off the VaR from the fitted curve.

In making use of a parametric approach, we therefore need to take account of both the statistical distribution and the type of data to which it applies.

Estimating VaR with Normally Distributed Profits/Losses

Suppose that we wish to estimate VaR under the assumption that P/L is normally distributed. In this case our VaR at the confidence level α is:

$$\alpha VaR = -\mu_{P/L} + \sigma_{P/L} z_{\alpha} \tag{1.8}$$

where z_{α} is the standard normal variate corresponding to α, and $\mu_{P/L}$ and $\sigma_{P/L}$ are the mean and standard deviation of P/L. Thus, z_{α} is the value of the standard normal variate such that α of the probability density mass lies to its left,

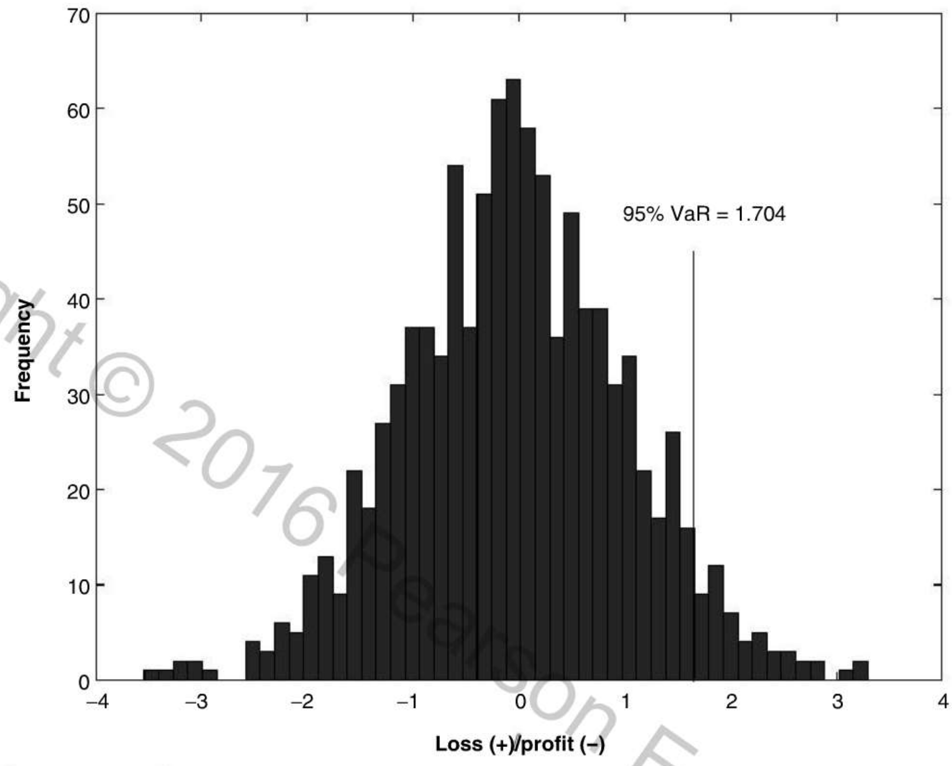

FIGURE 1-2 Historical simulation VaR.

Note: Based on 1000 random numbers drawn from a standard normal L/P distribution, and estimated with 'hsvarfigure' function.

and $1-\alpha$ of the probability density mass lies to its right. For example, if our confidence level is 95%, $z_\alpha = z_{0.95}$ will be 1.645.

In practice, $\mu_{P/L}$ and $\sigma_{P/L}$ would be unknown, and we would have to estimate VaR based on estimates of these parameters. Our VaR estimate, αVaR^e, would then be:

$$\alpha VaR^e = -m_{P/L} + s_{P/L} z_\alpha \tag{1.9}$$

where $m_{P/L}$ and $s_{P/L}$ are estimates of the mean and standard deviation of P/L.

Figure 1-4 shows the 95% VaR for a normally distributed P/L with mean 0 and standard deviation 1. Since the data are in P/L form, the VaR is indicated by the negative of the cut off point between the lower 5% and the upper 95% of P/L observations. The actual VaR is the negative of -1.645, and is therefore 1.645.

If we are working with normally distributed L/P data, then $\mu_{L/P} = -\mu_{P/L}$ and $\sigma_{L/P} = \sigma_{P/L}$, and it immediately follows that:

$$\alpha VaR = \mu_{L/P} + \sigma_{L/P} z_\alpha \tag{1.10a}$$

$$\alpha VaR^e = m_{L/P} + s_{L/P} z_\alpha \tag{1.10b}$$

Figure 1-5 illustrates the corresponding VaR. This figure gives the same information as Figure 1-4, but is a little more straightforward to interpret because the VaR is defined in units of losses (or 'lost money') rather than P/L. In this case, the VaR is given by the point on the *x*-axis that cuts off the top 5% of the pdf mass from the bottom 95% of pdf mass. If we prefer to work with the cumulative density function, the VaR is the *x*-value that corresponds to a cdf value of 95%. Either way, the VaR is again 1.645, as we would (hopefully) expect.

Example 1.2 VaR with Normal P/L

If P/L over some period is normally distributed with mean 10 and standard deviation 20, then (by Equation (1.8)) the 95% VaR is $-10 + 20z_{0.95} = -10 + 20 \times 1.645 = 22.9$. The corresponding 99% VaR is $-10 + 20z_{0.99} = -10 + 20 \times 2.326 = 36.52$.

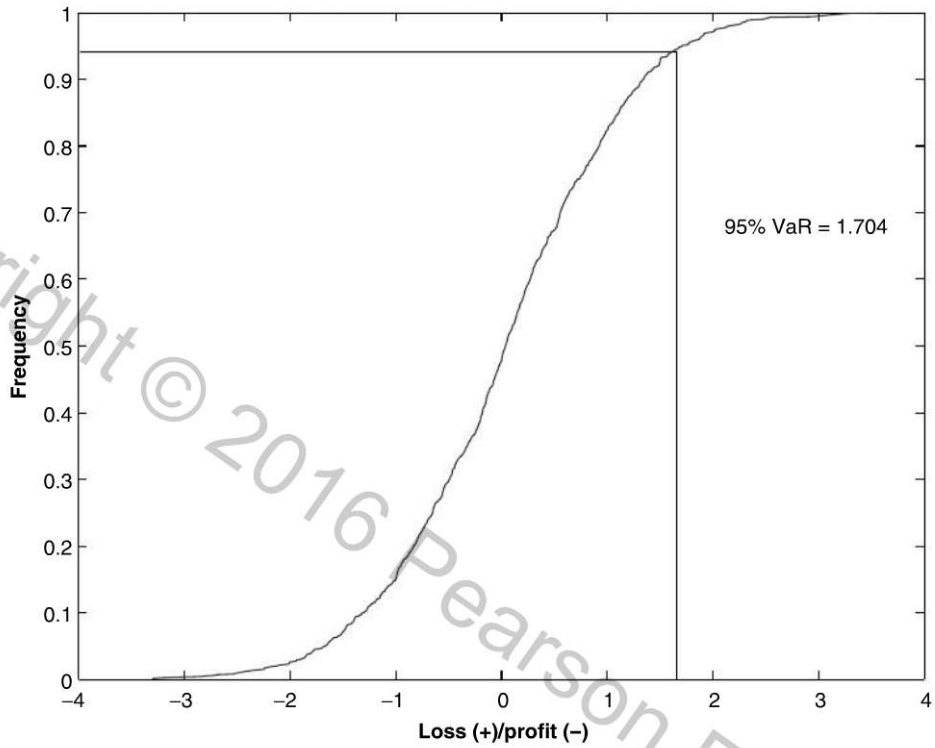

FIGURE 1-3 Historical simulation via an empirical cumulative frequency function.

Note: Based on the same data as Figure 1-2.

Estimating VaR with Normally Distributed Arithmetic Returns

We can also estimate VaR making assumptions about returns rather than P/L. Suppose then that we assume that arithmetic returns are normally distributed with mean μ_r and standard deviation σ_r. To derive the VaR, we begin by obtaining the critical value of r_t, r^*, such that the probability that r_t exceeds r^* is equal to our confidence level α. r^* is therefore:

$$r^* = \mu_r - \sigma_r z_\alpha \tag{1.11}$$

Since the actual return r_t is the loss/profit divided by the earlier asset value, P_{t-1}, it follows that:

$$r_t = \frac{P_t - P_{t-1}}{P_{t-1}} = -\frac{Loss_t}{P_{t-1}} \tag{1.12}$$

Substituting r^* for r_t then gives us the relationship between r^* and the VaR:

$$r_t^* = \frac{P_t^* - P_{t-1}}{P_{t-1}} = \frac{VaR}{P_{t-1}} \tag{1.13}$$

Substituting Equation (1.11) into Equation (1.13) and rearranging then gives us the VaR itself:

$$\alpha VaR = -(\mu_r - \sigma_r z_\alpha)P_{t-1} \tag{1.14}$$

Equation (1.14) will give us equivalent answers to our earlier VaR equations. For example, if we set $\alpha = 0.95$, $\mu_r = 0$, $\sigma_r = 1$ and $P_{t-1} = 1$, which correspond to our earlier illustrative P/L and L/P parameter assumptions, αVaR is 1.645: the three approaches give the same results, because all three sets of underlying assumptions are equivalent.

Example 1.3 VaR with Normally Distributed Arithmetic Returns

Suppose arithmetic returns r_t over some period are distributed as normal with mean 0.1 and standard deviation 0.25, and we have a portfolio currently worth 1. Then (by Equation (1.14)) the 95% VaR is $-0.1 + 0.25 \times 1.645 = 0.331$, and the 99% VaR is $-0.1 + 0.25 \times 2.326 = 0.482$.

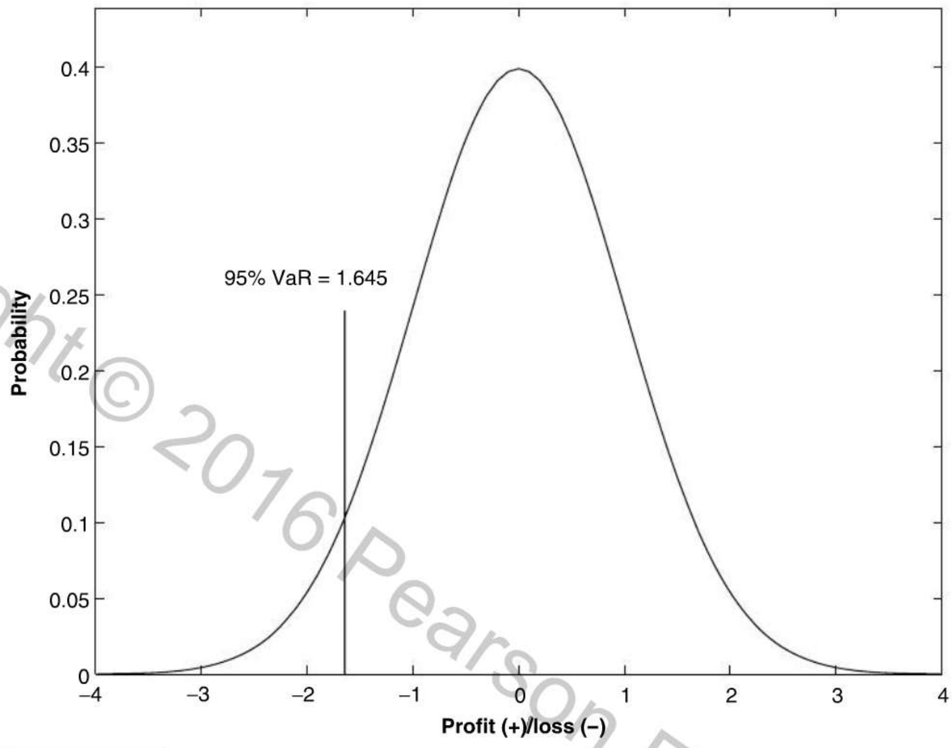

FIGURE 1-4 VaR with standard normally distributed profit/loss data.

Note: Obtained from Equation (1.9) with $\mu_{P/L} = 0$ and $\sigma_{P/L} = 1$. Estimated with the 'normalvarfigure' function.

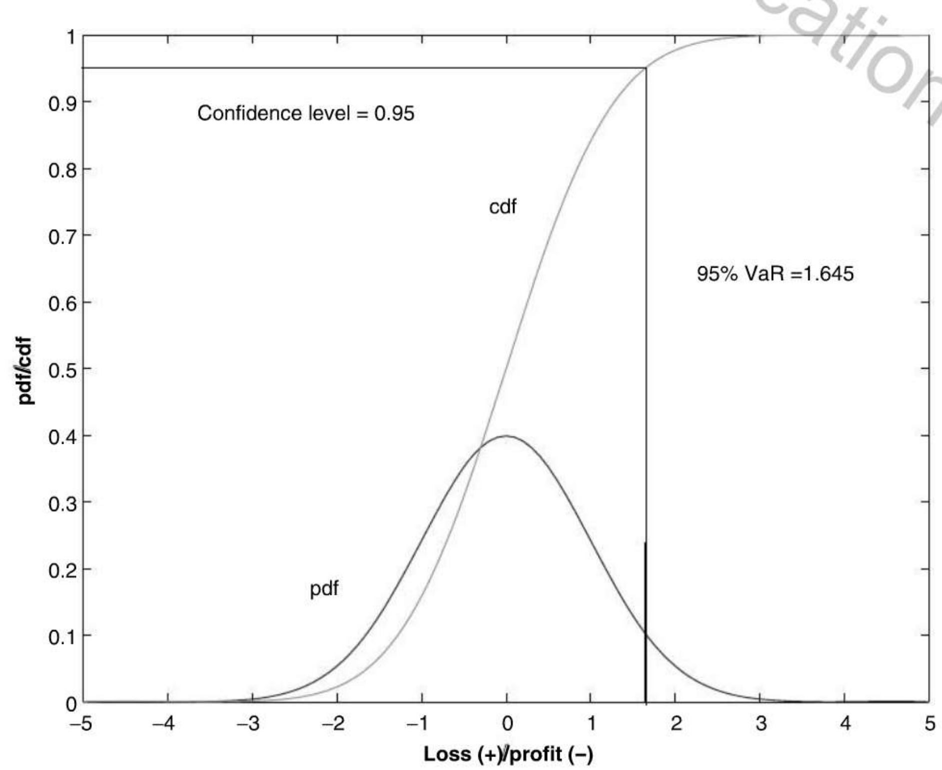

FIGURE 1-5 VaR with normally distributed loss/profit data.

Note: Obtained from Equation (1.10a) with $\mu_{L/P} = 0$ and $\sigma_{L/P} = 1$.

Estimating Lognormal VaR

Each of the previous approaches assigns a positive probability of the asset value, P_t, becoming negative, but we can avoid this drawback by working with geometric returns. Now assume that geometric returns are normally distributed with mean μ_R and standard deviation σ_R. If D_t is zero or reinvested continually in the asset itself (e.g., as with profits reinvested in a mutual fund), this assumption implies that the natural logarithm of P_t is normally distributed, or that P_t itself is lognormally distributed. The lognormal distribution is explained in Box 1-1, and a lognormal asset price is shown in Figure 1-6: observe that the price is always non-negative, and its distribution is skewed with a long right-hand tail.

Since the VaR is a loss, and since the loss is the difference between P_t (which is random) and P_{t-1} (which we can take here as given), then the VaR itself has the same distribution as P_t. Normally distributed geometric returns imply that the VaR is lognormally distributed.

If we proceed as we did earlier with the arithmetic return, we begin by deriving the critical value of R, R^*, such that the probability that $R > R^*$ is α:

$$R^* = \mu_R - \sigma_R z_\alpha \tag{1.15}$$

We then use the definition of the geometric return to unravel the critical value of P^* (i.e., the value of P_t corresponding to a loss equal to our VaR), and thence infer our VaR:

$$\begin{aligned} R^* &= \ln(P^*/P_{t-1}) = \ln P^* - \ln P_{t-1} \\ &\Rightarrow \ln P^* = R^* + \ln P_{t-1} \\ &\Rightarrow P^* = P_{t-1}\exp[R^*] = P_{t-1}\exp[\mu_R - \sigma_R z_\alpha] \\ &\Rightarrow \alpha VaR = P_{t-1} - P^* = P_{t-1}(1 - \exp[\mu_R - \sigma_R z_\alpha]) \end{aligned} \tag{1.16}$$

This gives us the lognormal VaR, which is consistent with normally distributed geometric returns.

The lognormal VaR is illustrated in Figure 1-7, based on the standardised (but typically unrealistic) assumptions that $\mu_R = 0$, $\sigma_R = 1$, and $P_{t-1} = 1$. In this case, the VaR at the 95% confidence level is 0.807. The figure also shows that the distribution of L/P is a reflection of the distribution of P_t shown earlier in Figure 1-6.

BOX 1-1 The Lognormal Distribution

A random variate X is said to be lognormally distributed if the natural log of X is normally distributed. The lognormal distribution can be specified in terms of the mean and standard deviation of ln X. Call these parameters μ and σ. The lognormal is often also represented in terms of m and σ, where m is the median of x, and $m = \exp(\mu)$.

The pdf of X can written:

$$\phi(x) = \frac{1}{x\sigma\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{\log(x)-\mu}{\sigma}\right)^2\right\}$$

for $x > 0$. Thus, the lognormal pdf is only defined for positive values of x and is skewed to the right as in Figure 1-6.

Let $\omega = \exp(\sigma^2)$ for convenience. The mean and variance of the lognormal can be written as:

$$\text{mean} = m\exp(\sigma^2/2) \quad \text{and} \quad \text{variance} = m^2\omega(\omega - 1)$$

Turning to higher moments, the skewness of the lognormal is

$$skewness = (\omega + 2)(\omega - 1)^{1/2}$$

and is always positive, which confirms the lognormal has a long right-hand tail. The kurtosis of the lognormal is

$$kurtosis = \omega^4 + 2\omega^3 + 3\omega^2 - 3$$

and therefore varies from a minimum of (just over) 3 to a potentially large value depending on the value of s.

Example 1.4 Lognormal VaR

Suppose that geometric returns R_t over some period are distributed as normal with mean 0.05, standard deviation 0.20, and we have a portfolio currently worth 1. Then (by Equation (1.16)) the 95% VaR is $1 - \exp(0.05 - 0.20 \times 1.645) = 0.244$. The corresponding 99% VaR is $1 - \exp(0.05 - 0.20 \times 2.326) = 0.340$. Observe that these VaRs are quite close to those obtained in Example 1.3, where the arithmetic return parameters were the same as the geometric return parameters assumed here.

Example 1.5 Lognormal VaR vs Normal VaR

Suppose that we make the empirically not too unrealistic assumptions that the mean and volatility of annualised returns are 0.10 and 0.40. We are interested in the 95% VaR at the 1-day holding period for a portfolio worth \$1. Assuming 250 trading days to a year, the daily return has a mean $0.1/250 = 0.00040$ and standard deviation $0.40/\sqrt{250} = 0.0253$. The normal 95% VaR is $-0.0004 + 0.0253 \times 1.645 = 0.0412$. If we assume a lognormal, then

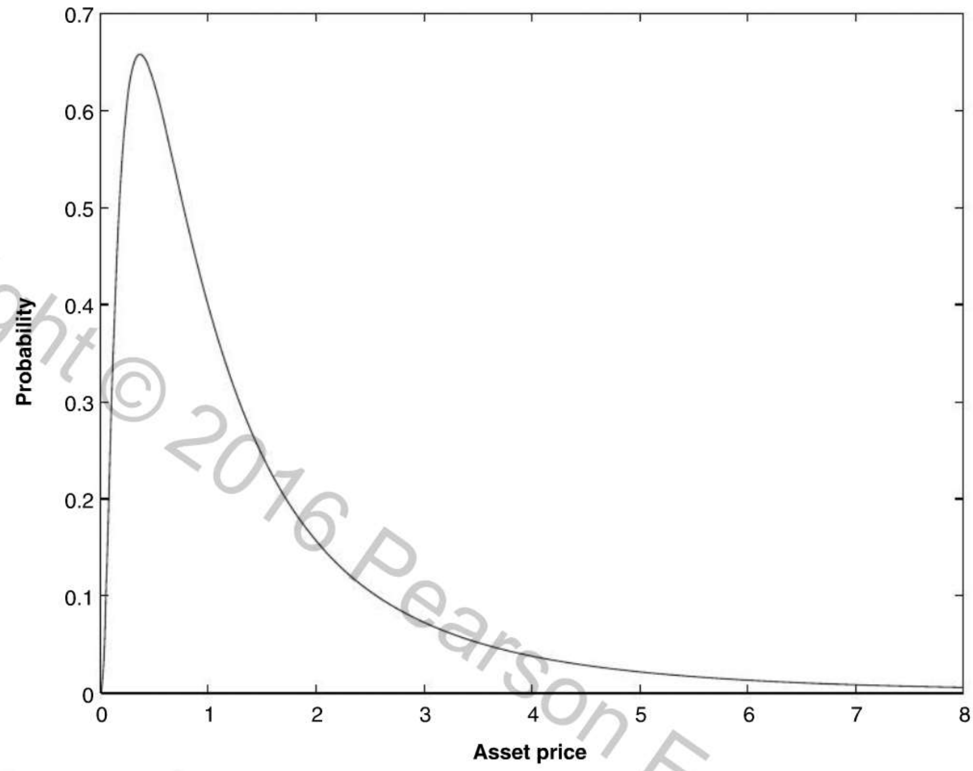

FIGURE 1-6 A lognormally distributed asset price.

Note: Estimated using the 'lognpdf' function in the Statistics Toolbox.

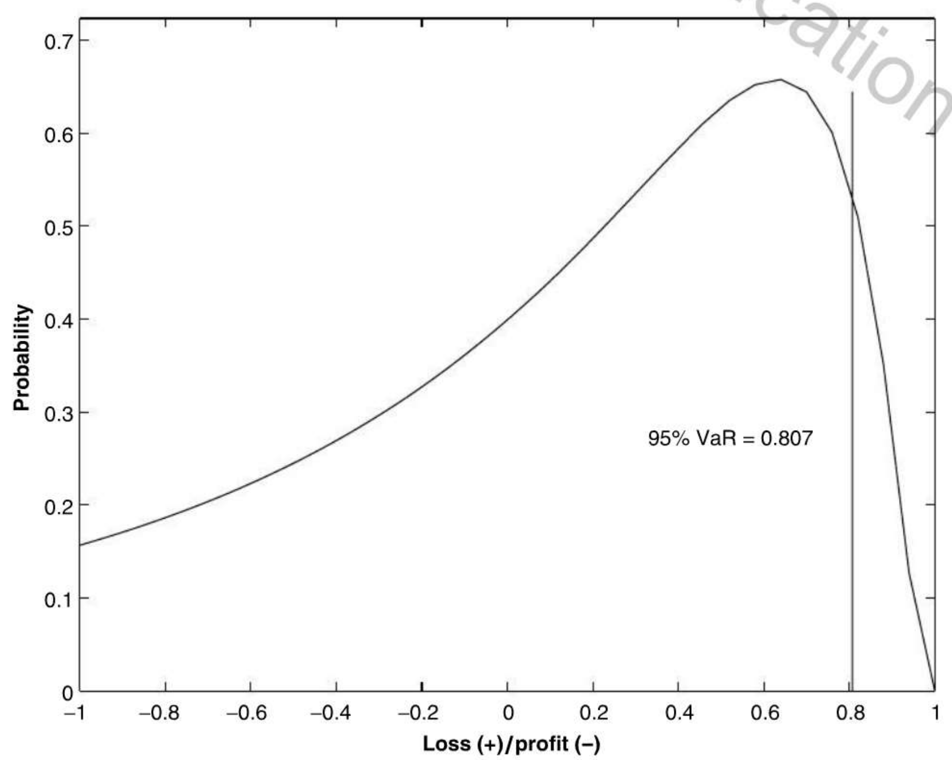

FIGURE 1-7 Lognormal VaR.

Note: Estimated assuming the mean and standard deviation of geometric returns are 0 and 1, and for an initial investment of 1. The figure is produced using the 'lognormalvarfigure' function.

the 95% VaR is $1 - \exp(0.0004 - 0.0253 \times 1.645) = 0.0404$. The normal VaR is 4.12% and the lognormal VaR is 4.04% of the value of the portfolio. This illustrates that normal and lognormal VaRs are much the same if we are dealing with short holding periods and realistic return parameters.

ESTIMATING COHERENT RISK MEASURES

Estimating Expected Shortfall

We turn now to the estimation of coherent risk measures, and the easiest of these to estimate is the expected shortfall (ES). The ES is the probability-weighted average of tail losses, and a normal ES is illustrated in Figure 1-8. In this case, the 95% ES is 2.063, corresponding to our earlier normal 95% VaR of 1.645.

The fact that the ES is a probability-weighted average of tail losses suggests that we can estimate ES as an average of 'tail VaRs'.[4] The easiest way to implement this approach is to slice the tail into a large number n of slices, each of which has the same probability mass, estimate the VaR associated with each slice, and take the ES as the average of these VaRs.

To illustrate the method, suppose we wish to estimate a 95% ES on the assumption that losses are normally distributed with mean 0 and standard deviation 1. In practice, we would use a high value of n and carry out the calculations on a spreadsheet or using appropriate software. However, to show the procedure manually, let us work with a very small n value of 10. This value gives us

[4] The obvious alternative is to seek a 'closed-form' solution, which we could use to estimate the ES, but ES formulas seem to be known only for a limited number of parametric distributions (e.g., elliptical, including normal, and generalised Pareto distributions), whereas the 'average-tail-VaR' method is easy to implement and can be applied to any 'well-behaved' ESs that we might encounter, parametric or otherwise.

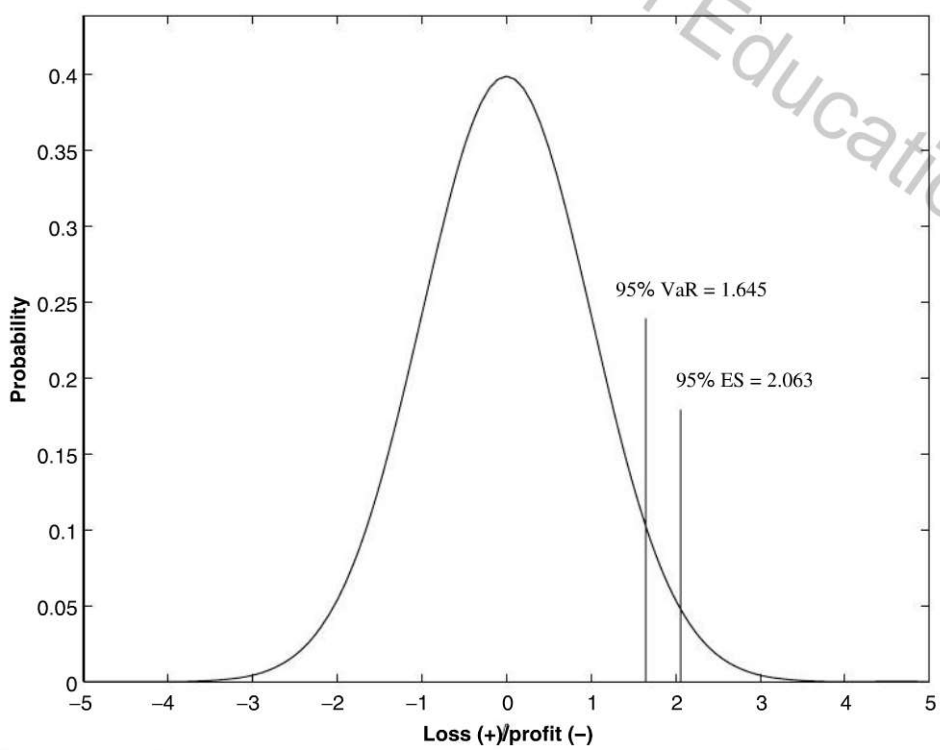

FIGURE 1-8 Normal VaR and ES.

Note: Estimated with the mean and standard deviation of P/L equal to 0 and 1 respectively, using the 'normalesfigure' function.

9 (i.e., $n - 1$) tail VaRs, or VaRs at confidence levels in excess of 95%. These VaRs are shown in Table 1-1, and vary from 1.6954 (for the 95.5% VaR) to 2.5758 (for the 99.5% VaR). Our estimated ES is the average of these VaRs, which is 2.0250.

Of course, in using this method for practical purposes, we would want a value of n large enough to give accurate results. To give some idea of what this might be, Table 1-2 reports some alternative ES estimates obtained using this procedure with varying values of n. These results show that the estimated ES rises with n, and gradually converges to the true value of 2.063. These results also show that our ES estimation procedure seems to be reasonably accurate even for quite small values of n. Any decent computer should therefore be able to produce accurate ES estimates quickly in real time.

TABLE 1-1 Estimating ES as a Weighted Average of Tail VaRs

Confidence level	Tail VaR
95.5%	1.6954
96.0%	1.7507
96.5%	1.8119
97.0%	1.8808
97.5%	1.9600
98.0%	2.0537
98.5%	2.1701
99.0%	2.3263
99.5%	2.5738
Average of tail VaRs	2.0250

Note: VaRs estimated assuming the mean and standard deviation of losses are 0 and 1, using the 'normalvar' function in the MMR Toolbox.

TABLE 1-2 ES Estimates as a Function of the Number of Tail Slices

Number of tail slices (n)	ES
10	2.0250
25	2.0433
50	2.0513
100	2.0562
250	2.0597
500	2.0610
1000	2.0618
2500	2.0623
5000	2.0625
10 000	2.0626
True value	2.0630

Note: VaRs estimated assuming the mean and standard deviation of losses are 0 and 1.

Estimating Coherent Risk Measures

Other coherent risk measures can be estimated using modifications of this 'average VaR' method. Recall that a coherent risk measure is a weighted average of the quantiles (denoted by q_p) of our loss distribution:

$$M_\phi = \int_0^1 \phi(p) q_p dp \tag{1.17}$$

where the weighting function or risk-aversion function $\phi(p)$ is specified by the user. The ES gives all tail-loss quantiles an equal weight, and other quantiles a weight of 0. Thus the ES is a special case of M_ϕ obtained by setting $\phi(p)$ to the following:

$$\phi(p) = \begin{cases} 0 & \text{if } p < \alpha \\ 1/(1-\alpha) & \quad p \geq \alpha \end{cases} \tag{1.18}$$

The more general coherent risk measure, M_ϕ, involves a potentially more sophisticated weighting function $\phi(p)$. We can therefore estimate any of these measures by replacing the equal weights in the 'average VaR' algorithm with the $\phi(p)$ weights appropriate to the risk measure being estimated.

To show how this might be done, suppose we have the exponential weighting function:

$$\phi_\gamma(p) = \frac{e^{-(1-p)/\gamma}}{\gamma(1-e^{-1/\gamma})} \tag{1.19}$$

and we believe that we can represent the degree of our risk-aversion by setting $\gamma = 0.05$. To illustrate the procedure manually, we continue to assume that losses are standard normally distributed and we set $n = 10$ (i.e., we divide the complete losses density function into 10 equal-probability slices). With $n = 10$, we have $n - 1 = 9$ (i.e., $n - 1$) loss quantiles or VaRs spanning confidence levels from 0.1 to 0.90. These VaRs are shown in the second column of Table 1-3, and vary from −1.2816 (for the 10% VaR) to 1.2816 (for the 90% VaR). The third column shows the $\phi(p)$ weights corresponding to each confidence level, and the fourth column shows the products of each VaR and corresponding weight. Our estimated exponential spectral risk measure is the $\phi(p)$-weighted average of the VaRs, and is therefore equal to 0.4228.

As when estimating the ES earlier, when using this type of routine in practice we would want a value of n large enough to give accurate results. Table 1-4 reports some alternative estimates obtained using this procedure with increasing values of n. These results show

TABLE 1-3 Estimating Exponential Spectral Risk Measure as a Weighted Average of VaRs

Confidence level (α)	αVaR	Weight $\phi(\alpha)$	$\phi(\alpha) \times \alpha VaR$
10%	−1.2816	0	0.0000
20%	−0.8416	0	0.0000
30%	−0.5244	0	0.0000
40%	−0.2533	0.0001	0.0000
50%	0	0.0009	0.0000
60%	0.2533	0.0067	0.0017
70%	0.5244	0.0496	0.0260
80%	0.8416	0.3663	0.3083
90%	1.2816	2.7067	3.4689
Risk measure = mean ($\phi(\alpha)$ times αVaR) =			0.4226

Note: VaRs estimated assuming the mean and standard deviation of losses are 0 and 1, using the 'normalvar' function in the MMR Toolbox. The weights $\phi(\alpha)$ are given by the exponential function (Equation (1.19)) with $\gamma = 0.05$.

TABLE 1-4 Estimates of Exponential Spectral Coherent Risk Measure as a Function of the Number of Tail Slices

Number of tail slices	Estimate of exponential spectral risk measure
10	0.4227
50	1.3739
100	1.5853
250	1.7338
500	1.7896
1000	1.8197
2500	1.8392
5000	1.8461
10 000	1.8498
50 000	1.8529
100 000	1.8533
500 000	1.8536

Note: VaRs estimated assuming the mean and standard deviation of losses are 0 and 1, using the 'normalvar' function in the MMR Toolbox. The weights $\phi(\alpha)$ are given by the exponential function (Equation (1.19)) with $\gamma = 0.05$.

that the estimated risk measure rises with n, and gradually converges to a value in the region of about 1.854. The estimates in this table indicate that we may need a considerably larger value of n than we did earlier to get results of the same level of accuracy. Even so, a good computer should still be able to produce accurate estimates of spectral risk measures fairly quickly.

TABLE 1-5 Estimated Risk Measures and Halving Errors

Number of tail slices	Estimated spectral risk measure	Halving error
100	1.5853	0.2114
200	1.7074	0.1221
400	1.7751	0.0678
800	1.8120	0.0368
1600	1.8317	0.0197
3200	1.8422	0.0105
6400	1.8477	0.0055
12 800	1.8506	0.0029
25 600	1.8521	0.0015
51 200	1.8529	0.0008

Note: VaRs estimated assuming the mean and standard deviation of losses are 0 and 1, using the 'normalvar' function in the MMR Toolbox. The weights $\phi(\alpha)$ are given by the exponential function (Equation (1.19)) with $\gamma = 0.05$.

When estimating ES or more general coherent risk measures in practice, it also helps to have some guidance on how to choose the value of n. Granted that the estimate does eventually converge to the true value as n gets large, one useful approach is to start with some small value of n, and then double n repeatedly until we feel the estimates have settled down sufficiently. Each time we do so, we halve the width of the discrete slices, and we can monitor how this 'halving' process affects our estimates. This suggests that we look at the 'halving error' ε_n given by:

$$\varepsilon_n = \hat{M}^{(n)} - \hat{M}^{(n/2)} \tag{1.20}$$

where $\hat{M}^{(n)}$ is our estimated risk measure based on n slices. We stop doubling n when ε_n falls below some tolerance level that indicates an acceptable level of accuracy. The process is shown in Table 1-5. Starting from an arbitrary value of 100, we repeatedly double n (so it becomes 200, 400, 800, etc.). As we do so, the estimated risk measure gradually converges, and the halving error gradually falls. So, for example, for $n = 6400$, the estimated risk measure is 1.8477, and the halving error is 0.0055. If we double n to 12,800, the estimated risk measure becomes 1.8506, and the halving error falls to 0.0029, and so on.

However, this 'weighted average quantile' procedure is rather crude, and (bearing in mind that the risk measure (Equation (1.17)) involves an integral) we can in principle expect to get substantial improvements in accuracy if we resorted to more 'respectable' numerical integration or quadrature methods. This said, the crude 'weighted average quantile' method actually seems to perform well for spectral exponential risk measures when compared against some of these alternatives, so one is not necessarily better off with the more sophisticated methods.[5]

Thus, the key to estimating any coherent risk measure is to be able to estimate quantiles or VaRs: the coherent risk measures can then be obtained as appropriately weighted averages of quantiles. From a practical point of view, this is extremely helpful as all the building blocks that go into quantile or VaR estimation—databases, calculation routines, etc.—are exactly what we need for the estimation of coherent risk measures as well. If an institution already has a VaR engine, then that engine needs only small adjustments to produce estimates of coherent risk measures: indeed, in many cases, all that needs changing is the last few lines of code in a long data processing system. The costs of switching from VaR to more sophisticated risk measures are therefore very low.

ESTIMATING THE STANDARD ERRORS OF RISK MEASURE ESTIMATORS

We should always bear in mind that any risk measure estimates that we produce are just that—estimates. We never know the true value of any risk measure, and an estimate is only as good as its precision: if a risk measure is very imprecisely estimated, then the estimator is virtually worthless, because its imprecision tells us that true value could be almost anything; on the other hand, if we know that an estimator is fairly precise, we can be confident that the true value is fairly close to the estimate, and the estimator has some value. Hence, we should always seek to supplement any risk estimates we produce with some indicator of their precision. This is a fundamental principle of good risk measurement practice.

[5] There is an interesting reason for this: the spectral weights give the highest loss the highest weight, whereas the quadrature methods such as the trapezoidal and Simpson's rules involve algorithms in which the two most extreme quantiles have their weights specifically cut, and this undermines the accuracy of the algorithm relative to the crude approach. However, there are ways round these sorts of problems, and in principle versions of the sophisticated approaches should give better results.

We can evaluate the precision of estimators of risk measures by means of their standard errors, or (generally better) by producing confidence intervals for them. In this chapter we focus on the more basic indicator, the standard error of a risk measure estimator.

Standard Errors of Quantile Estimators

We first consider the standard errors of quantile (or VaR) estimators. Following Kendall and Stuart,[6] suppose we have a distribution (or cumulative density) function $F(x)$, which might be a parametric distribution function or an empirical distribution function (i.e., a cumulative histogram) estimated from real data. Its corresponding density or relative-frequency function is $f(x)$. Suppose also that we have a sample of size n, and we select a bin width h. Let dF be the probability that $(k-1)$ observations fall below some value $q-h/2$, that one observation falls in the range $q \pm h/2$, and that $(n-k)$ observations are greater than $q+h/2$. dF is proportional to

$$\{F(q)\}^{k-1}f(q)dq\{1-F(q)\}^{n-k} \tag{1.21}$$

This gives us the frequency function for the quantile q not exceeded by a proportion k/n of our sample, i.e., the $100(k/n)$th percentile.

If this proportion is p, Kendall and Stuart show that Equation (1.21) is approximately equal to $p^{np}(1-p)^{n(1-p)}$ for large values of n. If ε is a very small increment to p, then

$$p^{np}(1-p)^{n(1-p)} \approx (p+\varepsilon)^{np}(1-p-\varepsilon)^{n(1-p)} \tag{1.22}$$

Taking logs and expanding, Equation (1.22) is itself approximately

$$(p+\varepsilon)^{np}(1-p-\varepsilon)^{n(1-p)} \approx -\frac{n\varepsilon^2}{2p(1-p)} \tag{1.23}$$

which implies that the distribution function dF is approximately proportional to

$$\exp\left(\frac{-n\varepsilon^2}{2p(1-p)}\right) \tag{1.24}$$

Integrating this out,

$$dF = \frac{1}{\sqrt{2\pi}\sqrt{p(1-p)/n}}\exp\left(\frac{-n\varepsilon^2}{2p(1-p)}\right)d\varepsilon \tag{1.25}$$

[6] Kendall and Stuart (1972), pp. 251–252.

which tells us that ε is normally distributed in the limit with variance $p(1-p)/n$. However, we know that $d\varepsilon = dF(q) = f(q)dq$, so the variance of q is

$$var(q) \approx \frac{p(1-p)}{n[f(q)]^2} \quad \textbf{(1.26)}$$

This gives us an approximate expression for the variance, and hence its square root, the standard error, of a quantile estimator q.

This expression shows that the quantile standard error depends on p, the sample size n and the pdf value $f(q)$. The way in which the (normal) quantile standard errors depend on these parameters is apparent from Figure 1-9. This shows that:

- The standard error falls as the sample size n rises.
- The standard error rises as the probabilities become more extreme and we move further into the tail—hence, the more extreme the quantile, the less precise its estimator.

In addition, the quantile standard error depends on the probability density function $f(.)$—so the choice of density function can make a difference to our estimates—and also on the bin width h, which is essentially arbitrary.

The standard error can be used to construct confidence intervals around our quantile estimates in the usual textbook way. For example, a 90% confidence interval for a quantile q is given by

$$\begin{aligned}&[q - 1.645se(q), q + 1.645se(q)] \\ &= \left[q - 1.645\frac{\sqrt{p(1-p)/n}}{f(q)}, q + 1.645\frac{\sqrt{p(1-p)/n}}{f(q)}\right]\end{aligned} \quad \textbf{(1.27)}$$

Example 1.6 Obtaining VaR Confidence Intervals Using Quantile Standard Errors

Suppose we wish to estimate the 90% confidence interval for a 95% VaR estimated on a sample of size of $n = 1000$ to be drawn from a standard normal distribution, based on an assumed bin width $h = 0.1$.

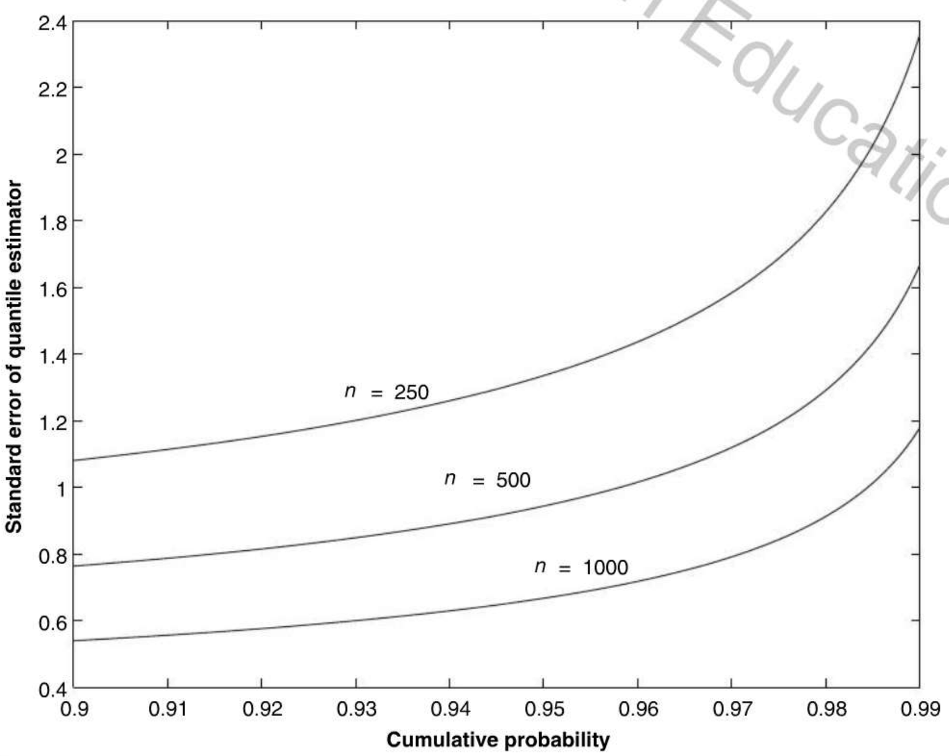

FIGURE 1-9 Standard errors of quantile estimators.

Note: Based on random samples of size n drawn from a standard normal distribution. The bin width h is set to 0.1.

We know that the 95% VaR of a standard normal is 1.645. We take this to be q in Equation (1.27), and we know that q falls in the bin spanning $1.645 \pm 0.1/2 = [1.595, 1.695]$. The probability of a loss exceeding 1.695 is 0.045, and this is also equal to p, and the probability of profit or a loss less than 1.595 is 0.9446. Hence $f(q)$, the probability mass in the q range, is $1 - 0.0450 - 0.9446 = 0.0104$. We now plug the relevant values into Equation (1.27) to obtain the 90% confidence interval for the VaR:

$$\left[\begin{array}{l} 1.645 - 1.645\dfrac{\sqrt{0.045(1-0.045)/1000}}{0.0104}, \\ 1.645 + 1.645\dfrac{\sqrt{0.045(1-0.045)/1000}}{0.0104} \end{array}\right] = [0.6081, 2.6819]$$

This is a wide confidence interval, especially when compared to the OS and bootstrap confidence intervals.

The confidence interval narrows if we take a wider bin width, so suppose that we now repeat the exercise using a bin width $h = 0.2$, which is probably as wide as we can reasonably go with these data. q now falls into the range $1.645 \pm 0.2/2 = [1.545, 1.745]$. p, the probability of a loss exceeding 1.745, is 0.0405, and the probability of profit or a loss less than 1.545 is 0.9388. Hence $f(q) = 1 - 0.0405 - 0.9388 = 0.0207$. Plugging these values into Equation (1.27) now gives us a new estimate of the 90% confidence interval:

$$\left[\begin{array}{l} 1.645 - 1.645\dfrac{\sqrt{0.0405(1-0.0405)/1000}}{0.0207}, \\ 1.645 + 1.645\dfrac{\sqrt{0.0405(1-0.0405)/1000}}{0.0207} \end{array}\right] = [1.1496, 2.1404]$$

This is still a rather wide confidence interval.

This example illustrates that although we can use quantile standard errors to estimate VaR confidence intervals, the intervals can be wide and also sensitive to the arbitrary choice of bin width.

The quantile-standard-error approach is easy to implement and has some plausibility with large sample sizes. However, it also has weaknesses relative to other methods of assessing the precision of quantile (or VaR) estimators—it relies on asymptotic theory and requires large sample sizes; it can produce imprecise estimators, or wide confidence intervals; it depends on the arbitrary choice of bin width; and the symmetric confidence intervals it produces are misleading for extreme quantiles whose 'true' confidence intervals are asymmetric reflecting the increasing sparsity of extreme observations as we move further out into the tail.

Standard Errors in Estimators of Coherent Risk Measures

We now consider standard errors in estimators of coherent risk measures. One of the first studies to examine this issue (Yamai and Yoshiba (2001b) did so by investigating the relative accuracy of VaR and ES estimators for Lévy distributions with varying α stability parameters. Their results suggested that VaR and ES estimators had comparable standard errors for near-normal Lévy distributions, but the ES estimators had much bigger standard errors for particularly heavy-tailed distributions. They explained this finding by saying that as tails became heavier, ES estimators became more prone to the effects of large but infrequent losses. This finding suggests the depressing conclusion that the presence of heavy tails might make ES estimators in general less accurate than VaR estimators.

Fortunately, there are grounds to think that such a conclusion might be overly pessimistic. For example, Inui and Kijima (2003) present alternative results showing that the application of a Richardson extrapolation method can produce ES estimators that are both unbiased and have comparable standard errors to VaR estimators.[7] Acerbi (2004) also looked at this issue and, although he confirmed that tail heaviness did increase the standard errors of ES estimators relative to VaR estimators, he concluded that the relative accuracies of VaR and ES estimators were roughly comparable in empirically realistic ranges.

However, the standard error of any estimator of a coherent risk measure will vary from one situation to another, and the best practical advice is to get into the habit of always estimating the standard error whenever one estimates the risk measure itself. Estimating the standard error of an estimator of a coherent risk measure is also relatively straightforward. One way to do so starts from recognition that a coherent risk measure is an L-estimator

[7] See Inui and Kijima (2003).

(i.e., a weighted average of order statistics), and L-estimators are asymptotically normal. If we take N discrete points in the density function, then as N gets large the variance of the estimator of the coherent risk measure (Equation (1.17)) is approximately

$$\sigma(M_\phi^{(N)}) \to \frac{2}{N}\int_{p<q}\phi(p)\phi(q)\frac{p(1-q)}{f(F^{-1}(p))f(F^{-1}(q))}dpdq$$
$$= \frac{2}{N}\int_{x<y}\phi(F(x))\phi(F(y))F(x)(1-F(y))dxdy \tag{1.28}$$

and this can be computed numerically using a suitable numerical integration procedure. Where the risk measure is the ES, the standard error becomes

$$\sigma(ES^{(N)}) \to \frac{1}{N\alpha^2}\int_0^{F^{-1}(\alpha)}\int_0^{F^{-1}(\alpha)}[\min(F(x),F(y)) - F(x)F(y)]dxdy \tag{1.29}$$

and used in conjunction with a suitable numerical integration method, this gives good estimates even for relatively low values of N.[8] If we wish to obtain confidence intervals for our risk measure estimators, we can make use of the asymptotic normality of these estimators to apply textbook formulas (e.g., such as Equation (1.27)) based on the estimated standard errors and centred around a 'good' best estimate of the risk measure.

An alternative approach to the estimation of standard errors for estimators of coherent risk measures is to apply a bootstrap: we bootstrap a large number of estimators from the given distribution function (which might be parametric or non-parametric, e.g., historical); and we estimate the standard error of the sample of bootstrapped estimators. Even better, we can also use a bootstrapped sample of estimators to estimate a confidence interval for our risk measure.

THE CORE ISSUES: AN OVERVIEW

Before proceeding to more detailed issues, it might be helpful to pause for a moment to take an overview of the structure, as it were, of the subject matter itself. This is very useful, as it gives the reader a mental frame of reference within which the 'detailed' material that follows can be placed. Essentially, there are three core issues, and all the material that follows can be related to these. They also have a natural sequence, so we can think of them as providing a roadmap that leads us to where we want to be.

Which risk measure? The first and most important is to choose the type of risk measure: do we want to estimate VaR, ES, etc.? This is logically the first issue, because we need to know *what* we are trying to estimate before we start thinking about *how* we are going to estimate it.

Which level? The second issue is the *level* of analysis. Do we wish to estimate our risk measure at the level of the portfolio as a whole or at the level of the individual positions in it? The former would involve us taking the portfolio as our basic unit of analysis (i.e., we take the portfolio to have a specified composition, which is taken as given for the purposes of our analysis), and this will lead to a *univariate* stochastic analysis. The alternative is to work from the position level, and this has the advantage of allowing us to accommodate changes in the portfolio composition within the analysis itself. The disadvantage is that we then need a *multivariate* stochastic framework, and this is considerably more difficult to handle: we have to get to grips with the problems posed by variance–covariance matrices, copulas, and so on, all of which are avoided if we work at the portfolio level. There is thus a trade-off: working at the portfolio level is more limiting, but easier, while working at the position level gives us much more flexibility, but can involve much more work.

Which method? Having chosen our risk measure and level of analysis, we then choose a suitable estimation method. To decide on this, we would usually think in terms of the classic 'VaR trinity':

- Non-parametric methods
- Parametric methods
- Monte Carlo simulation methods

Each of these involves some complex issues.

APPENDIX

Preliminary Data Analysis

When confronted with a new data set, we should *never* proceed straight to estimation without some preliminary analysis to get to know our data. Preliminary data analysis is useful because it gives us a feel for our data, and because it can highlight problems with our data set.

[8] See Acerbi (2004, pp. 200–201).

Remember that we never really know where our data come from, so we should always be a little wary of any new data set, regardless of how reputable the source might appear to be. For example, how do you know that a clerk hasn't made a mistake somewhere along the line in copying the data and, say, put a decimal point in the wrong place? The answer is that you don't, and never can. Even the most reputable data providers provide data with errors in them, however careful they are. Everyone who has ever done any empirical work will have encountered such problems at some time or other: the bottom line is that real data must always be viewed with a certain amount of suspicion.

Such preliminary analysis should consist of at least the first two and preferably all three of the following steps:

- The first and by far the most important step is to eyeball the data to see if they 'look right'—or, more to the point, we should eyeball the data to see if anything looks *wrong*. Does the pattern of observations look right? Do any observations stand out as questionable? And so on. The interocular trauma test is the most important test ever invented and also the easiest to carry out, and we should always perform it on any new data set.
- We should plot our data on a histogram and estimate the relevant summary statistics (i.e., mean, standard deviation, skewness, kurtosis, etc.). In risk measurement, we are particularly interested in any non-normal features of our data: skewness, excess kurtosis, outliers in our data, and the like. We should therefore be on the lookout for any evidence of non-normality, and we should take any such evidence into account when considering whether to fit any parametric distribution to the data.
- Having done this initial analysis, we should consider what kind of distribution might fit our data, and there are a number of useful diagnostic tools available for this purpose, the most popular of which are QQ plots—plots of empirical quantiles against their theoretical equivalents.

Plotting the Data and Evaluating Summary Statistics

To get to know our data, we should first obtain their histogram and see what might stand out. Do the data look normal, or non-normal? Do they show one pronounced peak, or more than one? Do they seem to be skewed? Do they have fat tails or thin tails? Are there outliers? And so on.

As an example, Figure 1-10 shows a histogram of 100 random observations. In practice, we would usually wish to work with considerably longer data sets, but a data set this small helps to highlight the uncertainties one often encounters in practice. These observations show a dominant peak in the centre, which suggests that they are probably drawn from a unimodal distribution. On the other hand, there may be a negative skew, and there are some large outlying observations on the extreme left of the distribution, which might indicate fat tails on at least the left-hand side. In fact, these particular observations are drawn from a Student-*t* distribution with 5 degrees of freedom, so in this case we know that the underlying true distribution is unimodal, symmetric and heavy tailed. However, we would not know this in a situation with 'real' data, and it is precisely because we do not know the distributions of real-world data sets that preliminary analysis is so important.

Some summary statistics for this data set are shown in Table 1-6. The sample mean (−0.099) and the sample mode differ somewhat (−0.030), but this difference is small relative to the sample standard deviation (1.363). However, the sample skew (−0.503) is somewhat negative and the sample kurtosis (3.985) is a little bigger than normal. The sample minimum (−4.660) and the sample maximum (3.010) are also not symmetric about the sample mean or mode, which is further evidence of asymmetry. If we encountered these results with 'real' data, we would be concerned about possible skewness and kurtosis. However, in this hypothetical case we know that the sample skewness is merely a product of sample variation, because we happen to know that the data are drawn from a symmetric distribution.

Depending on the context, we might also seriously consider carrying out some formal tests. For example, we might test whether the sample parameters (mean, standard deviation, etc.) are consistent with what we might expect under a null hypothesis (e.g., such as normality).

The underlying principle is very simple: since we *never* know the true distribution in practice, all we ever have to work with are *estimates* based on the *sample* at hand; it therefore behoves us to make the best use of the data we have, and to extract as much information as possible from them.

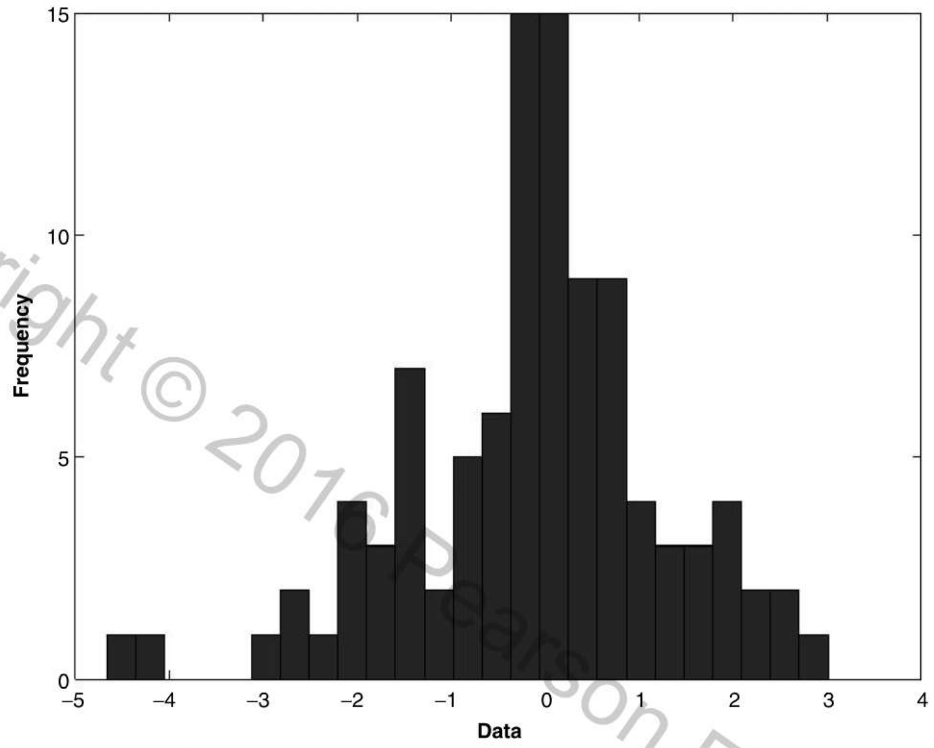

FIGURE 1-10 A histogram.

Note: Data are 100 observations randomly drawn from a Student-*t* with 5 degrees of freedom.

TABLE 1-6 Summary Statistics

Parameter	Value
Mean	−0.099
Mode	−0.030
Standard deviation	1.363
Skewness	−0.503
Kurtosis	3.985
Minimum	−4.660
Maximum	3.010
Number of observations	100

Note: Data are the same observations shown in Figure 1-10.

QQ Plots

Having done our initial analysis, it is often good practice to ask what distribution might fit our data, and a very useful device for identifying the distribution of our data is a quantile-quantile or QQ plot—a plot of the quantiles of the empirical distribution against those of some specified distribution. The shape of the QQ plot tells us a lot about how the empirical distribution compares to the specified one. In particular, if the QQ plot is linear, then the specified distribution fits the data, and we have identified the distribution to which our data belong. In addition, departures of the QQ from linearity in the tails can tell us whether the tails of our empirical distribution are fatter, or thinner, than the tails of the reference distribution to which it is being compared.

To illustrate, Figure 1-11 shows a QQ plot for a data sample drawn from a normal distribution, compared to a reference distribution that is also normal. The QQ plot is obviously close to linear: the central mass observations fit a linear QQ plot very closely, and the extreme tail observations somewhat less so. However, there is no denying that the overall plot is approximately linear. Figure 1-11 is a classic example of a QQ plot in which the empirical distribution matches the reference population.

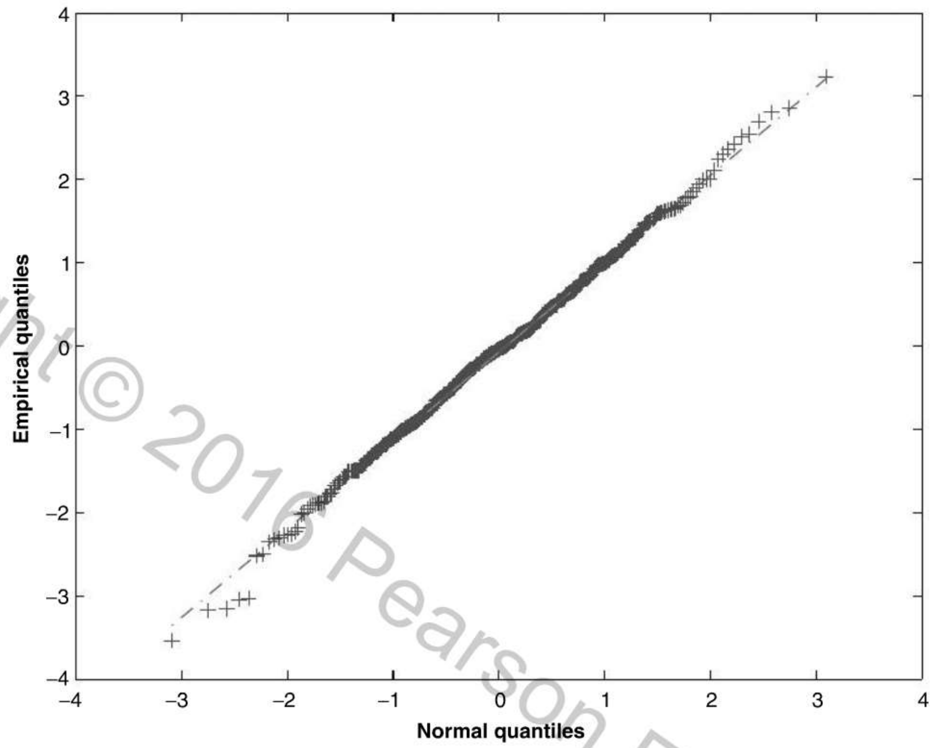

FIGURE 1-11 QQ plot: normal sample against normal reference distribution.

Note: The empirical sample is a random sample of 500 observations drawn from a standard normal. The reference distribution is standard normal.

By contrast, Figure 1-12 shows a good example of a QQ plot where the empirical distribution does not match the reference population. In this case, the data are drawn from a Student-*t* with 5 degrees of freedom, but the reference distribution is standard normal. The QQ plot is now clearly non-linear: although the central mass observations are close to linear, the tails show steeper slopes indicative of the tails being heavier than those of the reference distribution.

A QQ plot is useful in a number of ways. First, as noted already, if the data are drawn from the reference population, then the QQ plot should be linear. This remains true if the data are drawn from some linear transformation of the reference distribution (i.e., are drawn from the same distribution but with different location and scale parameters). We can therefore use a QQ plot to form a tentative view of the distribution from which our data might be drawn: we specify a variety of alternative distributions, and construct QQ plots for each. Any reference distributions that produce non-linear QQ plots can then be dismissed, and any distribution that produces a linear QQ plot is a good candidate distribution for our data.

Second, because a linear transformation in one of the distributions in a QQ plot merely changes the intercept and slope of the QQ plot, we can use the intercept and slope of a linear QQ plot to give us a rough idea of the location and scale parameters of our sample data. For example, the reference distribution in Figure 1-11 is a standard normal. The linearity of the QQ plot in this figure suggests that the data are normal, as mentioned already, but Figure 1-11 also shows that the intercept and slope are approximately 0 and 1 respectively, and this indicates that the data are drawn from a standard normal, and not just any normal. Such rough approximations give us a helpful yardstick against which we can judge more 'sophisticated' estimates of location and scale, and also provide useful initial values for iterative algorithms.

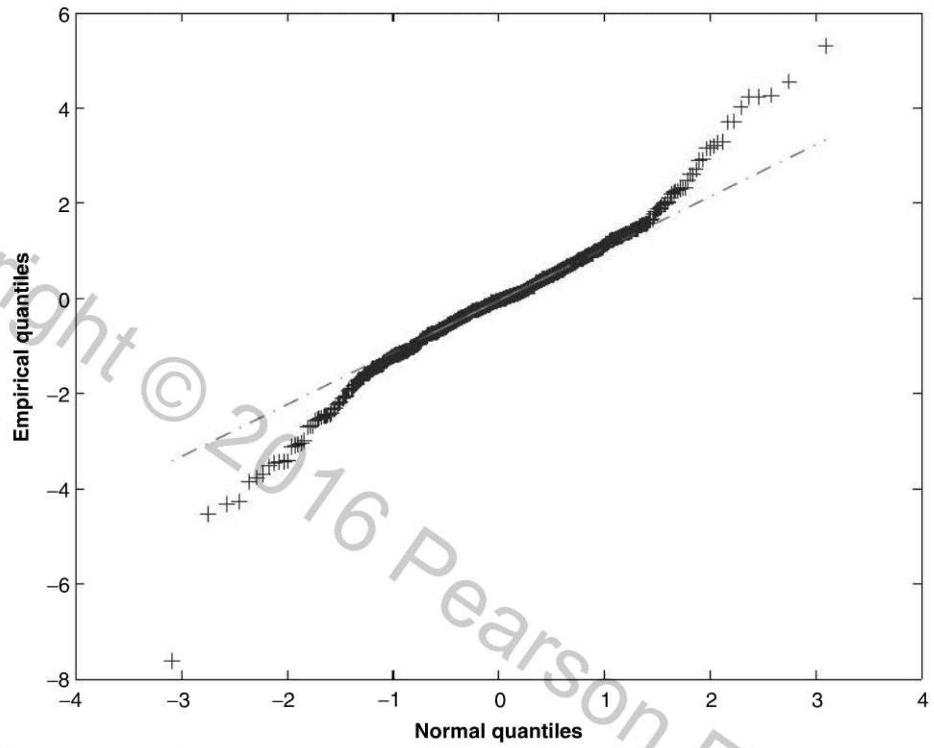

FIGURE 1-12 QQ plot: *t* sample against normal reference distribution.

Note: The empirical sample is a random sample of 500 observations drawn from Student-*t* with 5 degrees of freedom. The reference distribution is standard normal.

Third, if the empirical distribution has heavier tails than the reference distribution, the QQ plot will have steeper slopes at its tails, even if the central mass of the empirical observations are approximately linear. Figure 1-12 is a good case in point. A QQ plot where the tails have slopes different than the central mass is therefore suggestive of the empirical distribution having heavier, or thinner, tails than the reference distribution.

Finally, a QQ plot is good for identifying outliers (e.g., observations contaminated by large errors): such observations will stand out in a QQ plot, even if the other observations are broadly consistent with the reference distribution.[9]

[9] Another useful tool, especially when dealing with the tails, is the mean excess function (MEF): the expected amount by which a random variable X exceeds some threshold u, given that $X > u$. The usefulness of the MEF stems from the fact that each distribution has its own distinctive MEF. A comparison of the empirical MEF with the theoretical MEF associated with some specified distribution function therefore gives us an indication of whether the chosen distribution fits the tails of our empirical distribution. However, the results of MEF plots need to be interpreted with some care, because data observations become more scarce as X gets larger. For more on these and how they can be used, see Embrechts et. al. (1997, Chapters 3.4 and 6.2).

2017 Financial Risk Manager (FRM) Part II: Market Risk Measurement and Management, Sixth Edition by Global Association of Risk Professionals.

Non-parametric Approaches

2

Learning Objectives

After completing this reading you should be able to:

- Apply the bootstrap historical simulation approach to estimate coherent risk measures.
- Describe historical simulation using non-parametric density estimation.
- Compare and contrast the age-weighted, the volatility-weighted, the correlation-weighted, and the filtered historical simulation approaches.
- Identify advantages and disadvantages of non-parametric estimation methods.

Excerpt is Chapter 4 of Measuring Market Risk, *Second Edition, by Kevin Dowd.*

This chapter looks at some of the most popular approaches to the estimation of risk measures—the non-parametric approaches, which seek to estimate risk measures without making strong assumptions about the relevant (e.g., P/L) distribution. The essence of these approaches is that we try to let the P/L data speak for themselves as much as possible, and use the recent *empirical* (or in some cases simulated) distribution of P/L—not some assumed theoretical distribution—to estimate our risk measures. All non-parametric approaches are based on the underlying assumption that the near future will be sufficiently like the recent past that we can use the data from the recent past to forecast risks over the near future—and this assumption may or may not be valid in any given context. In deciding whether to use any non-parametric approach, we must make a judgment about the extent to which data from the recent past are likely to give us a good guide about the risks we face over the horizon period we are concerned with.

To keep the discussion as clear as possible, we will focus on the estimation of non-parametric VaR and ES. However, the methods discussed here extend very naturally to the estimation of coherent and other risk measures as well. These can be estimated using an 'average quantile' approach along the lines discussed in Chapter 1: we would select our weighting function $\phi(p)$, decide on the number of probability 'slices' n to take, estimate the associated quantiles, and take the weighted average using an appropriate numerical algorithm (see Chapter 1).[1] We can then obtain standard errors or confidence intervals for our risk measures using suitably modified forms.

In this chapter we begin by discussing how to assemble the P/L data to be used for estimating risk measures. We then discuss the most popular non-parametric approach-historical simulation (HS). Loosely speaking, HS is a histogram-based approach: it is conceptually simple, easy to implement, very widely used, and has a fairly good historical record. We focus on the estimation of VaR and ES, but as explained in the previous chapter, more general coherent risk measures can be estimated using appropriately weighted averages of any non-parametric VaR estimates. We then discuss refinements to basic HS using bootstrap and kernel methods, and the estimation of VaR or ES curves and surfaces. We will discuss how we can estimate confidence intervals for HS VaR and ES. Then we will address weighted HS—how we might weight our data to capture the effects of observation age and changing market conditions. These methods introduce parametric formulas (such as GARCH volatility forecasting equations) into the picture, and in so doing convert hitherto non-parametric methods into what are best described as semi-parametric methods. Such methods are very useful because they allow us to retain the broad HS framework while also taking account of ways in which we think that the risks we face over the foreseeable horizon period might differ from those in our sample period. Finally we review the main advantages and disadvantages of non-parametric and semi-parametric approaches, and offer some conclusions.

COMPILING HISTORICAL SIMULATION DATA

The first task is to assemble a suitable P/L series for our portfolio, and this requires a set of historical P/L or return observations on the positions in our current portfolio. These P/Ls or returns will be measured over a particular frequency (e.g., a day), and we want a reasonably large set of historical P/L or return observations over the recent past. Suppose we have a portfolio of n assets, and for each asset i we have the observed return for each of T subperiods (e.g., daily subperiods) in our historical sample period. If $R_{i,t}$ is the (possibly mapped) return on asset i in subperiod t, and if w_i is the amount currently invested in asset i, then the historically simulated portfolio P/L over the subperiod t is:

$$P/L_t = \sum_{i=1}^{n} w_i R_{i,t} \tag{2.1}$$

Equation (2.1) gives us a historically simulated P/L series for our current portfolio, and is the basis of HS VaR

[1] Nonetheless, there is an important caveat. This method was explained in Chapter 1 in an implicit context where the risk measurer could choose n, and this is sometimes not possible in a non-parametric context. For example, a risk measurer might be working with an n determined by the HS data set, and even where he/she has some freedom to select n, their range of choice might be limited by the data available. Such constraints can limit the degree of accuracy of any resulting estimated risk measures. However, a good solution to such problems is to increase the sample size by bootstrapping from the sample data. (The bootstrap is discussed further in Appendix 2 to this chapter).

and ES. This series will *not* generally be the same as the P/L *actually* earned on our portfolio—because our portfolio may have changed in composition over time or be subject to mapping approximations, and so on. Instead, the historical simulation P/L is the P/L we *would have* earned on our current portfolio had we held it throughout the historical sample period.[2]

As an aside, the fact that multiple positions collapse into one single HS P/L as given by Equation (2.1) implies that it is very easy for non-parametric methods to accommodate high dimensions—unlike the case for some parametric methods. With non-parametric methods, there are no problems dealing with variance-covariance matrices, curses of dimensionality, and the like. This means that non-parametric methods will often be the most natural choice for high-dimension problems.

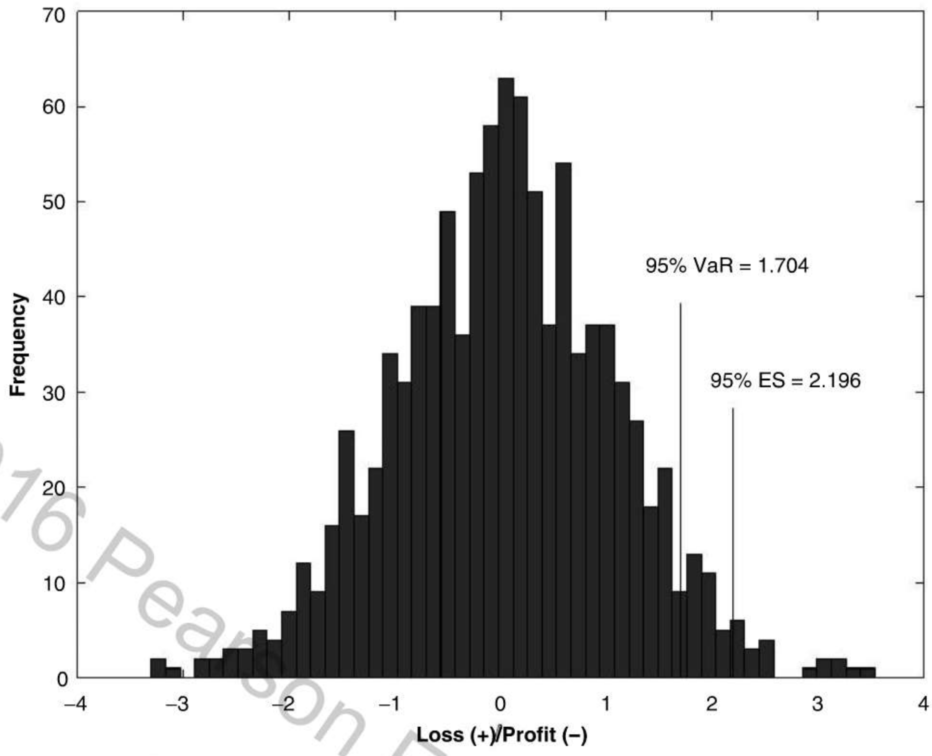

FIGURE 2-1 Basic historical simulation VaR and ES.

Note: This figure and associated VaR and ES estimates are obtained using the 'hsesfigure' function.

ESTIMATION OF HISTORICAL SIMULATION VaR AND ES

Basic Historical Simulation

Having obtained our historical simulation P/L data, we can estimate VaR by plotting the P/L (or L/P) on a simple histogram and then reading off the VaR from the histogram. To illustrate, suppose we have 1000 observations in our HS P/L series and we plot the L/P histogram shown in Figure 2-1. If these were daily data, this sample size would be equivalent to four years' daily data at 250 trading days to a year. If we take our confidence level to be 95%, our VaR is given by the *x*-value that cuts off the upper 5% of very high losses from the rest of the distribution. Given 1000 observations, we can take this value (i.e., our VaR) to be the 51st highest loss value, or 1.704.[3] The ES is then the average of the 50 highest losses, or 2.196.

The imprecision of these estimates should be obvious when we consider that the sample data set was drawn from a standard normal distribution. In this case the 'true' underlying VaR and ES are 1.645 and 2.063, and Figure 2-1 should (ideally) be normal. Of course, this imprecision underlines the need to work with large sample sizes where practically feasible.

[2] To be more precise, the historical simulation P/L is the P/L we would have earned over the sample period had we rearranged the portfolio at the end of each trading day to ensure that the amount left invested in each asset was the same as at the end of the previous trading day: we take out our profits, or make up for our losses, to keep the w_i constant from one end-of-day to the next.

[3] We can also estimate the HS VaR more directly (i.e., without bothering with the histogram) by using a spreadsheet function that gives us the 51st highest loss value (e.g., the 'Large' command in Excel), or we can sort our losses data with highest losses ranked first, and then obtain the VaR as the 51st observation in our sorted loss data. We could also take VaR to be any point between the 50th and 51st largest losses (e.g., such as their mid-point), but with a reasonable sample size (as here) there will seldom be much difference between these losses anyway. For convenience, we will adhere throughout to this convention of taking the VaR to be the highest loss observation outside the tail.

Bootstrapped Historical Simulation

One simple but powerful improvement over basic HS is to estimate VaR and ES from bootstrapped data. As explained in Appendix 2 to this chapter, a bootstrap procedure involves resampling from our existing data set with replacement. The bootstrap is very intuitive and easy to apply. A bootstrapped estimate will often be more accurate than a 'raw' sample estimate, and bootstraps are also useful for gauging the precision of our estimates. To apply the bootstrap, we create a large number of new samples, each observation of which is obtained by drawing at random from our original sample and replacing the observation after it has been drawn. Each new 'resampled' sample gives us a new VaR estimate, and we can take our 'best' estimate to be the mean of these resample-based estimates. The same approach can also be used to produce resample-based ES estimates—each one of which would be the average of the losses in each resample exceeding the resample VaR—and our 'best' ES estimate would be the mean of these estimates. In our particular case, if we take 1000 resamples, then our best VaR and ES estimates are (because of bootstrap sampling variation) about 1.669 and 2.114—and the fact that these are much closer to the known true values than our earlier basic HS estimates suggests that bootstraps estimates might be more accurate.

Historical Simulation Using Non-parametric Density Estimation

Another potential improvement over basic HS sometimes suggested is to make use of non parametric density estimation. To appreciate what this involves, we must recognise that basic HS does not make the best use of the information we have. It also has the practical drawback that it only allows us to estimate VaRs at discrete confidence intervals determined by the size of our data set. For example, if we have 100 HS P/L observations, basic HS allows us to estimate VaR at the 95% confidence level, but not the VaR at the 95.1% confidence level. The VaR at the 95% confidence level is given by the sixth largest loss, but the VaR at the 95.1% confidence level is a problem because there is no corresponding loss observation to go with it. We know that it should be greater than the sixth largest loss (or the 95% VaR), and smaller than the fifth largest loss (or the 96% VaR), but with only 100 observations there is no observation that corresponds to any VaR whose confidence level involves a fraction of 1%. With *n* observations, basic HS only allows us to estimate the VaRs associated with, at best, *n* different confidence levels.

Non-parametric density estimation offers a potential solution to both these problems. The idea is to treat our data as if they were drawings from some unspecified or unknown empirical distribution function. This approach also encourages us to confront potentially important decisions about the width of bins and where bins should be centred, and these decisions can sometimes make a difference to our results. Besides using a histogram, we can also represent our data using naïve estimators or, more generally, kernels, and the literature tells us that kernels are (or ought to be) superior. So, having assembled our 'raw' HS data, we need to make decisions on the widths of bins and where they should be centred, and whether to use a histogram, a naïve estimator, or some form of kernel. If we make good decisions on these issues, we can hope to get better estimates of VaR and ES (and more general coherent measures).

Non-parametric density estimation also allows us to estimate VaRs and ESs for any confidence levels we like and so avoid constraints imposed by the size of our data set. In effect, it enables us to draw lines through points on or near the edges of the 'bars' of a histogram. We can then treat the areas under these lines as a surrogate pdf, and so proceed to estimate VaRs for arbitrary confidence levels. The idea is illustrated in Figure 2-2. The left-hand side of this figure shows three bars from a histogram (or naïve estimator) close up. Assuming that the height of the histogram (or naïve estimator) measures relative frequency, then one option is to treat the histogram itself as a pdf. Unfortunately, the resulting pdf would be a strange one—just look at the corners of each bar—and it makes more sense to approximate the pdf by drawing lines through the upper parts of the histogram.

A simple way to do this is to draw in straight lines connecting the mid-points at the top of each histogram bar, as illustrated in the figure. Once we draw these lines, we can forget about the histogram bars and treat the area under the lines as if it were a pdf. Treating the area under the lines as a pdf then enables us to estimate VaRs at any confidence level, regardless of the size of our data set. Each possible confidence level would correspond to its

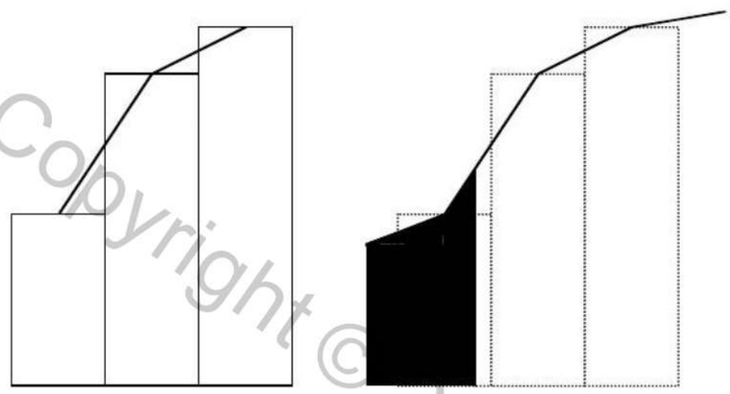

FIGURE 2-2 Histograms and surrogate density functions.

own tail similar to the shaded area shown in Figure 2-2(b), and we can then use a suitable calculation method to estimate the VaR (e.g., we can carry out the calculations on a spreadsheet or, more easily, by using a purpose-built function such as the 'hsvar' function in the MMR Toolbox).[4] Of course, drawing straight lines through the mid-points of the tops of histogram bars is not the best we can do: we could draw smooth curves that meet up nicely, and so on. This is exactly the point of non-parametric density estimation, the purpose of which is to give us some guidance on how 'best' to draw lines through the data points we have. Such methods are also straightforward to apply if we have suitable software.

Some empirical evidence by Butler and Schachter (1998) using real trading portfolios suggests that kernel-type methods produce VaR estimates that are a little different to those we would obtain under basic HS. However, their work also suggests that the different types of kernel methods produce quite similar VaR estimates, although to the extent that there are differences among them, they also found that the 'best' kernels were the adaptive Epanechinikov and adaptive Gaussian ones. To investigate these issues myself, I applied four standard kernel estimators—based on normal, box, triangular and Epanechinikov kernels—to the test data used in earlier examples, and found that each of these gave the same VaR estimate of 1.735. In this case, these different kernels produced the same VaR estimate, which is a little higher (and, curiously, a little less accurate) than the basic HS VaR estimate of 1.704 obtained earlier. Other results not reported here suggest that the different kernels can give somewhat different estimates with smaller samples, but again suggest that the exact kernel specification does not make a great deal of difference.

So although kernel methods are better in theory, they do not necessarily produce much better estimates in practice. There are also practical reasons why we might prefer simpler non-parametric density estimation methods over kernel ones. Although the kernel methods are theoretically better, crude methods like drawing straight-line 'curves' through the tops of histograms are more transparent and easier to check. We should also not forget that our results are subject to a number of sources of error (e.g., due to errors in P/L data, mapping approximations, and so on), so there is a natural limit to how much real fineness we can actually achieve.

Estimating Curves and Surfaces for VaR and ES

It is straightforward to produce plots of VaR or ES against the confidence level. For example, our earlier hypothetical P/L data yields the curves of VaR and ES against the confidence level shown in Figure 2-3. Note that the VaR curve is fairly unsteady, as it directly reflects the randomness of individual loss observations, but the ES curve is smoother, because each ES is an average of tail losses.

It is more difficult constructing curves that show how non-parametric VaR or ES changes with the holding period. The methods discussed so far enable us to estimate the VaR or ES at a single holding period equal to the frequency period over which our data are observed (e.g., they give us VaR or ES for a daily holding period if P/L is measured daily). In theory, we can then estimate VaRs or ESs for any other holding periods we wish by constructing

[4] The actual programming is a little tedious, but the gist of it is that if the confidence level is such that the VaR falls between two loss observations, then we take the VaR to be a weighted average of these two observations. The weights are chosen so that a vertical line drawn through the VaR demarcates the area under the 'curve' in the correct proportions, with α to one side and $1-\alpha$ to the other. The details can be seen in the coding for the 'hsvar' and related functions.

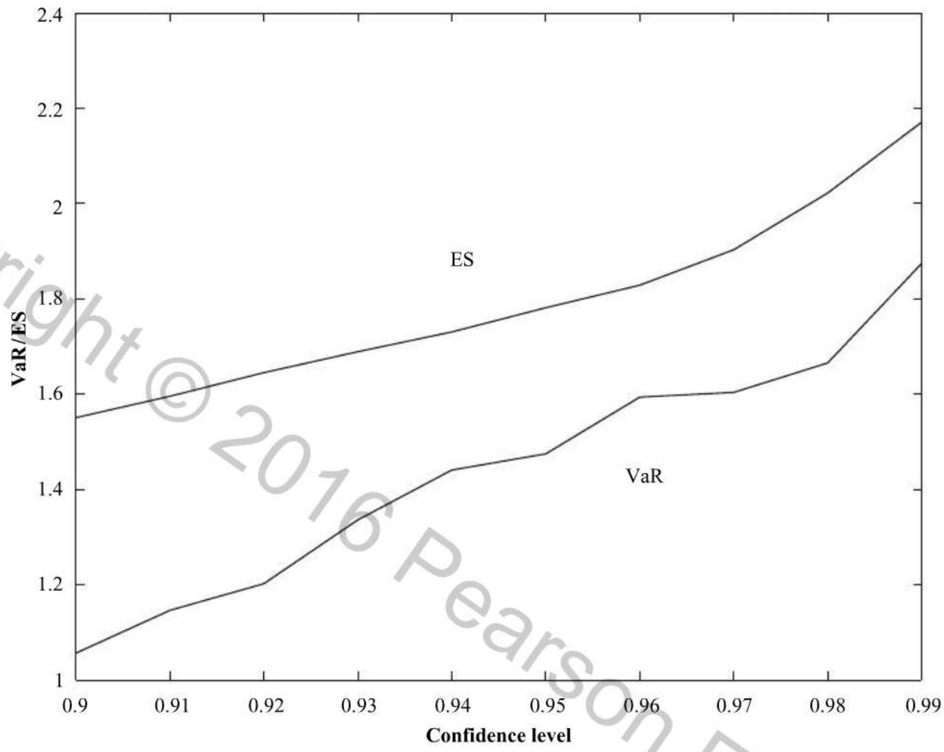

FIGURE 2-3 Plots of HS VaR and ES against confidence level.

Note: Obtained using the 'hsvaresplot2D_cl' function and the same hypothetical P/L data used in Figure 2-1.

a HS P/L series whose frequency matches our desired holding period: if we wanted to estimate VaR over a weekly holding period, say, we could construct a weekly P/L series and estimate the VaR from that. There is, in short, no theoretical problem as such with estimating HS VaR or ES over any holding period we like.

However, there is a major practical problem: as the holding period rises, the number of observations rapidly falls, and we soon find that we don't have enough data. To illustrate, if we have 1000 observations of daily P/L, corresponding to four years' worth of data at 250 trading days a year, then we have 1000 P/L observations if we use a daily holding period. If we have a weekly holding period, with five days to a week, each weekly P/L will be the sum of five daily P/Ls, and we end up with only 200 observations of weekly P/L; if we have a monthly holding period, we have only 50 observations of monthly P/L; and so on. Given our initial data, the number of effective observations rapidly falls as the holding period rises, and the size of the data set imposes a major constraint on how large the holding period can practically be. In any case, even if we had a very long run of data, the older observations might have very little relevance for current market conditions.

ESTIMATING CONFIDENCE INTERVALS FOR HISTORICAL SIMULATION VaR AND ES

The methods considered so far are good for giving point estimates of VaR or ES, but they don't give us any indication of the precision of these estimates or any indication of VaR or ES confidence intervals. However, there are methods to get around this limitation and produce confidence intervals for our risk estimates.[5]

[5] In addition to the methods considered in this section, we can also estimate confidence intervals for VaR using estimates of the quantile standard errors. However, as made clear there, such confidence intervals are subject to a number of problems, and the methods suggested here are usually preferable.

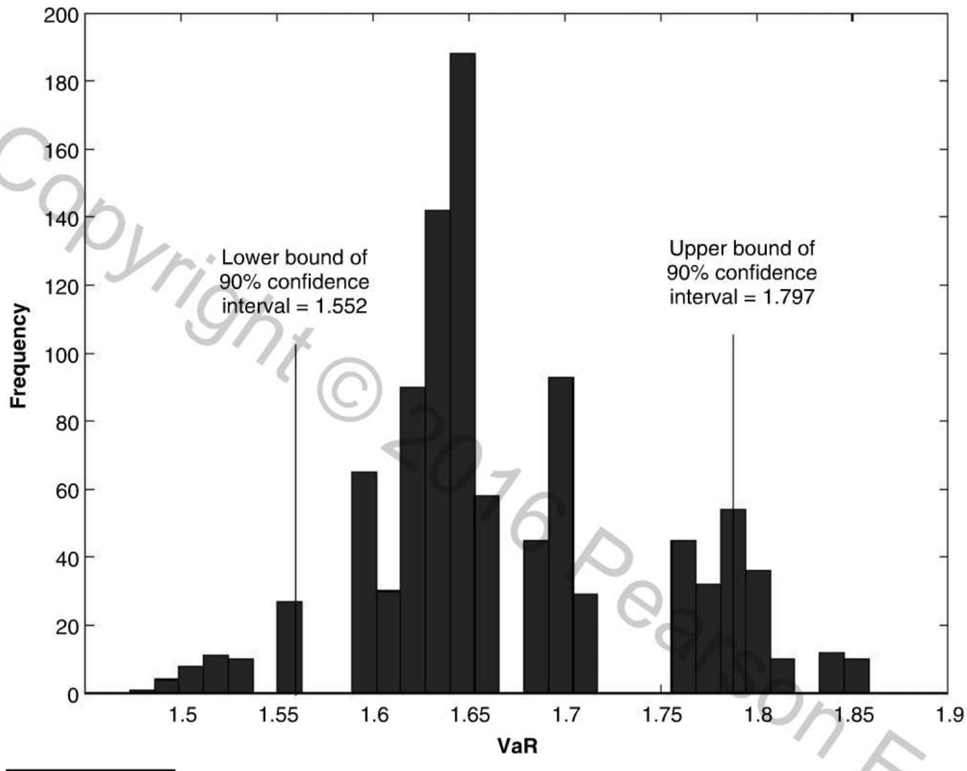

FIGURE 2-4 Bootstrapped VaR.

Note: Results obtained using the 'bootstrapvarfigure' function with 1000 resamples, and the same hypothetical data as in earlier figures.

The corresponding points of the ES distribution function can be obtained (using the 'hsesdfperc' function) by mapping from the VaR to the ES: we take a point on the VaR distribution function, and estimate the corresponding percentile point on the ES distribution function. Doing this gives us an estimated 90% confidence interval of [2.021, 2.224].[6]

An Order Statistics Approach to the Estimation of Confidence Intervals for HS VaR and ES

One of the most promising methods is to apply the theory of order statistics, explained in Appendix 1 to this chapter. This approach gives us, not just a VaR (or ES) estimate, but a complete VaR (or ES) distribution function from which we can read off the VaR (or ES) confidence interval. (The central tendency parameters (mean, mode, median) also give us alternative point estimates of our VaR or ES, if we want them.) This approach is (relatively) easy to programme and very general in its application.

Applied to our earlier P/L data, the OS approach gives us estimates (obtained using the 'hsvarpdfperc' function) of the 5% and 95% points of the 95% VaR distribution function—that is, the bounds of the 90% confidence interval for our VaR—of 1.552 and 1.797. This tells us we can be 90% confident that the 'true' VaR lies in the range [1.552, 1.797].

A Bootstrap Approach to the Estimation of Confidence Intervals for HS VaR and ES

We can also estimate confidence intervals using a bootstrap approach: we produce a bootstrapped histogram of resample-based VaR (or ES) estimates, and then read the confidence interval from the quantiles of this histogram. For example, if we take 1000 bootstrapped samples from our P/L data set, estimate the 95% VaR of each, and then plot them, we get the histogram shown in Figure 2-4. Using the basic percentile interval approach outlined in Appendix 2 to this chapter, the 90% confidence interval for our VaR is [1.554, 1.797]. The simulated histogram is surprisingly disjointed, although the bootstrap seems to give a relatively robust estimate of the confidence interval if we keep repeating the exercise.

We can also use the bootstrap to estimate ESs in much the same way: for each new resampled data set, we estimate the VaR, and then estimate the ES as the average of losses in excess of VaR. Doing this a large number of times gives us a large number of ES estimates, and we can plot them in the same way as the VaR estimates.

[6] Naturally, the order statistics approach can be combined with more sophisticated non-parametric density estimation approaches. Instead of applying the OS theory to the histogram or naive estimator, we could apply it to a more sophisticated kernel estimator, and thereby extract more information from our data. This approach has some merit and is developed in detail by Butler and Schachter (1998).

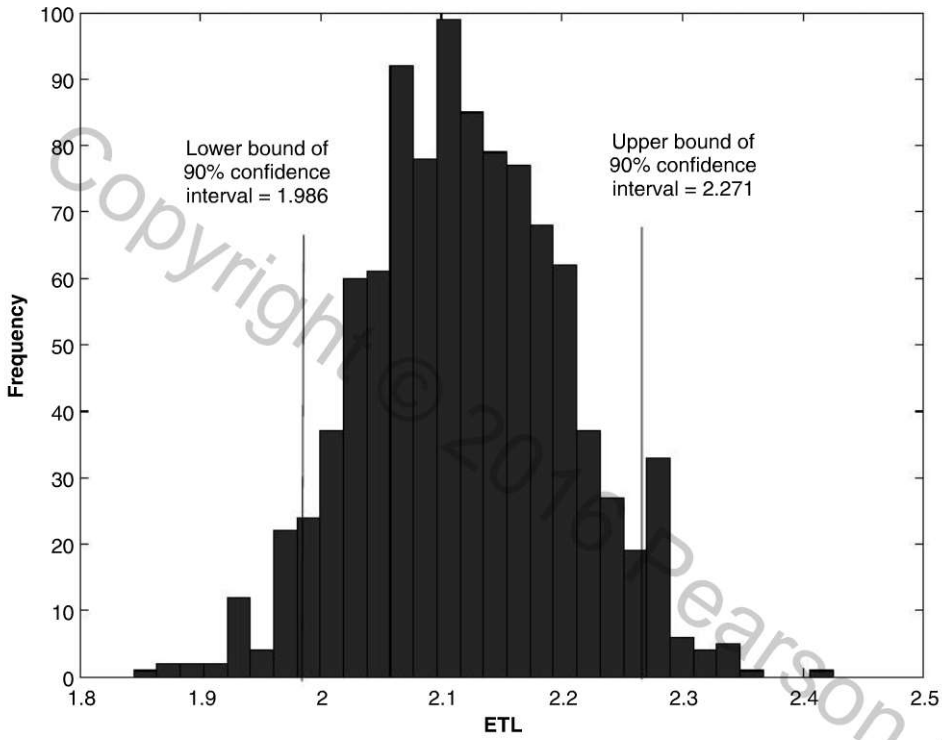

FIGURE 2-5 Bootstrapped ES.

Note: Results obtained using the 'bootstrapesfigure' function with 1000 resamples, and the same hypothetical data as in earlier figures.

TABLE 2-1 90% Confidence Intervals for Non-parametric VaR and ES

Approach	Lower bound	Upper bound
	95% VaR	
Order statistics	1.552	1.797
Bootstrap	1.554	1.797
	95% ES	
Order statistics	2.021	2.224
Bootstrap	1.986	2.271

Note: Bootstrap estimates based on 1000 resamples.

The histogram of bootstrapped ES values is shown in Figure 2-5, and is better behaved than the VaR histogram in the last figure because the ES is an average of tail VaRs. The 90% confidence interval for our ES is [1.986, 2.271].

It is also interesting to compare the VaR and ES confidence intervals obtained by the two methods. These are summarised in Table 2-1, and we can see that the OS and bootstrap approaches give very similar results. This suggests that either approach is likely to be a reasonable one to use in practice.

WEIGHTED HISTORICAL SIMULATION

One of the most important features of traditional HS is the way it weights past observations. Recall that $R_{i,t}$ is the return on asset i in period t, and we are implementing HS using the past n observations. An observation $R_{i,t-j}$ will therefore belong to our data set if j takes any of the values $1, \ldots, t-n$, where j is the age of the observation (e.g., so $j = 1$ indicates that the observation is 1 day old, and so on). If we construct a new HS P/L series, P/L_t, each day, our observation $R_{i,t-j}$ will first affect P/L_t, then affect P/L_{t+1}, and so on, and finally affect P/L_{t+n}: our return observation will affect each of the next n observations in our P/L series. Also, other things (e.g., position weights) being equal, $R_{i,t-j}$ will affect each P/L in exactly the same way. But after n periods have passed, $R_{i,t-j}$ will fall out of the data set used to calculate the current HS P/L series, and will thereafter have no effect on P/L. In short, our HS P/L series is constructed in a way that gives any observation the *same* weight on P/L provided it is less than n periods old, and *no* weight (i.e., a zero weight) if it is older than that.

This weighting structure has a number of problems. One problem is that it is hard to justify giving each observation in our sample period the same weight, regardless of age, market volatility, or anything else. A good example of the difficulties this can create is given by Shimko et. al. (1998). It is well known that natural gas prices are usually more volatile in the winter than in the summer, so a raw HS approach that incorporates both summer and winter observations will tend to average the summer and winter observations together. As a result, treating all observations as having equal weight will tend to underestimate true risks in the winter, and overestimate them in the summer.[7]

[7] If we have data that show seasonal volatility changes, a solution—suggested by Shimko et. al. (1998)—is to weight the data to reflect seasonal volatility (e.g., so winter observations get more weight, if we are estimating a VaR in winter).

The equal-weight approach can also make risk estimates unresponsive to major events. For instance, a stock market crash might have no effect on VaRs except at a very high confidence level, so we could have a situation where everyone might agree that risk had suddenly increased, and yet that increase in risk would be missed by most HS VaR estimates. The increase in risk would only show up later in VaR estimates if the stock market continued to fall in subsequent days—a case of the stable door closing only well after the horse had long since bolted. That said, the increase in risk *would* show up in ES estimates just after the first shock occurred—which is, incidentally, a good example of how ES can be a more informative risk measure than the VaR.[8]

The equal-weight structure also presumes that each observation in the sample period is equally likely and independent of the others over time. However, this 'iid' assumption is unrealistic because it is well known that volatilities vary over time, and that periods of high and low volatility tend to be clustered together. The natural gas example just considered is a good case in point.

It is also hard to justify why an observation should have a weight that suddenly goes to zero when it reaches age n. Why is it that an observation of age $n - 1$ is regarded as having a lot of value (and, indeed, the same value as any more recent observation), but an observation of age n is regarded as having no value at all? Even old observations usually have some information content, and giving them zero value tends to violate the old statistical adage that we should never throw information away.

This weighting structure also creates the potential for ghost effects—we can have a VaR that is unduly high (or low) because of a small cluster of high loss observations, or even just a single high loss, and the measured VaR will continue to be high (or low) until n days or so have passed and the observation has fallen out of the sample period. At that point, the VaR will fall again, but the fall in VaR is only a ghost effect created by the weighting structure and the length of sample period used.

We now address various ways in which we might 'adjust' our data to overcome some of these problems and take account of ways in which current market conditions might differ from those in our sample. These fall under the broad heading of 'weighted historical simulation' and can be regarded as semi-parametric methods because they combine features of both parametric and non-parametric methods.

Age-weighted Historical Simulation

One such approach is to weight the relative importance, of our observations by their age, as suggested by Boudoukh, Richardson and Whitelaw (BRW: 1998). Instead of treating each observation for asset i as having the same implied probability as any other (i.e., $1/n$), we could weight their probabilities to discount the older observations in favour of newer ones. Thus, if $w(1)$ is the probability weight given to an observation 1 day old, then $w(2)$, the probability given to an observation 2 days old, could be $\lambda w(1)$; $w(3)$ could be $\lambda^2 w(1)$; and so on. The λ term is between 0 and 1, and reflects the exponential rate of decay in the weight or value given to an observation as it ages: a λ close to 1 indicates a slow rate of decay, and a λ far away from 1 indicates a high rate of decay. $w(1)$ is set so that the sum of the weights is 1, and this is achieved if we set $w(1) = (1 - \lambda)/(1 - \lambda^n)$. The weight given to an observation i days old is therefore:

$$w(i) = \frac{\lambda^{i-1}(1-\lambda)}{1-\lambda^n} \tag{2.2}$$

and this corresponds to the weight of $1/n$ given to any in-sample observation under basic HS.

Our core information—the information inputted to the HS estimation process—is the paired set of P/L values and associated probability weights. To implement age-weighting, we merely replace the old equal weights 1/n with the age-dependent weights *w(i)* given by (2.4). For example, if we are using a spreadsheet, we can order our P/L observations in one column, put their weights *w(i)* in the next column, and go down that column until we reach our desired percentile. Our VaR is then the negative of the corresponding value in the first column. And if our desired percentile falls between two percentiles, we can take our

[8] However, both VaR and ES suffer from a related problem. As Pritsker (2001, p. 5) points out, HS fails to take account of useful information from the upper tail of the P/L distribution. If the stock experiences a series of large falls, then a position that was long the market would experience large losses that should show up, albeit later, in HS risk estimates. However, a position that was short the market would experience a series of large profits, and risk estimates at the usual confidence levels would be completely unresponsive. Once again, we could have a situation where risk had clearly increased—because the fall in the market signifies increased volatility, and therefore a significant chance of losses due to large rises in the stock market—and yet our risk estimates had failed to pick up this increase in risk.

VaR to be the (negative of the) interpolated value of the corresponding first-column observations.

This age-weighted approach has four major attractions. First, it provides a nice generalisation of traditional HS, because we can regard traditional HS as a special case with zero decay, or $\lambda \rightarrow 1$. If HS is like driving along a road looking only at the rear-view mirror, then traditional equal-weighted HS is only safe if the road is straight, and the age-weighted approach is safe if the road bends gently.

Second, a suitable choice of λ can make the VaR (or ES) estimates more responsive to large loss observations: a large loss event will receive a higher weight than under traditional HS, and the resulting next-day VaR would be higher than it would otherwise have been. This not only means that age-weighted VaR estimates are more responsive to large losses, but also makes them better at handling clusters of large losses.

Third, age-weighting helps to reduce distortions caused by events that are unlikely to recur, and helps to reduce ghost effects. As an observation ages, its probability weight gradually falls and its influence diminishes gradually over time. Furthermore, when it finally falls out of the sample period, its weight will fall from $\lambda^n w(1)$ to zero, instead of from $1/n$ to zero. Since $\lambda^n w(1)$ is less than $1/n$ for any reasonable values of λ and n, then the shock—the ghost effect—will be less than it would be under equal-weighted HS.

Finally, we can also modify age-weighting in a way that makes our risk estimates more efficient and effectively eliminates any remaining ghost effects. Since age-weighting allows the impact of past extreme events to decline as past events recede in time, it gives us the option of letting our sample size grow over time. (Why can't we do this under equal-weighted HS? Because we would be stuck with ancient observations whose information content was assumed never to date.) Age-weighting allows us to let our sample period grow with each new observation, so we never throw potentially valuable information away. This would improve efficiency and eliminate ghost effects, because there would no longer be any 'jumps' in our sample resulting from old observations being thrown away.

However, age-weighting also reduces the effective sample size, other things being equal, and a sequence of major profits or losses can produce major distortions in its implied risk profile. In addition, Pritsker shows that even with age-weighting, VaR estimates can still be insufficiently responsive to changes in underlying risk.[9] Furthermore, there is the disturbing point that the BRW approach is ad hoc, and that except for the special case where $\lambda = 1$, we cannot point to any asset-return process for which the BRW approach is theoretically correct.

Volatility-weighted Historical Simulation

We can also weight our data by volatility. The basic idea—suggested by Hull and White (HW; 1998b)—is to update return information to take account of recent changes in volatility. For example, if the current volatility in a market is 1.5% a day, and it was only 1% a day a month ago, then data a month old understate the changes we can expect to see tomorrow, and this suggests that historical returns would underestimate tomorrow's risks; on the other hand, if last month's volatility was 2% a day, month-old data will overstate the changes we can expect tomorrow, and historical returns would overestimate tomorrow's risks. We therefore adjust the historical returns to reflect how volatility tomorrow is believed to have changed from its past values.

Suppose we are interested in forecasting VaR for day T. Let $r_{t,i}$ be the historical return in asset i on day t in our historical sample, $\sigma_{t,i}$ be the historical GARCH (or EWMA) forecast of the volatility of the return on asset i for day t, made at the end of day $t-1$, and $\sigma_{T,i}$ be our most recent forecast of the volatility of asset i. We then replace the returns in our data set, $r_{t,i}$, with volatility-adjusted returns, given by:

$$r^*_{t,i} = \left(\frac{\sigma_{T,i}}{\sigma_{t,i}}\right) r_{t,i} \tag{2.3}$$

Actual returns in any period t are therefore increased (or decreased), depending on whether the current forecast of volatility is greater (or less than) the estimated volatility for period t. We now calculate the HS P/L using

[9] If VaR is estimated at the confidence level α, the probability of an HS estimate of VaR rising on any given day is equal to the probability of a loss in excess of VaR, which is of course $1-\alpha$. However, if we assume a standard GARCH(1,1) process and volatility is at its long-run mean value, then Pritsker's proposition 2 shows that the probability that HSVaR should increase is about 32% (Pritsker (2001, pp. 7-9)). In other words, most of the time HS VaR estimates should increase (i.e., when risk rises), they fail to.

Equation (2.3) instead of the original data set $r_{t,i}$, and then proceed to estimate HS VaRs or ESs in the traditional way (i.e., with equal weights, etc.).[10]

The HW approach has a number of advantages relative to the traditional equal-weighted and/or the BRW age-weighted approaches:

- It takes account of volatility changes in a natural and direct way, whereas equal-weighted HS ignores volatility changes and the age-weighted approach treats volatility changes in a rather arbitrary and restrictive way.
- It produces risk estimates that are appropriately sensitive to current volatility estimates, and so enables us to incorporate information from GARCH forecasts into HS VaR and ES estimation.
- It allows us to obtain VaR and ES estimates that can exceed the maximum loss in our historical data set: in periods of high volatility, historical returns are scaled upwards, and the HS P/L series used in the HW procedure will have values that exceed actual historical losses. This is a major advantage over traditional HS, which prevents the VaR or ES from being any bigger than the losses in our historical data set.
- Empirical evidence presented by HW indicates that their approach produces superior VaR estimates to the BRW one.

The HW approach is also capable of various extensions. For instance, we can combine it with the age-weighted approach if we wished to increase the sensitivity of risk estimates to large losses, and to reduce the potential for distortions and ghost effects. We can also combine the HW approach with OS or bootstrap methods to estimate confidence intervals for our VaR or ES—that is, we would work with order statistics or resample with replacement from the HW-adjusted P/L, rather than from the traditional HS P/L.

Correlation-weighted Historical Simulation

We can also adjust our historical returns to reflect changes between historical and current correlations. Correlation-weighting is a little more involved than volatility-weighting. To see the principles involved, suppose for the sake of argument that we have already made any volatility-based adjustments to our HS returns along Hull-White lines, but also wish to adjust those returns to reflect changes in correlations.[11]

To make the discussion concrete, we have m positions and our (perhaps volatility adjusted) $1 \times m$ vector of historical returns $\mathbf{R}$ for some period t reflects an $m \times m$ variance-covariance matrix Σ. Σ in turn can be decomposed into the product $\sigma\mathbf{C}\sigma^{\mathbf{T}}$, where σ is an $m \times m$ diagonal matrix of volatilities (i.e., so the ith element of σ is the ith volatility σ_i and the off-diagonal elements are zero), $\sigma^{\mathbf{T}}$ is its transpose, and $\mathbf{C}$ is the $m \times m$ matrix of historical correlations. $\mathbf{R}$ therefore reflects an historical correlation matrix $\mathbf{C}$, and we wish to adjust $\mathbf{R}$ so that they become $\bar{\mathbf{R}}$ reflecting a current correlation matrix $\bar{\mathbf{C}}$. Now suppose for convenience that both correlation matrices are positive definite. This means that each correlation matrix has an $m \times m$ 'matrix square root', $\mathbf{A}$ and $\bar{\mathbf{A}}$ respectively, given by a Choleski decomposition (which also implies that they are easy to obtain). We can now write $\mathbf{R}$ and $\bar{\mathbf{R}}$ as matrix products of the relevant Choleski matrices and an uncorrelated noise process ε:

$$\mathbf{R} = \mathbf{A}\varepsilon \tag{2.4a}$$

$$\bar{\mathbf{R}} = \bar{\mathbf{A}}\varepsilon \tag{2.4b}$$

We then invert Equation (2.4a) to obtain $\varepsilon = \mathbf{A}^{-1}\mathbf{R}$, and substitute this into (Equation 2.4b) to obtain the correlation-adjusted series $\bar{\mathbf{R}}$ that we are seeking:

$$\bar{\mathbf{R}} = \bar{\mathbf{A}}\mathbf{A}^{-1}\mathbf{R} \tag{2.5}$$

The returns adjusted in this way will then have the currently prevailing correlation matrix $\mathbf{C}$ and, more generally, the currently prevailing covariance matrix $\bar{\Sigma}$. This approach is a major generalisation of the HW approach, because it gives us a weighting system that takes account of correlations as well as volatilities.

Example 2.1 Correlation-weighted HS

Suppose we have only two positions in our portfolio, so $m = 2$. The historical correlation between our two positions is 0.3, and we wish to adjust our historical returns $\mathbf{R}$ to reflect a current correlation of 0.9.

[10] Naturally, volatility weighting presupposes that one has estimates of the current and past volatilities to work with.

[11] The correlation adjustment discussed here is based on a suggestion by Duffie and Pan (1997).

If a_{ij} is the i, jth element of the 2×2 matrix $\mathbf{A}$, then applying the Choleski decomposition tells us that

$$a_{11} = 1, \quad a_{12} = 0, \quad a_{21} = \rho, \quad a_{22} = \sqrt{1-\rho^2}$$

where $\rho = 0.3$. The matrix $\bar{\mathbf{A}}$ is similar except for having $\rho = 0.9$. Standard matrix theory also tells us that

$$\mathbf{A}^{-1} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} a_{22}, -a_{12} \\ -a_{21}, a_{11} \end{bmatrix}$$

Substituting these into Equation (2.5), we find that

$$\begin{aligned}
\bar{\mathbf{R}} = \bar{\mathbf{A}}\mathbf{A}^{-1}\mathbf{R} &= \begin{bmatrix} 1, 0 \\ 0.9, \sqrt{1-0.9^2} \end{bmatrix} \frac{1}{\sqrt{1-0.3^2}} \begin{bmatrix} \sqrt{1-0.3^2}, 0 \\ -0.3, 1 \end{bmatrix} \mathbf{R} \\
&= \frac{1}{\sqrt{1-0.3^2}} \begin{bmatrix} \sqrt{1-0.3^2}, 0 \\ 0.9\sqrt{1-0.3^2} - 0.3\sqrt{1-0.9^2}, \sqrt{1-0.9^2} \end{bmatrix} \mathbf{R} \\
&= \begin{bmatrix} 1, 0 \\ 0.7629, 0.4569 \end{bmatrix} \mathbf{R}
\end{aligned}$$

Filtered Historical Simulation

Another promising approach is filtered historical simulation (FHS).[12] This is a form of semi-parametric bootstrap which aims to combine the benefits of HS with the power and flexibility of conditional volatility models such as GARCH. It does so by bootstrapping returns within a conditional volatility (e.g., GARCH) framework, where the bootstrap preserves the non-parametric nature of HS, and the volatility model gives us a sophisticated treatment of volatility.

Suppose we wish to use FHS to estimate the VaR of a single-asset portfolio over a 1-day holding period. The first step in FHS is to fit, say, a GARCH model to our portfolio-return data. We want a model that is rich enough to accommodate the key features of our data, and Barone-Adesi and colleagues recommend an asymmetric GARCH, or AGARCH, model. This not only accommodates conditionally changing volatility, volatility clustering, and so on, but also allows positive and negative returns to have differential impacts on volatility, a phenomenon known as the leverage effect. The AGARCH postulates that portfolio returns obey the following process:

$$r_t = \mu + \varepsilon_t \tag{2.6a}$$

$$\sigma_t^2 = \omega + \alpha(\varepsilon_{t-1} + \gamma)^2 + \beta\sigma_{t-1}^2 \tag{2.6b}$$

The daily return in Equation (2.6a) is the sum of a mean daily return (which can often be neglected in volatility estimation) and a random error ε_t. The volatility in Equation (2.6b) is the sum of a constant and terms reflecting last period's 'surprise' and last period's volatility, plus an additional term γ that allows for the surprise to have an asymmetric effect on volatility, depending on whether the surprise term is positive or negative.

The second step is to use the model to forecast volatility for each of the days in a sample period. These volatility forecasts are then divided into the realised returns to produce a set of standardised returns. These standardised returns should be independently and identically distributed (iid), and therefore be suitable for HS.

Assuming a 1-day VaR holding period, the third stage involves bootstrapping from our data set of standardised returns: we take a large number of drawings from this data set, which we now treat as a sample, replacing each one after it has been drawn, and multiply each random drawing by the AGARCH forecast of tomorrow's volatility. If we take M drawings, we therefore get M simulated returns, each of which reflects current market conditions because it is scaled by today's forecast of tomorrow's volatility.

Finally, each of these simulated returns gives us a possible end-of-tomorrow portfolio value, and a corresponding possible loss, and we take the VaR to be the loss corresponding to our chosen confidence level.[13]

We can easily modify this procedure to encompass the obvious complications of a multi asset portfolio or a longer holding period. If we have a multi-asset portfolio, we would fit a multivariate GARCH (or AGARCH) to the set or vector of asset returns, and we would standardise this vector of asset returns. The bootstrap would then select, not just a standardised portfolio return for some chosen past (daily) period, but the standardised vector of asset returns for the chosen past period. This is important because it means that our simulations would keep any correlation structure present in the raw returns. The

[12] This approach is suggested in Barone-Adesi et. al. (1998), Barone-Adesi et. al. (1999), Barone-Adesi aud Giannopoulos (2000) and in other papers by some of the same authors.

[13] The FHS approach can also be extended easily to allow for the estimation of ES as well as VaR. For more on how this might be done, see Giannopoulos and Tunaru (2004).

bootstrap thus maintains existing correlations, without our having to specify an explicit multivariate pdf for asset returns.

The other obvious extension is to a longer holding period. If we have a longer holding period, we would first take a drawing and use Equation (2.8) to get a return for tomorrow; we would then use this drawing to update our volatility forecast for the day after tomorrow, and take a fresh drawing to determine the return for that day; and we would carry on in the same manner—taking a drawing, updating our volatility forecasts, taking another drawing for the next period, and so on—until we had reached the end of our holding period. At that point we would have enough information to produce a single simulated P/L observation; and we would repeat the process as many times as we wished in order to produce the histogram of simulated P/L observations from which we can estimate our VaR.

FHS has a number of attractions: (i) It enables us to combine the non-parametric attractions of HS with a sophisticated (e.g., GARCH) treatment of volatility, and so take account of changing market volatility conditions. (ii) It is fast, even for large portfolios. (iii) As with the earlier HW approach, FHS allows us to get VaR and ES estimates that can exceed the maximum historical loss in our data set. (iv) It maintains the correlation structure in our return data without relying on knowledge of the variance-covariance matrix or the conditional distribution of asset returns. (v) It can be modified to take account of autocorrelation or past cross-correlations in asset returns. (vi) It can be modified to produce estimates of VaR or ES confidence intervals by combining it with an OS or bootstrap approach to confidence interval estimation.[14] (vii) There is evidence that FHS works well.[15]

ADVANTAGES AND DISADVANTAGES OF NON-PARAMETRIC METHODS

Advantages

In drawing our discussion to a close, it is perhaps a good idea to summarise the main advantages and disadvantages of non-parametric approaches. The advantages include:

- Non-parametric approaches are intuitive and conceptually simple.
- Since they do not depend on parametric assumptions about P/L, they can accommodate fat tails, skewness, and any other non-normal features that can cause problems for parametric approaches.
- They can in theory accommodate any type of position, including derivatives positions.
- There is a widespread perception among risk practitioners that HS works quite well empirically, although formal empirical evidence on this issue is inevitably mixed.
- They are (in varying degrees, fairly) easy to implement on a spreadsheet.
- Non-parametric methods are free of the operational problems to which parametric methods are subject when applied to high-dimensional problems: no need for covariance matrices, no curses of dimensionality, etc.
- They use data that are (often) readily available, either from public sources (e.g., Bloomberg) or from in-house data sets (e.g., collected as a by-product of marking positions to market).
- They provide results that are easy to report and communicate to senior managers and interested outsiders (e.g., bank supervisors or rating agencies).
- It is easy to produce confidence intervals for non-parametric VaR and ES.

[14] The OS approach would require a set of paired P/L and associated probability observations, so we could apply this to FHS by using a P/L series that had been through the FHS filter. The bootstrap is even easier, since FHS already makes use of a bootstrap. If we want *B* bootstrapped estimates of VaR, we could produce, say, 100**B* or 1000**B* bootstrapped P/L values; each set of 100 (or 1000) P/L series would give us one HS VaR estimate, and the histogram of *M* such estimates would enable us to infer the bounds of the VaR confidence interval.

[15] Barone-Adesi and Giannopoulos (2000), p. 17. However, FHS does have problems. In his thorough simulation study of FHS, Pritsker (2001, pp. 22–24) comes to the tentative conclusions that FHS VaR might not pay enough attention to extreme observations or time-varying correlations, and Barone-Adesi and Giannopoulos (2000, p. 18) largely accept these points. A partial response to the first point would be to use ES instead of VaR as our preferred risk measure, and the natural response to the second concern is to develop FHS with a more sophisticated past cross-correlation structure. Pritsker (2001, p. 22) also presents simulation results that suggest that FHS-VaR tends to underestimate 'true' VaR over a 10-day holding period by about 10%, but this finding conflicts with results reported by Barone-Adesi et. al. (2000) based on real data. The evidence on FHS is thus mixed.

- Non-parametric approaches are capable of considerable refinement and potential improvement if we combine them with parametric 'add-ons' to make them semi-parametric: such refinements include age-weighting (as in BRW), volatility-weighting (as in HW and FHS), and correlation-weighting.

Disadvantages

Perhaps their biggest potential weakness is that their results are very (and in most cases, completely) dependent on the historical data set.[16] There are various other related problems:

- If our data period was unusually quiet, non-parametric methods will often produce VaR or ES estimates that are too low for the risks we are actually facing; and if our data period was unusually volatile, they will often produce VaR or ES estimates that are too high.
- Non-parametric approaches can have difficulty handling shifts that take place during our sample period. For example, if there is a permanent change in exchange rate risk, it will usually take time for the HS VaR or ES estimates to reflect the new exchange rate risk. Similarly, such approaches are sometimes slow to reflect major events, such as the increases in risk associated with sudden market turbulence.
- If our data set incorporates extreme losses that are unlikely to recur, these losses can dominate non-parametric risk estimates even though we don't expect them to recur.
- Most (if not all) non-parametric methods are subject (to a greater or lesser extent) to the phenomenon of ghost or shadow effects.
- In general, non-parametric estimates of VaR or ES make no allowance for plausible events that might occur, but did not actually occur, in our sample period.
- Non-parametric estimates of VaR and ES are to a greater or lesser extent constrained by the largest loss in our historical data set. In the simpler versions of HS, we cannot extrapolate from the largest historical loss to anything larger that might conceivably occur in the future. More sophisticated versions of HS can relax this constraint, but even so, the fact remains that non-parametric estimates of VaR or ES are still constrained by the largest loss in a way that parametric estimates are not. This means that such methods are not well suited to handling extremes, particularly with small- or medium-sized samples.

However, we can often ameliorate these problems by suitable refinements. For example, we can ameliorate volatility, market turbulence, correlation and other problems by semi-parametric adjustments, and we can ameliorate ghost effects by age-weighting our data and allowing our sample size to rise over time.

There can also be problems associated with the length of the sample window period. We need a reasonably long window to have a sample size large enough to get risk estimates of acceptable precision, and as a broad rule of thumb, most experts believe that we usually need at least a couple of year's worth of daily observations (i.e., 500 observations, at 250 trading days to the year), and often more. On the other hand, a very long window can also create its own problems. The longer the window:

- the greater the problems with aged data;
- the longer the period over which results will be distorted by unlikely-to-recur past events, and the longer we will have to wait for ghost effects to disappear;
- the more the news in current market observations is likely to be drowned out by older observations—and the less responsive will be our risk estimates to current market conditions; and
- the greater the potential for data-collection problems. This is a particular concern with new or emerging market instruments, where long runs of historical data don't exist and are not necessarily easy to proxy.

CONCLUSIONS

Non-parametric methods are widely used and in many respects highly attractive approaches to the estimation of financial risk measures. They have a reasonable track record and are often superior to parametric approaches based on simplistic assumptions such as normality. They are also capable of considerable refinement to deal with

[16] There can also be problems getting the data set. We need time series data on all current positions, and such data are not always available (e.g., if the positions are in emerging markets). We also have to ensure that data are reliable, compatible, and delivered to the risk estimation system on a timely basis.

some of the weaknesses of more basic non-parametric approaches. As a general rule, they work fairly well if market conditions remain reasonably stable, and are capable of considerable refinement. However, they have their limitations and it is often a good idea to supplement them with other approaches. Wherever possible, we should also complement non-parametric methods with stress testing to gauge our vulnerability to 'what if' events. We should never rely on non-parametric methods alone.

APPENDIX 1

Estimating Risk Measures with Order Statistics

The theory of order statistics is very useful for risk measurement because it gives us a practical and accurate means of estimating the distribution function for a risk measure—and this is useful because it enables us to estimate confidence intervals for them.

Using Order Statistics to Estimate Confidence Intervals for VaR

If we have a sample of n P/L observations, we can regard each observation as giving an estimate of VaR at an implied confidence level. For example, if $n = 1000$, we might take the 95% VaR as the negative of the 51st smallest P/L observation, we might take the 99% VaR as the negative of the 11th smallest, and so on. We therefore take the α VaR to be equal to the negative of the rth lowest observation, where r is equal to $100(1 - \alpha) + 1$. More generally, with n observations, we take the VaR as equal to the negative of the rth lowest observation, where $r = n(1 - \alpha) + 1$.

The rth order statistic is the rth lowest (or, alternatively, highest) in a sample of n observations, and the theory of order statistics is well established in the statistical literature. Suppose our observations $x_1, x_2, \ldots, x_n$ come from some known distribution (or cumulative density) function $F(x)$, with rth order statistic $x_{(r)}$. Now suppose that $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$. The probability that j of our n observations do not exceed a fixed value x must obey the following binomial distribution:

$$\Pr\{j \text{ observations} \leq x\} = \binom{n}{j}\{F(x)\}^j\{1 - F(x)\}^{n-j} \quad (2.7)$$

It follows that the probability that at least r observations in the sample do not exceed x is also a binomial:

$$G_r(x) = \sum_{j=r}^{n}\binom{n}{j}\{F(x)\}^j\{1 - F(x)\}^{n-j} \quad (2.8)$$

$G_r(x)$ is therefore the distribution function of our order statistic and, hence, of our quantile or VaR.[17]

This VaR distribution function provides us with estimates of our VaR *and* of its associated confidence intervals. The median (i.e., 50 percentile) of the estimated VaR distribution function gives us a natural 'best' estimate of our VaR, and estimates of the lower and upper percentiles of the VaR distribution function give us estimates of the bounds of our VaR confidence interval. This is useful, because the calculations are accurate and easy to carry out on a spreadsheet. Equation (2.8) is also very general and gives us confidence intervals for *any* distribution function $F(x)$, parametric (normal, t, etc.) or empirical.

To use this approach, all we need to do is specify $F(x)$ (as normal, t, etc.), set our parameter values, and use Equation (2.8) to estimate our VaR distribution function.

To illustrate, suppose we want to apply the order-statistics (OS) approach to estimate the distribution function of a standard normal VaR. We then assume that $F(x)$ is standard normal and use Equation (2.8) to estimate three key parameters of the VaR distribution: the median or 50 percentile of the estimated VaR distribution, which can be interpreted as an OS estimate of normal VaR; and the 5 and 95 percentiles of the estimated VaR distribution, which can be interpreted as the OS estimates of the bounds of the 90% confidence interval for standard normal VaR.

Some illustrative estimates for the 95% VaR are given in Table 2-2. To facilitate comparison, the table also shows the estimates of standard normal VaR based on the conventional normal VaR formula as explained in Chapter 1. The main results are:

- The confidence interval—the gap between the 5 and 95 percentiles—is quite wide for low values of n, but narrows as n gets larger.
- As n rises, the median of the estimated VaR distribution converges to the conventional estimate.
- The confidence interval is (in this case, a little) wider for more extreme VaR confidence levels than it is for the more central ones.

[17] See, e.g., Kendall and Stuart (1973), p. 348, or Reiss (1989), p. 20.

TABLE 2-2 Order Statistics Estimates of Standard Normal 95% VaRs and Associated Confidence Intervals

(a) As n varies

No. of observations	100	500	1000	5000	10 000
Lower bound of confidence interval	1.267	1.482	1.531	1.595	1.610
Median of VaR distribution	1.585	1.632	1.639	1.644	1.644
Standard estimate of VaR	1.645	1.645	1.645	1.645	1.645
Upper bound of confidence interval	1.936	1.791	1.750	1.693	1.679
Width of interval/median	42.2%	18.9%	13.4%	6.0%	4.2%

(b) As VaR confidence level varies (with n = 500)

VaR confidence level	0.90	0.95	0.99
Lower bound of confidence interval	1.151	1.482	2.035
Median of VaR distribution	1.274	1.632	2.279
Standard estimate of VaR	1.282	1.645	2.326
Upper bound of confidence interval	1.402	1.791	2.560
Width of interval/median of interval	19.7%	18.9%	23.0%

Notes: The confidence interval is specified at a 90% level of confidence, and the lower and upper bounds of the confidence interval are estimated as the 5 and 95 percentiles of the estimated VaR distribution (Equation (2.8)).

The same approach can also be used to estimate the percentiles of other VaR distribution functions. If we wish to estimate the percentiles of a non-normal parametric VaR, we replace the normal distribution function $F(x)$ by the non-normal equivalent—the t-distribution function, the Gumbel distribution function, and so on. We can also use the same approach to estimate the confidence intervals for an empirical distribution function (i.e., for historical simulation VaR), where $F(x)$ is some empirical distribution function.

Conclusions

The OS approach provides an ideal method for estimating the confidence intervals for our VaRs and ESs. In particular, the OS approach is:

- Completely general, in that it can be applied to any parametric or non-parametric VaR or ES.
- Reasonable even for relatively small samples, because it is not based on asymptotic theory—although it is also the case that estimates based on small samples will also be less accurate, precisely because the samples are small.
- Easy to implement in practice.

The OS approach is also superior to confidence-interval estimation methods based on estimates of quantile standard errors (see Chapter 1), because it does not rely on asymptotic theory and/or force estimated confidence intervals to be symmetric (which can be a problem for extreme VaRs and ESs).

APPENDIX 2

The Bootstrap

The bootstrap is a simple and useful method for assessing uncertainty in estimation procedures. Its distinctive feature is that it replaces mathematical or statistical analysis with simulation-based resampling from a given data set. It therefore provides a means of assessing the accuracy of parameter estimators without having to resort to strong parametric assumptions or closed-form confidence-interval formulas. The roots of the bootstrap go back a couple of centuries, but the idea only took off in the last three decades after it was developed and popularised by the work of Bradley Efron. It was Efron, too, who first gave it its name, which refers to the phrase 'to pull oneself up by one's bootstraps'. The bootstrap is a form of statistical 'trick', and is therefore very aptly named.

The main purpose of the bootstrap is to assess the accuracy of parameter estimates. The bootstrap is ideally suited for this purpose, as it can provide such estimates without having to rely on potentially unreliable assumptions (e.g., assumptions of normality or large samples).[18] The bootstrap is also easy to use because it does not require the user to engage in any difficult mathematical or

[18] The bootstrap is also superior to the jackknife, which was often used for similar purposes before the advent of powerful computers. The jackknife is a procedure in which we construct a large number of subsamples from an original sample by taking the original sample and leaving one observation out at a time. For each such subsample, we estimate the parameter of interest, and the jackknife estimator is the average of the subsample-based estimators. The jackknife can also be regarded as an approximation to the bootstrap, but it can provide a very poor approximation when the parameter estimator is a non-smooth function of the data. The bootstrap is therefore more reliable and easier to implement.

statistical analysis. In any case, such traditional methods only work in a limited number of cases, whereas the bootstrap can be applied more or less universally. So the bootstrap is easier to use, more powerful and (as a rule) more reliable than traditional means of estimating confidence intervals for parameters of interest. In addition, the bootstrap can be used to provide alternative 'point' estimates of parameters as well.[19]

Limitations of Conventional Sampling Approaches

The bootstrap is best appreciated by considering the limitations of conventional sampling approaches. Suppose we have a sample of size n drawn from a population. The parameters of the population distribution are unknown—and, more likely than not, so too is the distribution itself. We are interested in a particular parameter θ, where θ might be a mean, variance (or standard deviation), quantile, or some other parameter. The obvious approach is to estimate θ using a suitable sample estimator—so if θ is the mean, our estimator $\hat{\theta}$ would be the sample mean, if θ is the variance, our estimator $\hat{\theta}$ would be based on some sample variance, and so on. Obtaining an estimator for θ is therefore straightforward, but how do we obtain a confidence interval for it?

To estimate confidence intervals for θ using traditional closed-form approaches requires us to resort to statistical theory, and the theory available is of limited use. For example, suppose we wish to obtain a confidence interval for a variance. If we assume that the underlying distribution is normal, then we know that $(n-1)\hat{\sigma}^2/\sigma^2$ is distributed as χ^2 with $n-1$ degrees of freedom, and this allows us to obtain a confidence interval for σ^2. If we denote the α point of this distribution as $\chi^2_{\alpha,n-1}$, then the 90% confidence interval for $(n-1)\hat{\sigma}^2/\sigma^2$ is:

$$\left[\chi^2_{0.05,n-1}, \chi^2_{0.95,n-1}\right] \tag{2.9}$$

This implies that the 90% confidence interval for σ^2 is:

$$\left[\frac{(n-1)\hat{\sigma}^2}{\chi_{0.95,n-1}}, \frac{(n-1)\hat{\sigma}^2}{\chi_{0.05,n-1}}\right] \tag{2.10}$$

[19] The bootstrap also has other uses too. For example, it can be used to relax and check assumptions, to give quick approximations and to check the results obtained using other methods.

On the other hand, if we cannot assume that the underlying distribution is normal, then obtaining a confidence interval for σ^2 can become very difficult: the problem is that although we can estimate σ^2 itself, under more general conditions we would often not know the distribution of σ^2, or have expressions for standard errors, and we cannot usually obtain closed-form confidence intervals without them.

We can face similar problems with other parameters as well, such as medians, correlations, and tail probabilities.[20] So in general, closed-form confidence intervals are of limited applicability, and will not apply to many of the situations we are likely to meet in practice.

The Bootstrap and Its Implementation

The bootstrap frees us of this type of limitation, and is also much easier to implement. It enables us to estimate a confidence interval for any parameter that we can estimate, regardless of whether we have any formulas for the distribution function for that parameter or for the standard error of its estimator. The bootstrap also has the advantage that it comes with less baggage, in the sense that the assumptions needed to implement the bootstrap are generally less demanding than the assumptions needed to estimate confidence intervals using more traditional (i.e., closed-form) methods.

The basic bootstrap procedure is very simple.[21] We start with a given original sample of size n.[22] We now draw a new random sample of the same size from this original sample, taking care to replace each chosen observation back in the sample pool after it has been drawn. (This random sampling, or resampling, is the very heart of the

[20] However, in the case of quantiles, we can use order statistics to write down their distribution functions.

[21] This application of the bootstrap can be described as a non-parametric one because we bootstrap from a given data sample. The bootstrap can also be implemented parametrically, where we bootstrap from the assumed distribution. When used in parametric mode, the bootstrap provides more accurate answers than textbook formulas usually do, and it can provide answers to problems for which no textbook formulas exist. The bootstrap can also be implemented semi-parametrically and a good example of this is the FRS approach.

[22] In practice, it might be possible to choose the value of n, but we will assume for the sake of argument that n is given.

bootstrap. It requires that we have a uniform random number generator to select a random number between 1 and n, which determines the particular observation that is chosen each time.) When constructing the new sample, known as a resample, we would typically find that some observations get chosen more than once, and others don't get chosen at all: so the resample would typically be different from the original one, even though every observation included in it was drawn from the original sample. Once we have our resample, we use it to estimate the parameter we are interested in. This gives us a resample estimate of the parameter. We then repeat the 'resampling' process again and again, and obtain a set of B resample parameter estimates. This set of B resample estimates can also be regarded as a bootstrapped sample of parameter estimates.

We can then use the bootstrapped sample to estimate a confidence interval for our parameter θ. For example, if each resample i gives us a resample estimator $\hat{\theta}^B(i)$ we might construct a simulated density function from the distribution of our $\hat{\theta}^B(i)$ values and infer the confidence intervals from its percentile points. If our confidence interval spans the central $1 - 2\alpha$ of the probability mass, then it is given by:

$$\textit{confidence interval} = [\hat{\theta}^B_\alpha, \hat{\theta}^B_{1-\alpha}] \quad \textbf{(2.11)}$$

where $\hat{\theta}^B_\alpha$ is the α quantile of the distribution of bootstrapped $\hat{\theta}^B(i)$ values. This 'percentile interval' approach is very easy to apply and does not rely on any parametric theory, asymptotic or otherwise.

Nonetheless, this basic percentile interval approach is limited itself, particularly if parameter estimators are biased. It is therefore often better to use more refined percentile approaches, and perhaps the best of these is the bias-corrected and accelerated (or BC_a) approach, which generates a substantial improvement in both theory and practice over the basic percentile interval approach. To use this approach we replace the α and $1 - \alpha$ subscripts in Equation (2.11) with α_1 and α_2, where

$$\alpha_1 = \Phi\left(\hat{z}^0 + \frac{\hat{z}^0 + z_\alpha}{1 - \hat{a}(\hat{z}^0 + z_\alpha)}\right),\ \alpha_2 = \Phi\left(\hat{z}^0 + \frac{\hat{z}^0 + z_{1-\alpha}}{1 - \hat{a}(\hat{z}^0 + z_{1-\alpha})}\right) \quad \textbf{(2.12)}$$

If the parameters $\hat{a}$ and $\hat{z}^0$ are zero, this BC_a confidence interval will coincide with the earlier percentile interval. However, in general, they will not be 0, and we can think of the BC_a method as correcting the end-points of the confidence interval. The parameter $\hat{a}$ refers to the rate of change of the standard error of $\hat{\theta}$ with respect to the true parameter θ, and it can be regarded as a correction for skewness. This parameter can be estimated from the following, which would be based on an initial bootstrap or jackknife exercise:

$$\hat{a} = \frac{\Sigma_{i=1}^M(\hat{\theta} - \hat{\theta}^B(i))^3}{6\left[\Sigma_{i=1}^M(\hat{\theta} - \hat{\theta}^B(i))^2\right]^{3/2}} \quad \textbf{(2.13)}$$

The parameter $\hat{z}^0$ can be estimated as the standard normal inverse of the proportion of bootstrap replications that is less than the original estimate $\hat{\theta}$. The BC_a method is therefore (relatively) straightforward to implement, and it has the theoretical advantages over the percentile interval approach of being both more accurate and of being transformation-respecting, the latter property meaning that if we take a transformation of θ (e.g., if θ is a variance, we might wish to take its square root to obtain the standard deviation), then the BC_a method will automatically correct the end-points of the confidence interval of the transformed parameter.[23]

We can also use a bootstrapped sample of parameter estimates to provide an alternative point estimator of a parameter that is often superior to the raw sample estimator $\hat{\theta}$. Given that there are B resample estimators, we can take our bootstrapped point estimator $\hat{\theta}^B$ as the sample mean of our B $\hat{\theta}^B(i)$ values:[24]

$$\hat{\theta}^B = \frac{1}{B}\sum_{i=1}^{B}\hat{\theta}^B(i) \quad \textbf{(2.14)}$$

Relatedly, we can also use a bootstrap to estimate the bias in an estimator. The bias is the difference between the expectation of an estimator and the quantity estimated (i.e., the bias equals $E[\hat{\theta}] - \theta$)), and can be estimated by plugging Equation (2.14) and a basic sample estimator $\hat{\theta}$ into the bias equation:

$$\textit{bias} = E[\hat{\theta}] - \theta \Rightarrow \textit{estimated bias} = \hat{\theta}^B - \hat{\theta} \quad \textbf{(2.15)}$$

We can use an estimate of bias for various purposes (e.g., to correct a biased estimator, to correct prediction errors,

[23] For more on BC_a and other refinements to the percentile interval approach, see Efron and Tibshirani (1993, Chapters 14 and 22) or Davison and Hinkley (1997, Chapter 5).

[24] This basic bootstrap estimation method can also be supplemented by variance-reduction methods (e.g., importance sampling) to improve accuracy at a given computational cost. See Efron and Tibshirani (1993, Chapter 23) or Davison and Hinkley (1997, Chapter 9).

etc.). However, the bias can have a (relatively) large standard error. In such cases, correcting for the bias is not always a good idea, because the bias-corrected estimate can have a larger standard error than the unadjusted, biased, estimate.

The programs to compute bootstrap statistics are easy to write and the most obvious price of the bootstrap, increased computation, is no longer a serious problem.[25]

Standard Errors of Bootstrap Estimators

Naturally, bootstrap estimates are themselves subject to error. Typically, bootstrap estimates have little bias, but they often have substantial variance. The latter comes from basic sampling variability (i.e., the fact that we have a sample of size n drawn from our population, rather than the population itself) and from resampling variability (i.e., the fact that we take only B bootstrap resamples rather than an infinite number of them). The estimated standard error for $\hat{\theta}$, $\hat{s}_B$, can be obtained from:

$$\hat{s}_B = \left(\frac{1}{B}\sum_{i=1}^{B}(\hat{\theta}^B(i) - \hat{\theta}^B)^2\right)^{1/2} \quad \textbf{(2.16)}$$

where $\hat{\theta}^B = (1/B)\Sigma_{i=1}^{B}\hat{\theta}^B(i)$. $\hat{s}_B$ is of course also easy to estimate. However, $\hat{s}_B$ is itself variable, and the variance of $\hat{s}_B$ is:

$$\text{var}(\hat{s}_B) = \text{var}[E(\hat{s}_B)] + E[\text{var}(\hat{s}_B)] \quad \textbf{(2.17)}$$

Following Efron and Tibshirani (1993, Chapter 19), this can be rearranged as:

$$\text{var}(\hat{s}_B) = \text{var}[\hat{m}_2^{1/2}] + E\left[\frac{\hat{m}_2}{4B}\left(\frac{\hat{m}_4}{\hat{m}_2^2} - 1\right)\right] \quad \textbf{(2.18)}$$

where $\hat{m}_i$ is the ith moment of the bootstrap distribution of the $\hat{\theta}^B(i)$. In the case where θ is the mean, Equation (2.18) reduces to:

$$\text{var}(\hat{s}_B) = \frac{\hat{m}_4/\hat{m}_2 - \hat{m}_2}{4n^2} + \frac{\sigma^2}{2nB} + \frac{\sigma^2(\hat{m}_4/\hat{m}_2^2 - 3)}{4n^2B} \quad \textbf{(2.19)}$$

If the distribution is normal, this further reduces to:

$$\text{var}(\hat{s}_B) = \frac{\sigma^2}{2n^2}\left(1 + \frac{n}{B}\right) \quad \textbf{(2.20)}$$

We can then set B to reduce var($\hat{s}_B$) to a desired level, and so achieve a target level of accuracy in our estimate of $\hat{s}_B$. However, these results are limited, because Equation (2.19) only applies to the mean and Equation (2.20) presupposes normality as well.

We therefore face two related questions: (a) how we can estimate var($\hat{s}_B$) in general? and (b) how can we choose B to achieve a given level of accuracy in our estimate of $\hat{s}_B$? One approach to these problems is to apply brute force: we can estimate var($\hat{s}_B$) using a jackknife-after-bootstrap (in which we first bootstrap the data and then estimate var($\hat{s}_B$) by jackknifing from the bootstrapped data), or by using a double bootstrap (in which we estimate a sample of bootstrapped $\hat{s}_B$ values and then estimate their variance). We can then experiment with different values of B to determine the values of these parameters needed to bring var($\hat{s}_B$) down to an acceptable level.

If we are more concerned about the second problem (i.e., how to choose B), a more elegant approach is the following, suggested by Andrews and Buchinsky (1997). Suppose we take as our 'ideal' the value of $\hat{s}_B$ associated with an infinite number of resamples, i.e., $\hat{s}_\infty$. Let τ be a target probability that is close to 1, and let *bound* be a chosen bound on the percentage deviation of $\hat{s}_B$ from $\hat{s}_\infty$. We want to choose $B = B(bound, \tau)$ such that the probability that $\hat{s}_B$ is within the desired bound is τ:

$$\Pr\left[100\left|\frac{\hat{s}_B - \hat{s}_\infty}{\hat{s}_B}\right| \leq bound\right] = \tau \quad \textbf{(2.21)}$$

If B is large, then the required number of resamples is approximately

$$B \approx \frac{2500(\kappa - 1)\chi_\tau^2}{bound^2} \quad \textbf{(2.22)}$$

However, this formula is not operational because κ, the kurtosis of the distribution of $\hat{\theta}^B$, is unknown. To get around this problem, we replace κ with a consistent estimator of κ, and this leads Andrews and Buchinsky to suggest the following three-step method to determine B:

- We initially assume that $\kappa = 3$, and plug this into Equation (2.22) to obtain a preliminary value of B, denoted by B_0, where

$$B_0 = \text{int}\left(\frac{5000\chi_\tau^2}{bound^2}\right) \quad \textbf{(2.23)}$$

  and where int(a) refers to the smallest integer greater than or equal to a.

[25] An example of the bootstrap approach applied to VaR is given earlier in this chapter discussing the bootstrap point estimator and bootstrapped confidence intervals for VaR.

- We simulate B_0 resamples, and estimate the sample kurtosis of the bootstrapped $\hat{\theta}^B$ values, $\hat{\kappa}$.
- We take the desired number of bootstrap resamples as equal to $\max(B_0, B_1)$, where

$$B_1 \approx \frac{2500(\hat{\kappa} - 1)\chi_{\tau}^2}{bound^2} \quad \textbf{(2.24)}$$

- This method does not directly tell us what the variance of $\hat{s}_B$ might be, but we already know how to estimate this in any case. Instead, this method gives us something more useful: it tells us how to set B to achieve a target level of precision in our bootstrap estimators, and (unlike Equations (2.19) and (2.20)) it applies for any parameter θ and applies however $\hat{\theta}^B$ is distributed.[26]

Time Dependency and the Bootstrap

Perhaps the main limitation of the bootstrap is that standard bootstrap procedures presuppose that observations are independent over time, and they can be unreliable if this assumption does not hold. Fortunately, there are various ways in which we can modify bootstraps to allow for such dependence:

- If we are prepared to make parametric assumptions, we can model the dependence parametrically (e.g., using a GARCH procedure). We can then bootstrap from the residuals, which should be independent. However, this solution requires us to identify the underlying stochastic model and estimate its parameters, and this exposes us to model and parameter risk.
- An alternative is to use a block approach: we divide sample data into non-overlapping blocks of equal length, and select a block at random. However, this approach can 'whiten' the data (as the joint observations spanning different blocks are taken to be independent), which can undermine our results. On the other hand, there are also various methods of dealing with this problem (e.g., making block lengths stochastic, etc.) but these refinements also make the block approach more difficult to implement.
- A third solution is to modify the probabilities with which individual observations are chosen. Instead of assuming that each observation is chosen with the same probability, we can make the probabilities of selection dependent on the time indices of recently selected observations: so, for example, if the sample data are in chronological order and observation i has just been chosen, then observation $i + 1$ is more likely to be chosen next than most other observations.

[26] This three-step method can also be improved and extended. For example, it can be improved by correcting for bias in the kurtosis estimator, and a similar (although more involved) three-step method can be used to achieve given levels of accuracy in estimates of confidence intervals as well. For more on these refinements, see Andrews and Buchinsky (1997).

Backtesting VaR

3

Learning Objectives

After completing this reading you should be able to:

- Define backtesting and exceptions, and explain the importance of backtesting VaR models.
- Explain the significant difficulties in backtesting a VaR model.
- Verify a model based on exceptions or failure rates.
- Define and identify type I and type II errors.
- Explain the need to consider conditional coverage in the backtesting framework.
- Describe the Basel rules for backtesting.

Excerpt is Chapter 6 of Value-at-Risk: The New Benchmark for Managing Financial Risk, *Third Edition, by Philippe Jorion.*

> Disclosure of quantitative measures of market risk, such as value-at-risk, is enlightening only when accompanied by a thorough discussion of how the risk measures were calculated and how they related to actual performance.
>
> —Alan Greenspan (1996)

Value-at-risk (VaR) models are only useful insofar as they predict risk reasonably well. This is why the application of these models always should be accompanied by validation. *Model validation* is the general process of checking whether a model is adequate. This can be done with a set of tools, including backtesting, stress testing, and independent review and oversight.

This chapter turns to backtesting techniques for verifying the accuracy of VaR models. *Backtesting* is a formal statistical framework that consists of verifying that actual losses are in line with projected losses. This involves systematically comparing the history of VaR forecasts with their associated portfolio returns.

These procedures, sometimes called *reality checks*, are essential for VaR users and risk managers, who need to check that their VaR forecasts are well calibrated. If not, the models should be reexamined for faulty assumptions, wrong parameters, or inaccurate modeling. This process also provides ideas for improvement and as a result should be an integral part of all VaR systems.

Backtesting is also central to the Basel Committee's ground-breaking decision to allow internal VaR models for capital requirements. It is unlikely the Basel Committee would have done so without the discipline of a rigorous backtesting mechanism. Otherwise, banks may have an incentive to understate their risk. This is why the backtesting framework should be designed to maximize the probability of catching banks that willfully understate their risk. On the other hand, the system also should avoid unduly penalizing banks whose VaR is exceeded simply because of bad luck. This delicate choice is at the heart of statistical decision procedures for backtesting.

This chapter first provides an actual example of model verification and discusses important data issues for the setup of VaR backtesting, then presents the main method for backtesting, which consists of counting deviations from the VaR model. It also describes the supervisory framework by the Basel Committee for backtesting the internal-models approach. Finally, practical uses of VaR backtesting are illustrated.

SETUP FOR BACKTESTING

VaR models are only useful insofar as they can be demonstrated to be reasonably accurate. To do this, users must check systematically the validity of the underlying valuation and risk models through comparison of predicted and actual loss levels.

When the model is perfectly calibrated, the number of observations falling outside VaR should be in line with the confidence level. The number of exceedences is also known as the number of *exceptions*. With too many exceptions, the model underestimates risk. This is a major problem because too little capital may be allocated to risk-taking units; penalties also may be imposed by the regulator. Too few exceptions are also a problem because they lead to excess, or inefficient, allocation of capital across units.

An Example

An example of model calibration is described in Figure 3-1, which displays the fit between actual and forecast daily VaR numbers for Bankers Trust. The diagram shows the absolute value of the daily profit and loss (P&L) against the 99 percent VaR, defined here as the *daily price volatility*.[1] The graph shows substantial time variation in the VaR measures, which reflects changes in the risk profile of the bank. Observations that lie above the diagonal line indicate days when the absolute value of the P&L exceeded the VaR.

Assuming symmetry in the P&L distribution, about 2 percent of the daily observations (both positive and negative) should lie above the diagonal, or about 5 data points in a year. Here we observe four exceptions. Thus the model seems to be well calibrated. We could have observed, however, a greater number of deviations simply owing to bad luck. The question is: At what point do we reject the model?

[1] Note that the graph does not differentiate losses from gains. This is typically the case because companies usually are reluctant to divulge the extent of their trading losses. This illustrates one of the benefits of VaR relative to other methods, namely, that by taking the absolute value, it hides the direction of the positions.

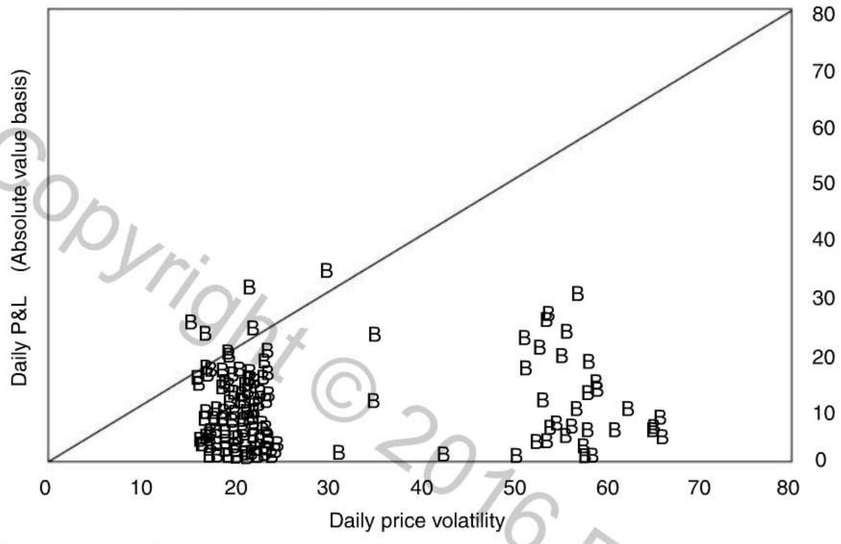

FIGURE 3-1 Model evaluation: Bankers Trust.

Which Return?

Before we even start addressing the statistical issue, a serious data problem needs to be recognized. VaR measures assume that the current portfolio is "frozen" over the horizon. In practice, the trading portfolio evolves dynamically during the day. Thus the actual portfolio is "contaminated" by changes in its composition. The *actual return* corresponds to the actual P&L, taking into account intraday trades and other profit items such as fees, commissions, spreads, and net interest income.

This contamination will be minimized if the horizon is relatively short, which explains why backtesting usually is conducted on daily returns. Even so, intraday trading generally will increase the volatility of revenues because positions tend to be cut down toward the end of the trading day. Counterbalancing this is the effect of fee income, which generates steady profits that may not enter the VaR measure.

For verification to be meaningful, the risk manager should track both the actual portfolio return R_t and the hypothetical return R_t^* that most closely matches the VaR forecast. The *hypothetical return* R_t^* represents a frozen portfolio, obtained from fixed positions applied to the actual returns on all securities, measured from close to close.

Sometimes an approximation is obtained by using a *cleaned return*, which is the actual return minus all non-mark-to-market items, such as fees, commissions, and net interest income. Under the latest update to the *market-risk amendment*, supervisors will have the choice to use either hypothetical or cleaned returns.[2]

Since the VaR forecast really pertains to R^*, backtesting ideally should be done with these hypothetical returns. Actual returns do matter, though, because they entail real profits and losses and are scrutinized by bank regulators. They also reflect the true ex post volatility of trading returns, which is also informative. Ideally, both actual and hypothetical returns should be used for backtesting because both sets of numbers yield informative comparisons. If, for instance, the model passes backtesting with hypothetical but not actual returns, then the problem lies with intraday trading. In contrast, if the model does not pass backtesting with hypothetical returns, then the modeling methodology should be reexamined.

MODEL BACKTESTING WITH EXCEPTIONS

Model backtesting involves systematically comparing historical VaR measures with the subsequent returns. The problem is that since VaR is reported only at a specified confidence level, we expect the figure to be exceeded in some instances, for example, in 5 percent of the observations at the 95 percent confidence level. But surely we will not observe exactly 5 percent exceptions. A greater percentage could occur because of bad luck, perhaps 8 percent. At some point, if the frequency of deviations becomes too large, say, 20 percent, the user must conclude that the problem lies with the model, not bad luck, and undertake corrective action. The issue is how to make this decision. This *accept or reject decision* is a classic statistical decision problem.

At the outset, it should be noted that this decision must be made at some confidence level. The choice of this level for the *test*, however, is not related to the quantitative level p selected for VaR. The decision rule may involve,

[2] See BCBS (2005b).

for instance, a 95 percent confidence level for backtesting VaR numbers, which are themselves constructed at some confidence level, say, 99 percent for the Basel rules.

Model Verification Based on Failure Rates

The simplest method to verify the accuracy of the model is to record the *failure rate*, which gives the proportion of times VaR is exceeded in a given sample. Suppose a bank provides a VaR figure at the 1 percent left-tail level ($p = 1 - c$) for a total of T days. The user then counts how many times the actual loss exceeds the previous day's VaR. Define N as the number of exceptions and N/T as the failure rate. Ideally, the failure rate should give an *unbiased* measure of p, that is, should converge to p as the sample size increases.

We want to know, at a given confidence level, whether N is too small or too large under the null hypothesis that $p = 0.01$ in a sample of size T. Note that this test makes no assumption about the return distribution. The distribution could be normal, or skewed, or with heavy tails, or time-varying. We simply count the number of exceptions. As a result, this approach is fully *nonparametric*.

The setup for this test is the classic testing framework for a sequence of success and failures, also called *Bernoulli trials*. Under the null hypothesis that the model is correctly calibrated, the number of exceptions x follows a *binomial* probability distribution:

$$f(x) = \binom{T}{x} p^x (1-p)^{T-x} \quad \textbf{(3.1)}$$

We also know that x has expected value of $E(x) = pT$ and variance $V(x) = p(1 - p)T$. When T is large, we can use the central limit theorem and approximate the binomial distribution by the normal distribution

$$z = \frac{x - pT}{\sqrt{p(1-p)T}} \approx N(0,1) \quad \textbf{(3.2)}$$

which provides a convenient shortcut. If the decision rule is defined at the two-tailed 95 percent test confidence level, then the cutoff value of $|z|$ is 1.96. Box 3-1 illustrates how this can be used in practice.

This binomial distribution can be used to test whether the number of exceptions is acceptably small. Figure 3-2 describes the distribution when the model is calibrated correctly, that is, when $p = 0.01$ and with 1 year of data, $T = 250$. The graph shows that under the null, we would observe more than

BOX 3-1 J.P. Morgan's Exceptions

In its 1998 annual report, the U.S. commercial bank J.P. Morgan (JPM) explained that

> In 1998, daily revenue fell short of the downside (95 percent VaR) band . . . on 20 days, or more than 5 percent of the time. Nine of these 20 occurrences fell within the August to October period.

We can test whether this was bad luck or a faulty model, assuming 252 days in the year. Based on Equation (3.2), we have $z = (x - pT)/\sqrt{p(1-p)T} = (20 - 0.05 \times 252)/\sqrt{0.05(0.95)252} = 2.14$. This is larger than the cutoff value of 1.96. Therefore, we reject the hypothesis that the VaR model is unbiased. It is unlikely (at the 95 percent test confidence level) that this was bad luck.

The bank suffered too many exceptions, which must have led to a search for a better model. The flaw probably was due to the assumption of a normal distribution, which does not model tail risk adequately. Indeed, during the fourth quarter of 1998, the bank reported having switched to a "historical simulation" model that better accounts for fat tails. This episode illustrates how backtesting can lead to improved models.

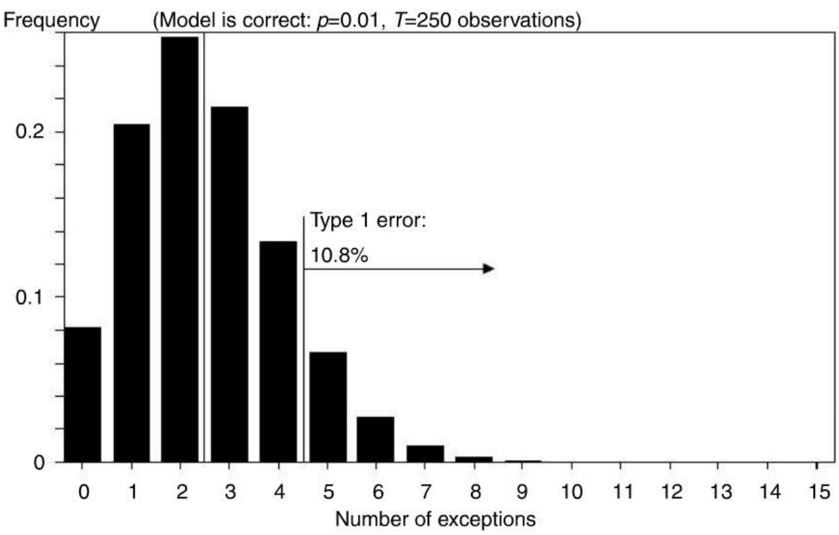

FIGURE 3-2 Distribution of exceptions when model is correct.

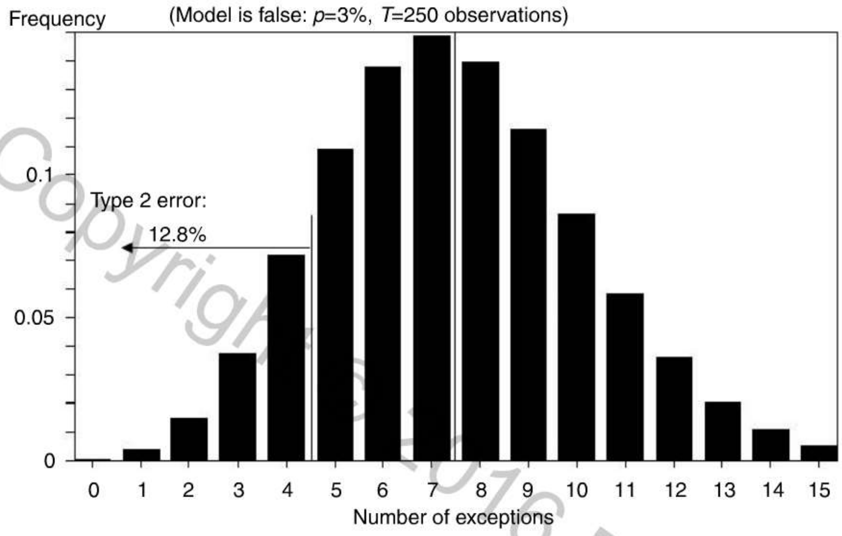

FIGURE 3-3 Distribution of exceptions when model is incorrect.

TABLE 3-1 Decision Errors

	Model	
Decision	**Correct**	**Incorrect**
Accept	OK	Type 2 error
Reject	Type 1 error	OK

TABLE 3-2 Model Backtesting, 95 Percent Nonrejection Test Confidence Regions

		Nonrejection Region for Number of Failures N		
Probability level p	**VaR Confidence Level c**	**T = 252 Days**	**T = 510 Days**	**T = 1000 Days**
0.01	99%	$N < 7$	$1 < N < 11$	$4 < N < 17$
0.025	97.5%	$2 < N < 12$	$6 < N < 21$	$15 < N < 36$
0.05	95%	$6 < N < 20$	$16 < N < 36$	$37 < N < 65$
0.075	92.5%	$11 < N < 28$	$27 < N < 51$	$59 < N < 92$
0.10	90%	$16 < N < 36$	$38 < N < 65$	$81 < N < 120$

Note: N is the number of failures that could be observed in a sample size T without rejecting the null hypothesis that p is the correct probability at the 95 percent level of test confidence.

Source: Adapted from Kupiec (1995).

four exceptions 10.8 percent of the time. The 10.8 percent number describes the probability of committing a *type 1* error, that is, rejecting a correct model.

Next, Figure 3-3 describes the distribution of number of exceptions when the model is calibrated incorrectly, that is, when $p = 0.03$ instead of 0.01. The graph shows that we will not reject the incorrect model more than 12.8 percent of the time. This describes the probability of committing a *type 2* error, that is, not rejecting an incorrect model.

When designing a verification test, the user faces a trade-off between these two types of error. Table 3-1 summarizes the two states of the world, correct versus incorrect model, and the decision. For backtesting purposes, users of VaR models need to balance type 1 errors against type 2 errors. Ideally, one would want to set a low type 1 error rate and then have a test that creates a very low type 2 error rate, in which case the test is said to be *powerful*. It should be noted that the choice of the confidence level for the decision rule is not related to the quantitative level p selected for VaR. This confidence level refers to the decision rule to reject the model.

Kupiec (1995) develops approximate 95 percent confidence regions for such a test, which are reported in Table 3-2. These regions are defined by the tail points of the log-likelihood ratio:

$$LR_{uc} = -2\ln[(1-p)^{T-N}p^{N}] + 2\ln\{[1-(N/T)]^{T-N}(N/T)^{N}\} \quad (3.3)$$

which is asymptotically (i.e., when T is large) distributed chi-square with one degree of freedom under the null hypothesis that p is the true probability. Thus we would reject the null hypothesis if $LR > 3.841$. This test is equivalent to Equation (3.2) because a chi-square variable is the square of a normal variable.

In the JPM example, we had $N = 20$ exceptions over $T = 252$ days, using $p = 95$ percent VaR confidence level. Setting these numbers into Equation (3.3) gives $LR_{uc} = 3.91$. Therefore, we reject unconditional coverage, as expected.

For instance, with 2 years of data ($T = 510$), we would expect to observe $N = pT = 1$ percent times $510 = 5$ exceptions. But the VaR user will not be able to reject the null hypothesis as long as N is within the $[1 < N < 11]$ confidence interval. Values of N greater than or equal to 11 indicate that the VaR is too low or that the model understates the probability of large losses. Values of N less than or equal to 1 indicate that the VaR model is overly conservative.

The table also shows that this interval, expressed as a proportion N/T, shrinks as the sample size increases. Select, for instance, the $p = 0.05$ row. The interval for $T = 252$ is $[6/252 = 0.024, 20/252 = 0.079]$; for $T = 1000$, it is $[37/1000 = 0.037, 65/1000 = 0.065]$. Note how the interval shrinks as the sample size extends. With more data, we should be able to reject the model more easily if it is false.

The table, however, points to a disturbing fact. For small values of the VaR parameter p, it becomes increasingly difficult to confirm deviations. For instance, the nonrejection region under $p = 0.01$ and $T = 252$ is $[N < 7]$. Therefore, there is no way to tell if N is abnormally small or whether the model systematically overestimates risk. Intuitively, detection of systematic biases becomes increasingly difficult for low values of p because the exceptions in these cases are very rare events.

This explains why some banks prefer to choose a higher VaR confidence level, such as $c = 95$ percent, in order to be able to observe sufficient numbers of deviations to validate the model. A multiplicative factor then is applied to translate the VaR figure into a safe capital cushion number. Too often, however, the choice of the confidence level appears to be made without regard for the issue of VaR backtesting.

The Basel Rules

This section now turns to a detailed analysis of the Basel Committee rules for backtesting. While we can learn much from the Basel framework, it is important to recognize that regulators operate under different constraints from financial institutions. Since they do not have access to every component of the models, the approach is perforce implemented at a broader level. Regulators are also responsible for constructing rules that are comparable across institutions.

The Basel (1996a) rules for backtesting the internal-models approach are derived directly from this failure rate test. To design such a test, one has to choose first the type 1 error rate, which is the probability of rejecting the model when it is correct. When this happens, the bank simply suffers bad luck and should not be penalized unduly. Hence one should pick a test with a low type 1 error rate, say, 5 percent (depending on its cost). The heart of the conflict is that, inevitably, the supervisor also will commit type 2 errors for a bank that willfully cheats on its VaR reporting.

The current verification procedure consists of recording daily exceptions of the 99 percent VaR over the last year. One would expect, on average, 1 percent of 250, or 2.5 instances of exceptions over the last year.

The Basel Committee has decided that up to four exceptions are acceptable, which defines a "green light" zone for the bank. If the number of exceptions is five or more, the bank falls into a "yellow" or "red" zone and incurs a progressive penalty whereby the multiplicative factor k is increased from 3 to 4, as described in Table 3-3. An incursion into the "red" zone generates an automatic penalty.

Within the "yellow" zone, the penalty is up to the supervisor, depending on the reason for the exception. The Basel Committee uses the following categories:

- *Basic integrity of the model.* The deviation occurred because the positions were reported incorrectly or because of an error in the program code.
- *Model accuracy could be improved.* The deviation occurred because the model does not measure risk with enough precision (e.g., has too few maturity buckets).

TABLE 3-3 The Basel Penalty Zones

Zone	Number of Exceptions	Increase in k
Green	0 to 4	0.00
Yellow	5	0.40
	6	0.50
	7	0.65
	8	0.75
	9	0.85
Red	10+	1.00

- *Intraday trading.* Positions changed during the day.
- *Bad luck.* Markets were particularly volatile or correlations changed.

The description of the applicable penalty is suitably vague. When exceptions are due to the first two reasons, the penalty "should" apply. With the third reason, a penalty "should be considered." When the deviation is traced to the fourth reason, the Basel document gives no guidance except that these exceptions should "be expected to occur at least some of the time." These exceptions may be excluded if they are the "result of such occurrences as sudden abnormal changes in interest rates or exchange rates, major political events, or natural disasters." In other words, bank supervisors want to keep the flexibility to adjust the rules in turbulent times as they see fit.

The crux of the backtesting problem is separating back luck from a faulty model, or balancing type 1 errors against type 2 errors. Table 3-4 displays the probabilities of obtaining a given number of exceptions for a correct model (with 99 percent coverage) and incorrect model (with only 97 percent coverage). With five exceptions or more, the cumulative probability, or type 1 error rate, is 10.8 percent. This is rather high to start with. In the current framework, one bank out of 10 could be penalized even with a correct model.

Even worse, the type 2 error rate is also very high. Assuming a true 97 percent coverage, the supervisor will give passing grades to 12.8 percent of banks that have an incorrect model. The framework therefore is not very powerful. And this 99 versus 97 percent difference in VaR coverage is economically significant. Assuming a normal distribution, the true VaR would be 23.7 percent times greater than officially reported, which is substantial.

The lack of power of this framework is due to the choice of the high VaR confidence level (99 percent) that generates too few exceptions for a reliable test. Consider instead the effect of a 95 percent VaR confidence level. (To ensure that the amount of capital is not affected, we could use a larger multiplier k.) We now have to decide on the cutoff number of exceptions to have a type 1 error rate similar to the Basel framework. With an average of 13 exceptions per year, we choose to reject the model if the number of exceptions exceeds 17, which corresponds to a type 1 error of 12.5 percent. Here we controlled the error rate so that it is close to the 10.8 percent for the Basel framework. But now the probability of a type 2 error is lower, at 7.4 percent only.[3] Thus, simply changing the VaR confidence level from 99 to 95 percent sharply reduces the probability of not catching an erroneous model.

Another method to increase the power of the test would be to increase the number of observations. With $T = 1000$, for instance, we would choose a cutoff of 14 exceptions, for a type 1 error rate of 13.4 percent and a type 2 error rate of 0.03 percent, which is now very small. Increasing the number of observations drastically improves the test.

TABLE 3-4 Basel Rules for Backtesting, Probabilities of Obtaining Exceptions ($T = 250$)

		Coverage = 99% Model Is Correct		Coverage = 97% Model Is Incorrect		
Zone	Number of Exceptions N	Probability $P(X = N)$	Cumulative (Type 1) (Reject) $P(X \geq N)$	Probability $P(X = N)$	Cumulative (Type 2) (Do not reject) $P(X < N)$	Power (Reject) $P(X \geq N)$
Green	0	8.1	100.0	0.0	0.0	100.0
	1	20.5	91.9	0.4	0.0	100.0
	2	25.7	71.4	1.5	0.4	99.6
	3	21.5	45.7	3.8	1.9	98.1
Green	4	13.4	24.2	7.2	5.7	94.3
Yellow	5	6.7	10.8	10.9	12.8	87.2
	6	2.7	4.1	13.8	23.7	76.3
	7	1.0	1.4	14.9	37.5	62.5
	8	0.3	0.4	14.0	52.4	47.6
Yellow	9	0.1	0.1	11.6	66.3	33.7
Red	10	0.0	0.0	8.6	77.9	21.1
	11	0.0	0.0	5.8	86.6	13.4

[3] Assuming again a normal distribution and a true VaR that is 23.7 percent greater than the reported VaR, for an alternative coverage of 90.8 percent.

Conditional Coverage Models

So far the framework focuses on *unconditional coverage* because it ignores conditioning, or time variation in the data. The observed exceptions, however, could cluster or "bunch" closely in time, which also should invalidate the model.

With a 95 percent VaR confidence level, we would expect to have about 13 exceptions every year. In theory, these occurrences should be evenly spread over time. If, instead, we observed that 10 of these exceptions occurred over the last 2 weeks, this should raise a red flag. The market, for instance, could experience increased volatility that is not captured by VaR. Or traders could have moved into unusual positions or risk "holes." Whatever the explanation, a verification system should be designed to measure proper *conditional coverage*, that is, conditional on current conditions. Management then can take the appropriate action.

Such a test has been developed by Christoffersen (1998), who extends the LR_{uc} statistic to specify that the deviations must be serially independent. The test is set up as follows: Each day we set a deviation indicator to 0 if VaR is not exceeded and to 1 otherwise. We then define T_{ij} as the number of days in which state j occurred in one day while it was at i the previous day and π_i as the probability of observing an exception conditional on state i the previous day. Table 3-5 shows how to construct a table of conditional exceptions.

TABLE 3-5 Building an Exception Table: Expected Number of Exceptions

	Conditional: Day Before — No Exception	Conditional: Day Before — Exception	Unconditional
Current day			
No exception	$T_{00} = T_0(1 - \pi_0)$	$T_{10} = T_1(1 - \pi_1)$	$T(1 - \pi)$
Exception	$T_{01} = T_0(\pi_0)$	$T_{11} = T_1(\pi_1)$	$T(\pi)$
Total	T_0	T_1	$T = T_0 + T_1$

If today's occurrence of an exception is independent of what happened the previous day, the entries in the second and third columns should be identical. The relevant test statistic is

$$LR_{ind} = -2\ln[(1-\pi)^{(T_{00}+T_{10})}\pi^{(T_{01}+T_{11})}] + 2\ln[(1-\pi_0)^{T_{00}}\pi_0^{T_{01}}(1-\pi_1)^{T_{10}}\pi_1^{T_{11}}] \quad (3.4)$$

Here, the first term represents the maximized likelihood under hypothesis that exceptions are independent across days, or $\pi = \pi_0 = \pi_1 = (T_{01} + T_{11})/T$. The second term is the maximized likelihood for the observed data.

The combined test statistic for conditional coverage then is

$$LR_{cc} = LR_{uc} + LR_{ind} \quad (3.5)$$

Each component is independently distributed as $\chi^2(1)$ asymptotically. The sum is distributed as $\chi^2(2)$. Thus we would reject at the 95 percent test confidence level if $LR > 5.991$. We would reject independence alone if $LR_{ind} > 3.841$.

As an example, assume that JPM observed the following pattern of exceptions during 1998. Of 252 days, we have 20 exceptions, which is a fraction of $\pi = 7.9$ percent. Of these, 6 exceptions occurred following an exception the previous day. Alternatively, 14 exceptions occurred when there was none the previous day. This defines conditional probability ratios of $\pi_0 = 14/232 = 6.0$ percent and $\pi_1 = 6/20 = 30.0$ percent. We seem to have a much higher probability of having an exception following another one. Setting these numbers into Equation (3.4), we find $LR_{ind} = 9.53$. Because this is higher than the cutoff value of 3.84, we reject independence. Exceptions do seem to cluster abnormally. As a result, the risk manager may want to explore models that allow for time variation in risk.

Extensions

We have seen that the standard exception tests often lack power, especially when the VaR confidence level is high and when the number of observations is low. This has led to a search for improved tests.

The problem, however, is that statistical decision theory has shown that this exception test is the most powerful among its class. More effective tests would have to focus on a different hypothesis or use more information.

For example, Crnkovic and Drachman (1996) developed a test focusing on the entire probability distribution, based on the *Kuiper statistic*. This test is still nonparametric but is more powerful. However, it uses other information than the VaR forecast at a given confidence level. Another approach is to focus on the time period between exceptions, called *duration*. Christoffersen and Pelletier (2004) show that duration-based tests can be more powerful than the standard test when risk is time-varying.

Finally, backtests could use parametric information instead. If the VaR is obtained from a multiple of the standard deviation, the risk manager could test the fit between the realized and forecast volatility. This would lead to more powerful tests because more information is used. Another useful avenue would be to backtest the portfolio components as well. From the viewpoint of the regulator, however, the only information provided is the daily VaR, which explains why exception tests are used most commonly nowadays.

	Conditional Day Before		
	No Exception	**Exception**	**Unconditional**
Current day			
No exception	218	14	232
Exception	14	6	20
Total	232	20	252

APPLICATIONS

Berkowitz and O'Brien (2002) provide the first empirical study of the accuracy of internal VaR models, using data reported to U.S. regulators. They describe the distributions of P&L, which are compared with the VaR forecasts. Generally, the P&L distributions are symmetric, although they display fatter tails than the normal. Stahl et. al. (2006) also report that, although the components of a trading portfolio could be strongly nonnormal, aggregation to the highest level of a bank typically produces symmetric distributions that resemble the normal.

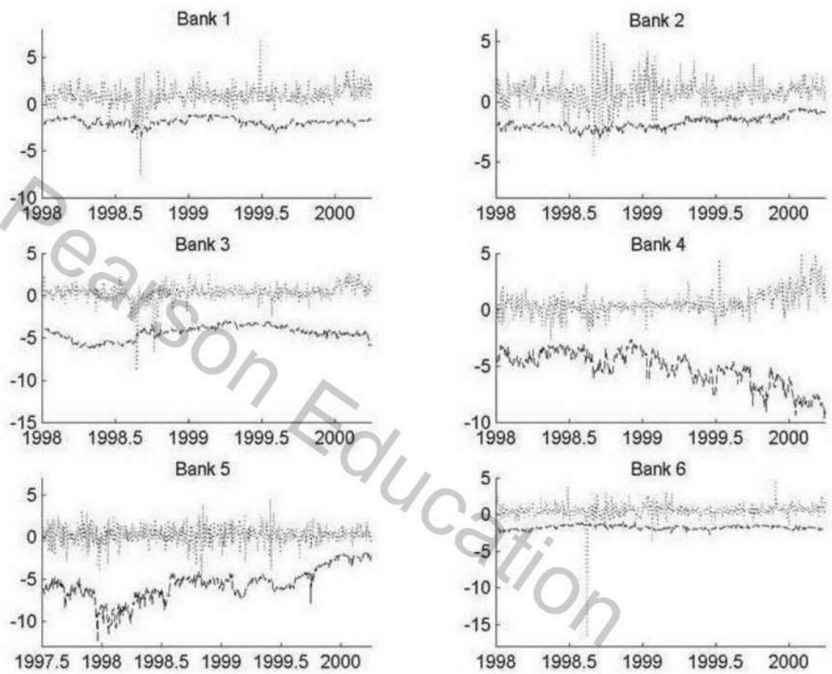

FIGURE 3-4 Bank VaR and trading profits.

Figure 3-4 plots the time series of P&L along with the daily VaR (the lower lines) for a sample of six U.S. commercial banks. With approximately 600 observations, we should observe on average 6 violations, given a VaR confidence level of 99 percent.

It is striking to see the abnormally small number of exceptions, even though the sample includes the turbulent 1998 period. Bank 4, for example, has zero exceptions over this sample. Its VaR is several times greater than the magnitude of extreme fluctuations in its P&L. Indeed, for banks 3 to 6, the average VaR is at least 60 percent higher than the actual 99th percentile of the P&L distribution. Thus banks report VaR measures that are *conservative*, or too large relative to their actual risks. These results are surprising because they imply that the banks' VaR and hence their market-risk charges are too high. Banks therefore allocate too much regulatory capital to their trading activities. Box 3-2 describes a potential explanation, which is simplistic.

Perhaps these observations could be explained by the use of actual instead of hypothetical returns.[4] Or maybe the

[4] Including fees increases the P&L, reducing the number of violations. Using hypothetical income, as currently prescribed in the European Union, could reduce this effect. Jaschke, Stahl, and Stehle (2003) compare the VaRs for 13 German banks and find that VaR measures are, on average, less conservative than for U.S. banks. Even so, VaR forecasts are still too high.

BOX 3-2 No Exceptions

The CEO of a large bank receives a daily report of the bank's VaR and P&L. Whenever there is an exception, the CEO calls in the risk officer for an explanation.

Initially, the risk officer explained that a 99 percent VaR confidence level implies an average of 2 to 3 exceptions per year. The CEO is never quite satisfied, however. Later, tired of going "upstairs," the risk officer simply increases the confidence level to cut down on the number of exceptions.

Annual reports suggest that this is frequently the case. Financial institutions routinely produce plots of P&L that show no violation of their 99 percent confidence VaR over long periods, proclaiming that this supports their risk model.

models are too simple, for example failing to account for diversification effects. Yet another explanation is that capital requirements are currently not binding. The amount of economic capital U.S. banks currently hold is in excess of their regulatory capital. As a result, banks may prefer to report high VaR numbers to avoid the possibility of regulatory intrusion. Still, these practices impoverish the informational content of VaR numbers.

CONCLUSIONS

Model verification is an integral component of the risk management process. Backtesting VaR numbers provides valuable feedback to users about the accuracy of their models. The procedure also can be used to search for possible improvements.

Due thought should be given to the choice of VaR quantitative parameters for backtesting purposes. First, the horizon should be as short as possible in order to increase the number of observations and to mitigate the effect of changes in the portfolio composition. Second, the confidence level should not be too high because this decreases the effectiveness, or power, of the statistical tests.

Verification tests usually are based on "exception" counts, defined as the number of exceedences of the VaR measure. The goal is to check if this count is in line with the selected VaR confidence level. The method also can be modified to pick up bunching of deviations.

Backtesting involves balancing two types of errors: rejecting a correct model versus accepting an incorrect model. Ideally, one would want a framework that has very high power, or high probability of rejecting an incorrect model. The problem is that the power of exception-based tests is low. The current framework could be improved by choosing a lower VaR confidence level or by increasing the number of data observations.

Adding to these statistical difficulties, we have to recognize other practical problems. Trading portfolios do change over the horizon. Models do evolve over time as risk managers improve their risk modeling techniques. All this may cause further structural instability.

Despite all these issues, backtesting has become a central component of risk management systems. The methodology allows risk managers to improve their models constantly. Perhaps most important, backtesting should ensure that risk models do not go astray.

2017 Financial Risk Manager (FRM) Part II: Market Risk Measurement and Management, Sixth Edition by Global Association of Risk Professionals.

4

VaR Mapping

Learning Objectives

After completing this reading you should be able to:

- Explain the principles underlying VaR mapping, and describe the mapping process.
- Explain how the mapping process captures general and specific risks.
- Differentiate among the three methods of mapping portfolios of fixed income securities.
- Summarize how to map a fixed income portfolio into positions of standard instruments.
- Describe how mapping of risk factors can support stress testing.
- Explain how VaR can be used as a performance benchmark.
- Describe the method of mapping forwards, forward rate agreements, interest rate swaps, and options.

Excerpt is Chapter 11 of Value at Risk: The New Benchmark for Managing Financial Risk, *Third Edition, by Philippe Jorion.*

The second [principle], to divide each of the difficulties under examination into as many parts as possible, and as might be necessary for its adequate solution.

—René Descartes

Whichever value-at-risk (VaR) method is used, the risk measurement process needs to simplify the portfolio by *mapping* the positions on the selected risk factors. Mapping is the process by which the current values of the portfolio positions are replaced by exposures on the risk factors.

Mapping arises because of the fundamental nature of VaR, which is portfolio measurement at the highest level. As a result, this is usually a very large-scale aggregation problem. It would be too complex and time-consuming to model all positions individually as risk factors. Furthermore, this is unnecessary because many positions are driven by the same set of risk factors and can be aggregated into a small set of exposures without loss of risk information. Once a portfolio has been mapped on the risk factors, any of the three VaR methods can be used to build the distribution of profits and losses.

This chapter illustrates the mapping process for major financial instruments. First we review the basic principles behind mapping for VaR. We then proceed to illustrate cases where instruments are broken down into their constituent components. We will see that the mapping process is instructive because it reveals useful insights into the risk drivers of derivatives. The next sections deal with fixed-income securities and linear derivatives. We cover the most important instruments, forward contracts, forward rate agreements, and interest-rate swaps. Then we describe nonlinear derivatives, or options.

MAPPING FOR RISK MEASUREMENT

Why Mapping?

The essence of VaR is aggregation at the highest level. This generally involves a very large number of positions, including bonds, stocks, currencies, commodities, and their derivatives. As a result, it would be impractical to consider each position separately (see Box 4-1). Too many computations would be required, and the time needed to measure risk would slow to a crawl.

BOX 4-1 Why Mapping?

"J.P. Morgan Chase's VaR calculation is highly granular, comprising more than 2.1 million positions and 240,000 pricing series (e.g., securities prices, interest rates, foreign exchange rates)." (Annual report, 2004)

Fortunately, mapping provides a shortcut. Many positions can be simplified to a smaller number of positions on a set of elementary, or *primitive*, risk factors. Consider, for instance, a trader's desk with thousands of open dollar/euro forward contracts. The positions may differ owing to different maturities and delivery prices. It is unnecessary, however, to model all these positions individually. Basically, the positions are exposed to a single major risk factor, which is the dollar/euro spot exchange rate. Thus they could be summarized by a single aggregate exposure on this risk factor. Such aggregation, of course, is not appropriate for the pricing of the portfolio. For risk measurement purposes, however, it is perfectly acceptable. This is why risk management methods can differ from pricing methods.

Mapping is also the only solution when the characteristics of the instrument change over time. The risk profile of bonds, for instance, changes as they age. One cannot use the history of prices on a bond directly. Instead, the bond must be mapped on yields that best represent its current profile. Similarly, the risk profile of options changes very quickly. Options must be mapped on their primary risk factors. Mapping provides a way to tackle these practical problems.

Mapping as a Solution to Data Problems

Mapping is also required in many common situations. Often a complete history of all securities may not exist or may not be relevant. Consider a mutual fund with a strategy of investing in *initial public offerings* (IPOs) of common stock. By definition, these stocks have no history. They certainly cannot be ignored in the risk system, however. The risk manager would have to replace these positions by exposures on similar risk factors already in the system.

Another common problem with global markets is the time at which prices are recorded. Consider, for

BOX 4-2 Market Timing and Stale Prices

In September 2003, New York Attorney General Eliot Spitzer accused a number of investment companies of allowing *market timing* into their funds. Market timing is a short-term trading strategy of buying and selling the same funds.

Consider, for example, our portfolio of Japanese and U.S. stocks, for which prices are set in different time zones. The problem is that U.S. investors can trade up to the close of the U.S. market. *Market timers* could take advantage of this discrepancy by rapid trading. For instance, if the U.S. market moves up following good news, it is likely the Japanese market will move up as well the following day. Market timers would buy the fund at the stale price and resell it the next day.

Such trading, however, creates transactions costs that are borne by the other investors in the fund. As a result, fund companies usually state in their prospectus that this practice is not allowed. In practice, Eliot Spitzer found out that many mutual-fund companies had encouraged market timers, which he argued was fraudulent. Eventually, a number of funds settled by paying more than $2 billion.

This practice can be stopped in a number of ways. Many mutual funds now impose short-term redemption fees, which make market timing uneconomical. Alternatively, the cutoff time for placing trades can be moved earlier.

instance, a portfolio or mutual funds invested in international stocks. As much as 15 hours can elapse from the time the market closes in Tokyo at 1:00 A.M. EST (3:00 P.M. in Japan) to the time it closes in the United States at 4:00 P.M. As a result, prices from the Tokyo close ignore intervening information and are said to be *stale*. This led to the mutual-fund scandal of 2003, which is described in Box 4-2.

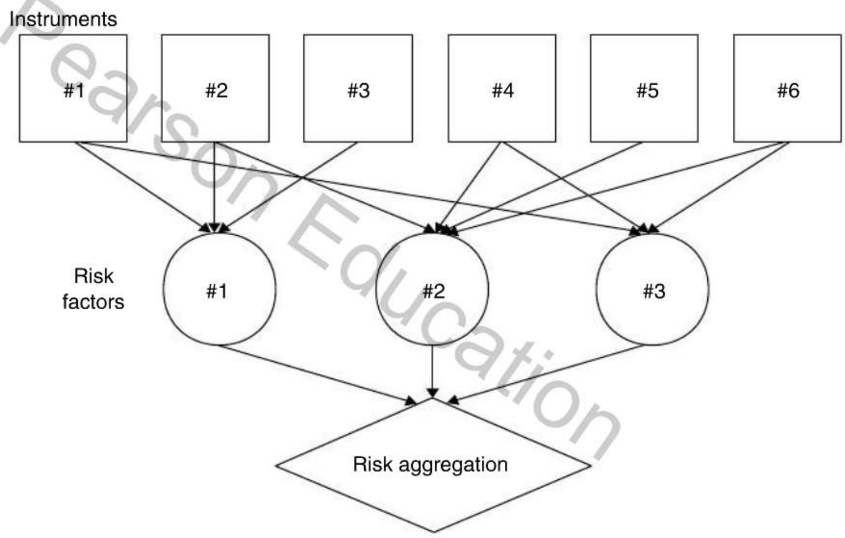

FIGURE 4-1 Mapping instruments on risk factors.

For risk managers, stale prices cause problems. Because returns are not synchronous, daily correlations across markets are too low, which will affect the measurement of portfolio risk.

One possible solution is mapping. For instance, prices at the close of the U.S. market can be estimated from a regression of Japanese returns on U.S. returns and using the forecast value conditional on the latest U.S. information. Alternatively, correlations can be measured from returns taken over longer time intervals, such as weekly. In practice, the risk manager needs to make sure that the data-collection process will lead to meaningful risk estimates.

The Mapping Process

Figure 4-1 illustrates a simple mapping process, where six instruments are mapped on three risk factors. The first step in the analysis is marking all positions to market in current dollars or whatever reference currency is used. The market value for each instrument then is allocated to the three risk factors.

Table 4-1 shows that the first instrument has a market value of V_1, which is allocated to three exposures, x_{11}, x_{12}, and x_{13}. If the current market value is not fully allocated to the risk factors, it must mean that the remainder is allocated to cash, which is not a risk factor because it has no risk.

Next, the system allocates the position for instrument 2 and so on. At the end of the process, positions

TABLE 4-1 Mapping Exposures

		Exposure on Risk Factor		
	Market Value	**1**	**2**	**3**
Instrument 1	V_1	x_{11}	x_{12}	x_{13}
Instrument 2	V_2	x_{21}	x_{22}	x_{23}
⋮	⋮	⋮	⋮	⋮
Instrument 6	V_6	x_{61}	x_{62}	x_{63}
Total portfolio	V	$x_1=\Sigma_{i=1}^{6}x_{i1}$	$x_2=\Sigma_{i=1}^{6}x_{i2}$	$x_3=\Sigma_{i=1}^{6}x_{i3}$

are summed for each risk factor. For the first risk factor, the dollar exposure is $x_1 = \Sigma_{i=1}^{6} x_{i1}$. This creates a vector x of three exposures that can be fed into the risk measurement system.

Mapping can be of two kinds. The first provides an exact allocation of exposures on the risk factors. This is obtained for derivatives, for instance, when the price is an exact function of the risk factors. As we shall see in the rest of this chapter, the partial derivatives of the price function generate *analytical* measures of exposures on the risk factors.

Alternatively, exposures may have to be *estimated*. This occurs, for instance, when a stock is replaced by a position in the stock index. The exposure then is estimated by the slope coefficient from a regression of the stock return on the index return.

General and Specific Risk

This brings us to the issue of the choice of the set of primitive risk factors. This choice should reflect the trade-off between better quality of the approximation and faster processing. More factors lead to tighter risk measurement but also require more time devoted to the modeling process and risk computation.

The choice of primitive risk factors also influences the size of specific risks. *Specific risk* can be defined as risk that is due to issuer-specific price movements, after accounting for general market factors. Hence the definition of specific risk depends on that of general market risk. The Basel rules have a separate charge for specific risk.[1]

[1] Typically, the charge is 4 percent of the position value for equities and unrated debt, assuming that the banks' models do not incorporate specific risks.

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 Ri is regressed on the return on the stock market index R_m, that is,

$$R_i = \alpha_i + \beta_i R_m + \epsilon_i \tag{4.1}$$

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

$$R_p = \sum_{i=1}^{N} w_i R_i = \sum_{i=1}^{N} w_i \beta_i R_m + \sum_{i=1}^{N} w_i \epsilon_i \tag{4.2}$$

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

$$\beta_p = \sum_{i=1}^{N} w_i \beta_i \tag{4.3}$$

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

Next, we decompose the variance R_p in Equation (4.2) and find

$$V(R_p) = (\beta_p^2)V(R_m) + \sum_{i=1}^{N} w_i^2 \sigma_{\epsilon i}^2 \tag{4.4}$$

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

$$dW = \sum_{i=1}^{N} \text{DVBP}_i dy_i = \sum_{j=1}^{J} \text{DVBP}_j dz_j + \sum_{k=1}^{K} \text{DVBP}_k ds_k + \sum_{i=1}^{N} \text{DVBP}_i d\epsilon_i \tag{4.5}$$

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

$$V(dW) = \text{general risk} + \sum_{i=1}^{N} \text{DVBP}_i^2 V(d\epsilon_i) \tag{4.6}$$

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.

MAPPING FIXED-INCOME PORTFOLIOS

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.

As an example, Table 4-2 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 4-3. VaR then is \$200 × 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. Table 4-3 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) × (2.733 − 2) = 1.351 percent for this hypothetical zero. With a \$200 million portfolio, the duration-based VaR is \$200 × 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

$$\text{VaR} = \alpha\sqrt{x'\Sigma x} = \sqrt{(x \times V)'R(x \times V)} \tag{4.7}$$

TABLE 4-2 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

TABLE 4-3 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

TABLE 4-4 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

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

Table 4-4 shows how to compute the portfolio VaR using cash-flow mapping. The second column reports the cash flows x from Table 4-2. 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 $x \times V$, which represents the individual VaRs.

With perfect correlation across all zeros, 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 earlier.

Stress Test

Table 4-3 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 4-3 of 0.4696, a 95 percent probability move would be for the zero to fall to $0.9615 \times (1 - 0.4696/100) = 0.9570$. 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 4-3 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 across different sectors of the yield curve.

TABLE 4-5 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

Benchmarking

Next, we provide a practical fixed-income compute VaR in relative terms, that is, relative to a performance benchmark. Table 4-5 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.

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

$$\text{Tracking error VaR} = \alpha\sqrt{(x - x_0)'\Sigma(x - x_0)} \tag{4.8}$$

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

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

which is in line with the explanatory power of the first factor in the variance decomposition.

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.

MAPPING LINEAR DERIVATIVES

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:

$$f_t = S_t e^{-y\tau} - Ke^{-r\tau} \quad \textbf{(4.9)}$$

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 (4.9), we have

$$F_t = (S_t e^{-y\tau})e^{r\tau} \quad \textbf{(4.10)}$$

This allows us to rewrite Equation (4.9) as

$$f_t = F_t e^{-r\tau} - Ke^{-r\tau} = (F_t - K)e^{-r\tau} \quad \textbf{(4.11)}$$

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 4-6 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 (4.9), accounting for the fact that the quoted interest rates are discretely compounded, as

TABLE 4-6 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				

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

$$= \$1.2589 - \$1.2589 = 0$$

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 4-7. 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 (4.9) with respect to the risk factors, we have

$$df = \frac{\partial f}{\partial S}dS + \frac{\partial f}{\partial r^*}dr^* + \frac{\partial f}{\partial r}dr \qquad \textbf{(4.12)}$$

$$= e^{-r^*\tau}dS - Se^{-r^*\tau}\tau dr^* + Ke^{-r\tau}\tau dr$$

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

$$df = (Se^{-r^*\tau})\frac{dS}{S} + (Se^{-r^*\tau})\frac{dP^*}{P^*} - (Ke^{-r\tau})\frac{dP}{P} \qquad \textbf{(4.13)}$$

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 4-7 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

TABLE 4-7 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

TABLE 4-8 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	

and EUR fixed-income components. As before, the total VaR will be driven primarily by the currency component.

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 (4.11) directly, where the main driver of the value of the contract is the current forward price for this commodity. Table 4-8 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.

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 (4.11), we have

$$df = \frac{\partial f}{\partial F}dF = e^{-r\tau}dF = (e^{-r\tau}F)\frac{dF}{F} \qquad \textbf{(4.14)}$$

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

$$\text{VaR} = \$43{,}743{,}000 \times 14.05/100 = \$6{,}146{,}000$$

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

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,

$$(1 + R_2\tau_2) = (1 + R_1\tau_1)\,[1 + F_{1,2}(\tau_2 - \tau_1)] \qquad \textbf{(4.15)}$$

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 × 12 FRA = long 6-month bill
+ short 12-month bill

Table 4-9 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 4-9 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 4-9 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, $\lvert x \rvert V$	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

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 4-4. The analysis is detailed in Table 4-10.

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

TABLE 4-10 Computing the VaR of a $100 Million Interest-Rate Swap (monthly VaR at 95 percent level)

	Cash Flows					
Term (Year)	Fixed	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

TABLE 4-11 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

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 4-11. 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 × 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 × 5 forward contract.

Table 4-11 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 4-10 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.

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

$$c = c(S,K,\tau,r,r^*,\sigma) = Se^{-r^*\tau}N(d_1) - Ke^{-r\tau}N(d_2) \quad \textbf{(4.16)}$$

where $N(d)$ is the cumulative normal distribution function with arguments

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

[2] For a systematic approach to pricing derivatives, see the excellent book by Hull (2005).

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,

$$dc = \frac{\partial c}{\partial S}dS + \frac{1}{2}\frac{\partial^2 c}{\partial S^2}dS^2 + \frac{\partial c}{\partial r^*}dr^* + \frac{\partial c}{\partial r}dr + \frac{\partial c}{\partial \sigma}d\sigma + \frac{\partial c}{\partial t}dt \quad \textbf{(4.17)}$$
$$= \Delta dS + \tfrac{1}{2}\Gamma dS^2 + \rho^* dr^* + \rho dr + \Lambda d\sigma + \Theta dt$$

The advantage of the BS model is that it leads to closed-form solutions for all these partial derivatives. Table 4-12 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

$$\Delta = e^{-r^*\tau}N(d_1) \quad \textbf{(4.18)}$$

This is related to the cumulative normal density function. Figure 4-2 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.

For instance, if the risk horizon is 1 day, the worst down move in the spot price is $-\alpha S\sigma\sqrt{T} = -1.645 \times \$100 \times 0.20\sqrt{1/252} = -\$2.08$, leading to a worst price of \$97.92. With a 90-day option, delta changes from 0.536 to 0.452 only.

TABLE 4-12 Derivatives for a European Call

Parameters: S = \$100, σ = 20%, r = 5%, r^* = 3%, τ = 3 months

	Variable	Unit	Exercise Price 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

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,

$$dc = \Delta dS + \rho^* dr^* + \rho dr$$
$$= [e^{-r^*\tau}N(d_1)]dS + [-Se^{-r^*\tau}\tau N(d_1)]dr^* + [Ke^{-r\tau}\tau N(d_2)]dr$$
$$= [Se^{-r^*\tau}N(d_1)]\frac{dS}{S} + [Se^{-r^*\tau}N(d_1)]\frac{dP^*}{P^*} - [Ke^{-r\tau}N(d_2)]\frac{dP}{P}$$
$$= x_1\frac{dS}{S} + x_2\frac{dP^*}{P^*} + x_3\frac{dP}{P} \quad \textbf{(4.19)}$$

This formula bears a striking resemblance to that for foreign currency forwards, as in Equation (4.13). 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 (4.9).

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

Long option = long Δ asset + short $(\Delta S - c)$ 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

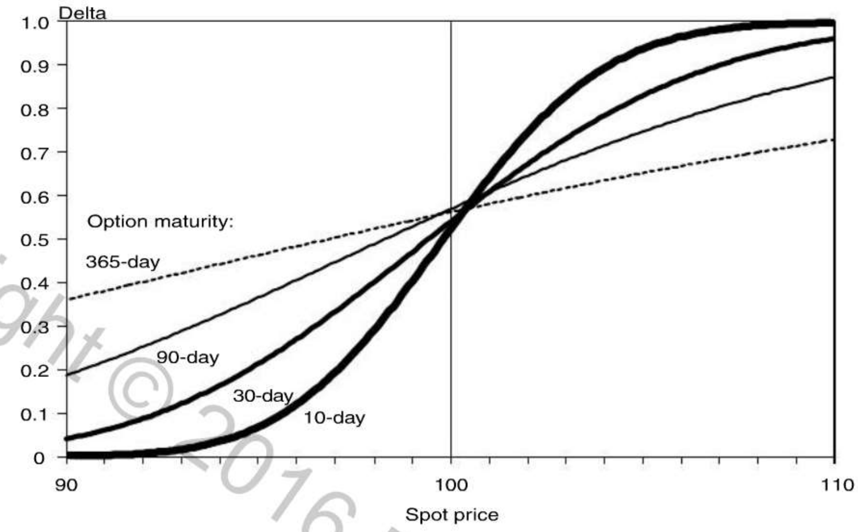

FIGURE 4-2 Delta as a function of the risk factor.

exposures can be combined with simulations of the underlying risk factors to generate a risk distribution.

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.

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.

2017 Financial Risk Manager (FRM) Part II: Market Risk Measurement and Management, Sixth Edition by Global Association of Risk Professionals.

Messages from the Academic Literature on Risk Management for the Trading Book

5

Learning Objectives

After completing this reading you should be able to:

- Explain the following lessons on VaR implementation: time horizon over which VaR is estimated, the recognition of time varying volatility in VaR risk factors, and VaR backtesting.
- Describe exogenous and endogenous liquidity risk and explain how they might be integrated into VaR models.
- Compare VaR, expected shortfall, and other relevant risk measures.
- Compare unified and compartmentalized risk measurement.
- Compare the results of research on "top-down" and "bottom-up" risk aggregation methods.
- Describe the relationship between leverage, market value of asset, and VaR within an active balance sheet management framework.

Excerpt is reprinted by permission from the Basel Committee on Banking Supervision.

2017 Financial Risk Manager (FRM) Part II: Market Risk Measurement and Management, Sixth Edition by Global Association of Risk Professionals.

INTRODUCTION

This report summarises the findings of a working group (the "group") that surveyed the academic literature that is relevant to a fundamental review of the regulatory framework of the trading book. This joint working group embraced members of the Trading Book Group and of the Research Task Force of the Basel Committee on Banking Supervision. This report summarises its main findings. It reflects the views of individual contributing authors, and should not be construed as representing specific recommendations or guidance by the Basel Committee for national supervisors or financial institutions.

The report builds on and extends previous work by the Research Task Force on the interaction of market and credit risk (see Basel Committee on Banking Supervision (2009a)). The literature review was complemented by feedback from academic experts at a workshop hosted by the Deutsche Bundesbank in April 2010 and reflects the state of the literature at this point in time.

The key findings of the group are presented in the executive summary. The structure of the remaining report is as follows:

We address fundamental issues of a sometimes highly technical nature in current VaR-based approaches to risk measurement. More specifically, we give an overview of implementation issues including questions on the necessity of including time-variation in volatility, the appropriate time horizon over which risk is measured and backtesting of VaR. Capturing market liquidity in a VaR framework is the key question addressed in the second section. Then, we look at the pros and cons of VaR as a metric for risk and consider alternative metrics put forward in the literature. Important aspects for the future evolution of stress tests are addressed next.

The last two sections include management aspects, such as inter-risk aggregation and the borderline between the banking and trading books (which is discussed only briefly). They also expand the scope of this review by including macro-prudential aspects, such as systemic risk and pro-cyclicality. This section is concerned with an integrated versus a compartmentalised approach to risk measurement, which has become particularly important since the recent financial crisis revealed that a focus on market risk alone may provide distorted results for a trading book. This topic draws heavily on the findings of the former working group of the Research Task Force on the interaction of market and credit risk (see Basel Committee on Banking Supervision (2009a)). The last section looks at the relations between and among risk measurement, systemic risk, and potential pro-cyclical effects of risk measurement.

SELECTED LESSONS ON VaR IMPLEMENTATION

Overview

In this section we review the academic and industry literature on VaR implementation issues, as it pertains to regulatory capital calculation. The three categories of implementation issues reviewed are: (1) time horizon over which VaR is estimated; (2) the recognition of time-varying volatility in VaR risk factors; and (3) VaR backtesting. With respect to (1), we find that the appropriate VaR horizon varies across positions and depends on the position's nature and liquidity. For regulatory capital purposes, the horizon should be long, and yet the common square-root of time scaling approach for short horizon VaR (e.g., one-day VaR) may generate biased long horizon VaR (e.g., ten-day VaR) estimates. Regarding (2), we find that while many trading book risk factors exhibit time-varying volatility, there are some concerns that regulatory VaR may suffer from instability and pro-cyclicality if VaR models incorporate time-varying volatility. We also sketch several approaches to incorporate time-varying volatility in VaR. As for (3), we survey the literature on VaR backtesting and discuss several regulatory issues including whether VaR should be backtested using actual or hypothetical P&L, and whether the banks' common practice of backtesting one-day VaR provides sufficient support for their ten-day, regulatory VaR.

It is worthwhile to note that some issues related to time horizons and time-varying volatility, and to a lesser extent backtesting, also pertain to risk measures other than VaR, such as Expected Shortfall (ES). A discussion of these alternative risk measures is contained in this chapter.

Time Horizon for Regulatory VaR

One of the fundamental issues in using VaR for regulatory capital is the horizon over which VaR is calculated. The 1998 Market Risk Amendment (MRA) sets this horizon to be ten days, and it allows ten-day VaR to be estimated using square-root of time scaling of one-day VaR. This approach raises three questions: (1) Is ten days an appropriate horizon? (2) Does VaR estimation based on time scaling of daily VaRs produce accurate risk measures? (3) What role do intra-horizon risks (i.e., P&L fluctuations within ten days) play, and should such risks be taken into account in the capital framework?

Is Ten Days an Appropriate Horizon?

There seems to be consensus among academics and the industry that the appropriate horizon for VaR should depend on the characteristics of the position. In the academic literature, Christoffersen and Diebold (2000) and Christoffersen, Diebold and Schuermann (1998) both assert that the relevant horizon will likely depend on where the portfolio lies in the firm (e.g., trading desk vs. CFO) and asset class (e.g., equity vs. fixed income), and the appropriate horizon should be assessed on an application-by-application basis. From this perspective, it appears that an across-the-board application of ten-day VaR horizon is not optimal. Indeed, one of the motivations for the Incremental Risk Charge (IRC) is to capture certain risks of credit related products at a longer horizon than ten days.

Although the literature suggests that it may be preferable to allow the risk horizon to vary across positions, Finger (2009), for instance, points out that there is no conceptual or statistical framework to justify the aggregation of a ten-day VaR and a one-year IRC. Danielsson (2002) adds that, if the purpose of VaR is to protect against losses during a liquidity crisis, the ten-day horizon at 99% refers to an event that happens roughly 25 times a decade, while a liquidity crisis is "unlikely to happen even once a decade. Hence the probability and problem are mismatched." In addition, even for the same financial product, the appropriate horizon may not be constant, because trade execution strategies depends on time-varying parameters, like transaction costs, expected price volatility, and risk aversion (Almgren and Chriss (2001), Engle and Ferstenberg (2006), Huberman and Stanzl (2005)). In addition, variation in risk aversion over the business cycle can be especially important in shortening the optimal trading horizon, potentially generating larger losses than those observable under more favourable conditions.

Danielsson (2002) questions the suitability of a ten-day horizon if VaR is to protect against a liquidity crisis, because a ten-day horizon implies a higher frequency of liquidity crisis than is observable in the data. Other authors have similarly suggested that the appropriate VaR horizon should depend on the economic purpose of VaR.[1] Smithson and Minton (1996), for instance, claim that nearly all risk managers believe a one-day horizon is valid for trading purposes but disagree on the appropriate horizon for long-term solvency or capital. Finger (2009) notes that there is "a tension between the regulatory risk horizon and the horizon at which banks manage their trading portfolios," and that the Market Risk Amendment (MRA) rules represent a compromise between regulatory and trading horizons through the introduction of the sixty-day moving average and backtesting multiplier mechanisms.

The computation of VaR over longer horizons introduces the issue of how to account for time variation in the composition of the portfolios, especially for institutions that make markets for actively traded assets like currencies (Diebold, Hickman, Inoue and Schuermann (1998)). A common solution is to sidestep the problem of changes to portfolio composition by calculating VaR at short horizons and scaling up the results to the desired time period using the square-root of time. While simple to implement, this choice may compromise the accuracy of VaR because, as discussed in the next section, tail risk is likely to be underestimated (Bakshi and Panayotov (2010)). A second way to tackle the problem is to focus directly on calculating the portfolio VaR over the relevant horizon of interest (Hallerbach (2003)). These approaches may have limited value if the composition of the portfolio changes rapidly. Furthermore, data limitations make it challenging to study the P&L of newly traded assets. A third solution is to extend VaR models by incorporating a prediction of future trading activity, as noted by Diebold et. al. (1998): "To understand the risk over a longer horizon, we need not only robust statistical models for the underlying market price volatility, but also robust behavioural models for changes in trading positions."

Christoffersen and Diebold (2000) aptly characterised the issue of the optimal VaR horizon as "an obvious question with no obvious answer." Voices from the industry have suggested that a horizon longer than ten days may be necessary for regulatory capital purposes. It was also suggested that

[1] For example, if VaR is expected to reduce the probability of bankruptcy, the horizon would line up with the time a bank needs to raise additional capital. If the focus is on losses while a position is being offloaded, the appropriate horizon would be more strictly related to asset characteristics.

combining the liquidity horizon of individual positions with a constant level of risk may be an appropriate avenue.

Is Square-Root of Time Scaling a Good Idea?

Under a set of restrictive assumptions[2] on risk factors, long horizon VaR can be calculated as short horizon VaR scaled by the square root of time, if the object of interest is unconditional VaR (Kaufman (2004), McNeil, Frey and Embrechts (2005) and Danielsson and Zigrand (2006)). Unfortunately, the assumptions that justify square root of time scaling are rarely verified for financial risk factors, especially at high frequencies. Furthermore, risk management and capital computation are more often interested in assessing potential losses *conditional* on current information, and scaling today's VaR by the square root of time ignores time variation in the distribution of losses. We have not found any evidence in support of square-root of time scaling for conditional VaRs.

The accuracy of square-root of time scaling depends on the statistical properties of the data generating process of the risk factors. Diebold et. al. (1998) show that, if risk factors follow a GARCH(1,1) process, scaling by the square-root of time *over-estimates* long horizon volatility and consequently VaR is over-estimated. Similar conclusions are drawn by Provizionatou, Markose and Menkens (2005). In contrast to the results that assume that risk factors exhibit time-varying volatility, Danielsson and Zigrand (2006) find that, when the underlying risk factor follows a jump diffusion process, scaling by the square root of time systematically *under-estimates* risk and the downward bias tends to increase with the time horizon. While these results argue against square-root of time scaling, it is important to acknowledge that we were not able to find immediate alternatives to square-root of time scaling in the literature. Therefore, the practical usefulness of square-root of time scaling should be recognised.[3]

[2] Specifically, the risk factors have to be normally distributed with zero mean, and be independently and identically distributed ("IID") across time.

[3] A concept related to square-root of time scaling is the scaling of VaR to higher confidence levels. Although we were unable to find literature on this topic, we recognize that this is an important issue particularly in situations when there are inadequate data points for one to accurately estimate risks deep into the tail. Some banks use certain reference densities (e.g., Student's t with six degrees of freedom) to conduct such scaling.

Is Intra-Horizon Risk Important?

Bakshi and Panayotov (2010) discuss intra-horizon VaR (VaR-I), a risk measure that combines VaR over the regulatory horizon with P&L fluctuations over the short term, with a particular focus on models that incorporate jumps in the price process. The rationale behind intra-horizon VaR is that the maximum cumulative loss, as distinct from the end-of-period P&L, exerts a distinct effect on the capital of a financial institution. Bakshi and Panayotov (2010) suggest that VaR-I "can be important when traders operate under mark-to-market constraints and, hence, sudden losses may trigger margin calls and otherwise adversely affect the trading positions." Daily VaR does carry information on high frequency P&L but, as noted by Kritzman and Rich (2002), "Knowledge of the VaR on a daily basis does not reveal the extent to which losses may accumulate over time." Bakshi and Panayotov (2010) find that taking intra-horizon risk into account generates risk measures consistently higher than standard VaR, up to multiples of VaR, and the divergence is larger for derivative exposures.

Time-Varying Volatility in VaR

It is a stylised fact that certain asset classes, such as equities and interest rates, exhibit time-varying volatility. Accounting for time-varying volatility in VaR models has been one of the most actively studied VaR implementation issues. This section explores this topic, focusing on large and complex trading portfolios.

Is It Necessary to Incorporate Time-Varying Volatilities and Correlations?

The industry seems to think so since many firms advocate the use of fast reacting measures of risk such as exponential time-weighted measures of volatility. The reason given is that such VaR models provide early warnings of changing market conditions and may perform better in backtesting. The academic literature has also observed that time-varying volatility in financial risk factors is important to VaR, dating back to the 1996 RiskMetrics Technical document (J.P. Morgan (1996)). Pritsker (2006) showed theoretically that using historical simulation VaR without incorporating time-varying volatility can dangerously under-estimate risk, when the true underlying risk factors exhibit time-varying volatility.

In contrast, some have argued that, depending on the purpose of VaR, capturing time-varying volatility in VaR may not be necessary, or may even be inappropriate. Christoffersen and Diebold (2000) observe that volatility forecastability decays quickly with time horizon for most equity, fixed income and foreign exchange assets. The implication is that capturing time-varying volatility may not be as important when the VaR horizon is long, compared to when the VaR horizon is relatively short. There are also concerns about pro-cyclicality and instability implications associated with regulatory VaRs that capture time-varying volatility. Dunn (2009), for instance, states that there is a "contradiction between the requirement for a risk sensitive metric to capture variations in volatility and correlation, and the regulatory requirement for a stable and forward looking basis for computing capital, that is not pro-cyclical." In reference to modelling time-varying volatility in VaR, it wrote, "Some firms mentioned a trade-off in this issue, and that for some purposes such as capital allocation, risk measures with more stable properties that reflected longer historical norms were desirable."

In summary, incorporating time-varying volatility in VaR appears to be necessary given that it is prevalent in many financial risk factors. Furthermore, many financial instruments are now priced with models with stochastic volatility features. It is logical that VaR models are constructed to account for these statistical properties. However, using VaR with time-varying volatility for regulatory capital raises the concerns of volatile and potentially pro-cyclical regulatory standards.

Methods to Incorporate Time-Varying Volatility in VaR for Large, Complex Portfolios

Beginning with J.P. Morgan (1996), the Exponentially Weighted Moving Average (EWMA) approach has been regarded as one of the industry standards for incorporating time-varying volatility in VaR. EWMA is a constrained version of an IGARCH (1,1) model, and in the case of RiskMetrics the parameter in IGARCH was set to 0.97. An alternative and simpler approach is to weight historical data according to the weights introduced by Boudoukh, Richardson and Whitelaw (1998), where an observation from i days ago receives a weight of

$$w(i) = \frac{\theta^{i}(1-\theta)}{1-\theta^{n}}$$

Here n is the total number of days in the historical window, and θ is a number between zero and one which controls the rate of memory decay. An even simpler approach is to compute VaR with historical simulation using a short and frequently updated time series. Dunn (2009) has suggested that this method captures time-varying volatility quite well. Using simulations, Pritsker (2006) has shown that the approach of Boudoukh et. al. (1998) is not sensitive enough to pick up volatility changes. He advocated the use of Filtered Historical Simulation (FHS), first introduced by Barone-Adesi, Giannopoulos and Vosper (1999). Broadly speaking, FHS is based on the idea that risk factors should first be filtered through a GARCH model. The volatility is then updated using the model, and adhered to the filtered risk factors to constructed VaR.

Naturally, considerations should be given to how the above method can be applied to portfolios with large numbers of positions or risk factors. Barone-Adesi et. al. (1999) outlined a position-by-position FHS approach. They recommended filtering each risk factor separately, and building volatility forecasts for each factor. Analogously, EWMA and the weights introduced by Boudoukh et. al. (1998) can be applied the same way. However, weighting or filtering risk factors separately implicitly assumes that the correlation structure across risk factors does not change over time. Pritsker (2006) has pointed out that time-varying correlation is an important source of risk. Indeed, the recent crisis has highlighted the fact that correlations among many risk factors change significantly over time. One would need to be careful in handling time-varying volatilities as well as correlations.

Multivariate GARCH models such as the BEKK model of Engle and Kroner (1995), or the DCC model of Engle (2002) can be used to estimate time-varying volatilities as well as correlations. However, such multivariate GARCH models are difficult to estimate when there are a large number of risk factors. Some recent advances in the literature allow one to estimate a multivariate GARCH-type model when there are a large number of risk factors. For instance, Engle, Shephard and Sheppard (2007) proposed to average likelihoods before estimating the GARCH model with maximum likelihood. Engle and Kelly (2009) imposes a restriction on the correlation structure that helps facilitate estimation in large dimensions, but still allow correlations to change over time. Finally, Aramonte, Rodriguez and Wu (2010) estimates VaR for large portfolios comprising stocks and bonds by first reducing the dimension of risk factors using dynamic factor models, and then estimating a time-varying volatility model. The resulting VaR estimates are shown to out-perform historical simulation and FHS based on filtering risk factors one-by-one.

All in all, incorporating time-varying volatility in VaR measures is not straight forward when there are many risk factors. Time-varying correlations should be taken into account. Rather than using more involved methods, the industry appears to be taking less burdensome alternatives, such as using simple weighting of observations, or shortening the data window used to estimate VaR. These approaches compromise on accuracy, but are computationally attractive for large and complex portfolios. The recent academic literature offers promise that some of the sophisticated empirical methodologies may soon become practical for large complex portfolios.

Backtesting VaR Models

As with any type of modelling, a VaR model must be validated. In particular, backtesting has been the industry standard for validating VaR models. This section reviews some backtesting methodologies suggested by the literature, and some issues pertaining to the application of such methodologies.

Backtesting Approaches

Banks typically draw inference on the performance of VaR models using backtesting exceptions (sometimes also known as backtesting "breaches" or "violations"). For regulatory capital, the MRA imposes a multiplier on VaR depending on the number of backtesting exceptions the bank experiences.

While the MRA does not require banks to statistically test whether VaR has the correct number of exceptions, formal statistical inference is always desirable and many alternatives have been proposed in the literature. Kupiec (1995) introduced the unconditional coverage likelihood ratio tests as inference tools for whether the VaR model generated the correct number of exceptions. This methodology is simple to implement, but has two drawbacks. First, as pointed out by Kupiec (1995 and 2005), when the number of trading days used in VaR evaluation is limited (e.g., one year or approximately 250 trading days), or when the confidence level is high (e.g., 99% as in regulatory VaR), such tests have low power. This is not surprising, since one would expect only a small number of backtesting exceptions in most cases. Building a statistic out of a handful of exceptions, then, may induce high variance in the test statistic itself and the result may be sensitive to an incremental exception. Second, given that this test only counts exceptions, its power may be improved by considering other aspects of the data such as the grouping of exceptions in time.

Christoffersen (1998) has proposed a conditional backtesting exception test that accounts for the timing as well as the number of exceptions. The test is based on the fact that when the VaR model has conditionally the correct number of exceptions, then indicator variables representing the exceptions are IID[4] Bernoulli random variables. This test, however, may still be exposed to the low power problem. To this end, Berkowitz, Christoffersen and Pelletier (2010) provided a suite of conditional tests that have good power. These tests are based on the intuition of Christoffersen (1998) (i.e., correct conditional exceptions results in IID and Bernoulli exception indicators) but derive inferences from autocorrelation, spectral, and hazard rate tests.

Aside from backtesting based on the number of exceptions, a natural measure of VaR performance is the magnitude of the exceptions. Lopez (1999), for instance, formalised this idea by introducing a quadratic loss function where loss is the difference between actual P&L and VaR, when an exception occurs. Some papers, including Pritsker (2006) and Shang (2009), also consider the use of Mean-Squared-Error (MSE) as a measure of VaR performance in backtesting. Typically, one would measure the MSE between the 'true VaR' and the VaR estimate based on the model. Clearly, this method is not directly applicable to observed portfolio P&Ls, since the true VaR is never known. Nonetheless, it can be a useful validation method prior to putting a VaR model into production: one can define data generating processes mimicking those imposed by front office pricing models, simulate position P&L enough times to construct a P&L distribution, and find the 'true VaR' based on this simulated distribution.

[4] IID: independently and identically distributed.

Then, the VaR model can be applied to the generated data, and the difference between 'true VaR' and estimated VaR can be analysed.

Backtesting Issues

An important and yet ambiguous issue for backtesting is which P&L series to compare to VaR. Broadly speaking, the estimated VaR can be compared to either actual P&L (i.e., the actual portfolio P&L at the VaR horizon), or hypothetical P&L (i.e., P&L constructed based on the portfolio for which VaR was estimated). To complicate matters further, actual P&L may sometimes contain commissions and fees, which are not directly related to trading and trading risk. Franke, Härdle and Hafner (2008) and Berry (2009) described the relative merits of actual and hypothetical backtesting: actual backtesting has little value if the portfolio has changed drastically since VaR was estimated, but is simple to implement; hypothetical backtesting would make an 'apples-to-apples' comparison, but it comes with significant implementation burden given that hypothetical portfolios need to be constructed.

Another issue is the appropriate backtesting horizon. Banks typically backtest one-day ahead VaR and use it as a validation of the regulatory VaR, which is ten-day. The problem here is clear: a good one-day VaR (as validated by backtesting) does not necessarily imply a good ten-day VaR, and vice versa. Ten-day backtesting may not be ideal either, given the potentially large portfolio shifts that may take place within ten days. In that case, actual P&L backtesting in particular may not be very informative. While we were unable to find literature on this particular issue, it remains an important policy question.

Conclusions

We have reviewed the literature on a number of VaR implementation issues, including the appropriate time horizon, time-variation in the volatility of risk factors, and backtesting. We find that the optimal way of addressing these points is idiosyncratic to the problem under consideration. For instance, when estimating long horizon VaR by scaling the short horizon counterpart by the square root of time, one may overestimate VaR if the underlying P&L process exhibits time-varying volatility, but underestimate VaR if the process has jumps.

Incorporating time-varying volatility in VaR measures appears to be important to make models more realistic although it is not straight forward when there are many risk factors. The recent academic literature offers promise in this direction. While many trading book risk factors have time-varying volatility, models that incorporate this feature may, however, generate pro-cyclical VaR and also be unstable, not least because of estimation issues.

In addition, the choice of whether to evaluate a VaR model on the basis of hypothetical or actual backtesting may be affected by the characteristics of the portfolio. Indeed, actual backtesting is less informative when the composition of the portfolio has recently changed. On the other hand, while hypothetical backtesting provides a more consistent comparison, it may impose substantial computational burdens because it requires reconstructing the history of the portfolio on the basis of its current composition.

INCORPORATING LIQUIDITY

Overview

Discussing the challenging issue of how to incorporate market liquidity into a VaR model requires first of all a distinction between *exogenous* and *endogenous* liquidity. This distinction is made from the point of view of the bank, rather than in general equilibrium terms (Bangia, Diebold, Schuermann and Stroughair (1999a) and Bervas (2006)), More specifically, *exogenous* liquidity refers to the transaction cost for trades of average size, while *endogenous* liquidity is related to the cost of unwinding portfolios large enough that the bid-ask spread cannot be taken as given, but is affected by the trades themselves.

Bangia et. al. (1999a) give a graphical representation of *exogenous* and *endogenous* liquidity that is reproduced in Figure 5-1. Below a certain size, transactions may be traded at the bid/ask price quoted in the market (*exogenous* liquidity), and above this size, the transaction will be done at a price below the initial bid or above the initial ask, depending on the sign of the trade (*endogenous* liquidity).

The exogenous component of liquidity risk corresponds to the average transaction costs set by the market for

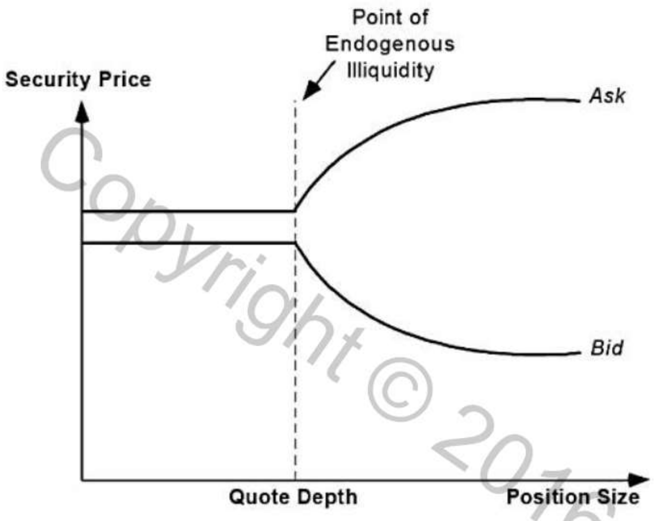

FIGURE 5-1 Effect of position size on liquidation value.

Source: Bangia et. al. (1999a).

standard transaction sizes. The endogenous component corresponds to the impact on prices of the liquidation of a position in a relatively tight market, or more generally when all market participants react in the same way, and therefore applies to orders that are large enough to move market prices (Bangia et. al. (1999a), Bervas (2006)). Exogenous liquidity risk, corresponding to the normal variation of bid/ask spreads across instruments can be, from a theoretical point of view, easily integrated into a VaR framework. Endogenous risk, corresponding to the impact on market prices of the liquidation of a position, or of collective portfolio adjustments, is more difficult to include in a VaR computation. Its impact, however, may be very significant, especially for many complex derivatives held in trading books of large institutions.

One way to incorporate liquidity risk into VaR measures is to include new VaR risk factors that can be used to model liquidity risks. This approach is feasible only when the parameters can be deduced from market data. Liquidity reserves taken by banks on their trading portfolio according to accounting standards correspond, more or less, to reserves for exogenous liquidity. In order to integrate this risk in the VaR computation, Bangia et. al. (1999a) propose to integrate the variability of the bid/offer spread for average size transactions as a risk factor.

To take into account endogenous liquidity in the value-at-risk is more difficult, as it is not even really taken into account in the valuation of trading portfolios, but its impact on both valuation and VaR should be significant. Academic literature on the subject—portfolio valuation and VaR computation—is quite rich, but very little application has been made in particular because endogenous liquidity reserves could be considered as not compliant to accounting standards.

In the following section, we first describe how, following existing literature, exogenous liquidity might be integrated into VaR measures. We then review several aspects of endogenous liquidity risk, and detail how this risk could be integrated in portfolio valuation and VaR computation. At last, we discuss on the choice of the VaR horizon when taking into account liquidity risk.

Exogenous Liquidity

For the trading portfolio, following IAS rules, only exogenous liquidity risk will be taken into account in the valuation of cash assets and derivatives. Bangia et. al. (1999a) propose adding the bid/offer spread to characterise exogenous liquidity as a risk factor.

Their method poses that the relative spread, S = (Ask-Bid)/Mid-price, has sample mean and variance $\hat{\mu}$ and $\hat{\sigma}^2$. If the 99% quantile of the normalised distribution of S is $\hat{q}_{0.99}$, then the Cost of Liquidity is defined as

$$CoL_t = P_t\left(\frac{\hat{\mu} + \hat{q}_{0.99}\hat{\sigma}}{2}\right)$$

where P_t is today's value of the position. CoL_t is added to VaR to form a liquidity-adjusted VaR.

Endogenous Liquidity: Motivation

Adverse market conditions can generate a flight to liquid and high-quality assets, which reduces the ability to unwind positions in thinly-traded, low-quality assets. The effect can be compounded when the inventory of market makers becomes imbalanced, thus reducing their willingness to further accommodate sell trades, and when risk management standards for traders become tighter, reducing the probability of finding a counterparty.

Margin requirements are also a source of variation in the response of assets' prices and liquidity to fundamental shocks, because higher margins increase the probability of binding funding constraints. While the choice of margin requirements is endogenous to a security's liquidity, assets with identical payoffs can have different prices depending on margin requirements and the opportunity cost of capital.

The trading activities associated with hedging may also have an impact on the dynamics of the underlying assets. For example, delta hedging an option position entails buying the asset when its price goes up, and selling it when the price goes down: if the size of these adjustments is not negligible with respect to the volumes traded on the underlying, this strategy will increase upward and downward price movements.

Such effects will be particularly important when:

- the underlying asset is not very liquid,
- the size of the positions of the investors hedging an option is important with respect to the market,
- large numbers of small investors follow the same hedging strategy,
- the market for the underlying of the derivative is subject to asymmetric information, which magnifies the sensitivity of prices to clusters of similar trades (Gennotte and Leland (1990)).

In particular, on some specific markets driven by exotic options (e.g., Power Reverse Dual Callable, some CPPI[5] strategies, etc.), even if a bank's trading book positions are small with respect to the market, this bank may be exposed to losses due to endogenous liquidity. When many other banks have the same kind of positions and none has an opposite position,[6] all these banks will have to adjust their hedging portfolio in the same way at the same time, and will then influence the market dynamics, and thus its small position may then be exposed to a significant liquidity cost.

The implications of derivative hedging have been extensively studied and derivative hedging has been identified as a potential explanation for the relation between implied volatilities and strike prices that can be observed on option markets (the so-called volatility smile). This literature includes the work of Platen and Schweizer (1998), Sircar and Papanicolaou (1998), Schönbucher and Wilmott (2000) and Subramanian (2008).

Endogenous Liquidity and Market Risk for Trading Portfolios

Several authors have studied the implications of endogenous liquidity risk for portfolio valuation and on value-at-risk measures (Jarrow and Protter (2005), Rogers and Singh (2005)). In general, these authors define an optimal liquidation strategy in a finite (or infinite time horizon) model and deduce from this strategy the market value of the portfolio which is equal to the expectation of its liquidation price. The associated VaR measure, defined as a confidence interval around this expected price, implicitly incorporates market and liquidity risks.

Some studies suggest that endogenous liquidity costs should be added to position returns before carrying out VaR calculations. To that end, Bervas (2006) suggests to incorporate Kyle's Lambda or Amihud's (2002) illiquidity ratio in returns. Both measures are based on the relationship between returns and volume. Wu (2009) applies the illiquidity cost of Amihud (2002) to stock returns and calculates the sum as "liquidity-adjusted returns." VaR is then estimated by applying a GARCH type model to the adjusted returns. Francois-Heude and Van Wynendaele (2001) suggest an approach that modifies the model of Bangia et. al. (1999a) by using average weighted bid-ask spreads, with weights based on volume. Berkowitz (2000b) proposes to incorporate price impact of an immediate liquidation via the concept of elasticity of demand. Jarrow and Subramanian (2001) modify the mean and variance that appears in the standard parametric VaR formula to incorporate means and variances of liquidation time and liquidity discount. Botha (2008) extended Jarrow and Subramanian (2001) to the two assets portfolio level. Other notable papers include Le Saout (2002) and Hisata and Yamai (2000). Finally, Acerbi and Scandolo (2008) explore the impact of market and funding liquidity on portfolio prices and risk measure. The authors revisit the coherent measures of risk criteria introduced by Artzner et. al. (1999). They explain how these criteria should be interpreted; in particular they study liquidity models that lead to solutions for the valuation of portfolios constituted of analytically tractable assets.

[5] CPPI: Constant Proportion Portfolio Insurance.

[6] The clients of these banks may be investors who do not dynamically hedge their positions.

The liquidity risk adjustments proposed in the academic literature, for the most part, have not been applied to the trading books of banks. One reason for this may be that the suggested valuation methods are not necessarily compliant with actual accounting standards.[7] Another reason academic proposals have been slow to be adopted may be the difficulty of estimating model liquidity parameters, especially for OTC products. Indeed, the necessary data are not always available, and some of these parameters may be subjective. But recent discussions in academic circles regarding OTC transaction reporting could contribute to solve this problem.

A study of the impact of endogenous liquidity on the valuation of exotic derivatives, similar to the contributions of exogenous liquidity, would be especially welcome. When significant market movements materialise, traders will adjust their hedging strategies which may have an impact on the market dynamics if the volumes they have to trade are significant. Such effect has been suggested as a possible explanation for the significant trading losses that some banks have experienced during the last financial crisis.

Some authors have integrated liquidity risk with market and credit risk. For example, in order to evaluate a portfolio, Zheng (2006) studies optimal liquidation strategies, taking into account market and liquidity risk, together with the probability of default of an issuer or of a counterparty. Stange and Kaserer (2008) suggest calculating liquidity-adjusted VaR conditional on the market value of a position by incorporating bid-ask spread liquidity adjustments in returns; Qi and Ng (2009) discuss intraday liquidity risk and its impact on VaR.

Adjusting the VaR Time Horizon to Account for Liquidity Risk

The recent financial crisis has provided examples where a change in market liquidity conditions alters the *liquidity horizon*, i.e., the time required to unwind a position without unduly affecting the underlying instrument prices (including in a stressed market). This finding was already addressed in previous work of the Research Task Force (see Basel Committee on Banking Supervision (2009a)) and it is consistent with the literature.

Lawrence and Robinson (1997), for example, suggest that the application of a unique horizon to all positions by ignoring their size and level of liquidity is undesirable. They suggest determining the temporal horizon by the size of the position and the liquidity of the market. Haberle and Persson (2000) propose a method based on the fraction of daily volume that can be liquidated without significant impact on the market price, which can be interpreted as holding the horizon fixed and determining how much can be liquidated during that horizon. The method of Jarrow and Subramanian (2001) is also relevant in this context as it requires an estimate of the average liquidation time.

Previous work of the Research Task Force suggests an *interdependence* between risk assessment and liquidity horizon: On the one hand the exposures of banks to market risk and credit risk may vary with a risk horizon that is set dependent on market liquidity. If liquidity decreases, for example, the risk horizon lengthens and the exposure to credit risk typically increases. On the other hand, liquidity conditions are also affected by perceptions of market and credit risk. A higher estimate of credit risk for example, may adversely affect the willingness to trade and thereby market liquidity (see Basel Committee on Banking Supervision (2009a)).

Liquidation horizons vary over the business cycle, increasing during times of market stress. Besides transaction costs or the size of the position relative to the market, a trade execution strategy also depends on factors like expected price volatility and risk aversion (Huberman and Stanzl (2005)). If, for instance, risk aversion increases during a crisis, an investor may choose to trade more rapidly than during normal times, thus generating higher losses than those observable under favourable economic conditions.

Conclusions

Both exogenous and endogenous liquidity risks are important; endogenous liquidity risk is particularly relevant for exotic/complex trading positions. While exogenous

[7] For example, IAS 39 specify in AG72: "The appropriate quoted market price for an asset held or liability to be issued is usually the current bid price. . . . The fair value of a portfolio of financial instruments is the product of the number of units of the instrument and its quoted market price," and in AG75: "The objective of using a valuation technique is to establish what the transaction price would have been on the measurement date in an arm's length exchange motivated by normal business considerations."

liquidity is partially incorporated in the valuation of trading portfolios, endogenous liquidity is typically not, even though its impact may be substantial. Although endogenous liquidity risk is especially relevant under stress conditions, portfolios may be subject to significant endogenous liquidity costs under *all* market conditions, depending on their size or on the positions of other market participants.

The academic literature suggests as a first step to adjust valuation methods in order to take endogenous liquidity risk into account. Then a VaR integrating liquidity risk could be computed. Notwithstanding academic findings on this topic, in practice, the ability to model exogenous and endogenous liquidity may be constrained by limited data availability, especially for OTC instruments.

RISK MEASURES

Overview

This section compares selected risk measures that appear to be relevant for risk management purposes either today or in the future. The alternative measures considered include VaR, expected shortfall and spectral measures of risk. The key features used to decide among alternative risk measurement approaches include ease of calculation, numerical stability, the possibility to calculate risk contributions of individual assets to portfolio risk, backtesting possibilities, incentives created for risk managers, and, linked to the latter, the relation between risk measures and regulators' objectives. Although few financial institutions currently make use of VaR alternatives, those that do are often considered as technologically leading in the industry.

In the literature, risk measures are usually defined as functions of *random variables* (portfolio losses or returns in most cases). This seems to be a trivial aspect but is actually a substantial restriction because it binds the analysis to one point of time; while this time horizon can be varied, a *joint* analysis of a portfolio's losses at several times, which may be important for asset/liabilities management, is excluded. Risk measures being a function of random loss variables also means that these variables are not an attribute of risk measures; the probability distributions of the variables are specified in a preceding step, and the analysis of risk measures is not an analysis of whether the random variables are correctly specified.

In our discussion of alternative measures we focus on VaR because of its high relevance to the industry today and on Expected Shortfall and Spectral Measures because of their advantages and hence a potentially growing importance in the future. Other risk measures, such as variance or upper-tail moments are briefly sketched for completeness.

VaR

Concept of VaR and Its Problems

VaR has become a standard measure used in financial risk management due to its conceptual simplicity, computational facility, and ready applicability. Given some random loss L and a confidence level α, $VaR_\alpha(L)$ is defined as the quantile of L at the probability α. The quantile is not necessarily unique if there are regions where the loss distribution function F_L does not grow. For these cases, McNeil et. al. (2005) define the VaR as the smallest, i.e., most optimistic quantile:

$$VaR_\alpha(L) = \inf\{l : F_L(l) \geq \alpha\}$$

Despite its prevalence in risk management and regulation, VaR has several conceptual problems. Artzner, Delbaen, Eber and Heath (1999) point out that VaR measures only quantiles of losses, and thus disregards any loss beyond the VaR level. As a consequence, a risk manager who strictly relies on VaR as the only risk measure may be tempted to avoid losses within the confidence level while increasing losses beyond the VaR level. This incentive sharply contrasts with the interests of regulators since losses beyond the VaR level are associated with cases where regulators or deposit insurers have to step in and bear some of the bank's losses. Hence, VaR provides the risk manager with incentives to neglect the severity of those losses that regulators are most interested in.

Neglecting the severity of losses in the tail of the distribution also has a positive flipside: it makes backtesting easier or possible in the first place simply because empirical quantiles are per se robust to extreme outliers, unlike typical estimators of the expected shortfall, e.g., (see below).

VaR is criticised for not being a *coherent* risk measure, which means that VaR lacks an axiomatic foundation as proposed by Artzner et. al. (1999). They set out the following four consistency rules. A risk measure R is called *coherent* if it satisfies the following axioms.

- Subadditivity (diversification) $R(L_1 + L_2) \leq R(L_1) + R(L_2)$
- Positive homogeneity (scaling) $R(\lambda L) = \lambda R(L)$, for every $\lambda > 0$
- Monotonicity $R(L_1) < R(L_2)$ if $L_1 < L_2$
- Transition property $R(L + a) < R(L) - a$

VaR is not coherent because it may violate the subadditivity criterion. For why subadditivity indeed makes sense we quote from McNeil et. al. (2005):

- "Subadditivity reflects the idea that risk can be reduced by diversification, . . . the use of non-subadditive risk measures in a Markowitz-type portfolio optimisation problem may lead to optimal portfolios that are very concentrated and that would be deemed quite risky by normal economic standards.
- If a regulator uses a non-subadditive risk measure in determining the regulatory capital for a financial institution, that institution has an incentive to legally break up into various subsidiaries in order to reduce its regulatory capital requirements. . . .
- Subadditivity makes decentralisation of risk-management systems possible. Consider as an example two trading desks with positions leading to losses L_1 and L_2. Imagine that a risk manager wants to ensure that $R(L)$, the risk of the overall loss $L = L_1 + L_2$, does not exceed some number M. If he uses a subadditive risk measure R, he may simply choose bounds M_1 and M_2 such that $M_1 + M_2 \leq M$ and impose on each of the desks the constraint that $R(L_i) \leq M_i$; subadditivity then ensures automatically that $R(L) \leq M_1 + M_2 \leq M$."

Remark 1: Related to non-coherency of VaR, Basak and Shapiro (2001) create an example where VaR-based risk management may possibly be problematic. They analyse optimal, dynamic portfolio and wealth/consumption policies of utility maximising investors who must also manage market-risk exposure using VaR. They find that VaR risk managers often optimally choose a larger exposure to risky assets than non-VaR risk managers and consequently incur larger losses when losses occur.

Remark 2: At first glance, subadditivity and positive homogeneity may not appear as meaningful concepts when risk measures are applied to counterparty credit risk (CCR) or other types of credit risk. For example, assume there is CCR involved with some position in the trading book. Doubling the position can *more than* double the CCR simply because not only the exposure doubles but also because the position becoming extremely profitable can *make* the counterparty go bankrupt. This appears to contradict the postulate of positive homogeneity which claims $R(2L) = 2R(L)$. However, it is not that positive homogeneity is wrong for CCR but rather that this idea reflects a misunderstanding of risk measures as functions of positions. Generally, the risk of the doubled position will not be $2L$ but rather a random variable with a wider probability distribution. Similar effects are possible for subadditivity; the issue is related to the Unified versus Compartmentalised Risk Measurement section on whether a compartmentalised measurement of risk is appropriate. For instance, there may exist two positions which cause individual risks L_1 and L_2, respectively, if held alone, but the risk of holding both positions may be more severe than $L_1 + L_2$ for similar reasons as in the above example. Knowing this, one might question the subadditivity property as such because it requires $R(L_1 + L_2) \leq R(L_1) + R(L_2)$. However, not subadditivity is to blame but a potential misunderstanding of $L_1 + L_2$ as the risk of holding both positions together. These considerations imply two lessons:

- It may be problematic to assume that a vector of assets linearly maps into the associated vector of random losses.
- If a "risk measure" is defined as a composed mapping *from positions* (via the loss variable) to numbers, this mapping is generally not coherent. Assuming coherence can lead to an underestimation of risk.

Is VaR Failing Subadditivity Relevant in Practice?

The favourite textbook example of VaR violating subadditivity is constructed with the help of two large losses the probability of which is lower than the confidence level of the VaR. When measured separately, each loss can have zero VaR but when aggregated, the probability that either of the losses occurs may exceed the confidence level so that the VaR of the aggregated loss is positive.

As the textbook example relies on jumps in the loss distribution one might conjecture that VaR works properly if loss distributions are smooth or if discrete losses are superimposed by sufficiently large smooth ones. Whether this intuition is correct ultimately depends on the situation and particularly on the tail thickness of the loss distributions:

- McNeil et. al. (2005) present an example of a *continuous* two-dimensional loss distribution in which VaR violates subadditivity. While this is alarming in that it does not build on the abovementioned textbook ingredients, the example is still rather artificial.

- If the joint distribution of risk factors is elliptical (multivariate normal, e.g.), VaR is subadditive; see McNeil et. al. (2005), Theorem 6.8.
- Gourier, Farkas, and Abbate (2009) give an example where the sum of some fat-tailed, continuously distributed, and independent (!) random variables has a larger VaR than the sum of individual VaRs. The example is rather exotic as one of the random variables has infinite mean. While VaR fails in that case, it must be conceded that there is no coherent *and* practicable alternative at all because any coherent risk measure must be infinite then.[8]
- Danielsson, Jorgensen, Samorodnitsky, Sarma, and de Vries (2005) prove that VaR is subadditive for a sufficiently high confidence level if the total loss has finite mean. Note, however, that this is not an "all clear" signal but an asymptotic result only. Generally it may happen that subadditivity is only achieved for impractically high confidence levels.
- Degen, Embrechts, and Lambrigger (2007) restrict their analysis to a parametric class of distributions but gain valuable insight into the interplay of tail thickness, confidence level, and subadditivity. For example, they find the 99%-VaR to be superadditive even for very moderate tail indices above 6, which means that moments of order 6 and lower may exist.[9] These are realistic cases in market risk. The dependence structure between the individual losses generally aggravates the problem but has surprisingly low impact in the cases considered.

[8] Gourier et. al. (2009) refer to Delbaen (2002) who shows in Theorem 13 that, given a continuous distribution and some continuity of the risk measure, any coherent risk measure larger or equal than the α-VaR cannot fall short of the α-expected shortfall; the latter is already infinite in Gourier's example so that no useful coherent measure of that risk can exist.

[9] The higher the tail index, the thinner is the tail. Degen et. al. (2007) consider random variables the distribution tail of which is as thick as that of a transform $\exp(gZ + 0{,}5hZ^2)$ of a standard normal Z; e.g., $g = 2.3$ and $h = 0.25$ make the VaR super-additive; the tail index is 4 in this example.

To sum up, while the literature provides us with conditions that assure VaR is subadditive (and thus coherent), these conditions are generally *not fulfilled* in the market risk context; for example, Balaban, Ouenniche and Politou (2005) estimate tail indices between 1 and 2 for UK stock index returns over holding periods between 1 and 10 days, meaning that these tails are substantially heavier than necessary for assuring the subadditivity of VaR in general.

Expected Shortfall

Expected shortfall (ES) is the most well-known risk measure following VaR. It is conceptually intuitive and has firm theoretical backgrounds; see, e.g., Dunn (2009), Artzner et. al. (1999), Acerbi and Tasche (2002), Sy (2006), and Yamai and Yoshiba (2005). Therefore, it is now preferred to VaR by an increasing number of risk managers in the industry.

ES corrects three shortcomings of VaR. First, ES does account for the severity of losses beyond the confidence threshold. This property is especially important for regulators, who are, as discussed above, concerned about exactly these losses. Second, it is always subadditive and coherent. Third, it mitigates the impact that the particular choice of a single confidence level may have on risk management decisions, while there is seldom an objective reason for this choice.

To define ES, let L be a random loss with distribution function F_L and $\alpha \in (0,1)$ a confidence level (close to 1). Recall that the α-VaR is defined as the α-quantile of F_L. The ES at level α is defined by

$$ES_\alpha \equiv \frac{1}{1-\alpha}\int_\alpha^1 VaR_u(L)du \qquad \textbf{(5.1)}$$

and can thus be understood as an average of all VaRs from level α up to 1. ES is a coherent risk measure—and so subadditive. It is continuous in α and thus avoids cliff effects that may appear when the distribution has discrete components.

If the loss distribution is continuous, there is an even more intuitive representation:

$$ES_\alpha = E(L|L \geq VaR_\alpha), \qquad \textbf{(5.2)}$$

i.e., ES is then the expected loss conditional on this loss belonging to the $100(1-\alpha)$ percent worst losses. This measure has several other names like tail conditional expectation (TCE) or conditional VaR (CVaR). It is the key to simulations-based calculations of ES but care has to be taken as it does not always coincide with ES, and it is also not necessarily subadditive. The technical problem arises if the distribution function jumps from a value below the VaR confidence level to a value above it. Then, a correction term must be introduced into (2) to reconcile it with the correct ES from (1); see Acerbi and Tasche (2002).

The calculation of ES and the marginal contributions of assets to portfolio ES is more challenging than the corresponding calculations for VaR, especially for high confidence levels, because a formula for the α-quantiles of the loss distribution is often missing. Simulations need to be done in most cases. Since the introduction of expected shortfall, substantial progress has been made on computational issues, mainly through the application of importance sampling techniques (Kalkbrener, Lotter, and Overbeck (2004), Egloff, Leippold and Jöhri (2005), or Kalkbrener, Kennedy, and Popp (2007)). Research suggests that computational techniques have advanced to a point that expected shortfall is a viable risk management option for financial institutions.

Remark 3: In Remark 1, it is noted that utility optimisation using a VaR constraint can lead to perverse investment decisions. Risk measures which control the first moment of a random variable (such as ES) have been proposed to overcome this problem. However, recently Wylie, Zhang, and Siu (2010) showed that in the context of hedging both ES and VaR can give rise to discontinuous hedging behaviour that can lead investors to take extremely high-risk positions even when apparently minimising the risk measures.

Backtesting ES

Intuitively, backtesting ES is more complicated and/or less powerful than backtesting VaR because the robust statistic given by the number of VaR violations, as the most common VaR backtest statistic, must be replaced by something that accounts for the *magnitude* of VaR exceedances so that ES backtests by nature have to cope with the size of outliers.

Whether specialised ES backtests are good or not, one simple option is always available: during an ES calculation, the VaR at the same α can be generated as a by-product with low additional effort. One can backtest this VaR with traditional methods; if the VaR is rejected, the corresponding ES calculation can hardly be correct. Of course, VaR backtest acceptance does not guarantee the correctness of the ES calculation, and this would be true even if the VaR backtest were always right.

Some backtests verify if the VaR correctly adjusts for changes in risk dynamics ("conditional coverage"; see Berkowitz and O'Brien (2002)). According to Pritsker (2006), they exploit the fact that exceedances of a correctly calculated VaR "should not help forecast future exceedances. Therefore, the autocorrelation function of the VaR exceedances should be equal to 0 at all lags." It is hard to decide whether ES or VaR is verified with these backtests because VaR exceedances are the very constituents of ES.

Because backtests that are strictly focused on some historical estimator of the risk measure, like the number of VaR violations, often have low power, several authors propose to backtest the whole distribution (or at least the tail), for instance by transforming loss realisations with the forecasted loss distribution: If the latter is correct, the transformed sample must be equally distributed on [0,1]. This hypothesis can be tested (Berkowitz (2001)). While not all backtests of this kind could be used in regulation,[10] Kerkhof and Melenberg (2004) follow this approach to develop test statistics directly applicable to VaR and ES. The test statistic for the ES involves, besides the forecasted ES and VaR, also the calculation of the ES of the *squared* loss, which would be a tolerable extra effort in practice.

Kerkhof and Melenberg (2004) show that their backtest statistics for ES perform better than those for VaR. They also derive regulatory multiplication factors for their backtests and conclude that "the resulting regulatory capital scheme using expected shortfall compares favourably to the current Basel Accord backtesting scheme." It is important to notice that, according to Kerkhof and Melenberg (2004), a comparison of an α-ES with an α-VaR is not "fair" in the context of economic or regulatory capital. Since $ES_\alpha \geq VaR_\alpha$ for the same confidence level α, they lower the confidence level α' for the ES such that $ES(\alpha') \approx VaR(\alpha)$. The intuition is that a regulator would require roughly the same amount of capital for a fixed portfolio, irrespective of the risk measure in use.

[10] Some of these tests require that the bank fully specifies the loss distribution in the tail, not just the risk measure (ES or VaR). While this should not be a problem for bank internal purposes, a fully specified tail distribution would entail a fairly complex interface between bank and supervisor.

This aspect is important not only in the context of backtesting but also when estimation errors for ES and VaR are compared. Yamai and Yoshiba (2005) find ES estimates of fat (generalised Pareto distributed) tailed losses to be much more volatile than their VaR counterparts but they compare ES and VaR at the same confidence level. A comparison in the spirit of Kerkhof and Melenberg (2004) seems not to have been conducted so far but could easily be done.

Wong (2008) suggests another backtest statistic for ES that accounts for the small samples of VaR exceedances. The statistic is derived for normally distributed losses and turns out to perform very well under these assumptions. The test is also powerful in detecting non-normal VaR exceedances. For the case that a bank models non-normal losses when calculating the ES, Wong suggests to derive adapted saddle-point approximations for the estimator's distribution or to use the sample transform as used by Berkowitz (2001) and Kerkhof and Melenberg (2004). These results are promising, but in the context of banking regulation it must be taken into account that Wong's backtest would require that banks provide more information than they currently do for regulatory backtests. At present, past returns are compared with reported VaRs. With Wong's backtest, the bank would also have to report its estimates of tail thickness, which is potentially involved with weird incentives. For instance, banks might, keeping minimum capital constant, be tempted to rely on certain tail distributions under which Wong's backtest has particular low power so that it is difficult to provide firm evidence of wrong risk reporting. Whether such concerns are substantial is left to future research.

Spectral Risk Measures

Spectral risk measures (SRM) are a promising generalisation of ES (Acerbi (2002)). While the α-ES assigns equal weight to all β-VaRs with $\beta \geq \alpha$ but zero to all others, an SRM allows these weights to be chosen more freely. This is implemented by a weight function $w:[0,1] \rightarrow [0, \infty)$ that integrates to 1. An SRM is formally defined as

$$SRM = \int_0^1 w(u) VaR_u(L) du.$$

Expected shortfall is a special case of spectral measure, where $w(u) = (1-\alpha)^{-1} 1_{\{\alpha \leq u \leq 1\}}$. The definition of SRM is restricted to functions w that increase over [0,1], which ensures that the risk measure is coherent. This restriction also implies that larger losses are taken more seriously than smaller losses and thus the function w establishes a relationship to risk aversion. The intuition is that a financial institution is not very risk averse for small losses, which can be absorbed by income, but becomes increasingly risk averse to larger losses. As there may be a level of loss where employing additional capital to absorb yet higher loss is no longer desirable, such losses should be given the highest weights from a regulator's angle because often the public would have to bear such losses. Intuitively, a weight function that increases can also be thought of as marginal costs that rise while losses become increasingly rare, i.e., large.

Another advantage of SRM over ES (and VaR, a fortiori) is that they are not bound to a single confidence level. Rather, one can choose w to grow continuously with losses and thereby make the risk measure react to changes in the loss distribution more smoothly than the ES, and avoid the risk that an atom in the distribution being slightly above or below the confidence level has large effects.

If the underlying risk model is simulation-based, the additional effort to calculate an SRM as opposed to the ES seems negligible; the simulated VaR realisation are just differently weighed (Acerbi (2002)).

In spite of their theoretical advantages, SRMs other than ES are still seldom used in practice.[11] However, insurers use the closely related concept of *distortion measures* (see the next section). Prominent examples such as the measure based on the Wang transformation (see next page) are also SRMs.

Remark 4: Leaving aside that w must be increasing to meet the definition of SRM, VaR is a limiting case of spectral risk measures: for instance, the sequence of SRMs based on the weight functions $w_n(u) \equiv 0.5n 1_{\{\alpha - n^{-1} \leq u < \alpha + n^{-1}\}}$ converges to the α-VaR.

Other Risk Measures

There also are a number of other risk measures which are briefly introduced in this subsection.

[11] At least one reputable risk consulting company reports it is currently implementing an SRM-based risk management system for some of its clients.

Distortion risk measures: These measures are used in actuarial risk measurement. The definition is very general; both spectral risk measures (including ES) and the VaR are nested. To define distortion risk measures, let D be any distribution function on [0,1] that is right-continuous and increasing with $D(0) = 0$ and $D(1) = 1$. This D is called the *distortion function*. A *distortion risk measure* of loss L is defined as

$$DM(L) \equiv \int_0^1 VaR_u(L)dD(u).$$

Each spectral risk measure is clearly a distortion risk measure; to see this, recall that the weight function w integrates to 1 and observe that the SRM and the distortion measure defined by the antiderivative $D(u) = \int_0^u w(s)ds$ are identical.

Distortion risk measures are not necessarily coherent; the definition allows for distortion functions with a non-monotonous derivative (this is just the weight function of the corresponding SRM), whereas Acerbi (2002) has shown that the monotonicity of w is also necessary for the risk measure to be coherent.[12]

The VaR has a representation as a distortion risk measure by $D_{VaR}(u) = 1_{\{u \geq \alpha\}}$.

The Wang transform (Wang (2001)) $D_\theta^{Wang}(u) = \Phi(\Phi^{-1}(u) + \log\theta)$, where Φ denotes the Gaussian distribution function and $\theta < 1$, is an interesting distortion function. The corresponding risk measure is also a spectral risk measure because the first derivative of D_θ^{Wang} is strictly increasing. Hence the Wang transform indeed implements risk aversion over the whole range of losses but particularly in the tail. It has been applied to the pricing of catastrophe insurance contracts and exotic option pricing where Black-Scholes assumptions cannot be applied.

Variance: The variance is historically the most important risk measure and widely used in practice. It has many desirable properties but at least two drawbacks from a regulatory perspective. McNeil et. al. (2005) state "if we want to work with variance, we have to assume that the second moment of the loss distribution exists. . . . [V]ariance is a good measure of risk only for distributions which are (approximately) symmetric. . . . However, in many areas of risk management, we deal with highly skewed distributions."

The **mean deviation**, defined as $MD(L) \equiv \mathbf{E}|L - \mathbf{E}L|$, can do without second moments but suffers from the same problems with skewed distributions as the variance. It is less accessible to analytical treatment than the variance and therefore rarely used as a risk measure.

Upper partial moments (see McNeil et. al. (2005)): Given a loss distribution F_L, an exponent $k \geq 0$ and a reference point q, which could be some VaR, the upper partial moment $UPM(k,q)$ is defined as

$$UPM(k,q) = \int_q^\infty (l-q)^k \, dF_L(l).$$

Hence, for $k > 1$ an UPM measures losses beyond the threshold q with increasing weight. It is therefore related to spectral risk measures in spirit but not equivalent in analytic terms. The higher k is, the more conservative is the UPM. For $k = 1$ and continuous loss distributions, there is a close relationship with expected shortfall:

$$UPM(1, VaR_\alpha) = (1-\alpha)(ES_\alpha - VaR_\alpha).$$

Left-tail measure: In a similar vein of mean deviation and lower (upper) partial moment, Wu and Xiao (2002) propose a *left-tail measure*, defined as the conditional standard deviation of VaR exceedances, i.e.,

$$LTM \equiv \sqrt{E\left\{\left[L - E(L \mid L \geq VaR_\alpha)\right]^2 \mid L \geq VaR_\alpha\right\}}$$

Wu and Xiao (2002) show that the left-tail measure is useful particularly for the measurement of non-normal tail risks. This risk measure has several undesirable features such as a lack of coherency and a heavy burden of calculation.

Conclusions

While VaR has been criticised for its lack of coherence, until recently it was unclear whether this flaw is relevant for real asset portfolios, particularly for risks in the trading book. Degen et. al. (2007) have shown that the lack of coherence can be an important problem for trading book risk measurement. A risk measurement based on VaR is thus not necessarily conservative.

The ES avoids the major flaws of VaR but its fundamental difference from VaR—that it accounts for the magnitude

[12] Wang (2001) claims all smooth distortion measures are coherent. This is wrong as subadditivity is missing in general. Wang (2001) means to build on Wang, Young and Panjer (1997) which, however, state that a distortion measure is subadditive *if it is convex* (in our notation). The latter is correct and conforms to Acerbi (2002).

of losses beyond a threshold—is an equally important advantage. By this, it aligns the interests of bank managers and owners to those of the public much better than VaR.

Much of the criticism of ES that has been brought forward in defence of VaR could be refuted. Advanced simulation techniques have helped to make ES calculations stable enough, and ES and VaR backtests have similar power, if compared on the basis that both risk measures have roughly the same value.

Spectral risk measures are a promising generalisation of expected shortfall. The main advantages are improved smoothness and the intuitive link to risk aversion. If the underlying risk model is simulations-based, the additional calculation effort as opposed to ES seems negligible.

STRESS TESTING PRACTICES FOR MARKET RISK

Overview

VaR limitations have been highlighted by the recent financial turmoil. Financial industry and regulators now regard stress tests as no less important than VaR methods for assessing a bank's risk exposure. A new emphasis on stress testing exercises derives also from the amended Basel II framework which requires banks to compute a valid stressed VaR number.

A stress test can be defined as a risk management tool used to evaluate the potential impact on portfolio values of unlikely, although plausible, events or movements in a set of financial variables (Lopez (2005)). They are designed to explore the tails of the distribution of losses beyond the threshold (typically 99%) used in value-at-risk (VaR) analysis.

However, stress testing exercises often are designed and implemented on an ad hoc compartmentalised basis, and the results of stress tests are not integrated with the results of traditional market risk (or VaR) models. The absence of an integrated framework creates problems for risk managers, who have to choose which set of risk exposures are more reliable. There is also the related problem that traditional stress testing exercises typically remain silent on the likelihood of stress-test scenarios.

A survey of stress testing practices conducted by the Basel Committee in 2005 showed that most stress tests are designed around a series of scenarios based either on historical events, hypothetical events, or some combination of the two. Such methods have been criticised by Berkowitz (2000a). Without using a risk model the probability of each scenario is unknown, making its importance difficult to evaluate. There is also the possibility that many extreme yet plausible scenarios are not even considered.

Berkowitz proposed the integration of stress testing into formal risk modelling by assigning probabilities to stress-test scenarios. The resulting risk estimates incorporate both traditional market risk estimates and the outcomes of stress tests, as well as the probabilities of each. Therefore, they provide an integrated set of risk indicators and estimates to work with.

Incorporating Stress Testing into Market-Risk Modelling

Traditional stress testing exercises can be classified into three main types, which differ in how the scenarios are constructed:

1. historical scenarios;
2. predefined or set-piece scenarios where the impact on P/L of adverse changes in a series of given risk factors is simulated;
3. mechanical-search stress tests, based on automated routines to cover prospective changes in risk factors, then the P/L is evaluated under each set of risk-factor changes, and the worst-case results are reported.

All these approaches depend critically on the choice of scenarios. A related problem is that the results of stress tests are difficult to interpret because they give no idea of the probabilities of the events concerned (Berkowitz (2000a)). These criticisms can be addressed by integrating stress testing into the market risk modelling process and assigning probabilities to the scenarios used in stress testing. Once scenarios are put in probabilistic form, a unified and coherent risk measurement system is obtained rather than two incompatible ones and backtesting procedures can be applied to impose some (albeit limited) check on scenarios. Inevitably, the choice of scenarios will remain subjective, but even there, the need to assign probabilities to scenarios will impose some discipline on risk management.

Several authors have developed an integrated approach to stress testing including Kupiec (1998) who examines cross-market effects resulting from a market shock and Aragones et. al. (2001) who incorporated hypothetical stress events into an Extreme Value Theory (EVT) framework.

Alexander and Sheedy (2008) analysed the problem of determining the most suitable risk model in which to conduct a stress test. Obviously if the model is mis-specified, their approach is vulnerable to a considerable degree of model risk. Hence a significant part of their research is supported through backtests, which are designed to reduce the model risk in risk models that are used for stress testing. They conduct backtests for eight risk models, including both conditional and unconditional models and four possible return distributions. Their backtesting experiment suggests that unconditional historical simulation, currently the most popular VaR methodology in the industry according to Perignon and Smith (2006), is likely to be mis-specified and is therefore unsuited for stress testing purposes.

Breuer et. al. (2009) define an operational definition to three requirements which the Basel Committee specifies for stress tests: plausibility and severity of stress scenarios as well as suggestiveness of risk-reducing actions. The basic idea of their approach is to define a suitable region of plausibility in terms of the risk-factor distribution and search systematically for the scenario with the worst portfolio loss over this region. One key innovation of their approach compared with the existing literature is the solution of two open problems. They suggest a measure of plausibility that is not dependent to the problem of dimensional dependence of maximum loss and they derive a way to consistently deal with situations where some but not all risk factors are stressed. They show that setting the non-stressed risk factors to their conditional expected value given the value of the stressed risk factors, the procedure first suggested by Kupiec (1998), maximises plausibility among the various approaches used in the literature. Furthermore, Breuer et. al. (2010b) propose a new method for analyzing multi-period stress scenarios for portfolio credit risk more systematically than in the current practice of macro stress testing. This method quantifies the plausibility of scenarios by considering the distance of the stress scenario from an average scenario. For a given level of plausibility their method searches systematically for the most adverse scenario for the given portfolio.

Finally, as a general point, it must be underlined that for the purposes of calculating the P&L impact of stress shock-factors it is generally assumed that the shock occurs instantaneously, i.e., that traders have no opportunity to re-hedge or adjust their positions, and it is ignored the impact of declining tenors for, for example, futures and options contracts. Apart from simplifying the calculations, such an assumption could be unreasonable in some cases given the practical experience of the actions of traders during historical events, and it may generate inconsistent results by amplifying the magnitude of the losses. Such issues have not yet been addressed in the literature.

Stressed VaR

The pressing technical issue now facing financial institutions that intend to comply with the amended Basel II framework is to understand how to calculate a valid stressed VaR number. After the revisions of July 2009, banks have to calculate a VaR using the risk engine it normally uses but "with model inputs calibrated to historical data from a continuous 12-month period of significant financial stress relevant to the bank's portfolio" (Basel Committee on Banking Supervision (2009b)).

An over-simplistic interpretation of this specification might be to increase the assumed volatilities of the securities in a portfolio. This would have the effect of lengthening the tails of the Gaussian (normal) loss distributions that underlie all standard VaR calculations.

However, in order to calculate stressed VaR accurately it is also necessary to stress the correlation matrix used in all VaR methodologies. It is a repeated observation that during times of extreme volatility, such as occurs during every market crash, correlations are dramatically perturbed relative to their 'normal' historical values. In general, most correlations tend to increase during market crises, asymptotically approaching 1.0 during periods of complete meltdown, such as occurred in 1987, 1998 and 2008.

One possibility is to adopt the conditional stress test approach of Kupiec (1998). In this approach, the risk factor distributions are conditional on an extreme value realisation of one or more of the risk factors. Conditional on a large move of at least one factor, the conditional factor covariance matrix exhibits much higher correlations among the remaining factors. In this approach, the apparent shift in the correlation structure is a consequence of conditioning the distribution on a large factor shock. The unconditional correlations remain unchanged. Analysing a large number of stress test results for currency portfolios over the Asian currency crisis period, Kupiec shows that the conditional stress test process performs extremely well as very few stress test violations are recorded during this crisis period.

An alternative approach to conditional correlation is to stress the unconditional correlation matrix of the risk factors. Unfortunately, this approach is not as straightforward as the conditional correlation approach or stretching the tails of the loss distributions. The VaR calculation engine requires a correlation matrix that satisfies the mathematical property of positive definiteness, which is a way of saying that all of the correlations are internally consistent with each other. Noisy or erroneous historical price data can result in matrices that are not positive definite. Perturbing the correlation matrix, which is necessary for a true stressed VaR calculation, may result in correlation matrices that also violate the internal consistency requirement. If the matrix is not positive definite the VaR calculus will fail, so methods have to be devised to modify the stressed matrix until it becomes positive definite. Kupiec (1998) discusses some practical methods that can be used to address this problem.

Besides these technical issues one may also more fundamentally consider concepts that are not covered by the current regulatory definition of stressed VaR. A more sophisticated approach might include not only linear transforms of multivariate normal risk factors but also employing 'fat-tailed' distributions to model the extreme loss events more accurately. Examples of those 'extreme value theory' distributions are the Gumbel, Generalised Pareto, Weibull, Fréchet, and the Tukey g&h distributions.

However, one should keep in mind that the stressed VaR is from a theoretical perspective an imperfect solution—its purpose is to reflect that current market conditions may not lead to an accurate assessment of the risk in a more stressful environment. Extreme value theory distributions may already incorporate extreme market conditions and could in principle make a stressed VaR redundant. In general, these distributions are flexible enough to obtain very good fits but serious robustness issues arise instead, as regulators and risk managers had to learn in the context of operational risk, for instance.

Conclusions

More recent research advocates the integration of stress testing into the risk modelling framework. This would overcome drawbacks of reconciling stand-alone stress test results with standard VaR model output.

Progress has also been achieved in theoretical research on the selection of stress scenarios. In one approach, for example, the "optimal" scenario is defined by the maximum loss event in a certain region of plausibility of the risk factor distribution.

The regulatory "stressed VaR" approach is still too recent to have been analyzed in the academic literature. Certain methods that could be meaningful in this context can be identified in the earlier literature on stress testing. Employing fat-tailed distributions for the risk factors and replacing the standard correlation matrix with a stressed one are two examples.

UNIFIED VERSUS COMPARTMENTALISED RISK MEASUREMENT

Overview

In this section, we survey the academic literature on the implications of modelling the aggregate risks present across a bank's trading and banking books using either a compartmentalised approach—namely, the sum of risks measured separately—or a unified approach that considers the interaction between these risks explicitly. Finally, we survey the recent literature on the systemic implications of the current regulatory capital requirements that aggregate capital requirements across risk types.

In many financial institutions, aggregate economic capital needs are calculated using a two step procedure. First, capital is calculated for individual risk types, most prominently for credit, market and operational risk. In a second step, the stand-alone economic capital requirements are added up to obtain the overall capital requirement for the bank.

The Basel framework for regulatory capital uses a similar idea. As discussed by Cuenot, Masschelein, Pritsker, Schuermann and Siddique (2006), the Basel framework is based on a "building block" approach such that a bank's regulatory capital requirement is the sum of the capital requirements for each of the defined risk categories (i.e., market, credit and operational risk), which are calculated separately within the formulas and rules that make up Pillar 1. Capital requirements for other risk categories are determined by the supervisory process that fits within Pillar 2; see Figure 5-2 which is reproduced from Cuenot et. al. (2006). This approach is therefore often referred to as a non-integrated approach to risk measurement. An integrated approach would, by contrast, calculate capital for all the risks borne by a bank simultaneously in one single step and accounting for possible correlations and interactions, as opposed to adding up compartmentalised risk calculations.

Pressure to reconsider the regulatory compartmentalised approach came mainly from the financial industry, where it has been frequently argued that a procedure that simply adds up economic capital estimates across portfolios ignores diversification benefits. These alleged benefits have been estimated to be between 10 and 30% for banks (see Brockmann and Kalkbrener (2010)).

Capital diversification arguments and estimates of potential capital savings are partially supported in the academic literature. More recently this view and the estimates have been fundamentally challenged by the Basel Committee (Basel Committee on Banking Supervision (2009)) and by Breuer et. al. (2010a). These papers have pointed out that nonlinear interaction between risk categories may even lead to compounding effects. This fact questions whether the compartmentalised approach will in general give a conservative and prudent upper bound for economic capital.

Is this a merely academic debate or does it have practical implications for reform considerations related to the trading book? In this section, we survey the main arguments and give a brief review of the main papers and their findings. We then discuss policy implications that might be relevant for a discussion related to potential future reform related to the trading book.

Aggregation of Risk: Diversification versus Compounding Effects

Diversification is a term from portfolio theory referring to the mix of a variety of investments within a portfolio. Since different investments will develop differently in the future with value losses in some investment offset by value gains in another investment, the overall portfolio risk is reduced through the spreading of risk. In a similar way, the assets of a bank can be thought of as an overall portfolio that can be divided into subportfolios. If risk analysis is done by looking at risk measures at the level of the subportfolios and the risk measures are added up, the intuition of diversification suggests that we should arrive at a conservative risk measure for the bank as a whole.

So, what is wrong with this straightforward intuition about diversification between market, credit and other risk categories? The flaw in the intuition lies in the fact that it is usually not possible to divide the overall portfolio of a bank into subportfolios purely consisting of market, credit and operational risk; these risk categories are too intertwined in a modern financial institution to possibly separate in a meaningful way. In short, we cannot construct

	Banking book	Trading book
Pillar 1	Credit risk	
	Counterparty credit risk	
		Interest rate risk (general and specific)
		Equity risk (general and specific)
	Foreign exchange risk	
	Commodity risk	
	Operational risk	
Pillar 2	Interest rate risk	
	Concentration risk	
	Stress tests	
	Other risks (liquidity, residual, business...)	

FIGURE 5-2 Overview of risk categories relevant for banking book and trading book in Pillar 1 and Pillar 2.

Source: Cuenot et. al. (2006).

a subportfolio of risk factors. It is therefore incorrect to think of the banking book as a subportfolio of the overall bank portfolio for which only credit risk is relevant. It is also incorrect to view the trading book as another subportfolio related solely to market risk.

A simple way to summarise this argument is to consider a portfolio of loans. The interest rate risk related to such a portfolio is usually counted as a market risk, and this risk affects the bank's refinancing costs and the revaluation of these loans. If the interest rate risk is borne by the creditors in some way, this market risk suddenly may transform into a credit risk for the bank. So, do assets with a value that fluctuates with interest rates belong in a subportfolio for market risk or in a subportfolio of credit risk? They clearly belong to both, because each loan has a market risk component as well as a credit risk component simultaneously. Trading book positions with counterparty risks or positions related to carry trades fall into the same category.

Breuer et. al. (2010a) consider portfolios of foreign currency loans, which are loans denominated in a foreign currency extended to domestic creditors with income in domestic currency. The credit risk in these portfolios is always a function of the market risk (i.e., exchange rate movements), and the risk of each position in a foreign currency loan portfolio has simultaneously a credit and a market risk component. Adding up capital and hoping for an upper bound amounts to ignoring possible "malign risk interactions" as they are called in Breuer et. al. (2010a). This issue has been known in the market risk literature for a long time as "wrong way risk." Wrong way risk is the risk arising from the problem that the value of a trading position is inversely correlated with the default risk of some counterparty.

From these examples, we see that a formation of subportfolios along the lines of risk factors—and for that matter across banking and trading books—is usually not possible. Breuer et. al. (2010a) indeed show that the ability to form subportfolios along the lines of risk categories is a sufficient condition for diversification effects to occur. Since we can in general not form such subportfolios, we must anticipate the possibility that there can be risk compounding effects between the banking and the trading book. In short, while the intuition of diversification is inviting, it does not apply to the interaction of banking and trading books since there are in general no pure subportfolios of market, credit or operational risks.

This insight is important because it demonstrates that "diversification effects" that are derived from papers using a so-called "top-down" approach are often assuming what they want to derive. By construction, the assumption of splitting up the bank portfolio into subportfolios according to market, credit and operational risk assumes that this can indeed be done. If such a split were possible, it follows from the results in Breuer et. al. (2010a) that diversification effects must occur necessarily.

To estimate the quantitative dimension of the problem, we therefore must focus on papers working with a "bottom-up" approach. We also need to examine the results of papers based on the "top-down" approach that assumes risk separability at the beginning of the analysis. In this section, we survey several key papers that use either of these risk aggregation methods. As part of this literature survey, we provide a summary of recent papers that estimate the range and magnitude of these differences between compartmentalised and unified risk measures. Our proposed measure is a simple ratio of these two measures, as used in other papers, such as Breuer et. al. (2010a). In that paper, the authors adopt the term "inter-risk diversification index" for the ratio; see also the related measure in Alessandri and Drehmann (2010). Ratio values greater than one indicate risk compounding, and values less than one indicate risk diversification. In the summary tables later in this chapter, we list the various papers, the portfolio analysed, the risk measures used, the horizon over which the risks are measured, and these risk ratios.

Papers Using the "Bottom-Up" Approach

As mentioned above, a common assumption of most current risk measurement models is that market and credit risks are separable and can be addressed independently. Yet, as noted as early as Jarrow and Turnbull (2000), economic theory clearly does not support this simplifying assumption.

While the reasons behind this common assumption are mostly operational in nature, some studies have used numerical simulation techniques to generate results. For example, Barnhill and Maxwell (2002) examine the economic value of a portfolio of risky fixed income securities, which they define as a function of changes in the risk-free interest rate, bond spreads, exchange rates, and the credit quality of the bond issuers. They develop a numerical simulation methodology for assessing the VaR of such a portfolio when all of these risks are correlated. Barnhill et. al. (2000) use this methodology to examine capital ratios for a representative South African bank. However, in these studies, the authors do not examine the differing values of their chosen risk measures using a unified risk measurement approach versus a compartmentalised approach that sums the independent risk measures.

The study by Jobst, Mitra and Zenios (2006) provides some analysis along these lines. The authors construct a simulation model, based on Jobst and Zenios (2001), in which the risk underlying the future value of a bond portfolio is decomposed into:

- the risk of a borrower's rating change (including default);
- the risk that credit spreads will change; and
- the risk that risk-free interest rates will change.

Note that the first item is more narrowly defined to represent the portfolio's credit risk, while the last item is more narrowly defined to represent the portfolio's market risk. However, the middle item is sensitive to both risks and challenges the notion that market and credit risk can be readily separated in this analysis. The authors use portfolios of US corporate bonds and one-year VaR and CVaR risk measures at the 95%, 99% and 99.9% confidence levels for their analysis.

In their analysis, the authors generate risk measures under three sets of assumptions. To concentrate on the pure credit risk contributions to portfolio losses, they simulate only rating migration and default events as well as recovery rates, while assuming that future interest rates and credit spreads are deterministic. The authors then allow future credit spreads to be stochastically determined, and finally, they allow future interest rates to be stochastically determined. Note that the latter case provides an integrated or unified risk measurement, according to our definition for this survey.[13]

[13] Note, however, that the authors do not conduct an analysis of a market risk scenario (i.e., deterministic ratings and stochastic credit spreads and interest rates). Thus, we cannot examine their ratio of unified to compartmentalised risk measures as discussed above.

The authors' results are quite strong regarding the magnitude of the risk measures across risk types and credit ratings. For AAA-rated bonds, the authors find that the unified risk measures at all three tail percentiles are on the order of ten times the pure credit risk measures, since highly-rated bonds are unlikely to default. As the credit quality of the portfolio declines, the ratio between the unified risk measures and the risk measures for pure credit risk drops to just above one for C-rated bonds.

Table 5-1 presents a short summary of several papers for which we can directly examine the ratio of unified to compartmentalised risk measures for bottom-up models. As mentioned earlier, the recent work of Breuer et. al. (2008, 2010a) provides a leading example of how market and credit risk cannot be readily separated in a portfolio, a fact that complicates risk measurement and works to undermine the simple assumptions underling additive risk measures.

In Breuer et. al. (2010a), the authors present analysis of hypothetical loan portfolios for which the impact of market and credit risk fluctuations are not linearly separable. They argue that changes in aggregate portfolio value caused by market and credit risk fluctuations in isolation should sum up to the integrated change incorporating all risk interactions very rarely. The magnitude and direction of the discrepancy between these two types of risk assessments can vary broadly. For example, the authors examine a portfolio of foreign currency loans for which exchange rate fluctuations (i.e., market risk) affect the size of loan payments and hence the ability of the borrowers to repay the loan (i.e., credit risk). For their empirically calibrated example, they use expected shortfall at various tail percentiles as their risk measure and examine portfolios of BBB+ and B+ rated loans. Their analysis shows that changes in market and credit risks can cause compounding losses such that the sum of value changes from the individual risk factors are smaller than the value change due to accounting for integrated risk factors.

In particular, their reported inter-risk diversification index for expected shortfall increased sharply as the tail quantile decreased, which suggests that the sum of the two separate risk measures becomes much less useful as an approximation of the total integrated risk in the portfolio as we go further into the tail. These index values also increase for all but the most extreme tail percentiles as the original loan rating is lowered. The authors argue that this example presents evidence of a "malign interaction of market and credit risk which cannot be captured by providing separately for market risk and credit risk capital." The authors show a similar qualitative outcome for domestic currency loans (i.e., loans for which default probability are simply a function of interest rates), although the index values are much lower.

In Breuer et. al. (2008), the authors use a similar analytical framework to examine variable rate loans in which the interaction between market and credit risk can be analysed. In particular, they model the dependence of credit risk factors—such as the loans' default probabilities (PD), exposure at default (EAD), and loss-given-default (LGD)—on the interest rate environment. A key risk of variable rate loans is the danger of increased defaults triggered by adverse rate moves. For these loans, market and credit risk factors cannot be readily separated, and their individual risk measures cannot be readily aggregated back to a unified risk measure. They conduct a simulation study based on portfolios of 100 loans of equal size by borrowers rated B+ or BBB+ over a one-year horizon using the expected shortfall measure at various tail percentiles. They find that the ratio of unified expected shortfall to the sum of the separate expected shortfalls is slightly greater than one, suggesting that risk compounding effects can occur. Furthermore, these compounding effects are more pronounced for lower-rated loans and higher loan-to-value ratios.

In contrast to this work, the paper by Grundke (2005) lays out a bottom-up model that assumes the separability of interest rate risk (i.e., market risk) and credit spread risk (i.e., credit risk). The author examines a calibrated multi-factor credit risk model that accommodates various asset value correlations, correlations between credit spreads and other model factors, and distributional assumptions for innovations. The author examines hypothetical loan portfolios of varying credit quality over a three-year horizon, both with and without the joint modelling of interest rates and credit spreads. To assess the joint impact of interest rate and credit risk, the author uses forward market interest rates instead of separate interest rate and credit spread processes. Interestingly, the reported VaR measures at various tail percentiles lead to ratios of unified VaR measures to summed VaR measures that range widely from near zero to one, which seems to be due mainly to the separability of the interest rate risk (i.e., market risk) and credit spread risk (i.e., credit risk) in the model.

Kupiec (2007) proposes a single-factor, migration-style credit risk model that accounts for market risk. This modelling approach generates a portfolio loss distribution that accounts for the non-diversifiable elements of the interactions between market and credit risks. The integrated exposure distribution of the model is used to examine capital allocations at various thresholds. These integrated capital allocations are compared to the separated assessments. The results show that capital allocations derived from a unified risk measure importantly alter the estimates of the minimum capital needed to achieve a given target solvency margin. The capital amount could be larger or smaller than capital allocations estimated from compartmentalised risk measures. Regarding specifically the Basel II AIRB approach, the author argues that the results show that no further diversification benefit is needed for banking book positions since no market risk capital is required. Thus, Basel II AIRB capital requirements fall significantly short of the capital required by a unified risk measure.

Numerically speaking, the risk measure used in this study is the amount of capital that the unified and the compartmentalised capital approaches generate as the appropriate value to assure funding costs of a certain magnitude calibrated to historical funding rates for specific credit ratings. The hypothetical portfolios of interest are corporate loans with various rating categories represented in proportion to historical data. The author examines a wide variety of alternative separated approaches with which to calculate economic capital measures, ranging from three different alternative credit risk models to several methods for measuring market risk. Correspondingly, the range of inter-risk diversification index values is quite wide for the AAA- and BBB-rated portfolios, ranging from about 0.60 to almost 4.00. In summary, the author's capital calculations show that capital allocations derived from a unified market and credit risk measure can be larger or smaller than capital allocations that are estimated from aggregated compartmentalised risk measures.

The studies discussed above examine the different risk implications of a unified risk measurement approach relative to a compartmentalised approach for specific portfolios. In contrast, Drehmann et. al. (2010) examine a hypothetical bank calibrated to be representative of the UK banking system as a whole. Within their analytical framework, they do not explicitly assume that market and credit risk are separable. The authors decompose the total risk in their bank scenario analysis into:

- the impact of credit risk from non-interest rate factors,
- the impact of interest rate risk (excluding the effect of changes in interest rates on credit risk), and
- the impact of the interaction of credit risk and interest rate risk.

The latter is calculated as the difference between the total impact of the scenario shock and the sum of the first two components.

Their simulations confirm that interest rate risk and credit risk must be assessed jointly for the whole portfolio to gauge overall risk correctly. In particular, the authors find in their simulations that if banks gauged credit risk by solely monitoring their write-offs, aggregate risk would be underestimated in the short term since a rate increase would also lower its net interest income and profits. Correspondingly, the bank's aggregate risk would be overestimated in the long run as net interest income and profits recover while write-offs continue to rise.

Their main variable of interest is net profits over twelve quarters after their macroeconomic stress scenario hits their representative bank, although they also report separate measures of write-offs and net interest income. They report that the interaction between interest rate and credit risk accounts for about 60% of the decline in capital adequacy for their calibrated bank. While the decline in capital adequacy does not perfectly match our other risk measures, we can still think of the diversification index here as the ratio of the capital decline for the unified risk framework relative to the capital decline that would come from separate identification of market and credit risks. Given their reported numbers, that ratio here is 100% / (100% − 60%) = 2.5, which suggests a very clear contribution of this interaction to risk management concerns.

Following up on the work of Drehmann et. al. (2010), Alessandri and Drehmann (2010) develop an integrated economic capital model that jointly accounts for credit and interest rate risk in the banking book; i.e., where all exposures are held to maturity. Note that they explicitly examine repricing mismatches (and thus market and credit risks) that typically arise between a bank's assets and liabilities.

For a hypothetical, average UK bank with exposures to only the UK and US, they find that the difference between aggregated and unified economic capital levels is often significant but depends on various bank features, such as the granularity of assets, the funding structure or bank pricing behaviour. They derive capital for the banking book over a one year horizon. For credit and interest rate risk, they define unexpected losses and thus economic capital as the difference between VaR at the specified 99% confidence level and expected losses. Note that their measures of economic capital for just credit risk and just interest rate risk do not fully disentangle these risks as the credit risk measure incorporates the effects of higher interest rates on default probabilities and the latter the effect of higher credit risk on income. The key point is that the framework represents a plausible description of how current capital models for the banking book capture these risks.

The authors examine the ratio of unified economic capital to the sum of the component measures at three VaR quantiles. For the 95th percentile of portfolio losses, unified capital measure is near zero, and thus the ratio is nearly zero as well. For the 99th percentile, the ratio is quite small at 0.03, but the ratio rises quickly to just over 50% for the 99.9th percentile. Note, however, that this result still suggests that the compartmentalised approach is more conservative than the unified approach. The authors examine certain modifications of their assumptions—such as infinitely fine-grained portfolios to increase the correlation of portfolio credit risk with the macroeconomic factors, banking funding scenarios from all short-term debt that is frequently repriced to all long-term debt that is repriced only on a yearly basis—and find some interesting difference with the base case scenario. However, the lower integrated capital charge holds.

On balance, these authors conclude that the bank's capital is mismeasured if risk interdependencies are ignored. In particular, the addition of economic capital for interest rate and credit risk derived separately provides an upper bound relative to the integrated capital level. Two key factors determine this outcome. First, the credit risk in this bank is largely idiosyncratic and thus less dependent on the macroeconomic environment; and second, bank assets that are frequently repriced lead to a reduction in bank risk. Given that these conditions may be viewed as special cases, the authors recommend that "As a consequence, risk managers and regulators should work on the presumption that interactions between risk types may be such that the overall level of capital is higher than the sum of capital derived from risks independently."

Papers Using the "Top-Down" Approach

An alternative method for determining total firm risk, primarily for enterprise-wide risk management, is to aggregate risks calculated for different business lines or different risk types using so-called "top-down" approaches. An important difference is that top-down approaches always reference an institution as a whole, whereas bottom-up approaches can range from the portfolio level up to an institutional level. With respect to market and credit risk, the top-down approach explicitly assumes that the risks are separable and can be aggregated in some way. As outlined by Cuenot et. al. (2006), firms may compute their market and credit risk capital separately and aggregate the two risk types by imposing some form of correlation between them. The top-down approach thus does not require a common scenario across risk types, but because the correct form of aggregation is not known, the approach "loses the advantages of logical coherence." In addition, as suggested by Breuer et. al. (2008, 2010a), the assumption of separable risk will generally prevent the ability to gauge the degree of risk compounding that might be present and instead typically provide support for risk diversification.

The literature is unclear on whether the combination of financial business lines within one organisation leads to an increase or decrease in risk. The literature as surveyed by Saunders and Walters (1994) and Stiroh (2004) suggests mixed results. However, as surveyed by Kuritzkes, Schuermann and Weiner (2003), several studies, including their own, suggest that reductions in economic capital arise from the combination of banking and insurance firms. The papers surveyed in Table 5-2 and below find this result as well for various risk combinations at the firm level.

For example, Dimakos and Aas (2004) decompose the joint risk distribution for a Norwegian bank with an insurance subsidiary into a set of conditional probabilities and impose sufficient conditional independence that only pairwise dependence remains; the total risk is then just the sum of the conditional marginals (plus the unconditional credit risk, which serves as their anchor). Their simulations indicate that total risk measured using near tails (95%–99%) is about 10%–12% less than the sum of the individual risks. In terms of our proposed ratio, the value ranges from 0.88 to 0.90. Using the far tail (99.97%), they find that total risk is often overestimated by more than 20% using the additive method. In terms of our proposed ratio of unified risk measure to the sum of the compartmentalised risk measures, its value would be 0.80.

Similarly, Kuritzkes et. al. (2003) examine the unified risk profile of a "typical banking-insurance conglomerate" using the simplifying assumption of joint normality across the risk types, which allows for a closed-form solution. They use a broad set of parameters to arrive at a range of risk aggregation and diversification results for a financial conglomerate. Based on survey data for Dutch banks on the correlations between losses within specific risk categories, their calculations of economic capital at the 99.9% level is lower for the unified, firm-level calculation than for the sum of the risk-specific, compartmentalised calculations. The ratio of these two quantities ranges from 0.72 through 0.85 based on correlation assumptions across market, credit and operational risk.

Rosenberg and Schuermann (2006) conduct a more detailed, top-down analysis of a representative large, internationally active bank that uses copulas to construct the joint distribution of losses. The copula technique combines the marginal loss distributions for different business lines or risk types into a joint distribution for all risk types and takes account of the interactions across risk types based on assumptions. Using a copula, parametric or nonparametric marginals with different tail shapes can be combined into a joint risk distribution that can span a range of dependence types beyond correlation, such as tail dependence. The aggregation of market, credit and operational risk requires knowledge of the marginal distributions of the risk components as well as their relative weights. Rosenberg and Schuermann assign inter-risk correlations and specify a copula, such as the Student-t copula, which captures tail dependence as a function of the degrees of freedom. They impose correlations of 50% for market and credit risk, and 20% for the other two correlations with operational risk; all based on triangulation with existing studies and surveys.[14]

Rosenberg and Schuermann find several interesting results, such as that changing the inter-risk correlation between market and credit risk has a relatively small impact on total risk compared to changes in the

[14] Note that different correlation values could lead to risk compounding, but it is not clear what those values might be and what values would be implied by the bottom-up exercises discussed here.

correlation of operational risk with the other risk types. The authors examine the sensitivity of their risk estimates to business mix, dependence structure, risk measure, and estimation method. Overall, they find that "assumptions about operational exposures and correlations are much more important for accurate risk estimates than assumptions about relative market and credit exposures or correlations." Comparing their VaR measures for the 0.1% tail to the sum of the three different VaR measures for the three risk types, they find diversification benefits in all cases. For our benchmark measure of the ratio between the unified risk measure and the compartmentalised risk measure, their results suggest values ranging from 0.42 to 0.89. They found similar results when the expected shortfall (ES) measure was used.

Note that the authors state that the sum of the separate risk measures is always the most conservative and overestimates risk, "since it fixes the correlation matrix at unity, when in fact the empirical correlations are much lower." While the statement of imposing unit correlation is mathematically correct, it is based on the assumption that the risk categories can be linearly separated. If that assumption were not correct, as suggested by papers cited above, the linear correlations could actually be greater than one and lead to risk compounding.

Finally, Kuritzkes and Schuermann (2007) examine the distribution of earnings volatility for US bank holding companies with at least $1 billion in assets over the period from 1986.Q2 to 2005.Q1; specially, they examine the 99.9% tail of this distribution. Using a decomposition methodology based on the definition of net income, the authors find that market risk accounts for just 5% of total risk at the 99.9% level, while operational risk accounts for 12% of total risk. Using their risk measure of the lower tail of the earnings distribution, as measured by the return on risk-weighted assets, their calculations suggest that the ratio of the integrated risk measure to the sum of the disaggregated risk measures ranges from 0.53 through 0.63.

Conclusions

Academic studies have generally found that at a high level of aggregation, such as at the holding company level, the ratio of the risk measures for the unified approach to that of the separated approach is often less than one, i.e., risk diversification is prevalent and ignored by the separated approach. However, this approach often assumes that diversification is present. At a lower level of aggregation, such as at the portfolio level, this ratio is also often found to be less than one, but important examples arise in which risk compounding (i.e., a ratio greater than one) is found. These results suggest, at a minimum, that the assumption of risk diversification cannot be applied without questioning, especially for portfolios subject to both market and credit risk, regardless of where they reside on the balance sheet.

Recent literature on the systemic implications of the current regulatory capital requirements that aggregate capital requirements across risk types suggests that this compartmentalised approach can—at least in general—be argued to contribute to the amplification of systemic risk, which is counter to its intentions.

In terms of policy implications, the academic literature suggests that if we are able to divide risk types easily across the trading book and the banking book (as is assumed in the top-down studies), diversification benefits appear to be certain, and aggregation of capital requirements across the books is conservative. However, recent studies have shown that if this risk separation cannot be done completely, simple aggregation of compartmentalised measures may not be conservative and, in fact, may understate the total risk. Such an outcome would clearly be undesirable as the necessary amount of capital could be underestimated by a significant margin.

These conclusions seem to directly question whether separate capital requirements for the trading and banking books provide a reasonable path to setting the appropriate level of capital for the entire firm. If we retained the different capital treatments, attempts could be made to fully detail each type of risk within each book, and the subsequent aggregation might then be considered conservative. However, performing such an analysis within the current and traditional separation between a trading and a banking book would require important changes in operational procedures. An alternative approach might be to develop a system of book keeping and risk allocation that does not artificially assign positions into different books when its risk characteristics are interrelated.

RISK MANAGEMENT AND VALUE-AT-RISK IN A SYSTEMIC CONTEXT

Overview

In this section, we survey the research literature on the systemic consequences of individual risk management systems and regulatory capital charges that rely on them. At the time when the Basel Committee implemented the MRA in 1996, risk management and banking regulation still was a subject that had received relatively little attention in the academic literature. Perhaps the most important change brought to the Basel framework by the MRA was the ability for banks to use their own quantitative risk models for determining the capital requirements for market risk.

Both conceptually and procedurally, this amendment was a significant departure from the previous regulatory approaches to determine bank capital. The conceptual innovation was that the notion of risk on which the new regulation relied was much closer to the notions of risk that were in use in the financial, economic and statistical research literature. Procedurally the amendment amounted to an official recognition that financial institutions themselves are in the best positions to assess their risk exposures. The new regulatory approach seemed to suggest that using and relying on this knowledge might be the best way to cope with methodological problems of risk assessment in a rapidly changing economic environment.

At the time of the amendment and in the years after, the academic literature on risk management and regulation largely accepted the conceptual reasoning behind the amendment and confined itself mostly to developing the technology of quantitative risk management itself. The discussion in the economics community remained sparse and largely sceptical.

Hellwig (1995, 1996) raised several important issues related to this new regulatory approach that did not take hold very much in the regulatory community but sound very modern in the current debate about the recent financial crises: Hellwig discussed incentive problems. Banks may find it desirable to bias their model development towards the goal of minimising capital. With hindsight, we know that the practice of determining capital based on VaR models helped large and international active banks to reduce greatly the amount of capital to be held against any given asset during the pre-crisis boom years. He also pointed out the difficulties related to using statistical techniques which work under the assumption of a stationary world in a non-stationary environment like financial markets. He also criticised the separation between market and credit risk while he acknowledged that quantitative models of integrated risk measurement are subject to the general problems outlined above.

During the discussion of the new Basel II framework, in May 2001, a group of academics at the Financial Markets Group (FMG) of the London School of Economics wrote a paper that raised a concern with respect to the use of value-at-risk that is more fundamental.[15] In the report's executive summary, there is a conclusion that calls into question the conceptual construction of the 1996 amendment: "The proposed regulations fail to consider the fact that risk is endogenous. Value-at-risk can destabilise and induce crashes when they would not otherwise occur."

In the current practice of risk management and regulation, these conclusions so far have only partly lead to a serious reconsideration of the framework initiated and extended more than a decade ago. In the current regulatory discussion, the general view seems to be that the conclusions from the financial crisis call for suitable expansions and amendments to the prevailing framework. In the meantime, the conclusions derived in the FMG paper have received more substantive underpinnings from academic research, both empirically and theoretically. The papers of Adrian and Shin (2008), the book of Shin (2008a) and joint work by Danielsson, Shin and Zigrand (2009) suggest that the use of value-at-risk models in regulation intended to function as a "fire extinguisher," function in practice rather like a "fire accelerant."[16] Rather than suggesting improving the VaR-based capital regulations by various refinements and amendments to the concepts in place, this literature suggests to abandon this approach and remove a VaR-based capital requirement from the regulatory framework. It should not be ignored, however,

[15] See Danielsson et. al. (2001).

[16] See Hellwig (2009).

that some of the new regulatory initiatives will likely dampen procyclical effects in the future. The *stressed VaR* introduced by the July 2009 revisions of the Market Risk Framework is a case in point: its calculation is based on estimates from bad historical periods of the economy and so acts rather "through the cycle." Admittedly, the stressed VaR is only one addend of total trading book capital.

Although the literature for this section generally refers to VaR as the risk measure at issue, it is important to bear in mind that the term VaR should be interpreted here in a wide sense since the results generally do not depend on this specific risk measure.

In the following we give a brief outline of the main arguments and explain the boom and bust amplification mechanism identified in this literature. We then go through some of the policy conclusions suggested by this analysis.

Intermediation, Leverage and Value-at-Risk: Empirical Evidence

Adrian and Shin (2010) empirically investigated the relationship between leverage and balance sheet size of the five major US investment banks shortly before the financial crises. All these institutions meanwhile left the broker-dealer sector, either because they were taken over or went bankrupt or were converted to bank holding companies. A major reason why these institutions are particularly interesting is because they all show a very clear picture of how financial intermediation works in a capital markets-based financial system with active balance sheet management through risk management systems.

When an intermediary actively manages its balance sheet, leverage becomes procyclical because risk models and economic capital require balance sheet adjustments as a response to changes in financial market prices and measured risks. This relationship follows from simple balance sheet mechanics. The following example is taken from Shin (2008a, pp. 24 ff.) Assume a balance sheet is given with 100 in assets and a liability side which consists of 90 in debt claims and 10 in equity shares. Leverage is defined as the ratio of total assets to equity, 10 in our example. If we assume more generally that the market value of assets is A and make the simplifying assumption that the value of debt stays roughly constant at 90 for small changes in A, we see that total leverage is given by:

$$L = \frac{A}{A - 90}$$

Leverage is thus related inversely to the market value of total assets. When net worth increases, because A is rising, leverage goes down, when net worth decreases, because A is falling, leverage increases.

Consider now what happens if an intermediary actively manages its balance sheet to maintain a constant leverage of 10. If asset prices rise by 1%, the bank can take on an additional amount of 9 in debt, its assets have grown to 110, its equity is 11, and the debt is 99. If asset values shrink by 1%, leverage rises. The bank can adjust its leverage by selling securities worth 9 and pay down a value 9 of debt to bring the balance sheet back to the targeted leverage ratio.

This kind of behaviour leads to a destabilising feedback loop, because it induces an *increase* in asset purchases as asset prices are rising and a *sale of assets* when prices are falling. Whereas the textbook market mechanism is self stabilising because the reaction to a price increase is a reduction in quantity demanded and an expansion in quantity supplied, and to a price decreases an expansion in quantity demanded and a contraction in quantity supplied, active balance sheet management reverses this self stabilising mechanism into a destabilising positive feedback loop.

Adrian and Shin (2010) document this positive relationship between total assets and leverage for all of the (former) big Wall Street investment banks. Furthermore, they produce econometric evidence that the balance sheet adjustments brought about by active risk management of financial institutions indeed has an impact on risk premiums and aggregate volatility in financial markets.

What Has All This to Do with VaR-Based Regulation?

Why would a bank target a constant leverage and what is the role of value-at-risk in all of this? The book of Shin (2008a) and the papers by Shin (2008b) and Adrian and Shin (2008) as well as by Danielsson, Shin and Zigrand (2009) explore this role in more detail.

If we consider the future value of bank assets *A* as a random variable, the value-at-risk (VaR) at a confidence level *c* is defined by

$$\Pr(A < A_0 - VaR) \leq 1 - c$$

The VaR is equal to the equity capital the firm must hold to be solvent with probability c. The economic capital is tied to the overall value-at-risk.

If a bank adjusts its balance sheet to target a ratio of value-at-risk to economic capital then bank capital to meet VaR is

$$K = \lambda \times VaR,$$

where λ is the proportion of capital to be held per total value-at-risk. This proportion may vary with time. Leverage is thus

$$L = \frac{A}{K} = \frac{1}{\lambda} \times \frac{A}{VaR}$$

Since VaR per value of assets is countercyclical, it directly follows that leverage is procyclical as the data in Adrian and Shin (2008) indeed show.[17]

The systemic consequences of this built-in risk limiting technology at the level of individual institutions works in the aggregate as an amplifier of financial boom and bust cycles. The mechanism by which the systemic amplification works is risk perception and the pricing of risk, even if all network effects and complex interconnectedness patterns in the financial system are absent.[18]

Consider intermediaries who run a VaR-based risk management system and start with a balance sheet consisting of risk-free debt and equity. Now an asset boom takes place, leading to an expansion in the values of securities. Since debt was risk-free to begin with, without any balance sheet adjustment, this leads to a pure expansion in equity. The VaR constraint is relaxed through the asset boom and creates new balance sheet capacity to take on more risky securities or increase its debt. The boom gets amplified by the portfolio decisions of the leveraged banking system.

Put differently, in a system of investors driven by a VaR constraint, investors' demand follows and amplifies the most recent price changes in the financial market. Price increases and balance sheet effects become intertwined through the active VaR-driven risk management of financial institutions.

Of course, the described mechanism also works on the way down. A negative shock drives down market values, tightening the VaR constraints of leveraged investors. These investors have to sell assets to reduce leverage to the new VaR constraint. By hardwiring VaR-driven capital management in banking regulation, a positive feedback loop with potent destabilising force both in booms and busts has been built into the financial system.

The mechanisms described in this section have been theoretically analysed in Shin (2008a), Danielsson, Shin, Zigrand (2009) theoretically and with explicit reference to value-at-risk. They are also central in the work of Geanakoplos (2009), although there the connection with VaR is not made explicit.

Conclusions

A literature stream on the systemic consequences of individual risk management systems as the basis of regulatory capital charges has found that the mechanical link between measured risks derived from risk models and historical data and regulatory capital charges can work as a systemic amplifier of boom and bust cycles.

The central mechanism that leads to this feedback loop works through the pricing of risk. In good times, when measured risks look benign, a financial institution that targets a regulatory capital requirement as a function of a model-based risk measure has slack capacity in its balance sheet that it can either use to buy additional risky assets or to increase its debt. This means that we have a mechanism where institutions are buying more risky assets when the price of these assets is rising and where they are buying less of these assets when prices are falling. The stabilising properties of the market mechanism are turned on their head. By this mechanic link of measured risk to regulatory capital a powerful amplifier of booms and busts is created at the system level counteracting the intention of the regulation to make the system as a whole safer.

It is important to recognise that while the current system may implement a set of rules that limit the risk taken at the level of individual institutions, the system may also enable institutions to take on more risk when times are

[17] This formal derivation of the procyclicality of VaR is taken directly from Shin (2008a).

[18] For this point, see also Geanakoplos (2009), who has shown in a theoretical model how risk-free debt may nevertheless give rise to fluctuations in leverage and risk pricing and thus create systemic spillover effects.

good and thereby lay the foundations for a subsequent crisis. The very actions that are intended to make the system safer may have the potential to generate systemic risk in the system.

These results question a regulatory approach that accepts industry risk models as an input to determine regulatory capital charges. This critique applies in particular to the use of VaR to determine regulatory capital for the trading book but it questions also an overall trend in recent regulation.

The amplifying mechanism identified in this section will be at work no matter how sophisticated VaR becomes, whether it is replaced by more sophisticated risk measures, like expected shortfall, or whether it goes beyond the naïve categorisation of risk classes (market, credit and operational) towards a more integrated risk measurement. These changes generally do not address the problems raised by the papers reviewed in this section. One exception is the *stressed VaR* introduced in July 2009. This new component of trading book capital acts more "through the cycle" than the "normal" VaR. Still some argue that what may be needed is a less mechanical approach to capital adequacy that takes into account a system-wide perspective on endogenous risk. The academic literature has identified many potential shortcomings in the currently regulatory approach for bank capital but it has yet to develop an alternative approach that simultaneously satisfies all the (sometimes conflicting) regulatory policy objectives.

References

Acerbi, C (2002): "Spectral measures of risk: a coherent representation of subjective risk aversion," *Journal of Banking and Finance,* vol 26, no 7, pp 1505-1518.

Acerbi, C and G Scandolo (2008): "Liquidity risk theory and coherent measures of risk," *Quantitative Finance,* vol 8, no 7, pp 681-692.

Acerbi, C and D Tasche (2002): "On the coherence of expected shortfall," *Journal of Banking and Finance,* vol 26, pp 1487-1503.

Adrian, T and H S Shin (2010): "Liquidity and leverage," *Journal of Financial Intermediation,* vol 19, no 3, pp 418-437.

—— (2008), "Financial intermediary leverage and value at risk," Federal Reserve Bank of New York, staff report no 338.

Alessandri, P and M Drehmann (2010): "An economic capital model integrating credit and interest rate risk in the banking book," *Journal of Banking and Finance,* vol 34, pp 730-742.

Alexander, C and E Sheedy (2008): "Developing a stress testing framework based on market risk models," *Journal of Banking and Finance,* vol 32, no 10, pp 2220-2236.

Almgren, R and N Chriss (2001): "Optimal execution of portfolio transactions," *Journal of Risk,* vol 3, pp 5-39.

Amihud, Y (2002): "Illiquidity and stock returns: cross-section and time-series effects," *Journal of Financial Markets,* pp 31-56.

Aragones, J, C Blanco and K Dowd (2001): "Incorporating stress tests into market risk modeling," *Derivatives Quarterly,* pp 44-49.

Aramonte, S, M Rodriguez and J Wu (2010): "Portfolio value-at-risk: a dynamic factor approach," Federal Reserve Board.

Artzner, P F, J Delbaen, J Eber and D Heath (1999): "Coherent measures of risk," *Mathematical Finance,* 203-228.

Bakshi, G and G Panayotov (2010): "First-passage probability, jump models, and intra-horizon risk," *Journal of Financial Economics,* vol 95, pp 20-40.

Balaban, E, J Ouenniche and D Politou (2005), "A note on return distribution of UK stock indices," *Applied Economics Letters,* vol 12, pp 573-576.

Bangia, A, F X Diebold, T Schuermann and J D Stroughair (1999a): "Modeling liquidity risk, with implication for traditional market risk measurement and management," Wharton working paper.

—— (1999b): "Liquidity on the outside," *Risk*, December, pp 68-73.

Barnhill, T and W Maxwell (2002): "Modeling correlated interest rate, exchange rate, and credit risk in fixed income portfolios," *Journal of Banking and Finance,* vol 26, pp 347-374.

Barnhill, T M, P Papapanagiotou and L Schumacher (2000): "Measuring integrated market and credit risks in bank portfolios: an application to a set of hypothetical banks operating in South Africa," IMF Working Paper #2000-212.

Barone-Adesi, G, F Bourgoin and K Giannopoulos (1998): "Don't look back," *Risk*, November, pp 100–103.

Basak, S and A Shapiro (2001): "Value-at-risk-based risk management: optimal policies and asset prices," *The Review of Financial Studies,* pp 371–405.

Basel Committee on Banking Supervision (2009a): *Findings on the interaction of market and credit risk,* Working Paper no 16, Basel.

—— (2009b): *Revisions to the Basel II market risk framework,* http://www.bis.org/publ/bcbs158.pdf, July.

Berkowitz, J (2000a): "A coherent framework for stress-testing," *Journal of Risk,* vol 2, pp 1–11.

—— (2000b): "Incorporating liquidity risk into value-at-risk models," Working Paper, University of Houston, September.

—— (2001): "Testing density forecasts, with applications to risk management," *Journal of Business and Economic Statistics,* vol 19, no 4, pp 465–474.

Berkowitz, J and J O'Brien (2002): "How accurate are value-at-risk models at commercial banks?," *Journal of Finance,* vol 57, pp 1093–1111.

Berkowitz, J, P F Christoffersen and D Pelletier (2010): "Evaluating value-at-risk models with desk-level data," *Management Science.*

Berry, R P (2009): "Back testing value-at-risk," *Investment Analytics and Consulting*, September.

Bervas, A (2006): "Market liquidity and its incorporation into risk management," Banque de France Financial Stability Review.

Botha, A (2008): "Portfolio liquidity-adjusted value at risk," SAJEMS NS, pp 203–216.

Boudoukh, J, M Richardson and R Whitelaw (1998): "The best of both worlds," *Risk*, May, pp 64–67.

Breuer, T, M Jandačka, K Rheinberger and M Summer (2008): "Compounding effects between market and credit risk: the case of variable rate loans," in A Resti (ed), *The Second Pillar in Basel II and the Challenge of Economic Capital,* London: Risk Books.

—— (2009): "How to find plausible, severe, and useful stress scenarios," *International Journal of Central Banking,* September.

—— (2010a): "Does adding up of economic capital for market and credit risk amount to a conservative risk estimate?," *Journal of Banking and Finance,* vol 34, pp 703–712.

Breuer, T, M Jandačka, J Mencia and M Summer (2010b): "A systematic approach to multi-period stress testing of portfolio credit risk," Bank of Spain Working Paper, June.

Brockmann, M and M Kalkbrener (2010): "On the aggregation of risk," *Journal of Risk,* vol 12, no 3.

Campbell, S D (2005): "A review of backtesting and backtesting procedures," FEDS Working Paper Series.

Christoffersen, P (1998): "Evaluating interval forecasts," *International Economic Review*, pp 841–862.

Christoffersen, P and F Diebold (2000): "How relevant is volatility forecasting for financial risk management?," *The Review of Economics and Statistics*, vol 82, no 1, pp 12–22.

Christoffersen, P, F Diebold and T Schuermann (1998): "Horizon problems and extreme events in financial risk management," *FRBNY Economic Policy Review*, October 1998, pp 109–118.

Crouhy, M, D Galai and R Mark (2003): *Risk management*, McGraw-Hill.

Cuenot, S, N Masschelein, M Pritsker, T Schuermann and A Siddique (2006): "Interaction of market and credit risk: framework and literature review," manuscript, Basel Committee Research Task Force Working Group.

Danielsson, J (2002): "The emperor has no clothes: limits to risk modelling," *Journal of Banking and Finance*, vol 26, pp 1273–1296.

Danielsson, J, P Embrechts, C Goodhart, C Keating, F Muennich, O Renault and H Shin (2001): *An academic response to Basel II.*

Danielsson, J, B N Jorgensen, G Samorodnitsky, M Sarma and C G de Vries (2005): "Subadditivity re-examined: the case for value-at-risk," FMG Discussion Papers, London School of Economics.

Danielsson, J and J Zigrand (2006): "On time-scaling of risk and the square-root-of-time rule," *Journal of Banking and Finance,* vol 30, pp 2701–2713.

Danielsson, J, H Shin and J Zigrand (2009): "Risk appetite and endogenous risk," mimeo, http://www.princeton.edu/~hsshin/www/riskappetite.pdf.

Degen, M, P Embrechts and D Lambrigger (2007): "The quantitative modeling of operational risk: between g-and-h and EVT," ASTIN Bulletin, vol 37, no 2, pp 265–291.

Delbaen, F (2002), "Coherent risk measures," lecture notes for University of Pisa lectures, draft.

Diebold, F, A Hickman, A Inoue and T Schuermann (1998), "Scale models," *Risk,* no 11, pp 104–107.

Dimakos, X and K Aas (2004): "Integrated risk modeling," *Statistical Modelling,* vol 4, no 4, pp 265–277.

Drehmann, M, S Sorensen and M Stringa (2010): "The integrated impact of credit and interest rate risk on banks: a dynamic framework and stress testing application," *Journal of Banking and Finance,* vol 34, pp 713–742.

Dunn, G (2009), "Modelling market risk," UK FSA.

Egloff, D, M Leippold and S Jöhri (2005): "Optimal importance sampling for credit portfolios with stochastic approximation," working paper, http://ssrn.com/abstract=1002631.

Engle, R F (2002): "Dynamic conditional correlation—a simple class of multivariate GARCH models," *Journal of Business and Economic Statistics,* pp 339–350.

Engle, R F and R Ferstenberg (2006): "Execution risk," NBER working paper 12165.

Engle, R F and B Kelly (2009): "Dynamic equicorrelation," Stern School of Business.

Engle, R F and K F Kroner (1995): "Multivariate simultaneous generalized ARCH," *Econometric Theory*, pp 122–150.

Engle, R F, N Shephard and K Sheppard (2007): "Fitting and testing vast dimensional time-varying covariance models," New York University.

Finger, C (2009), "IRC comments," *RiskMetrics Group Research Monthly,* February.

Francois-Heude, A and P Van Wynendaele (2001): "Integrating liquidity risk in a parametric intraday VaR framework," Université de Perpignan, France, Facultés Universitaires Catholiques de Mons, Belgium.

Franke, J, W K Härdle and C M Hafner (2008): "Value at risk and backtesting," in *Statistics of Financial Markets*, pp 321–332, Berlin, Heidelberg: Springer.

Geanakoplos, J (2009): "The leverage cycle," Cowles Foundation Discussion Paper no 1715R.

Gennotte, G and H Leland (1990): "Market liquidity, hedging, and crashes," *American Economic Review*, vol 80, no 5, pp 999–1021.

Gourier, E, W Farkas and D Abbate (2009): "Operational risk quantification using extreme value theory and copulas: from theory to practice," *Journal of Operational Risk,* vol 4, no 3, pp 3–26.

Grundke, P (2005): "Risk measurement with integrated market and credit portfolio models," *Journal of Risk*, vol 7, pp 63–94.

Häberle, R and P Persson (2000): "Incorporating market liqudity constraints in VaR," *Bankers Markets & Investors,* vol 44, 01/01/2000.

Hallerbach, W G (2003): "Decomposing portfolio value-at-risk: a general analysis," *Journal of Risk,* vol 5, no 2, pp 1–18.

Hellwig, M (1995), "Systemic aspects of risk management in banking and finance," *Swiss Journal of Economics and Statistics,* vol 131, no 4/2, pp 723–737.

—— (1996), "Capital adequacy rules as instruments for the regulation of banks," *Swiss Journal of Economics and Statistics*, vol 132, no 4/2, pp 609–612.

—— (2009): "Brandbeschleuniger im Finanzsystem," *Max Planck Research,* vol 2, pp 10–15.

Hisata, Y and Y Yamai (2000): "Research toward the practical application of liquidity risk evaluation methods," *Monetary and Economic Studies*, pp 83–128.

Huberman, G and W Stanzl (2005), "Optimal liquidity trading," *Review of Finance*, vol 9, no 2, pp 165–200.

International Accounting Standard 39, *Financial instruments: recognition and measurement,* last version 31 December 2008.

Jarrow, R and S Turnbull (2000): "The intersection of market and credit risk," *Journal of Banking and Finance*, vol 24, no 1–2, pp 271–299.

Jarrow, R and A Subramanian (2001): "The liquidity discount," *Mathematical Finance*, pp 447–474.

Jarrow, R and Ph Protter (2005): "Liquidity risk and risk measure computation," working paper.

Jobst, N J, G Mitra and S A Zenios (2006): "Integrating market and credit risk: a simulation and optimisation perspective," *Journal of Banking and Finance*, vol 30, pp 717–742.

Jobst, N J S A Zenios (2001): "The tail that wags the dog: integrating credit risk in asset portfolios," *Journal of Risk Finance,* vol 3, pp 31–43.

J.P. Morgan (1996): *RiskMetrics Technical Document,* http://www.riskmetrics.com/system/files/private/td4e.pdf.

Kalkbrener, M, H Lotter and L Overbeck (2004): "Sensible and efficient capital allocation for credit portfolios," *Risk*, January, pp 19–24.

Kalkbrener, M, A Kennedy and M Popp (2007): "Efficient calculation of expected shortfall contributions in large credit portfolios," *Journal of Computational Finance*, vol 11, no 2, pp 45–77.

Kaufman, R (2004), *Long-term risk management*, PhD thesis, ETH Zurich.

Kerkhof, J and B Melenberg (2004): "Backtesting for risk-based regulatory capital," *Journal of Banking and Finance,* vol 28, no 8, pp 1845–1865.

Kritzman, M and D Rich (2002): "The mismeasurement of risk," *Financial Analysts Journal*, vol 58, no 3, pp 91–99.

Kupiec, P (1998), "Stress testing in a value-at-risk framework," *Journal of Derivatives*, pp 7–24.

—— (1995), "Techniques for verifying the accuracy of risk measurement models," *Journal of Derivatives*, pp 73–84.

—— (2007), "An integrated structural model for portfolio market and credit risk," manuscript, Federal Deposit Insurance Corporation.

Kuritzkes, A and T Schuermann (2007): "What we know, don't know and can't know about bank risk: a view from the trenches," forthcoming in F X Diebold, N Doherty and R J Herring (eds), *The Known, The Unknown and The Unknowable in Financial Risk Management*, Princeton University Press.

Kuritzkes, A, T Schuermann and S M Weiner (2003): "Risk measurement, risk management and capital adequacy in financial conglomerates," Wharton Financial Institutions Center Working Paper #2003-02.

Lawrence, C and G Robinson (1997): "Liquidity, dynamic hedging and VaR," in: *Risk management for financial institutions,* Risk Publications, London, pp 63–72.

Le Saout, E (2002): "Intégration du risque de liquidité dans les modèles de valeur en risque," *Bankers Markets & Investors,* vol 61, November–December, pp 15–25.

Lopez, J (1999): "Methods for evaluating value-at-risk estimates," *Federal Reserve Bank of San Francisco Review*, 2, pp 3–17.

—— (2005), "Stress tests: useful complements to financial risk models," *FRBSF Economic Letter*, pp 119–124.

McNeil, A, R Frey and P Embrechts (2005): *Quantitative risk management,* Princeton.

Perignon, C and D Smith (2006): "The level and quality of value-at-risk disclosure by commercial banks," Simon Fraser University.

Platen, E and M Schweizer (1998): "On feedback effects from hedging derivatives," *Mathematical Finance,* vol 8, pp 67–84.

Pritsker, M (2006), "The hidden dangers of historical simulation," *Journal of Banking and Finance*, vol 30, no 2, pp 561–582.

Provizionatou, V, S Markose and O Menkens (2005): "Empirical scaling rules for value-at-risk," University of Essex.

Qi, J and W L Ng (2009): "Liquidity adjusted intraday value at risk," *Proceedings of the World Congress of Engineering.*

Rogers, L C G and S Singh (2005): "Option pricing in an illiquid market," Technical Report, University of Cambridge.

Rosenberg, J and T Schuermann (2006): "A general approach to integrated risk management with skewed, fat-tailed risks," *Journal of Financial Economics*, vol 79, pp 569–614.

Saunders, A and I Walter (1994): *Universal banking in the United States: what could we gain? What could we lose?,* Oxford University Press, New York.

Schönbucher, P J and P Wilmott (2000): "The feedback effects of hedging in illiquid markets," *SIAM Journal on Applied Mathematics*, vol 61, pp 232–272.

Shang, D (2009): "ARCH-based value-at-risk with heavy-tailed errors," London School of Economics.

Shin, H S (2008a): "Risk and liquidity," Clarendon Lectures, Oxford University Press, forthcoming.

—— (2008b), "Risk and liquidity in a system context," *Journal of Financial Intermediation*, vol 17, no 3, pp 315–329.

Sircar, K R and G Papanicolaou (1998): "Generalized Black-Scholes models accounting for increased market volatility from hedging strategies," *Applied Mathematical Finance*, vol 5, no 1, pp 45–82.

Smithson, C and L Minton (1996): "Value-at-risk," *Risk*, September, pp 38–39.

Stange, S and C Kaserer (2008): "Why and how to integrate liquidity risk into a VaR framework," CEFS Working Paper.

Stiroh, K J (2004): "Diversification in banking: Is noninterest income the answer?," *Journal of Money, Credit and Banking*, vol 36, no 5, pp 853–82.

Subramanian, A (2008): "Optimal liquidation by a large investor," *SIAM Journal of Applied Mathematics*, vol 68, no 4, pp 1168–1201.

Sy, W (2006): "On the coherence of VaR risk measures for Lévy Stable distributions," Australian Prudential Regulation Authority.

Wang, S (2001): "A risk measure that goes beyond coherence," University of Waterloo, Institute of Insurance and Pension.

Wang, S, V Young and H Panjer (1997): "Axiomatic characterization of insurance prices," *Insurance: Mathematics and Economics*, vol 21, pp 173–183.

Wong, W K (2008): "Backtesting trading risk of commercial banks using expected shortfall," *Journal of Banking and Finance*, vol 32, no 7, pp 1404–1415.

Wu, G and Z Xiao (2002): "An analysis of risk measures," *Journal of Risk*, vol 4, no 4, pp 53–75.

Wu, L (2009): "Incorporating liquidity risk in value-at-risk based on liquidity adjusted returns," Southwestern University of Economics and Finance.

Wylie, J, Q Zhang and T Siu (2010): "Can expected shortfall and value-at-risk be used to statistically hedge options?," *Quantitative Finance*, vol 10, no 6, pp 575–583.

Yamai, Y and Y Yoshiba (2005): "Value-at-risk versus expected shortfall: a practical perspective," *Journal of Banking and Finance*, vol 29, pp 997–1015.

Zheng, H (2006): "Interaction of credit and liquidity risks: modelling and valuation," *Journal of Banking and Finance*, vol 30, pp 391–407.

ANNEX

TABLE 5-1 Summary of "Bottom-Up" Risk Aggregation Papers in the Survey

Research Paper	Portfolio Analysed	Horizon	Risk Measure Used	Ratio of Unified Risk Measure to Sum of Compartmentalised Risk Measures
Breuer, Jandačka, Rheinberger and Summer (2010a)	Hypothetical portfolios of foreign-exchange denominated loans of rating: BBB+ B+	One year	Expected shortfall at: the 1% level the 0.1% level the 1% level the 0.1% level	 1.94 8.22 3.54 7.59
Breuer, Jandačka, Rheinberger and Summer (2008)	Hypothetical portfolios of variable rate loans of rating: BBB+ B+	One year	Expected shortfall at: the 1% level the 0.1% level the 1% level the 0.1% level	 1.11 1.16 1.06 1.10
Grundke (2005)	Hypothetical portfolios of loans with various credit ratings, asset value correlations, distributional assumptions, and correlations between the risk-free rate, credit spreads and firm asset returns	Three years	VaR at: the 1% level the 0.1% level	 0.07–0.97 0.09–1.00
Kupiec (2007)	Hypothetical portfolio of corporate loans with various rating categories calibrated to historical data	Six months	Portfolio losses at funding cost levels consistent with: AAA rating BBB rating	 0.60–3.65 0.61–3.81
Drehmann, Sorensen and Stringa (2010)	Hypothetical UK bank	Three years	Decline in capital over the horizon	2.5
Alessandri and Drehmann (2008)	Hypothetical UK bank	One year	Value-at-risk at: the 1% level the 0.1% level	 0.03 0.50

TABLE 5-2 Summary of "Top-Down" Risk Aggregation Papers in the Survey

Research Paper	Portfolio Analysed	Horizon	Risk Measure Used	Ratio of Unified Risk Measure to Sum of Compartmentalised Risk Measures
Dimakos and Aas (2004)	Norwegian financial conglomerate	—	Total risk exposure at: the 1% level the 0.1% level	 0.90 0.80
Rosenberg and Schuermann (2008)	Hypothetical, internationally-active financial conglomerate	One year	Value-at-risk based on a normal copula at the 0.1% level. (Note: similar results using expected shortfall.)	0.42–0.89 based on different correlation assumptions between market, credit and operational risk.
Kuritzkes, Schuermann and Weiner (2003)	Representative Dutch bank	One year	Economic capital	0.72–0.85 based on different correlation assumptions between market, credit and operational risk.
Kuritzkes and Schuermann (2007)	US banking system from 1986.Q2 through 2005.Q1	—	Tail quantile of the earnings distribution at: the 1% level the 0.1% level	 0.63 0.63

Some Correlation Basics: Properties, Motivation, Terminology

6

Learning Objectives

After completing this reading you should be able to:

- Describe financial correlation risk and the areas in which it appears in finance.
- Explain how correlation contributed to the global financial crisis of 2007 to 2009.
- Describe the structure, uses, and payoffs of a correlation swap.
- Estimate the impact of different correlations between assets in the trading book on the VaR capital charge.
- Explain the role of correlation risk in market risk and credit risk.
- Relate correlation risk to systemic and concentration risk.

Excerpt is Chapter 1 of Correlation Risk Modeling and Management *by Gunter Meissner.*

Behold, the fool saith, "Put not all thine eggs in the one basket."

—Mark Twain

In this chapter we introduce the basic concepts of financial correlations and financial correlation risk. We show that correlations are critical in many areas of finance such as investments, trading, and risk management, as well as in financial crises and in financial regulation. We also show how correlation risk relates to other risks in finance such as market risk, credit risk, systemic risk, and concentration risk.

WHAT ARE FINANCIAL CORRELATIONS?

Heuristically (meaning nonmathematically), we can define two types of financial correlations: static and dynamic.

Static financial correlations measure how two or more financial assets are associated within a certain time period.

Examples are:

- The classic value-at-risk (VaR) model. It answers the question: What is the maximum loss of correlated assets in a portfolio with a certain probability for a given time period? This time period can be 10 days as Base III requires, as well as shorter or longer (see later section for more on VaR and correlation).
- The original copula approach for collateralized debt obligations (CDOs). It measures the default correlations between all assets in the CDO for a certain time period, which is typically identical to the maturity date of the CDO.
- The binomial default correlation model of Lucas (1995), which is a special case of the Pearson correlation model. It measures the probability of two assets defaulting together within a short time period (see Chapter 8 for details).

Besides the static correlation concept, there are dynamic correlations:

Dynamic financial correlations measure how two or more financial assets move together in time.

Examples are:

- In practice, pairs trading, a type of statistical arbitrage, is performed. Let's assume the movements of assets *x* and *y* have been highly correlated in time. If now asset *x* performs poorly with respect to *y*, then asset *x* is bought and asset *y* is sold with the expectation that the gap will narrow.
- Within the deterministic correlation approaches, the Heston 1993 model correlates the Brownian motions dz_1 and dz_2 of assets 1 and 2. The core equation is $dz_1(t) = \rho\, dz_2(t) + \sqrt{(1-\rho^2)}\, dz_3(t)$ where dz_1 and dz_2 are correlated in time with correlation parameter ρ. See Chapter 8 for details.
- Correlations behave randomly and unpredictably. Therefore, it is a good idea to model them as a stochastic process. Stochastic correlation processes are by construction time dependent and can replicate correlation properties well.

Suddenly everything was highly correlated.

—*Financial Times,* April 2009

WHAT IS FINANCIAL CORRELATION RISK?

Financial correlation risk is the risk of financial loss due to adverse movements in correlation between two or more variables.

These variables can comprise any financial variables. For example, the positive correlation between Mexican bonds and Greek bonds can hurt Mexican bond investors if Greek bond prices decrease, which happened in 2012 during the Greek crisis. Or the negative correlation between commodity prices and interest rates can hurt commodity investors if interest rates rise. A further example is the correlation between a bond issuer and a bond insurer, which can hurt the bond investor (see the example displayed in Figure 6-1).

Correlation risk is especially critical in risk management. An increase in the correlation of asset returns increases the risk of financial loss, which is often measured by the value-at-risk (VaR) concept. For details see later section of this chapter. An increase in correlation is typical in a severe systemic crisis. For example, in the Great Recession from 2007 to 2009, financial assets and financial markets worldwide became highly correlated. Risk managers who had in their portfolios assets with negative or low correlations suddenly witnessed many of them decline together; hence asset correlations increased sharply. For more on systemic risk, see the section, "The Global Financial Crisis of 2007 to 2009 and Correlation," as well as Chapter 7, which displays empirical findings of correlations.

Correlation risk can also involve variables that are non-financial, such as economic or political events. For example, the correlation between the increasing sovereign debt and currency value can hurt an exporter, as occurred in Europe in 2012, where a decreasing euro hurt U.S. exporters. Geopolitical tensions as, for example, in the Middle East can hurt airline companies due to increasing oil prices, or a slowing gross domestic product (GDP) in the United States can hurt Asian and European exporters and investors, since economies and financial markets are correlated worldwide.

Let's look at correlation risk via an example of a credit default swap (CDS). A CDS is a financial product in which the credit risk is transferred from the investor (or CDS buyer) to a counterparty (CDS seller). Let's assume an investor has invested \$1 million in a bond from Spain. He is now worried about Spain defaulting and has purchased a credit default swap from a French bank, BNP Paribas. Graphically this is displayed in Figure 6-1.

The investor is protected against a default from Spain, since in case of default, the counterparty BNP Paribas will pay the originally invested \$1 million to the investor. For simplicity, let's assume the recovery rate and accrued interest are zero.

The value of the CDS, i.e., the fixed CDS spread S, is mainly determined by the default probability of the reference entity Spain. However, the spread S is also determined by the joint default correlation of BNP Paribas and Spain. If the correlation between Spain and BNP Paribas increases, the present value of the CDS for the investor will decrease and he will suffer a paper loss. The worst-case scenario is the joint default of Spain and BNP Paribas, in which case the investor will lose his entire investment in the Spanish bond of \$1 million.

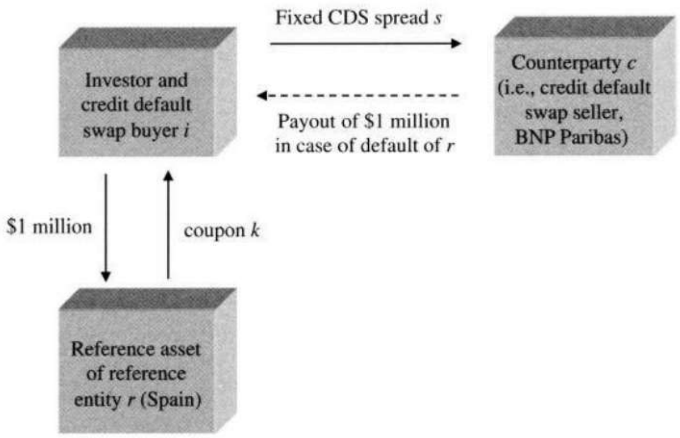

FIGURE 6-1 An investor hedging his Spanish bond exposure with a CDS.

In other words, the investor is exposed to default correlation risk between the reference asset r (Spain) and the counterparty c (BNP Paribas). Since both Spain and BNP Paribas are in Europe, let's assume that there is a positive default correlation between the two. In this case, the investor has wrong-way correlation risk or short wrong-way risk (WWR). Let's assume the default probability of Spain and BNP Paribas both increase. This means that the exposure to the reference entity Spain increases (since the CDS has a higher present value for the investor) *and* it is more unlikely that the counterparty BNP Paribas can pay the default insurance.

The magnitude of the correlation risk is expressed graphically in Figure 6-2.

From Figure 6-2 we observe that for a correlation of −0.3 and higher, the higher the correlation, the lower the CDS spread. This is because an increasing ρ means a higher probability of the reference asset and the counterparty defaulting together. In the extreme case of a perfect correlation of 1, the CDS is worthless. This is because if Spain defaults, so will the insurance seller BNP Paribas.

We also observe from Figure 6-2 that for a correlation from −1 to about −0.3, the CDS spread increases slightly. This seems counterintuitive at first. However, an increase in the negative correlation means a higher probability of either Spain *or* BNP Paribas defaulting. In the case of Spain defaulting, the CDS buyer will get compensated by BNP Paribas. However, if the insurance seller BNP Paribas defaults, the CDS buyer will lose his insurance and will have to repurchase it. This may have to be done at a higher cost. The cost will be higher if the credit quality of Spain has decreased since inception of the original CDS. For example, the CDS spread may have been 3% in the original CDS, but may have increased to 6% due to a credit deterioration of Spain. For more details on pricing CDSs with counterparty risk and the reference asset-counterparty correlation, see Kettunen and Meissner (2006).

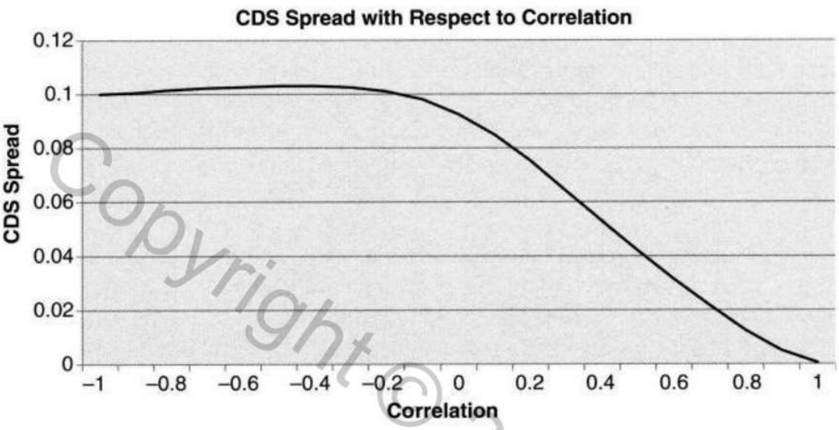

FIGURE 6-2 CDS spread *s* of a hedged bond purchase (as displayed in Figure 6-1) with respect to the default correlation between the reference entity *r* and the counterparty *c*.

We observe from Figure 6-2 that the dependencies between a variable (here the CDS spread) and correlation may be nonmonotonous; that is, the CDS spread sometimes increases and sometimes decreases if correlation increases.

MOTIVATION: CORRELATIONS AND CORRELATION RISK ARE EVERYWHERE IN FINANCE

Why study financial correlations? That's an easy one. Financial correlations appear in many areas in finance. We will briefly discuss five areas: (1) investments and correlation, (2) trading and correlation, (3) risk management and correlation, (4) the global financial crisis and correlation, and (5) regulation and correlation. Naturally, if an entity is exposed to correlation, this means that the entity has correlation risk (i.e., the risk of a change in the correlation).

Investments and Correlation

From our studies of the Nobel Prize–winning capital asset pricing model (CAPM) (Markowitz 1952; Sharpe 1964) we remember that an increase in diversification increases the return/risk ratio. Importantly, high diversification is related to low correlation. Let's show this in an example. Let's assume we have a portfolio of two assets, *X* and *Y*. They have performed as in Table 6-1.

Let's define the return of asset *X* at time *t* as x_t, and the return of asset *Y* at time *t* as y_t. A return is calculated as a percentage change, $(S_t - S_{t-1})/S_{t-1}$, where *S* is a price or a rate. The average return of asset *X* for the time frame 2009 to 2013 is μ_X = 29.03%; for asset *Y* the average return is μ_X = 20.07%. If we assign a weight to asset *X*, w_X, and a weight to asset *Y*, w_Y, the portfolio return is

$$\mu_P = w_X \mu_X + w_Y \mu_Y \quad \textbf{(6.1)}$$

where $w_X + w_Y = 1$.

The standard deviation of returns, called *volatility*, is derived for asset *X* with Equation (6.2):

$$\sigma_X = \sqrt{\frac{1}{n-1}\sum_{t=1}^{n}(x_t - \mu_X)^2} \quad \textbf{(6.2)}$$

where x_t is the return of asset *X* at time *t* and *n* is the number of observed points in time.

The volatility of asset *Y* is derived accordingly. Equation (6.2) can be computed with =stdev in Excel and std in MATLAB. From our example in Table 6-1, we find that σ_X = 44.51% and σ_Y = 47.58%.

Let's now look at the covariance. The covariance measures how two variables covary (i.e., move together). More precisely, the covariance measures the strength of the linear relationship between two variables. The covariance of returns for assets *X* and *Y* is derived with Equation (6.3):

$$\mathrm{Cov}_{XY} = \frac{1}{n-1}\sum_{t=1}^{n}(x_t - \mu_X)(y_t - \mu_Y) \quad \textbf{(6.3)}$$

TABLE 6-1 Performance of a Portfolio with Two Assets

Year	Asset X	Asset Y	Return of Asset X	Return of Asset Y
2008	100	200		
2009	120	230	20.00%	15.00%
2010	108	460	−10.00%	100.00%
2011	190	410	75.93%	−10.87%
2012	160	480	−15.79%	17.07%
2013	280	380	75.00%	−20.83%
		Average	29.03%	20.07%