VALUATION AND RISK MODELS

GARP
DIGITAL LIBRARY

"Quantifying Volatility in VaR Models"

Chapter 2 of

Understanding Market, Credit, and Operational Risk: The Value at Risk Approach

By Linda Allen, Jacob Boudoukh, and Anthony Saunders

CHAPTER TWO

QUANTIFYING VOLATILITY IN VaR MODELS

CHAPTER OUTLINE

2.1 THE STOCHASTIC BEHAVIOR OF RETURNS

2.1.1 Revisiting the assumptions

In the previous chapter we discussed the notion of VaR. Measuring VaR involves identifying the tail of the distribution of asset returns. One approach to the problem is to impose specific distributional assumptions on asset returns. This approach is commonly termed the *parametric approach*, requiring a specific set of distributional assumptions. As we saw in the previous chapter, if we are willing to make a specific parametric distributional assumption, for example, that asset returns are normally distributed, then all we need is to provide two parameters – the mean (denoted μ) and the standard deviation (denoted σ) of returns. Given those, we are able to fully characterize the distribution and comment on risk in any way required; in particular, quantifying VaR, percentiles (e.g., 50 percent, 98 percent, 99 percent, etc.) of a loss distribution.

The problem is that, in reality, asset returns tend to deviate from normality. While many other phenomena in nature are often well described by the Gaussian (normal) distribution, asset returns tend to deviate from normality in meaningful ways. As we shall see below in detail, asset returns tend to be:

- *Fat-tailed*: A fat-tailed distribution is characterized by having more probability weight (observations) in its tails relative to the normal distribution.
- *Skewed*: A skewed distribution in our case refers to the empirical fact that declines in asset prices are more severe than increases. This is in contrast to the symmetry that is built into the normal distribution.
- *Unstable*: Unstable parameter values are the result of varying market conditions, and their effect, for example, on volatility.

All of the above require a risk manager to be able to reassess distributional parameters that vary through time.

In what follows we elaborate and establish benchmarks for these effects, and then proceed to address the key issue of how to adjust our set of assumptions to be able to better model asset returns, and better predict extreme market events. To do this we use a specific dataset, allowing us to demonstrate the key points through the use of an example.

2.1.2 The distribution of interest rate changes

Consider a series of daily observations of interest rates. In the series described below we plot three-month US Treasury bill (T-bill) rates calculated by the Federal Reserve.[1] We use ten years of data and hence we have approximately 2,500 observations. For convenience let us assume we have 2,501 data points on interest rate levels, and hence 2,500 data points on daily interest rate changes. Figure 2.1 depicts the time series of the yield to maturity, fluctuating between 11 percent p.a. and 4 percent p.a. during the sample period (in this example, 1983–92).

The return on bonds is determined by interest rate *changes*, and hence this is the relevant variable for our discussion. We calculate daily interest changes, that is, the first difference series of observed yields. Figure 2.2 is a histogram of yield changes. The histogram is the result of 2,500 observations of daily interest rate changes from the above data set.

Using this series of 2,500 interest rate changes we can obtain the average interest rate change and the standard deviation of interest rate

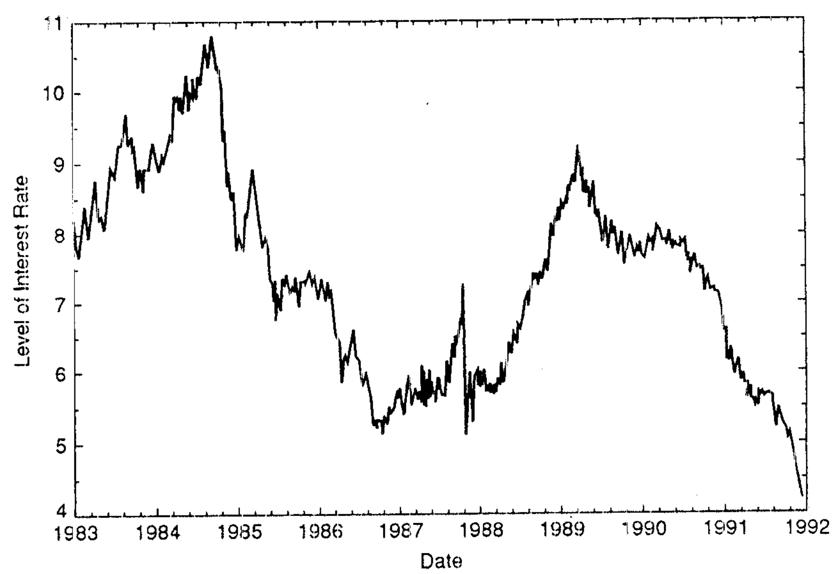

Figure 2.1 Three-month Treasury rates

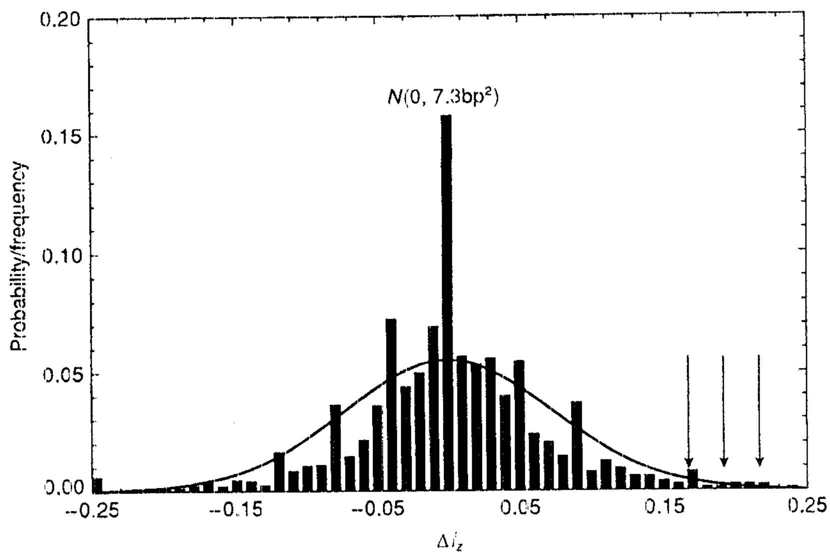

Figure 2.2 Three-month Treasury rate changes

changes over the period. The mean of the series is zero basis points per day. Note that the average daily change in this case is simply the last yield minus the first yield in the series, divided by the number of days in the series. The series in our case starts at 4 percent and ends at a level of 8 percent, hence we have a 400 basis point (bp) change over the course of 2,500 days, for an average change of approximately zero. Zero expected change as a forecast is, as we discussed in the previous chapter, consistent with the random walk assumption as well. The standard deviation of interest rate changes turns out to be 7.3bp/day.

Using these two parameters, figure 2.2 plots a normal distribution curve on the same scale of the histogram, with basis point changes on the X-axis and probability on the Y-axis. If our assumption of normality is correct, then the plot in figure 2.2 should resemble the theoretical normal distribution shown in figure 1.1. Observing figure 2.2 we find some important differences between the theoretical normal distribution using the mean and standard deviation from our data, and the empirical histogram plotted by actual interest rate changes. The difference is primarily the result of the "fat-tailed" nature of the distribution.

2.1.3 Fat tails

The term "fat tails" refers to the tails of one distribution *relative* to another reference distribution. The reference distribution here is the normal distribution. A distribution is said to have "fatter tails" than the normal distribution if it has a similar mean and variance, but different probability mass at the extreme tails of the probability distribution. The critical point is that the first two moments of the distribution, the mean and the variance, are the same.

This is precisely the case for the data in figure 2.2, where we observe the empirical distribution of interest rate changes. The plot includes a histogram of interest rate changes in different probability buckets. In addition to the histogram, and on the same plot, a normal distribution is also plotted, so as to compare the two distributions. The normal distribution has the same mean (zero) and the same volatility (7.3 basis points) as the empirical distribution.

We can observe "fat tail" effects by comparing the two distributions. There is extra probability mass in the empirical distribution relative to the normal distribution benchmark around zero, and there is a "missing" probability mass in the intermediate portions around the plus ten and minus ten basis point change region of the histogram. Although it is difficult to observe directly in figure 2.2, it is also the case that at the probability extremes (e.g., around 25bp and higher), there are more observations than the theoretical normal benchmark warrants. A more detailed figure focusing on the tails is presented later in this chapter.

This pattern, more probability mass around the mean and at the tails, and less around plus/minus one standard deviation, is precisely what we expect of a fat tailed distribution. Intuitively, a probability mass is taken from around the one standard deviation region, and distributed to the zero interest rate change and to the two extreme-change regions. This is done in such way so as to preserve the mean and standard deviation. In our case the mean of zero and the standard deviation of 7.3bp, are preserved by construction, because we plot the normal distribution benchmark given these two empirically determined parameters.

To illustrate the impact of fat tails, consider the following exercise. We take the vector of 2,500 observations of interest rate changes, and order this vector not by date but, instead, by the size of the interest rate change, in descending order. This ordered vector will have the larger interest rate increases at the top. The largest change may be, for example, an increase of 35 basis points. It will appear as entry number one of the ordered vector. The following entry will be the second largest change, say 33 basis points, and so on. Zero changes should be found around the middle of this vector, in the vicinity of the 1,250th entry, and large declines should appear towards the "bottom" of this vector, in entries 2,400 to 2,500.

If it were the case that, indeed, the distribution of interest rate changes were normal with a mean of zero and a standard deviation of 7.3 basis points, what would we expect of this vector, and, in particular, of the tails of the distribution of interest rate changes? In particular, what should be a one percentile (%) interest rate shock; i.e., an interest rate shock that occurs approximately once in every 100 days? For the standard normal distribution we know that the first percentile is delineated at 2.33 standard deviations from the mean. In our case, though, *losses* in asset values are related to *increases* in interest rates. Hence we examine the +2.33 standard deviation rather than the −2.33 standard deviation event (i.e., 2.33 standard deviations above the mean rather than 2.33 standard deviations below the mean). The +2.33 standard deviations event for the standard normal translates into an increase in interest rates of $\sigma \times 2.33$ or $7.3\text{bp} \times 2.33 = 17\text{bp}$. Under the assumption that interest rate changes are normal we should, therefore, see in 1 percent of the cases interest rate changes that are greater or equal to 17 basis points.

What do we get in reality? The empirical first percentile of the distribution of interest rate changes can be found as the 25th out of the 2,500 observations in the ordered vector of interest rate changes. Examining this entry in the vector we find an interest rate increase of 21 basis points. Thus, the empirical first percentile (21bp) does not conform to the theoretical 17 basis points implied by the normality assumption, providing a direct and intuitive example of the fat tailedness of the empirical distribution. That is, we find that the (empirical) tails of the actual distribution are fatter than the theoretical tails of the distribution.[2]

2.1.4 Explaining fat tails

The phenomenon of fat tails poses a severe problem for risk managers. Risk measurement, as we saw above, is focused on extreme events, trying to quantify the probability and magnitude of severe losses. The normal distribution, a common benchmark in many cases, seems to fail here. Moreover, it seems to fail precisely where we need it to work best – in the tails of the distributions. Since risk management is all about the tails, further investigation of the tail behavior of asset returns is required.

In order to address this issue, recall that the distribution we examine is the *unconditional distribution* of asset returns. By "unconditional" we mean that on any given day we assume the same distribution exists, regardless of market and economic conditions. This is in spite of the fact that there is information available to market participants about the distribution of asset returns at any given point in time which may be different than on other days. This information is relevant for an asset's *conditional distribution*, as measured by parameters, such as the conditional mean, conditional standard deviation (volatility), conditional skew and kurtosis. This implies two possible explanations for the fat tails: (i) conditional volatility is time-varying; and (ii) the conditional mean is time-varying. Time variations in either could, arguably, generate fat tails in the unconditional distribution, in spite of the fact that the conditional distribution is normal (albeit with different parameters at different points in time, e.g., in recessions and expansions).

Let us consider each of these possible explanations for fat tails. First, is it plausible that the fat tails observed in the unconditional distribution are due to time-varying conditional distributions? We will show that the answer is generally "no." The explanation is based on the implausible assumption that market participants know, or can predict in advance, future changes in asset prices. Suppose, for example, that interest rate changes are, in fact, normal, with a time-varying conditional mean. Assume further that the conditional mean of interest rate changes is known to market participants during the period under investigation, but is unknown to the econometrician. For simplicity, assume that the *conditional* mean can be +5bp/day on some days, and –5bp/day on other days. If the split between high mean and low mean days were 50–50, we would observe an unconditional mean change in interest rates of 0bp/day.

In this case when the econometrician or the risk manager approaches past data without the knowledge of the conditional means, he mistakes variations in interest rates to be due to volatility. Risk is overstated, and changes that are, in truth, distributed normally and are centered around plus or minus five basis points, are mistaken to be normal with a mean of zero. If this were the case we would have obtained a "mixture of normals" with varying means, that would appear to be, unconditionally, fat tailed.

Is this a likely explanation for the observed fat tails in the data? The answer is negative. The belief in efficient markets implies that asset prices reflect all commonly available information. If participants in the marketplace know that prices are due to rise over the next day, prices would have already risen today as traders would have traded on this information. Even detractors of market efficiency assumptions would agree that conditional means do not vary enough on a daily basis to make those variations a first order effect.

To verify this point consider the debate over the predictability of market returns. Recent evidence argues that the conditional risk premium, the expected return on the market over and above the risk free rate, varies through time in a predictable manner. Even if we assume this to be the case, predicted variations are commonly estimated to be between zero and 10 percent on an annualized basis. Moreover, variations in the expected premium are slow to change (the predictive variables that drive these variations vary slowly). If at a given point you believe the expected excess return on the market is 10 percent per annum rather than the unconditional value of, say, 5 percent, you predict, on a daily basis, a return which is 2bp different from the market's average premium (a 5 percent per annum difference equals approximately a return of 2bp/day). With the observed volatility of equity returns being around 100bp/day, we may view variations in the conditional mean as a second order effect.

The second possible explanation for the fat tail phenomenon is that volatility (standard deviation) is time-varying. Intuitively, one can make a compelling case against the assumption that asset return volatility is constant. For example, the days prior to important Federal announcements are commonly thought of as days with higher than usual uncertainty, during which interest rate volatility as well as equity return volatility surge. Important political events, such as the turmoil in the Gulf region, and significant economic events, such as the defaults of Russia and Argentina on their debts, are also associated with a spike in global volatility. Time-varying volatility may also be generated by regular, predictable events. For example, volatility in the Federal funds market increases dramatically on the last days of the reserve maintenance period for banks as well as at quarter-end in response to balance sheet window dressing.[5] Stochastic volatility is clearly a candidate explanation for fat tails, especially if the econometrician fails to use relevant information that generates excess volatility.[4]

2.1.5 Effects of volatility changes

How does time-varying volatility affect our distributional assumptions, the validity of the normal distribution model and our ability to provide a useful risk measurement system? To illustrate the problem and its potential solution, consider an illustrative example. Suppose interest rate changes do not fit the normal distribution model with a mean of zero and a standard deviation of 7.3 basis points per day. Instead, the true conditional distribution of interest rate changes is normal with a mean of zero but with a time-varying volatility that during some periods is 5bp/day and during other periods is 15bp/day.

This type of distribution is often called a "regime-switching volatility model." The regime switches from low volatility to high volatility, but is never in between. Assume further that market participants are aware of the state of the economy, i.e., whether volatility is high or low. The econometrician, on the other hand, does not have this knowledge. When he examines the data, oblivious to the true regime-switching distribution, he estimates an unconditional volatility of 7.3bp/day that is the result of the mixture of the high volatility and low volatility regimes. Fat tails appear only in the unconditional distribution. The conditional distribution is always normal, albeit with a varying volatility.[5]

Figure 2.3 provides a schematic of the path of interest rate volatility in our regime-switching example. The solid line depicts the true volatility, switching between 5bp/day and 15bp/day. The econometrician observes periods where interest rates change by as much as, say, 30 basis points. A change in interest rates of 30bp corresponds to a change of more than four standard deviations given that the estimated standard deviation is 7.3bp. According to the normal

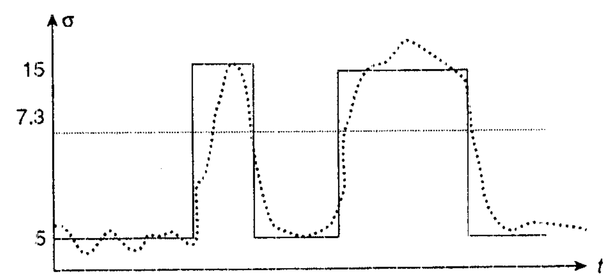

Figure 2.3 A schematic of actual and estimated volatility

Table 2.1 Tail event probability and odds under normality

No. of deviations Z	Prob($X < z$) (in %)	Odds (one in . . . days)
−1.50	6.68072	15
−2.00	2.27501	44
−2.50	0.62097	161
−3.00	0.13500	741
−3.50	0.02327	4,298
−4.00	0.00317	31,560
−4.50	0.00034	294,048
−5.00	0.00003	3,483,046

distribution benchmark, a change of four standard deviations or more should be observed very infrequently. More precisely, the probability that a truly random normal variable will deviate from the mean by four standard deviations or more is 0.003 percent. Putting it differently, the odds of seeing such a change are one in 31,560, or once in 121 years. Table 2.1 provides the number of standard deviations, the probability of seeing a random normal being less than or equal to this number of standard deviations, in percentage terms, and the odds of seeing such an event.

The risk manager may be puzzled by the empirical observation of a relatively high frequency of four or more standard deviation moves. His risk model, one could argue, based on an unconditional normal distribution with a standard deviation of 7.3bp, is of little use, since it under-predicts the odds of a 30bp move. In reality (in the reality of our illustrative example), the change of 30bp occurred, most likely, on a high volatility day. On a high volatility day a 30bp move is only a two standard deviation move, since interest rate changes are drawn from a normal distribution with a standard deviation of 15bp/day. The probability of a change in interest rates of two standard deviations or more, equivalent to a change of 30bp or more on high volatility days, is still low, but is economically meaningful. In particular, the probability of a 30bp move conditional on a high volatility day is 2.27 percent, and the odds are one in 44.

The dotted line in figure 2.3 depicts the estimated volatility using a volatility estimation model based on historical data. This is the typical picture for common risk measurement engines – the estimated volatility trails true volatility. Estimated volatility rises after having observed an increase, and declines having observed a decrease. The estimation error and estimation lag is a central issue in risk measurement, as we shall see in this chapter.

This last example illustrates the challenge of modern dynamic risk measurement. The most important task of the risk manager is to raise a "red flag," a warning signal that volatility is expected to be high in the near future. The resulting action given this information may vary from one firm to another, as a function of strategy, culture, appetite for risk, and so on, and could be a matter of great debate. The importance of the risk estimate as an input to the decision making process is, however, not a matter of any debate. The effort to improve risk measurement engines' dynamic prediction of risk based on market conditions is our focus throughout the rest of the chapter.

This last illustrative example is an extreme case of stochastic volatility, where volatility jumps from high to low and back periodically. This model is in fact quite popular in the macroeconomics literature, and more recently in finance as well. It is commonly known as regime switching.[6]

2.1.6 Can (conditional) normality be salvaged?

In the last example, we shifted our concept of normality. Instead of assuming asset returns are normally distributed, we now assume that asset returns are *conditionally normally distributed*. Conditional normality, with a time-varying volatility, is an economically reasonable description of the nature of asset return distributions, and may resolve the issue of fat tails observed in unconditional distributions.

This is the focus of the remainder of this chapter. To preview the discussion that follows, however, it is worthwhile to forewarn the reader that the effort is going to be, to an extent, incomplete. Asset returns are generally non-normal, both unconditionally as well as conditionally; i.e., fat tails are exhibited in asset returns regardless of the estimation method we apply. While the use of dynamic risk measurement models capable of adapting model parameters as a function of changing market conditions is important, these models do not eliminate all deviations from the normal distribution benchmark. Asset returns keep exhibiting asymmetries and unexpectedly large movements regardless of the sophistication of estimation models. Putting it more simply – large moves will always occur "out of the blue" (e.g., in relatively low volatility periods).

One way to examine conditional fat tails is by normalizing asset returns. The process of normalizations of a random normal variable is simple. Consider X, a random normal variable, with a mean of μ and a standard deviation σ,

$$X \sim N(\mu, \sigma^2).$$

A standardized version of X is

$$(X - \mu)/\sigma \sim N(0, 1).$$

That is, given the mean and the standard deviation, the random variable X less its mean, divided by its standard deviation, is distributed according to the standard normal distribution.

Consider now a series of interest rate changes, where the mean is assumed, for simplicity, to be always zero, and the volatility is re-estimated every period. Denote this volatility estimate by σ_t. This is the forecast for next period's volatility based on some volatility estimation model (see the detailed discussion in the next section). Under the normality assumption, interest rate changes are now conditionally normal

$$\Delta i_{t,t+1} \sim N(0, \sigma_t^2).$$

We can standardize the distribution of interest rate changes dynamically using our estimated conditional volatility σ_t, and the actual change in interest rate that followed $\Delta i_{t,t-1}$. We create a series of standardized variables.

$$\Delta i_{t,t+1}/\sigma_t \sim N(0, 1).$$

This series should be distributed according to the standard normal distribution. To check this, we can go back through the data, and with the benefit of hindsight put all pieces of data, drawn under the null assumption of conditional normality from a normal distribution with time-varying volatilities, on equal footing. If interest rate changes are, indeed, conditionally normal with a time-varying volatility, then the unconditional distribution of interest rate changes can be fat tailed. However, the distribution of interest rate changes standardized by their respective conditional volatilities should be distributed as a standard normal variable.

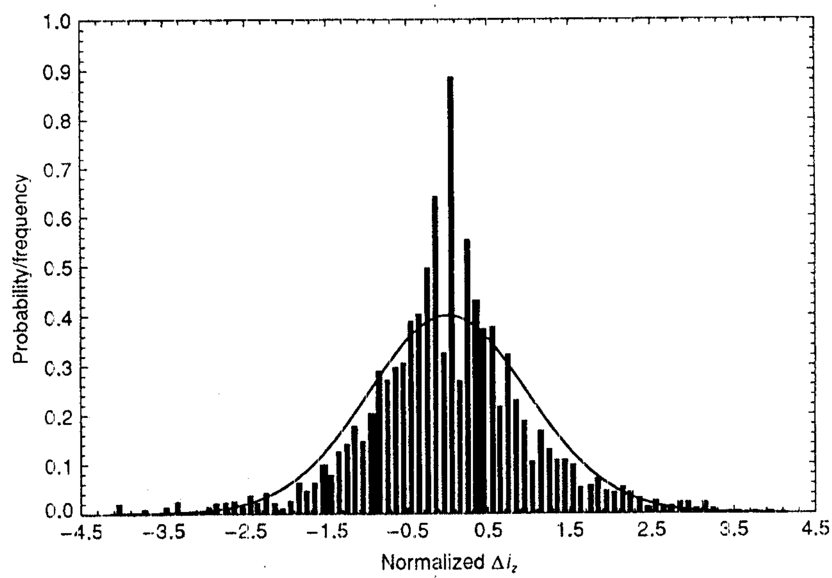

Figure 2.4 Standardized interest rate changes – empirical distribution relative to the $N(0, 1)$ benchmark

Figure 2.4 does precisely this. Using historical data we estimate conditional volatility.[7] We plot a histogram similar to the one in figure 2.2, with one exception. The X-axis here is not in terms of interest rate changes, but, instead, in terms of *standardized interest rate changes*. All periods are now adjusted to be comparable, and we may expect to see a "well-behaved" standard normal. Standardized interest rate changes are going to be well behaved on two conditions: (i) that interest rate changes are, indeed, conditionally normal; and (ii) that we accurately estimated conditional volatility, i.e., that we were able to devise a "good" dynamic volatility estimation mechanism. This joint condition can be formalized into a statistical hypothesis that can be tested.

Normalized interest rate changes, plotted in figure 2.4, provide an informal test. First note that we are not interested in testing for normality *per se*, since we are not interested in the entire distribution. We only care about our ability to capture tail behavior in asset returns – the key to dynamic risk measurement. Casual examination of figure 2.5, where the picture focuses on the tails of the conditional distribution, vividly shows the failure of the conditional normality model to describe the data. Extreme movements of *standardized* interest rate

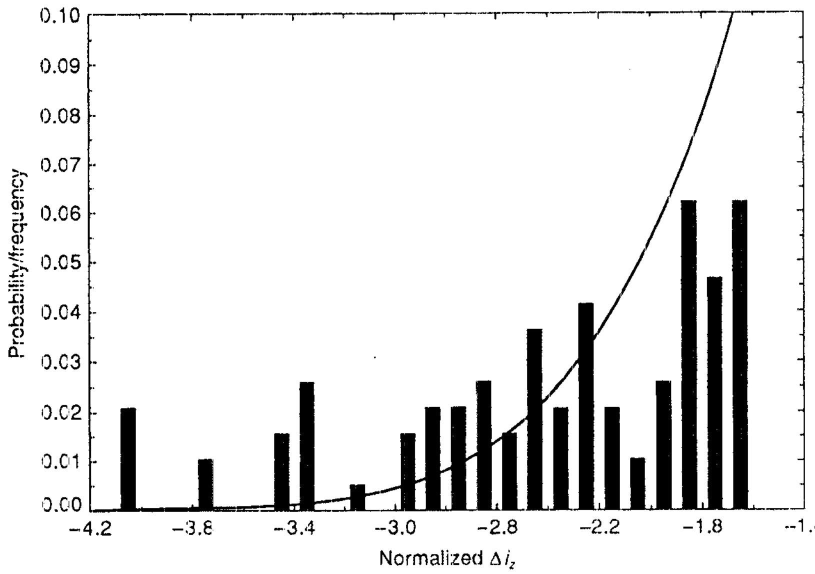

Figure 2.5 Tail standardized interest rate changes

movements – deviating from the conditional normality model – are still present in the data. Recall, though, that this is a failure of the joint model – conditional normality and the method for dynamic estimation of the conditional volatility.[8] In principle it is still possible that an alternative model of volatility dynamics will be able to capture the conditional distribution of asset returns better and that the conditional returns based on the alternative model will indeed be normal.

2.1.7 Normality cannot be salvaged

The result apparent in figure 2.5 holds true, however, to a varying degree, for most financial data series. Sharp movements in asset returns, even on a normalized basis, occur in financial data series no matter how we manipulate the data to estimate volatility. Conditional asset returns exhibit sharp movements, asymmetries, and other difficult-to-model effects in the distribution. This is, in a nutshell, the problem with all extant risk measurement engines. All VaR-based systems tend to encounter difficulty where we need them to perform best – at the tails. Similar effects are also present for the multivariate distribution of portfolios of assets – correlations as well tend to be unstable – hence making VaR engines often too conservative at the worst possible times.

This is a striking result with critical implications for the practice of risk management. The relative prevalence of extreme moves, even after adjusting for current market conditions, is the reason we need additional tools, over and above the standard VaR risk measurement tool. Specifically, the need for *stress testing* and *scenario analysis* is related directly to the failure of VaR-based systems.

Nevertheless, the study of conditional distributions is important. There is still important information in current market conditions, e.g., conditional volatility, that can be exploited in the process of risk assessment. In this chapter we elaborate on risk measurement and VaR methods. In the next chapter we augment our set of tools discussing stress testing and scenario analysis.

2.2 VaR ESTIMATION APPROACHES

There are numerous ways to approach the modeling of asset return distribution in general, and of tail behavior (e.g., risk measurement) in particular. The approaches to estimating VaR can be broadly divided as follows

- *Historical-based approaches.* The common attribute to all the approaches within this class is their use of historical time series data in order to determine the shape of the conditional distribution.
  - *Parametric approach.* The parametric approach imposes a specific distributional assumption on conditional asset returns. A representative member of this class of models is the conditional (log) normal case with time-varying volatility, where volatility is estimated from recent past data.
  - *Nonparametric approach.* This approach uses historical data directly, without imposing a specific set of distributional assumptions. Historical simulation is the simplest and most prominent representative of this class of models.
  - *Hybrid approach.* A combined approach.
- *Implied volatility based approach.* This approach uses derivative pricing models and current derivative prices in order to impute an implied volatility without having to resort to historical data. The

use of implied volatility obtained from the Black–Scholes option pricing model as a predictor of future volatility is the most prominent representative of this class of models.

2.2.1 Cyclical volatility

Volatility in financial markets is not only time-varying, but also sticky, or predictable. As far back as 1963, Mandelbrot wrote

> large changes tend to be followed by large changes – of either sign – and small changes by small changes. (Mandelbrot 1963)

This is a very useful guide to modeling asset return volatility, and hence risk. It turns out to be a salient feature of most extant models that use historical data. The implication is simple – since the magnitude (but not the sign) of recent changes is informative, the most recent history of returns on a financial asset should be most informative with respect to its volatility in the near future. This intuition is implemented in many simple models by placing more weight on recent historical data, and little or no weight on data that is in the more distant past.

2.2.2 Historical standard deviation

Historical standard deviation is the simplest and most common way to estimate or predict future volatility. Given a history of an asset's continuously compounded rate of returns we take a specific window of the K most recent returns. The data in hand are, hence, limited by choice to be $r_{t-1,t}, r_{t-2,t-1}, \ldots, r_{t-K,t-K+1}$. This return series is used in order to calculate the current/conditional standard deviation σ_t, defined as the square root of the conditional variance

$$\sigma_t^2 = ({r_{t-K,t-K+1}}^2 + \ldots + {r_{t-2,t-1}}^2 + {r_{t-1,t}}^2)/K.$$

This is the most familiar formula for calculating the variance of a random variable – simply calculating its "mean squared deviation." Note that we make an explicit assumption here, that the conditional mean is zero. This is consistent with the random walk assumption.

The standard formula for standard deviation[9] uses a slightly different formula, first demeaning the range of data given to it for calculation. The estimation is, hence, instead

$$\mu_t = (r_{t-K,t-K+1} + \ldots + r_{t-2,t-1} + r_{t-1,t})/K,$$

$$\sigma_t^2 = ((r_{t-K,t-K+1} - \mu_t)^2 + \ldots + (r_{t-2,t-1} - \mu_t)^2 + (r_{t-1,t} - \mu_t)^2)/(K - 1).$$

Note here that the standard deviation is the mean of the squared deviation, but the mean is taken by dividing by $(K - 1)$ rather than K. This is a result of a statistical consideration related to the loss of one degree of freedom because the conditional mean, μ_t, has been estimated in a prior stage. The use of $K - 1$ in the denominator guarantees that the estimator σ_t^2 is unbiased.

This is a minor variation that makes very little practical difference in most instances. However, it is worthwhile discussing the pros and cons of each of these two methods. Estimating the conditional mean μ_t from the most recent K days of data is risky. Suppose, for example, that we need to estimate the volatility of the stock market, and we decide to use a window of the most recent 100 trading days. Suppose further that over the past 100 days the market has declined by 25 percent. This can be represented as an average decline of 25bp/day (–2,500bp/100days = –25bp/day). Recall that the econometrician is trying to estimate the conditional mean and volatility that were known to market participants during the period. Using –25bp/day as μ_t, the conditional mean, and then estimating σ_t^2, implicitly assumes that market participants knew of the decline, and that their conditional distribution was centered around minus 25bp/day.

Since we believe that the decline was entirely unpredictable, imposing our priors by using $\mu_t = 0$ is a logical alternative. Another approach is to use the unconditional mean, or an expected change based on some other theory as the conditional mean parameter. In the case of equities, for instance, we may want to use the unconditional average return on equities using a longer period – for example 12 percent per annum, which is the sum of the average risk free rate (approximately 6 percent) plus the average equity risk premium (6 percent). This translates into an average daily increase in equity prices of approximately 4.5bp/day. This is a relatively small number that tends to make little difference in application, but has a sound economic rationale underlying its use.

For other assets we may want to use the forward rate as the estimate for the expected average change. Currencies, for instance, are expected to drift to equal their forward rate according to the expectations hypothesis. If the USD is traded at a forward premium of 2.5 percent p.a. relative to the Euro, a reasonable candidate for the mean parameter would be μ_t = 1bp/day. The difference here between 0bp and 1bp seems to be immaterial, but when VaR is estimated for longer horizons this will become a relevant consideration, as we discuss later.

2.2.3 Implementation considerations

The empirical performance of historical standard deviation as a predictor of future volatility is affected by statistical error. With respect to statistical error, it is always the case in statistics that "more is better". Hence, the more data available to us, the more precise our estimator will be to the true return volatility. On the other hand, we estimate standard deviation in an environment where we believe, a priori, that volatility itself is unstable. The stickiness of time variations in volatility are important, since it gives us an intuitive guide that recent history is more relevant for the near future than distant history.

In figure 2.6 we use the series of 2,500 interest rate changes in order to come up with a series of rolling estimates of conditional volatility. We use an estimation window K of different lengths in order to demonstrate the tradeoff involved. Specifically, three different

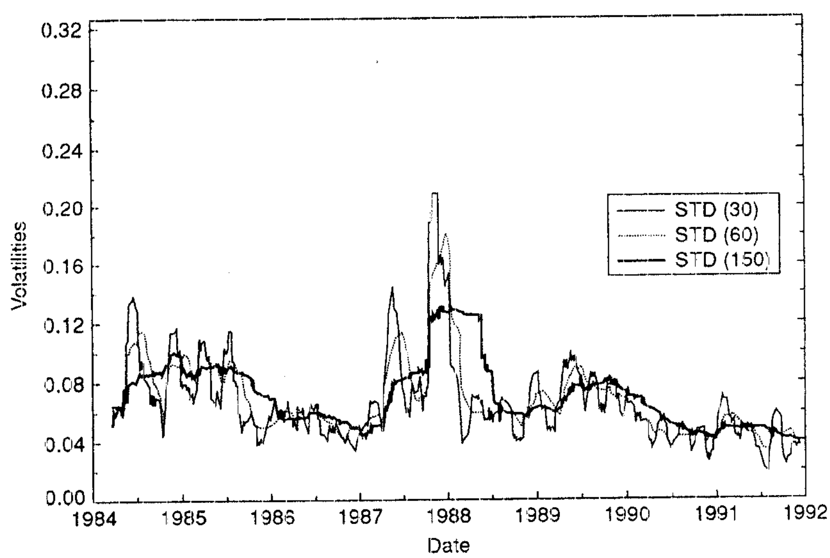

Figure 2.6 Time-varying volatility using historical standard deviation with various window lengths

window-lengths are used: $K = 30$, $K = 60$, and $K = 150$. On any given day we compare these three lookback windows. That is, on any given day (starting with the 151st day), we look back 30, 60, or 150 days and calculate the standard deviation by averaging the squared interest rate changes (and then taking a square root). The figure demonstrates the issues involved in the choice of K. First note that the forecasts for series using shorter windows are more volatile. This could be the result of a statistical error – 30 observations, for example, may provide only a noisy estimate of volatility. On the other hand, variations could be the result of true changes in volatility. The longer window length, $K = 150$ days, provides a relatively smoother series of estimators/forecasts, varying within a tighter range of 4–12 basis points per day. Recall that the unconditional volatility is 7.3bp/day. Shorter window lengths provide extreme estimators, as high as 22bp/day. Such estimators are three times larger than the unconditional volatility.

The effect of the statistical estimation error is particularly acute for small samples, e.g., $K = 30$. The STDEV estimator is particularly sensitive to extreme observations. To see why this is the case, recall that the calculation of STDEV involves an equally weighted average of *squared* deviations from the mean (here zero). Any extreme, perhaps non-normal, observation becomes larger in magnitude by taking it to the power of two. Moreover, with small window sizes each observation receives higher weight by definition. When a large positive or negative return is observed, therefore, a sharp increase in the volatility forecast is observed.

In this context it is worthwhile mentioning that an alternative procedure of calculating the volatility involves averaging absolute values of returns, rather than squared returns. This method is considered more robust when the distribution is non-normal. In fact it is possible to show that while under the normality assumption STDEV is optimal, when returns are non-normal, and, in particular, fat tailed, then the absolute squared deviation method may provide a superior forecast.

This discussion seems to present an argument that longer observation windows reduce statistical error. However, the other side of the coin is that small window lengths provide an estimator that is more adaptable to changing market conditions. In the extreme case where volatility does not vary at all, the longer the window length is, the more accurate our estimates. However, in a time varying volatility environment we face a tradeoff – short window lengths are less precise, due to estimation error, but more adaptable to innovations in volatility. Later in this chapter (in Section 2.2.4.2) we discuss the issue of benchmarking various volatility estimation models and describe simple optimization procedures that allow us to choose the most appropriate window length. Intuitively, for volatility series that are in and of themselves more volatile, we will tend to shorten the window length, and vice versa.

Finally, yet another important shortcoming of the STDEV method for estimating conditional volatility is the periodic appearance of large decreases in conditional volatility. These sharp declines are the result of extreme observations disappearing from the rolling estimation window. The STDEV methodology is such that when a large move occurs we use this piece of data for K days. Then, on day $K + 1$ it falls off the estimation window. The extreme return carries the same weight of ($100/K$) percent from day $t - 1$ to day $t - K$, and then disappears. From an economic perspective this is a counterintuitive way to describe memory in financial markets. A more intuitive description would be to incorporate a gradual decline in memory such that when a crisis occurs it is very relevant for the first week, affecting volatility in financial markets to a great extent, and then as time goes by it becomes gradually less important. Using STDEV with equal weights on observations from the most recent K days, and zero thereafter (further into the past) is counterintuitive. This shortcoming of STDEV is precisely the one addressed by the exponential smoothing approach, adopted by RiskMetrics™ in estimating volatility.

2.2.4 Exponential smoothing – RiskMetrics™ volatility

Suppose we want to use historical data, specifically, squared returns, in order to calculate conditional volatility. How can we improve upon our first estimate, STDEV? We focus on the issue of information decay and on giving more weight to more recent information and less weight to distant information. The simplest, most popular, approach is exponential smoothing. Exponential smoothing places exponentially declining weights on historical data, starting with an initial weight, and then declining to zero as we go further into the past.

The smoothness is achieved by setting a parameter λ, which is equal to a number greater than zero, but smaller than one, raised to a power (i.e., $0 < \lambda < 1$). Any such smoothing parameter λ, when raised to a high enough power, can get arbitrarily small. The sequence of numbers $\lambda^0, \lambda^1, \lambda^2, \ldots \lambda^i, \ldots$ has the desirable property that it starts with a finite number, namely λ^0 (= 1), and ends with a number that could become arbitrarily small (λ^i where i is large). The only problem with this sequence is that we need it to sum to 1 in order for it to be a weighting scheme.

In order to rectify the problem, note that the sequence is geometric, summing up to $1/(1-\lambda)$. For a smoothing parameter of 0.9 for example, the sum of $0.9^0, 0.9^1, 0.9^2, \ldots 0.9^i, \ldots$ is $1/(1-0.9) = 10$. All we need is to define a new sequence which is the old sequence divided by the sum of the sequence and the new sequence will then sum to 1. In the previous example we would divide the sequence by 10. More generally we divide each of the weights by $1/(1-\lambda)$, the sum of the geometric sequence. Note that dividing by $1/(1-\lambda)$ is equivalent to multiplying by $(1-\lambda)$. Hence, the old sequence $\lambda^0, \lambda^1, \lambda^2, \ldots \lambda^i, \ldots$ is replaced by the new sequence

$$(1-\lambda)\lambda^0,\ (1-\lambda)\lambda^1,\ (1-\lambda)\lambda^2, \ldots,\ (1-\lambda)\lambda^i, \ldots$$

This is a "legitimate" weighting scheme, since by construction it sums to one. This is the approach known as the RiskMetrics™ exponential weighting approach to volatility estimation.

The estimator we obtain for conditional variance is:

$$\sigma_t^2 = (1-\lambda)*(\lambda^0 r_{t-1,t}{}^2 + \lambda^1 r_{t-2,t-1}{}^2 + \lambda^2 r_{t-3,t-2}{}^2 + \ldots + \lambda^N r_{t-N-1,t-N}{}^2),$$

where N is some finite number which is the truncation point. Since we truncate after a finite number (N) of observations the sum of the series is not 1. It is, in fact, λ^N. That is, the sequence of the weights we drop, from the "$N+1$"th observation and thereafter, sum up to $\lambda^N/(1-\lambda)$. For example, take $\lambda = 0.94$:

Weight 1	$(1-\lambda)\lambda^0$	$= (1-0.94)$	$= 6.00\%$
Weight 2	$(1-\lambda)\lambda^1$	$= (1-0.94)*0.94$	$= 5.64\%$
Weight 3	$(1-\lambda)\lambda^2$	$= (1-0.94)*0.94^2$	$= 5.30\%$
Weight 4	$(1-\lambda)\lambda^3$	$= (1-0.94)*0.94^3$	$= 4.98\%$
...			
Weight 100	$(1-\lambda)\lambda^{99}$	$= (1-0.94)*0.94^{99}$	$= 0.012\%$

The residual sum of truncated weights is $0.94^{100}/(1-0.94) = 0.034$.

We have two choices with respect to this residual weight

1 We can increase N so that the sum of residual weight is small (e.g., $0.94^{200}/(1-0.94) = 0.00007$);

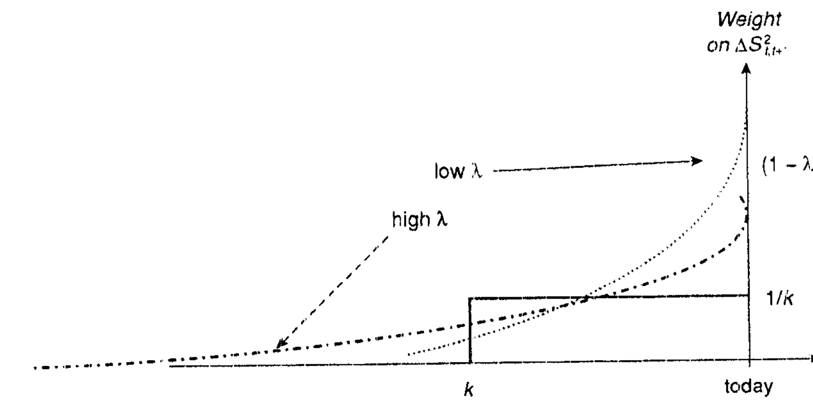

Figure 2.7 STDEV and exponential smoothing weighting schemes

2 or divide by the truncated sum of weights $(1 - \lambda^N)/(1 - \lambda)$ rather than the infinite sum $1/(1 - \lambda)$. In our previous example this would mean dividing by 16.63 instead of 16.66 after 100 observations.

This is a purely technical issue. Either is technically fine, and of little real consequence to the estimated volatility.

In figure 2.7 we compare RiskMetrics™ to STDEV. Recall the important commonalities of these methods

- both methods are parametric;
- both methods attempt to estimate conditional volatility;
- both methods use recent historical data;
- both methods apply a set of weights to past squared returns.

The methods differ only as far as the weighting scheme is concerned. RiskMetrics™ poses a choice with respect to the smoothing parameter λ, (in the example above, equal to 0.94) similar to the choice with respect to K in the context of the STDEV estimator. The tradeoff in the case of STDEV was between the desire for a higher precision, consistent with higher K's, and quick adaptability to changes in conditional volatility, consistent with lower K's. Here, similarly, a λ parameter closer to unity exhibits a slower decay in information's relevance with less weight on recent observations (see the dashed-dotted line in figure 2.7), while lower λ parameters provide a weighting scheme with more weight on recent observations, but effectively a smaller sample (see the dashed line in figure 2.7).

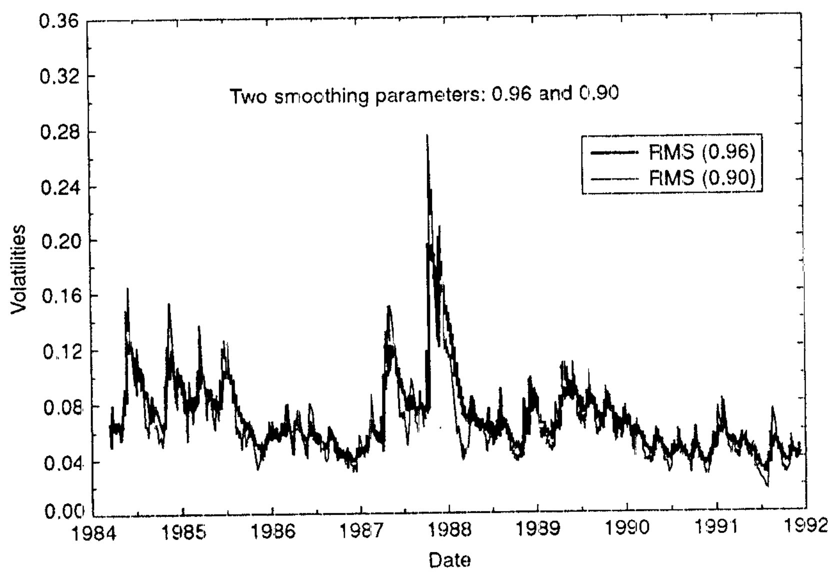

Figure 2.8 RiskMetrics™ volatilities

higher λ. A higher λ not only means more weight on recent observations, it also means that our current beliefs have not changed dramatically from what we believed to be true yesterday.

2.2.4.3 *The empirical performance of RiskMetrics™*

The intuitive appeal of exponential smoothing is validated in empirical tests. For a relatively large portion of the reasonable range for lambdas (most of the estimators fall above 0.90), we observe little visible difference between various volatility estimators. In figure 2.8 we see a series of rolling volatilities with two different smoothing parameters, 0.90 and 0.96. The two series are close to being superimposed on one another. There are extreme spikes using the lower lambda parameter, 0.9, but the choppiness of the forecasts in the back end that we observed with STDEV is now completely gone.[10]

2.2.4.4 *GARCH*

The exponential smoothing method recently gained an important extension in the form of a new time series model for volatility. In a sequence of recent academic papers Robert Engel and Tim Bollereslev introduced a new estimation methodology called GARCH, standing for General Autoregressive Conditional Heteroskedasticity. This sequence of relatively sophisticated-sounding technical terms essentially means that GARCH is a statistical time series model that enables the econometrician to model volatility as time varying and predictable. The model is similar in spirit to RiskMetrics™. In a GARCH(1,1) model the period t conditional volatility is a function of period $t-1$ conditional volatility and the return from $t-1$ to t squared.

$$\sigma_t^2 = a + br_{t-1,t}{}^2 + c\sigma_{t-1}{}^2,$$

where a, b, and c are parameters that need to be estimated empirically. The general version of GARCH, called GARCH(p,q), is

$$\sigma_t^2 = a + b_1 r_{t-1,t}{}^2 + b_2 r_{t-2,t-1}{}^2 + \ldots + b_p r_{t-p+1,t-p}{}^2 + c_1\sigma_{t-1}{}^2 + c_2\sigma_{t-2}{}^2 + \ldots + c_q\sigma_{t-q}{}^2,$$

allowing for p lagged terms on past returns squared, and q lagged terms on past volatility.

With the growing popularity of GARCH it is worth pointing out the similarities between GARCH and other methods, as well as the possible pitfalls in using GARCH. First note that GARCH(1,1) is a generalized case of RiskMetrics™. Put differently, RiskMetrics™ is a restricted case of GARCH. To see this, consider the following two constraints on the parameters of the GARCH(1,1) process:

$$a = 0, \quad b + c = 1.$$

Substituting these two restrictions into the general form of GARCH(1,1) we can rewrite the GARCH model as follows

$$\sigma_t^2 = (1 - c)r_{t-1,t}{}^2 + c\sigma_{t-1}{}^2.$$

This is identical to the recursive version of RiskMetrics™.

The two parameter restrictions or constraints that we need to impose on GARCH(1,1) in order to get the RiskMetrics™ exponential smoothing parameter imply that GARCH is more general or less restrictive. Thus, for a given dataset, GARCH should have better explanatory power than the RiskMetrics™ approach. Since GARCH offers more degrees of freedom, it will have lower error or better describe a given set of data. The problem is that this may not constitute a real advantage in practical applications of GARCH to risk management-related situations.

In reality, we do not have the full benefit of hindsight. The challenge in reality is to predict volatility out-of-sample, not in-sample. Within sample there is no question that GARCH would perform better, simply because it is more flexible and general. The application of GARCH to risk management requires, however, forecasting ability.

The danger in using GARCH is that estimation error would generate noise that would harm the out-of-sample forecasting power. To see this consider what the econometrician interested in volatility forecasting needs to do as time progresses. As new information arrives the econometrician updates the parameters of the model to fit the new data. Estimating parameters repeatedly creates variations in the model itself, some of which are true to the change in the economic environment, and some simply due to sampling variation. The econometrician runs the risk of providing less accurate estimates using GARCH relative to the simpler RiskMetrics™ model in spite of the fact that RiskMetrics™ is a constrained version of GARCH. This is because while the RiskMetrics™ methodology has just one fixed model – a lambda parameter that is a constant (say 0.94) – GARCH is chasing a moving target. As the GARCH parameters change, forecasts change with it, partly due to true variations in the model and the state variables, and partly due to changes in the model due to estimation error. This can create model risk.

Figure 2.9 illustrates this risk empirically. In this figure we see a rolling series of GARCH forecasts, re-estimated daily using a moving window

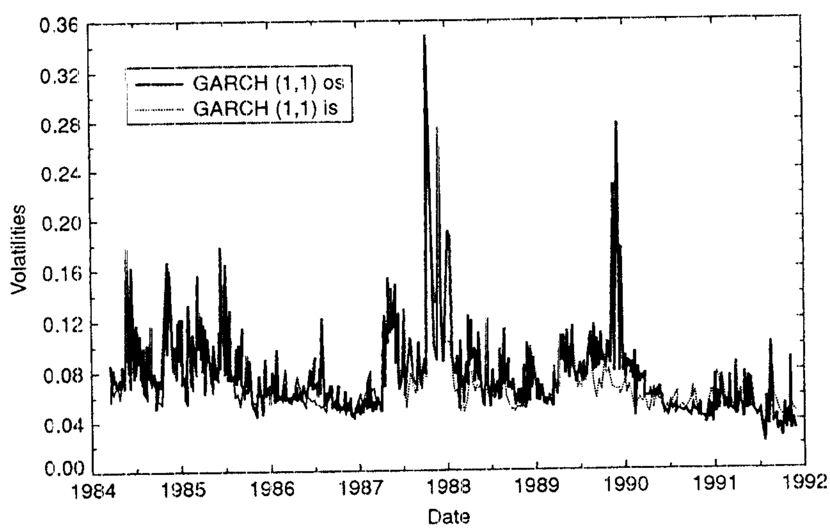

Figure 2.9 GARCH in- and out-of-sample

of 150 observations. The extreme variations in this series relative to a relatively smooth RiskMetrics™ volatility forecast series, that appears on the same graph, demonstrates the risk in using GARCH for forecasting volatility, using a short rolling window.[11]

2.2.5 Nonparametric volatility forecasting

2.2.5.1 Historical simulation

So far we have confined our attention to parametric volatility estimation methods. With parametric models we use all available data, weighted one way or another, in order to estimate parameters of a given distribution. Given a set of relevant parameters we can then determine percentiles of the distribution easily, and hence estimate the VaR of the return on an asset or a set of assets. Nonparametric methods estimate VaR, i.e., percentile of return distribution, directly from the data, without making assumptions about the entire distribution of returns. This is a potentially promising avenue given the phenomena we encountered so far – fat tails, skewness and so forth.

The most prominent and easiest to implement methodology within the class of nonparametric methods is historical simulation (HS). HS uses the data directly. The only thing we need to determine up front is the lookback window. Once the window length is determined, we order returns in descending order, and go directly to the tail of this ordered vector. For an estimation window of 100 observations, for example, the fifth lowest return in a rolling window of the most recent 100 returns is the fifth percentile. The lowest observation is the first percentile. If we wanted, instead, to use a 250 observations window, the fifth percentile would be somewhere between the 12th and the 13th lowest observations (a detailed discussion follows), and the first percentile would be somewhere between the second and third lowest returns.

This is obviously a very simple and convenient method, requiring the estimation of *zero* parameters (window size aside). HS can, in theory, accommodate fat tails skewness and many other peculiar properties of return series. If the "true" return distribution is fat tailed, this will come through in the HS estimate since the fifth observation will be more extreme than what is warranted by the normal distribution. Moreover, if the "true" distribution of asset returns is left skewed since market falls are more extreme than market rises, this will surface through the fact that the 5th and the 95th ordered observations will not be symmetric around zero.

This is all true in theory. With an infinite amount of data we have no difficulty estimating percentiles of the distribution directly. Suppose, for example, that asset returns are truly non-normal, and the correct model involves skewness. If we assume normality we also assume symmetry, and in spite of the fact that we have an infinite amount of data we suffer from model specification error – a problem which is insurmountable. With the HS method we could take, say, the 5,000th of 100,000 observations, a very precise estimate of the fifth percentile.

In reality, however, we do not have an infinite amount of data. What is the result of having to use a relatively small sample in practice? Quantifying the precision of percentile estimates using HS in finite samples is a rather complicated technical issue. The intuition is, however, straightforward. Percentiles around the median (the 50th percentile) are easy to estimate relatively accurately even in small samples. This is because every observation contributes to the estimation by the very fact that it is under or over the median.

Estimating extreme percentiles, such as the first or the fifth percentile, is much less precise in small samples. Consider, for example, estimating the fifth percentile in a window of 100 observations. The fifth percentile is the fifth smallest observation. Suppose that a crisis occurs and during the following ten trading days five new extreme declines were observed. The VaR using the HS method grows sharply. Suppose now that in the following few months no new extreme declines occurred. From an economic standpoint this is news – "no news is good news" is a good description here. The HS estimator of the VaR, on the other hand, reflects the same extreme tail for the following few months, until the observations fall out of the 100 day observation window. There is no updating for 90 days, starting from the ten extreme days (where the five extremes were experienced) until the ten extreme days start dropping out of the sample. This problem can become even more acute with a window of one year (250 observations) and a 1 percent VaR, that requires only the second and third lowest observations.

This problem arises because HS uses data very inefficiently. That is, out of a very small initial sample, focus on the tails requires throwing away a lot of useful information. Recall that the opposite holds true for the parametric family of methods. When the standard deviation is estimated, every data point contributes to the estimation. When

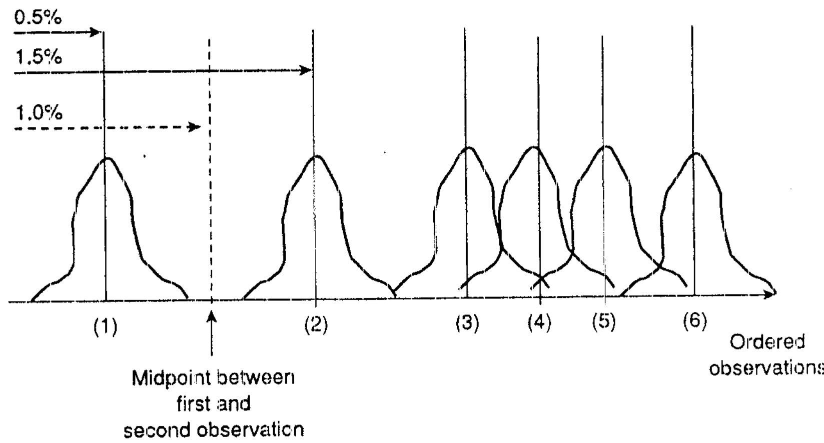

Figure 2.10 Historical simulation method

extremes are observed we update the estimator upwards, and when calm periods bring into the sample relatively small returns (in absolute value), we reduce the volatility forecast. This is an important advantage of the parametric method(s) over nonparametric methods – data are used more efficiently. Nonparametric methods' precision hinges on large samples, and falls apart in small samples.

A minor technical point related to HS is in place here. With 100 observations the first percentile could be thought of as the first observation. However, the observation itself can be thought of as a random event with a probability mass centered where the observation is actually observed, but with 50 percent of the weight to its left and 50 percent to its right. As such, the probability mass we accumulate going from minus infinity to the lowest of 100 observations is only ½ percent and not the full 1 percent. According to this argument the first percentile is somewhere in between the lowest and second lowest observation. Figure 2.10 clarifies the point.

Finally, it might be argued that we can increase the precision of HS estimates by using more data; say, 10,000 past daily observations. The issue here is one of regime relevance. Consider, for example, foreign exchange rates going back 10,000 trading days – approximately 40 years. Over the last 40 years, there have been a number of different exchange rate regimes in place, such as fixed exchange rates under Bretton Woods. Data on returns during periods of fixed exchange rates would have no relevance in forecasting volatility under floating exchange rate regimes. As a result, the risk manager using conventional HS is often forced to rely on the relatively short time period relevant to current market conditions, thereby reducing the usable number of observations for HS estimation.

2.2.5.2 *Multivariate density estimation*

Multivariate density estimation (MDE) is a methodology used to estimate the joint probability density function of a set of variables. For example, one could choose to estimate the joint density of returns and a set of predetermined factors such as the slope of the term structure, the inflation level, the state of the economy, and so forth. From this distribution, the conditional moments, such as the mean and volatility of returns, conditional on the economic state, can be calculated.

The MDE volatility estimate provides an intuitive alternative to the standard set of approaches to weighting past (squared) changes in determining volatility forecasts. The key feature of MDE is that the weights are no longer a constant function of time as in RiskMetrics™ or STDEV. Instead, the weights in MDE depend on how the current state of the world compares to past states of the world. If the current state of the world, as measured by the state vector x_t, is similar to a particular point in the past, then this past squared return is given a lot of weight in forming the volatility forecast, *regardless of how far back in time it is*.

For example, suppose that the econometrician attempts to estimate the volatility of interest rates. Suppose further that according to his model the volatility of interest rates is determined by the level of rates – higher rates imply higher volatility. If today's rate is, say 6 percent, then the relevant history is any point in the past when interest rates were around 6 percent. A statistical estimate of current volatility that uses past data should place high weight on the magnitude of interest rate changes during such times. Less important, although relevant, are times when interest rates were around 5.5 percent or 6.5 percent, even less important although not totally irrelevant are times when interest rates were 5 percent or 7 percent, and so on. MDE devises a weighting scheme that helps the econometrician decide how far the relevant state variable was at any point in the past from its value today. Note that to the extent that relevant state variables are going to be autocorrelated, MDE weights may look, to an extent, similar to RiskMetrics™ weights.

The critical difficulty is to select the relevant (economic) state variables for volatility. These variables should be useful in describing the economic environment in general, and be related to volatility specifically. For example, suppose that the level of inflation is related to the level of return volatility, then inflation will be a good conditioning variable. The advantages of the MDE estimate are that it can be interpreted in the context of weighted lagged returns, and that the functional form of the weights depends on the true (albeit estimated) distribution of the relevant variables.

Using the MDE method, the estimate of conditional volatility is

$$\sigma_t^2 = \sum_{i=1,2,\ldots,K} \omega(x_{t-i}) r_{t-i}^2.$$

Here, x_{t-i} is the vector of variables describing the economic state at time $t - i$ (e.g., the term structure), determining the appropriate weight $\omega(x_{t-i})$ to be placed on observation $t - i$, as a function of the "distance" of the state x_{t-i} from the current state x_t. The relative weight of "near" relative to "distant" observations from the current state is measured via the kernel function.[12]

MDE is extremely flexible in allowing us to introduce dependence on state variables. For example, we may choose to include past squared returns as conditioning variables. In doing so the volatility forecasts will depend nonlinearly on these past changes. For example, the exponentially smoothed volatility estimate can be added to an array of relevant conditioning variables. This may be an important extension to the GARCH class of models. Of particular note, the estimated volatility is still based directly on past squared returns and thus falls into the class of models that places weights on past squared returns.

The added flexibility becomes crucial when one considers cases in which there are other relevant state variables that can be added to the current state. For example, it is possible to capture: (i) the dependence of interest rate volatility on the level of interest rates; (ii) the dependence of equity volatility on current implied volatilities; and (iii) the dependence of exchange rate volatility on interest rate spreads, proximity to intervention bands, etc.

There are potential costs in using MDE. We must choose a weighting scheme (a kernel function), a set of conditioning variables, and the number of observations to be used in estimating volatility. For our purposes, the bandwidth and kernel function are chosen objectively (using standard criteria). Though they may not be optimal choices, it is important to avoid problems associated with data snooping and overfitting. While the choice of conditioning variables is at our discretion and subject to abuse, the methodology does provide a considerable

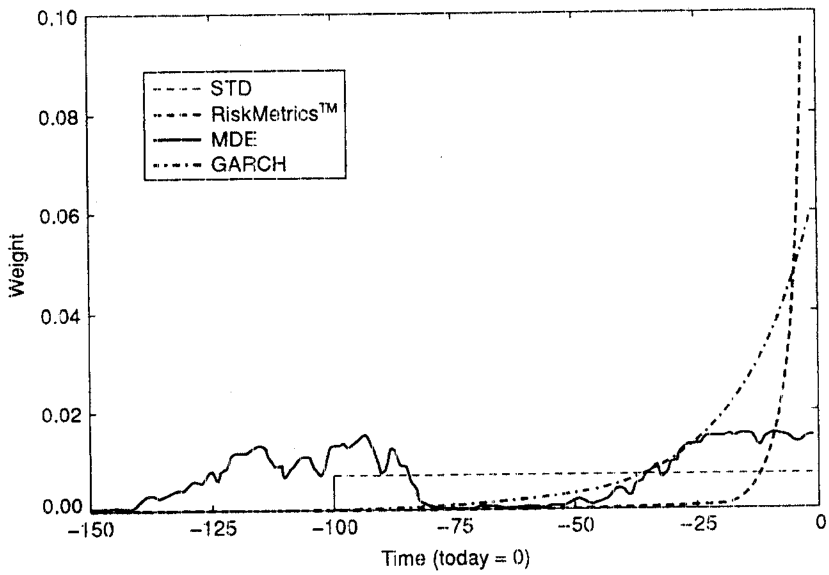

Figure 2.11 MDE weights on past returns squared

advantage. Theoretical models and existing empirical evidence may suggest relevant determinants for volatility estimation, which MDE can incorporate directly. These variables can be introduced in a straightforward way for the class of stochastic volatility models we discuss.

The most serious problem with MDE is that it is data intensive. Many data are required in order to estimate the appropriate weights that capture the joint density function of the variables. The quantity of data that is needed increases rapidly with the number of conditioning variables used in estimation. On the other hand, for many of the relevant markets this concern is somewhat alleviated since the relevant state can be adequately described by a relatively low dimensional system of factors.[13]

As an illustration of the four methodologies put together, figure 2.11 shows the weights on past squared interest rate changes as of a specific date estimated by each model. The weights for STDEV and RiskMetrics™ are the same in every period, and will vary only with the window length and the smoothing parameter. The GARCH(1,1) weighting scheme varies with the parameters, which are re-estimated every period, given each day's previous 150-day history. The date was selected at random. For that particular day, the GARCH parameter selected is $b = 0.74$. Given that this parameter is relatively low, it is not surprising that the weights decay relatively quickly. Figure 2.11 is particularly illuminating with respect to MDE. As with GARCH, the weights change over time. The weights are high for dates t through $t - 25$ (25 days prior) and then start to decay. The state variables chosen here for volatility are the level and the slope of the term structure, together providing information about the state of interest rate volatility (according to *our* choice). The weights decrease because the economic environment, as described by the interest rate level and spread, is moving further away from the conditions observed at date t. However, we observe an increase in the weights for dates $t - 80$ to $t - 120$. Economic conditions in this period (the level and spread) are similar to those at date t. MDE puts high weight on relevant information, regardless of how far in the past this information is.[14]

2.2.6 *A comparison of methods*

Table 2.2 compares, on a period-by-period basis, the extent to which the forecasts from the various models line up with realized future volatility. We define realized daily volatility as the average squared daily changes during the following (trading) week, from day $t + 1$ to day $t + 5$. Recall our discussion of the mean squared error. In order to benchmark various methods we need to test their accuracy vis-à-vis realized volatility – an unknown before and after the fact. If we used the realized squared return during the day following each volatility forecast we run into estimation error problems. On the other hand

Table 2.2 A comparison of methods

	STDEV	RiskMetrics™	MDE	GARCH
Mean	0.070	0.067	0.067	0.073
Std. Dev	0.022	0.029	0.024	0.030
Autocorr.	0.999	0.989	0.964	0.818
MSE	0.999	0.930	0.887	1.115
		Linear regression		
Beta	0.577	0.666	0.786	0.559
(s.e.)	(0.022)	(0.029)	(0.024)	(0.030)
R^2	0.100	0.223	0.214	0.172

if we measure realized volatility as standard deviation during the following month, we run the risk of inaccuracy due to over aggregation because volatility may shift over a month's time period. The tradeoff between longer and shorter horizons going forward is similar to the tradeoff discussed in section 2.2.3 regarding the length of the lookback window in calculating STDEV. We will use the realized volatility, as measured by mean squared deviation during the five trading days following each forecast. Interest rate changes are mean-adjusted using the sample mean of the previous 150-day estimation period.

The comparison between realized and forecasted volatility is done in two ways. First, we compare the out-of-sample performance over the entire period using the mean-squared error of the forecasts. That is, we take the difference between each model's volatility forecast and the realized volatility, square this difference, and average through time. This is the standard MSE formulation. We also regress realized volatility on the forecasts and document the regression coefficients and R^2s.

The first part of table 2.2 documents some summary statistics that are quite illuminating. First, while all the means of the volatility forecasts are of a similar order of magnitude (approximately seven basis points per day), the standard deviations are quite different, with the most volatile forecast provided by GARCH(1,1). This result is somewhat surprising because GARCH(1,1) is supposed to provide a relatively smooth volatility estimate (due to the moving average term). However, for rolling, out-of-sample forecasting, the variability of the parameter estimates from sample to sample induces variability in the forecasts. These results are, however, upwardly biased, since GARCH would commonly require much more data to yield stable parameter estimates. Here we re-estimate GARCH every day using a 150-day lookback period. From a practical perspective, this finding of unstable forecasts for volatility is a model disadvantage. In particular, to the extent that such numbers serve as inputs in setting time-varying rules in a risk management system (for example, by setting trading limits), smoothness of these rules is necessary to avoid large swings in positions.

Regarding the forecasting performance of the various volatility models, table 2.2 provides the mean squared error measure (denoted MSE). For this particular sample and window length, MDE minimizes the MSE, with the lowest MSE of 0.887. RiskMetrics™ (using $\lambda = 0.94$ as the smoothing parameter) also performs well, with an MSE of 0.930.

Note that this comparison involves just one particular GARCH model (i.e., GARCH(1,1)), over a short estimation window, and does not necessarily imply anything about other specifications and window lengths. One should investigate other window lengths and specifications, as well as other data series, to reach general conclusions regarding model comparisons. It is interesting to note, however, that, nonstationarity aside, exponentially smoothed volatility is a special case of GARCH(1,1) in sample, as discussed earlier. The results here suggest, however, the potential cost of the error in estimation of the GARCH smoothing parameters on an out-of-sample basis.

An alternative approach to benchmarking the various volatility-forecasting methods is via linear regression of realized volatility on the forecast. If the conditional volatility is measured without error, then the slope coefficient (or beta) should equal one. However, if the forecast is unbiased but contains estimation error, then the coefficient will be biased downwards. Deviations from one reflect a combination of this estimation error plus any systematic over- or underestimation. The ordering in this "horse race" is quite similar to the previous one. In particular, MDE exhibits the beta coefficient closest to one (0.786), and exponentially smoothed volatility comes in second, with a beta parameter of 0.666. The goodness of fit measure, the R^2 of each of the regressions, is similar for both methods.

2.2.7 The hybrid approach

The hybrid approach combines the two simplest approaches (for our sample), HS and RiskMetrics™, by estimating the percentiles of the return directly (similar to HS), and using exponentially declining weights on past data (similar to RiskMetrics™). The approach starts with ordering the returns over the observation period just like the HS approach. While the HS approach attributes equal weights to each observation in building the conditional empirical distribution, the hybrid approach attributes exponentially declining weights to historical returns. Hence, while obtaining the 1 percent VaR using 250 daily returns involves identifying the third lowest observation in the HS approach, it may involve more or less observations in the hybrid approach. The exact number of observations will depend on whether the extreme low returns were observed recently or further in the past. The weighting scheme is similar to the one applied in the exponential smoothing (EXP hence) approach.

The hybrid approach is implemented in three steps:

Step 1: Denote by $r_{t-1,t}$ the realized return from $t-1$ to t. To each of the most recent K returns: $r_{t-1,t}$, $r_{t-2,t-1}$, . . . , $r_{t-K,t-k-1}$ assign a weight $[(1-\lambda)/(1-\lambda^K)]$, $[(1-\lambda)/(1-\lambda^K)]\lambda$, . . . , $[(1-\lambda)/(1-\lambda^K)]\lambda^{K-1}$, respectively. Note that the constant $[(1-\lambda)/(1-\lambda^K)]$ simply ensures that the weights sum to one.

Step 2: Order the returns in ascending order.

Step 3: In order to obtain the x percent VaR of the portfolio, start from the lowest return and keep accumulating the weights until x percent is reached. Linear interpolation is used between adjacent points to achieve exactly x percent of the distribution.

Consider the following example, we examine the VaR of a given series at a given point in time, and a month later, assuming that no extreme observations were realized during the month. The parameters are $\lambda = 0.98$, $K = 100$.

The top half of table 2.3 shows the ordered returns at the initial date. Since we assume that over the course of a month no extreme

Table 2.3 The hybrid approach – an example

Order	Return	Periods ago	Hybrid weight	Hybrid cumul. weight	HS weight	HS cumul. weight
Initial date:						
1	−3.30%	3	0.0221	0.0221	0.01	0.01
2	−2.90%	2	0.0226	0.0447	0.01	0.02
3	−2.70%	65	0.0063	0.0511	0.01	0.03
4	−2.50%	45	0.0095	0.0605	0.01	0.04
5	−2.40%	5	0.0213	0.0818	0.01	0.05
6	−2.30%	30	0.0128	0.0947	0.01	0.06
25 days later:						
1	−3.30%	28	0.0134	0.0134	0.01	0.01
2	−2.90%	27	0.0136	0.0270	0.01	0.02
3	−2.70%	90	0.0038	0.0308	0.01	0.03
4	−2.50%	70	0.0057	0.0365	0.01	0.04
5	−2.40%	30	0.0128	0.0494	0.01	0.05
6	−2.30%	55	0.0077	0.0571	0.01	0.06

returns are observed, the ordered returns 25 days later are the same. These returns are, however, further in the past. The last two columns show the equally weighted probabilities under the HS approach. Assuming an observation window of 100 days, the HS approach estimates the 5 percent *VaR* to be 2.35 percent for both cases (note that VaR is the negative of the actual return). This is obtained using interpolation on the actual historical returns. That is, recall that we assume that half of a given return's weight is to the right and half to the left of the actual observation (see figure 2.10). For example, the −2.40 percent return represents 1 percent of the distribution in the HS approach, and we assume that this weight is split evenly between the intervals from the actual observation to points halfway to the next highest and lowest observations. As a result, under the HS approach, −2.40 percent represents the 4.5th percentile, and the distribution of weight leads to the 2.35 percent VaR (halfway between 2.40 percent and 2.30 percent).

In contrast, the hybrid approach departs from the equally weighted HS approach. Examining first the initial period, table 2.3 shows that the cumulative weight of the −2.90 percent return is 4.47 percent and 5.11 percent for the −2.70 percent return. To obtain the 5 percent VaR for the initial period, we must interpolate as shown in figure 2.10. We obtain a cumulative weight of 4.79 percent for the −2.80 percent return. Thus, the 5th percentile VaR under the hybrid approach for the initial period lies somewhere between 2.70 percent and 2.80 percent. We define the required VaR level as a linearly interpolated return, where the distance to the two adjacent cumulative weights determines the return. In this case, for the initial period the 5 percent VaR under the hybrid approach is:

$$2.80\% - (2.80\% - 2.70\%)*[(0.05 - 0.0479)/(0.0511 - 0.0479)]$$
$$= 2.73\%.$$

Similarly, the hybrid approach estimate of the 5 percent VaR 25 days later can be found by interpolating between the −2.40 percent return (with a cumulative weight of 4.94 percent) and −2.35 percent (with a cumulative weight of 5.33 percent, interpolated from the values on table 2.3). Solving for the 5 percent VaR:

$$2.35\% - (2.35\% - 2.30\%)*[(0.05 - 0.0494)/(0.0533 - 0.0494)]$$
$$= 2.34\%.$$

Thus, the hybrid approach initially estimates the 5 percent VaR as 2.73 percent. As time goes by and no large returns are observed, the VaR estimate smoothly declines to 2.34 percent. In contrast, the HS approach yields a constant 5 percent VaR over both periods of 2.35 percent, thereby failing to incorporate the information that returns were stable over the two month period. Determining which methodology is appropriate requires backtesting (see Appendix 2.1).

2.3 RETURN AGGREGATION AND VaR

Our discussion of the HS and hybrid methods missed one key point so far. How do we aggregate a number of positions into a single VaR number for a portfolio comprised of a number of positions? The answer to this question in the RiskMetrics™ and STDEV approaches is simple – under the assumption that asset returns are jointly normal, the return on a portfolio is also normally distributed. Using the variance–covariance matrix of asset returns we can calculate portfolio volatility and VaR. This is the reason for the fact that the RiskMetrics™ approach is commonly termed the Variance–Covariance approach (VarCov).

The HS approach needs one more step – missing so far from our discussion – before we can determine the VaR of a portfolio of positions. This is the aggregation step. The idea is simply to aggregate each period's historical returns, weighted by the relative size of the position. This is where the method gets its name – "simulation". We calculate returns using historical data, but using today's weights. Suppose for example that we hold today positions in three equity portfolios – indexed to the S&P 500 index, the FTSE index and the Nikkei 225 index – in equal amounts. These equal weights are going to be used to calculate the return we would have gained J days ago if we were to hold this equally weighted portfolio. This is regardless of the fact that our equity portfolio J days ago may have been completely different. That is, we pretend that the portfolio we hold today is the portfolio we held up to K days into the past (where K is our lookback window size) and calculate the returns that would have been earned.

From an implementation perspective this is very appealing and simple. This approach has another important advantage – note that we do not estimate any parameters whatsoever. For a portfolio involving N positions the VarCov approach requires the estimation of N volatilities and $N(N - 1)/2$ correlations. This is potentially a very large number, exposing the model to estimation error. Another important issue is related to the estimation of correlation. It is often argued that when markets fall, they fall together. If, for example, we see an abnormally large decline of 10 percent in the S&P index on a given day, we strongly believe that other components of the portfolio, e.g., the Nikkei position and the FTSE position, will also fall sharply. This is regardless of the fact that we may have estimated a correlation of, for example, 0.30 between the Nikkei and the other two indexes under more normal market conditions (see Longin and Solnik (2001)).

The possibility that markets move together at the extremes to a greater degree than what is implied by the estimated correlation parameter poses a serious problem to the risk manager. A risk manager using the VarCov approach is running the risk that his VaR estimate for the position is understated. At the extremes the benefits of diversification disappear. Using the HS approach with the initial aggregation step may offer an interesting solution. First, note that we do not need to estimate correlation parameters (nor do we need to estimate volatility parameters). If, on a given day, the S&P dropped 10 percent, the Nikkei dropped 12 percent and the FTSE dropped 8 percent, then an equally weighted portfolio will show a drop of 10 percent – the average of the three returns. The following step of the HS methods is to order the observations in ascending order and pick the fifth of 100 observations (for the 5 percent VaR, for example). If the tails are extreme, and if markets co-move over and above the estimated correlations, it will be taken into account through the aggregated data itself.

Figure 2.12 provides a schematic of the two alternatives. Given a set of historical data and current weights we can either use the variance-covariance matrix in the VarCov approach, or aggregate the returns and then order them in the HS approach. There is an obvious third alternative methodology emerging from this figure. We may estimate the volatility (and mean) of the vector of aggregated returns and assuming normality calculate the VaR of the portfolio.

Is this approach sensible? If we criticize the normality assumption we should go with the HS approach. If we believe normality we should take the VarCov approach. What is the validity of this intermediate approach of aggregating first, as in the HS approach, and only then assuming normality as in the VarCov approach? The answer lies in one of the most important theorems in statistics, the strong law of large numbers. Under certain assumptions it is the case that an average of a very large number of random variables will end up converging to a normal random variable.

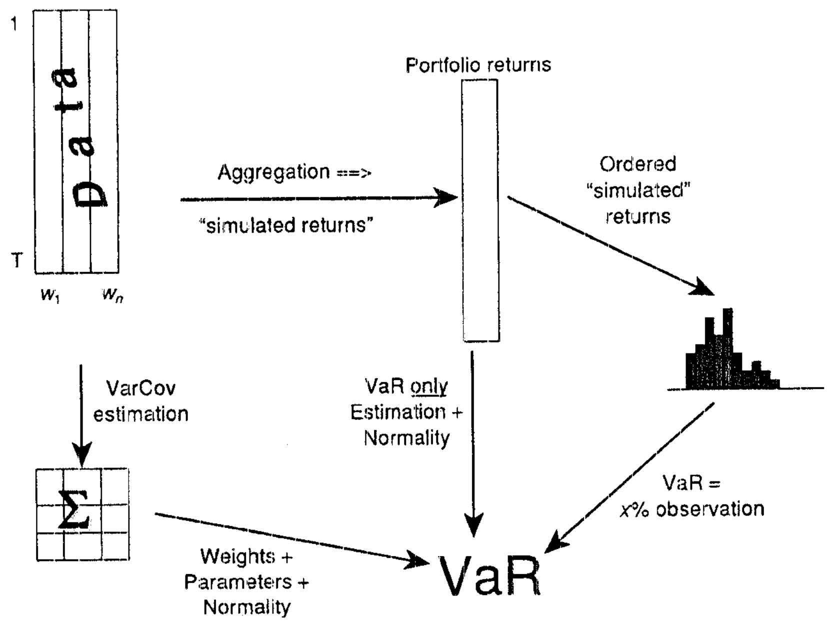

Figure 2.12 VaR and aggregation

It is, in principle, possible, for the specific components of the portfolio to be non-normal, but for the portfolio as a whole to be normally distributed. In fact, we are aware of many such examples. Consider daily stock returns for example. Daily returns on specific stocks are often far from normal, with extreme moves occurring for different stocks at different times. The aggregate, well-diversified portfolio of these misbehaved stocks, could be viewed as normal (informally, we may say the portfolio is more normal than its component parts – a concept that could easily be quantified and is often tested to be true in the academic literature). This is a result of the strong law of large numbers.

Similarly here we could think of normality being regained, in spite of the fact that the single components of the portfolio are non-normal. This holds only if the portfolio is well diversified. If we hold a portfolio comprised entirely of oil- and gas-related exposures, for example, we may hold a large number of positions that are all susceptible to sharp movements in energy prices.

This last approach – of combining the first step of aggregation with the normality assumption that requires just a single parameter estimate – is gaining popularity and is used by an increasing number of risk managers.

2.4 IMPLIED VOLATILITY AS A PREDICTOR OF FUTURE VOLATILITY

Thus far our discussion has focused on various methods that involve using historical data in order to estimate future volatility. Many risk managers describe managing risk this way as similar to driving by looking in the rear-view mirror. When extreme circumstances arise in financial markets an immediate reaction, and preferably even a preliminary indication, are of the essence. Historical risk estimation techniques require time in order to adjust to changes in market conditions. These methods suffer from the shortcoming that they may follow, rather than forecast risk events. Another worrisome issue is that a key assumption in all of these methods is stationarity; that is, the assumption that the past is indicative of the future.

Financial markets provide us with a very intriguing alternative – option-implied volatility. Implied volatility can be imputed from derivative prices using a specific derivative pricing model. The simplest example is the Black–Scholes implied volatility imputed from equity option prices. The implementation is fairly simple, with a few technical issues along the way. In the presence of multiple implied volatilities for various option maturities and exercise prices, it is common to take the at-the-money (ATM) implied volatility from puts and calls and extrapolate an average implied; this implied is derived from the most liquid (ATM) options. This implied volatility is a candidate to be used in risk measurement models in place of historical volatility. The advantage of implied volatility is that it is a forward-looking, predictive measure.

A particularly strong example of the advantage obtained by using implied volatility (in contrast to historical volatility) as a predictor of future volatility is the GBP currency crisis of 1992. During the summer of 1992, the GBP came under pressure as a result of the expectation that it should be devalued relative to the European Currency Unit (ECU) components, the deutschmark (DM) in particular (at the time the strongest currency within the ECU). During the weeks preceding the final drama of the GBP devaluation, many signals were present in the public domain. The British Central Bank raised the GBP interest rate. It also attempted to convince the Bundesbank to lower the DM

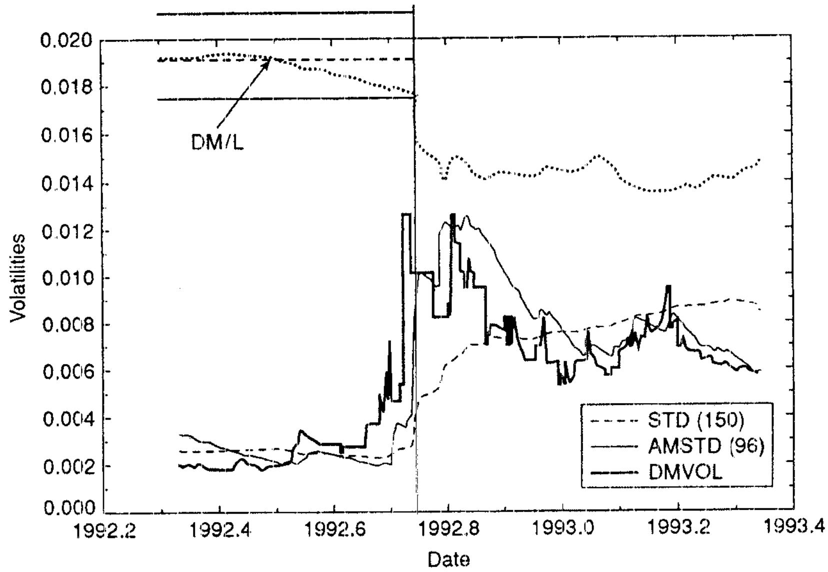

Figure 2.13 Implied and historical volatility: the GBP during the ERM crisis of 1992

interest rate, but to no avail. Speculative pressures reached a peak toward summer's end, and the British Central Bank started losing currency reserves, trading against large hedge funds such as the Soros fund.

The market was certainly aware of these special market conditions, as shown in figure 2.13. The top dotted line is the DM/GBP exchange rate, which represents our "event clock." The event is the collapse of the exchange rate. Figure 2.13 shows the Exchange Rate Mechanism (ERM) intervention bands. As was the case many times prior to this event, the most notable predictor of devaluation was already present – the GBP is visibly close to the intervention band. A currency so close to the intervention band is likely to be under attack by speculators on the one hand, and under intervention by the central banks on the other. This was the case many times prior to this event, especially with the Italian lira's many devaluations. Therefore, the market was prepared for a crisis in the GBP during the summer of 1992. Observing the thick solid line depicting option-implied volatility, the growing pressure on the GBP manifests itself in options prices and volatilities. Historical volatility is trailing, "unaware" of the pressure. In this case, the situation is particularly problematic since historical volatility happens to decline as implied volatility rises. The fall in historical volatility is due to the fact that movements close to the intervention band are bound to be smaller by the fact of the intervention bands' existence and the nature of intervention, thereby dampening the historical measure of volatility just at the time that a more predictive measure shows increases in volatility.

As the GBP crashed, and in the following couple of days, RiskMetrics™ volatility increased quickly (thin solid line). However, simple STDEV ($K = 50$) badly trailed events – it does not rise in time, nor does it fall in time. This is, of course, a particularly sharp example, the result of the intervention band preventing markets from fully reacting to information. As such, this is a unique example. Does it generalize to all other assets? Is it the case that implied volatility is a superior predictor of future volatility, and hence a superior risk measurement tool, relative to historical? It would seem as if the answer must be affirmative, since implied volatility can react immediately to market conditions. As a predictor of future volatility this is certainly an important feature.

Implied volatility is not free of shortcomings. The most important reservation stems from the fact that implied volatility is model-dependent. A misspecified model can result in an erroneous forecast. Consider the Black–Scholes option-pricing model. This model hinges on a few assumptions, one of which is that the underlying asset follows a continuous time lognormal diffusion process. The underlying assumption is that the volatility parameter is constant from the present time to the maturity of the contract. The implied volatility is supposedly this parameter. In reality, volatility is not constant over the life of the options contract. Implied volatility varies through time. Oddly, traders trade options in "vol" terms, the volatility of the underlying, fully aware that (i) this vol is implied from a constant volatility model, and (ii) that this very same option will trade tomorrow at a different vol, which will also be assumed to be constant over the remaining life of the contract.

Yet another problem is that at a given point in time, options on the same underlying may trade at different vols. An example is the *smile effect* – deep out of the money (especially) and deep in the money (to a lesser extent) options trade at a higher vol than at the money options.[15]

The key is that the option-pricing model provides a convenient nonlinear transformation allowing traders to compare options with different maturities and exercise prices. The true underlying process is not a lognormal diffusion with constant volatility as posited by the model. The underlying process exhibits stochastic volatility, jumps, and a non-normal conditional distribution. The vol parameter serves as a "kitchen-sink" parameter. The market converses in vol terms, adjusting for the possibility of sharp declines (the smile effect) and variations in volatility.

The latter effect – stochastic volatility, results in a particularly difficult problem for the use of implied volatility as a predictor of future volatility. To focus on this particular issue, consider an empirical exercise repeatedly comparing the 30-day implied volatility with the empirically measured volatility during the following month. Clearly, the forecasts (i.e., implied) should be equal to the realizations (i.e., measured return standard deviation) only on average. It is well understood that forecast series are bound to be smoother series, as expectations series always are relative to realization series. A reasonable requirement is, nevertheless, that implied volatility should be equal, on average, to realized volatility. This is a basic requirement of every forecast instrument – it should be unbiased.

Empirical results indicate, strongly and consistently, that implied volatility is, on average, greater than realized volatility. From a modeling perspective this raises many interesting questions, focusing on this empirical fact as a possible key to extending and improving option pricing models. There are, broadly, two common explanations. The first is a market inefficiency story, invoking supply and demand issues. This story is incomplete, as many market-inefficiency stories are, since it does not account for the presence of free entry and nearly perfect competition in derivative markets. The second, rational markets, explanation for the phenomenon is that implied volatility is greater than realized volatility due to stochastic volatility. Consider the following facts: (i) volatility is stochastic; (ii) volatility is a priced source of risk; and (iii) the underlying model (e.g., the Black–Scholes model) is, hence, misspecified, assuming constant volatility. The result is that the premium required by the market for stochastic volatility will manifest itself in the forms we saw above – implied volatility would be, on average, greater than realized volatility.

From a risk management perspective this bias, which can be expressed as $\sigma_{implied} = \sigma_{true} + Stoch.Vol.Premium$, poses a problem for the use of implied volatility as a predictor for future volatility. Correcting for this premium is difficult since the premium is unknown, and requires the "correct" model in order to measure precisely. The only thing we seem to know about this premium is that it is on average positive, since implied volatility is on average greater than historical volatility.

It is an empirical question, then, whether we are better off with historical volatility or implied volatility as the predictor of choice for future volatility. Many studies have attempted to answer this question with a consensus emerging that implied volatility is a superior estimate. This result would have been even sharper if these studies were to focus on the responsiveness of implied and historical to sharp increases in conditional volatility. Such times are particularly important for risk managers, and are the primary shortcoming associated with models using the historical as opposed to the implied volatility.

In addition to the upward bias incorporated in the measures of implied volatility, there is another more fundamental problem associated with replacing historical volatility with implied volatility measures. It is available for very few assets/market factors. In a covariance matrix of 400 by 400 (approximately the number of assets/markets that RiskMetrics™ uses), very few entries can be filled with implied volatilities because of the sparsity of options trading on the underlying assets. The use of implied volatility is confined to highly concentrated portfolios where implied volatilities are present. Moreover, recall that with more than one pervasive factor as a measure of portfolio risk, one would also need an implied correlation. Implied correlations are hard to come by. In fact, the only place where reliable liquid implied correlations could be imputed is in currency markets.[16]

As a result, implied volatility measures can only be used for fairly concentrated portfolios with high foreign exchange rate risk exposure. Where available, implied volatility can always be compared in real time to historical (e.g., RiskMetrics™) volatility. When implied volatilities get misaligned by more than a certain threshold level (say, 25 percent difference), then the risk manager has an objective "red light" indication. This type of rule may help in the decision making process of risk limit readjustment in the face of changing market conditions. In the discussion between risk managers and traders, the comparison of historical to implied can serve as an objective judge.[17]

2.5 LONG HORIZON VOLATILITY AND VaR

In many current applications, e.g., such as by mutual fund managers, there is a need for volatility and VaR forecasts for horizons longer than a day or a week. The simplest approach uses the "square root rule," discussed briefly in Chapter 1. Under certain assumptions, to be discussed below, the rule states that an asset's J-period return volatility is equal to the square root of J times the single period return volatility

$$\sigma(r_{t,t+J}) = \sqrt{(J)} \times \sigma(r_{t,t+1}).$$

Similarly for VaR this rule is

$$J\text{-period VaR} = \sqrt{(J)} \times 1\text{-period VaR}.$$

The rule hinges on a number of key assumptions. It is important to go through the proof of this rule in order to examine its limits. Consider, first, the multiperiod continuously compounded rate of return. For simplicity consider the two-period return:

$$r_{t,t+2} = r_{t,t+1} + r_{t+1,t+2}.$$

The variance of this return is

$$\text{var}(r_{t,t+2}) = \text{var}(r_{t,t+1}) + \text{var}(r_{t+1,t+2}) + 2^*\text{cov}(r_{t,t+1}, r_{t+1,t+2}).$$

Assuming:

$$A1: \quad \text{cov}(r_{t,t+1}, r_{t+1,t+2}) = 0,$$

$$A2: \quad \text{var}(r_{t,t+1}) = \text{var}(r_{t+1,t+2}),$$

we get

$$\text{var}(r_{t,t+2}) = 2^*\text{var}(r_{t,t+1}),$$

and hence

$$\text{STD}(r_{t,t+2}) = \sqrt{(2)}^*\text{STD}(r_{t,t+1}).$$

Which is the square root rule for two periods. The rule generalizes easily to the J period rule.

The first assumption is the assumption of non-predictability, or the random walk assumption. The term $\text{cov}(r_{t,t+1}, r_{t+1,t+2})$ is the autocovariance of returns. Intuitively the autocovariance being zero means that knowledge that today's return is, for example, positive, tells us nothing with respect to tomorrow's return. Hence this is also a direct result of the random walk assumption, a standard market efficiency

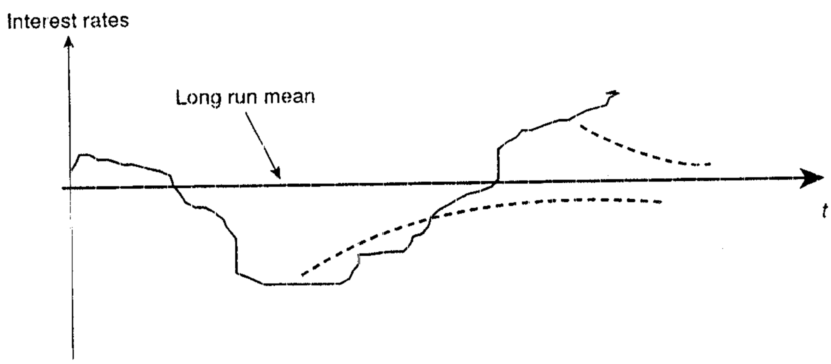

Figure 2.14 Mean reverting process

assumption. The second assumption states that the volatility is the same in every period (i.e., on each day).

In order to question the empirical validity of the rule, we need to question the assumptions leading to this rule. The first assumption of non-predictability holds well for most asset return series in financial markets. Equity returns are unpredictable at short horizons. The evidence contrary to this assertion is scant and usually attributed to luck. The same is true for currencies. There is some evidence of predictability at long horizons (years) for both, but the extent of predictability is relatively small. This is not the case, though, for many fixed-income-related series such as interest rates and especially spreads.

Interest rates and spreads are commonly believed to be predictable to varying degrees, and modeling predictability is often done through time series models accounting for autoregression. An autoregressive process is a stationary process that has a long run mean, an average level to which the series tends to revert. This average is often called the "Long Run Mean" (LRM). Figure 2.14 represents a schematic of interest rates and their long run mean. The dashed lines represent the expectations of the interest rate process. When interest rates are below their LRM they are expected to rise and vice versa.

Mean reversion has an important effect on long-term volatility. To understand the effect, note that the autocorrelation of interest rate changes is no longer zero. If increases and decreases in interest rates (or spreads) are expected to be reversed, then the serial covariance is negative. This means that the long horizon volatility is overstated using the zero-autocovariance assumption. *In the presence of mean reversion in*

Table 2.4 Long horizon volatility

Mean reversion	$\sqrt{J}$ rule using today's volatility
In returns	overstates true long horizon volatility
In return volatility	If today's vol. > LRM vol. then overstated If today's vol. < LRM vol. then understated

the underlying asset's long horizon, volatility is lower than the square root times the short horizon volatility.

The second assumption is that volatility is constant. As we have seen throughout this chapter, this assumption is unrealistic. Volatility is stochastic, and, in particular, autoregressive. This is true for almost all financial assets. Volatility has a long run mean – a "steady state" of uncertainty. Note here the important difference – most financial series have an unpredictable series of returns, and hence no long run mean (LRM), with the exception of interest rates and spreads. However, most volatility series are predictable, and do have an LRM.

When current volatility is above its long run mean then we can expect a decline in volatility over the longer horizon. Extrapolating long horizon volatility using today's volatility will overstate the true expected long horizon volatility. On the other hand, if today's volatility is unusually low, then extrapolating today's volatility using the square root rule may understate true long horizon volatility. The bias – upwards or downwards, hence, depends on today's volatility relative to the LRM of volatility. The discussion is summarized in table 2.4.

2.6 MEAN REVERSION AND LONG HORIZON VOLATILITY

Modeling mean reversion in a stationary time series framework is called the analysis of autoregression (AR). We present here an AR(1) model, which is the simplest form of mean reversion in that we consider only one lag. Consider a process described by the regression of the time series variable X_t:

$$X_{t+1} = a + bX_t + e_{t+1}.$$

This is a regression of a variable on its own lag. It is often used in financial modeling of time series to describe processes that are mean reverting, such as the real exchange rate, the price/dividend or price/earnings ratio, and the inflation rate. Each of these series can be modeled using an assumption about how the underlying process is predictable. This time series process has a finite long run mean under certain restrictions, the most important of which is that the parameter b is less than one. The expected value of X_t as a function of period t information is

$$E_t[X_{t+1}] = a + bX_t.$$

We can restate the expectations as follows

$$E_t[X_{t+1}] = (1 - b)*[a/(1 - b)] + bX_t.$$

Next period's expectations are a weighted sum of today's value, X_t, and the long run mean $a/(1 - b)$. Here b is the key parameter, often termed "the speed of reversion" parameter. If $b = 1$ then the process is a random walk – a nonstationary process with an undefined (infinite) long run mean, and, therefore, next period's expected value is equal to today's value. If $b < 1$ then the process is mean reverting. When X_t is above the LRM, it is expected to decline, and vice versa.

By subtracting X_t from the autoregression formula we obtain the "return", the change in X_t

$$\begin{aligned} X_{t+1} - X_t &= a + bX_t + e_{t+1} - X_t \\ &= a + (b - 1)X_t + e_{t+1} \end{aligned}$$

and the two period return is

$$\begin{aligned} X_{t+2} - X_t &= a + ab + b^2X_t + be_{t+1} + e_{t+2} - X_t \\ &= a(1 + b) + (b^2 - 1)X_t + be_{t+1} + e_{t+2}. \end{aligned}$$

The single period conditional variance of the rate of change is

$$\begin{aligned} \mathrm{var}_t(X_{t+1} - X_t) &= \mathrm{var}_t(a + bX_t + e_{t+1} - X_t) \\ &= \mathrm{var}_t(e_{t+1}) \\ &= \sigma^2. \end{aligned}$$

The volatility of e_{t+1} is denoted by σ. The two period volatility is market closes at 1:00 a.m. EST, fifteen hours earlier. Any information that is relevant for global interest rates (e.g., changes in oil prices) coming out after 1:00 a.m. EST and before 4:00 p.m. EST will influence today's interest rates in the US and tomorrow's interest rates in Japan.

Recall that the correlation between two assets is the ratio of their covariance divided by the product of their standard deviations

$$\begin{aligned}&\mathrm{corr}(\Delta i_{t,t+1}{}^{US}, \Delta i_{t,t+1}{}^{Jap})\\ &= \mathrm{cov}(\Delta i_{t,t+1}{}^{US}, \Delta i_{t,t+1}{}^{Jap})/\{\mathrm{STD}(\Delta i_{t,t+1}{}^{US})*\mathrm{STD}(\Delta i_{t,t+1}{}^{Jap})\}.\end{aligned}$$

Assume that the daily standard deviation is estimated correctly irrespective of the time zone. The volatility of close-to-close equites covers 24 hours in any time zone. However, the covariance term is *underestimated* due to the nonsynchronisity problem.

The problem may be less important for portfolios of few assets, but as the number of assets increases, the problem becomes more and more acute. Consider for example an equally weighted portfolio consisting of n assets, all of which have the same daily standard deviation, denoted σ and the same cross correlation, denoted ρ. The variance of the portfolio would be

$$\sigma_P^2 = (1/n)\sigma^2 + (1 - 1/n)\rho\sigma^2.$$

The first term is due to the own asset variances, and the second term is due to the cross covariance terms. For a large n, the volatility of the portfolio is $\rho\sigma^2$, which is the standard deviation of each asset scaled down by the correlation parameter. The bias in the covariance would translate one-for-one into a bias in the portfolio volatility.

For US and Japanese ten year zero coupon rate changes for example, this may result in an understatement of portfolio volatilities by up to 50 percent relative to their true volatility. For a global portfolio of long positions this will result in a severe understatement of the portfolio's risk. Illusionary diversification benefits will result in lower-than-true VaR estimates.

There are a number of solutions to the problem. One solution could be sampling both market open and market close quotes in order to make the data more synchronous. This is, however, costly because more data are required, quotes may not always be readily available and quotes may be imprecise. Moreover, this is an incomplete solution since some nonsynchronicity still remains. There are two other alternative avenues for amending the problem and correcting for the correlation in the covariance term. Both alternatives are simple and appealing from a theoretical and an empirical standpoint.

The first alternative is based on a natural extension of the random walk assumption. The random walk assumption assumes consecutive daily returns are independent. In line with the independence assumption, assume intraday independence – e.g., consecutive hourly returns – are independent. Assume further, for the purpose of demonstration, that the US rate is sampled without a lag, whereas the Japanese rate is sampled with some lag. That is, 4:00 p.m. EST is the "correct" time for accurate and up to the minute sampling, and hence a 1:00 a.m. EST. quote is stale. The true covariance is

$$
\begin{aligned}
&\mathrm{cov}^{tr}(\Delta i_{t,t+1}{}^{US}, \Delta i_{t,t-1}{}^{Jap}) \\
&= \mathrm{cov}^{obs}(\Delta i_{t,t+1}{}^{US}, \Delta i_{t,t+1}{}^{Jap}) + \mathrm{cov}^{obs}(\Delta i_{t,t+1}{}^{US}, \Delta i_{t+1,t+2}{}^{Jap}),
\end{aligned}
$$

a function of the contemporaneous observed covariance plus the covariance of today's US change with tomorrow's change in Japan.

The second alternative for measuring true covariance is based on another assumption in addition to the independence assumption; the assumption that the intensity of the information flow is constant intraday, and that the Japanese prices/rates are 15 hours behind US prices/rates. In this case

$$
\mathrm{cov}^{tr}(\Delta i_{t,t+1}{}^{US}, \Delta i_{t,t+1}{}^{Jap}) = [24/(24 - 15)]^{*}\mathrm{cov}^{obs}(\Delta i_{t,t+1}{}^{US}, \Delta i_{t,t+1}{}^{Jap}).
$$

The intuition behind the result is that we observe a covariance which is the result of a partial overlap, of only 9 out of 24 hours. If we believe the intensity of news throughout the 24 hour day is constant than we need to inflate the covariance by multiplying it by 24/9 = 2.66. This method may result in a peculiar outcome, that the correlation is greater than one, a result of the assumptions. This factor will transfer directly to the correlation parameter – the numerator of which increases by a factor of 2.66, while the denominator remains the same. The factor by which we need to inflate the covariance term falls as the level of nonsynchronicity declines. With London closing 6 hours prior to New York, the factor is smaller – 24/(24 – 6) = 1.33.

Both alternatives rely on the assumption of independence and simply extend it in a natural way from interday to intraday independence.

This concept is consistent, in spirit, with the kind of assumptions backing up most extant risk measurement engines. The first alternative relies only on independence, but requires the estimation of one additional covariance moment. The second alternative assumes in addition to independence that the intensity of news flow is constant throughout the trading day. Its advantage is that it requires no further estimation.[18]

2.8 SUMMARY

This chapter addressed the motivation for and practical difficulty in creating a dynamic risk measurement methodology to quantify VaR. The motivation for dynamic risk measurement is the recognition that risk varies through time in an economically meaningful and in a predictable manner. One of the many results of this intertemporal volatility in asset returns distributions is that the magnitude and likelihood of tail events changes though time. This is critical for the risk manager in determining prudent risk measures, position limits, and risk allocation.

Time variations are often exhibited in the form of fat tails in asset return distributions. One attempt is to incorporate the empirical observation of fat tails is to allow volatility to vary through time. Variations in volatility can create deviations from normality, but to the extent that we can measure and predict volatility through time we may be able to recapture normality in the conditional versions, i.e., we may be able to model asset returns as conditionally normal with time-varying distributions.

As it turns out, while indeed volatility is time-varying, it is not the case that extreme tails events disappear once we allow for volatility to vary through time. It is still the case that asset returns are, even conditionally, fat tailed. This is the key motivation behind extensions of standard VaR estimates obtained using historical data to incorporate scenario analysis and stress testing. This is the focus of the next chapter.

APPENDIX 2.1 BACKTESTING METHODOLOGY AND RESULTS

In Section 2.2, we discussed the MSE and regression methods for comparing standard deviation forecasts. Next, we present a more detailed discussion of the methodology for backtesting VaR methodologies. The dynamic VaR estimation algorithm provides an estimate of the x percent VaR for the sample period for each of the methods. Therefore, the probability of observing a return lower than the calculated VaR should be x percent:

$$\text{prob}[r_{t-1,t} < -VaR_t] = x\%.$$

There are a few attributes which are desirable for VaR_t. We can think of an indicator variable I_t, which is 1 if the VaR is exceeded, and 0 otherwise. There is no direct way to observe whether our VaR estimate is precise; however, a number of different indirect measurements will, together, create a picture of its precision.

The first desirable attribute is *unbiasedness*. Specifically, we require that the VaR estimate be the x percent tail. Put differently, we require that the average of the indicator variable I_t should be x percent:

$$\text{avg}[I_t] = x\%.$$

This attribute alone is an insufficient benchmark. To see this, consider the case of a VaR estimate which is constant through time, but is also highly precise unconditionally (i.e., achieves an average VaR probability which is close to x percent). To the extent that tail probability is cyclical, the occurrences of violations of the VaR estimate will be "bunched up" over a particular state of the economy. This is a very undesirable property, since we require dynamic updating which is sensitive to market conditions.

Consequently, the second attribute which we require of a VaR estimate is that extreme events do not "bunch up." Put differently, a VaR estimate should increase as the tail of the distribution rises. If a large return is observed today, the VaR should rise to make the probability of another tail event exactly x percent tomorrow. In terms of the indicator variable, I_t, we essentially require that I_t be independently and identically distributed (i.i.d.). This requirement is similar to saying that the VaR estimate should provide a filter to transform a serially dependent return volatility and tail probability into a serially independent I_t series.

The simplest way to assess the extent of independence here is to examine the empirical properties of the tail event occurrences, and compare them to the theoretical ones. Under the null that I_t is independent over time

$$\text{corr}[I_{t-s}*I_t] = 0 \quad \forall s,$$

that is, the indicator variable should not be autocorrelated at any lag. Since the tail probabilities that are of interest tend to be small, it is very difficult to make a distinction between pure luck and persistent error in the above test for any individual correlation. Consequently, we consider a joint test of whether the first five daily autocorrelations (one trading week) are equal to zero.

Note that for both measurements the desire is essentially to put all data periods on an equal footing in terms of the tail probability. As such, when we examine a number of data series for a given method, we can aggregate across data series, and provide an average estimate of the unbiasedness and the independence of the tail event probabilities. While the different data series may be correlated, such an aggregate improves our statistical power.

The third property which we examine is related to the first property – the biasedness of the VaR series, and the second property – the autocorrelation of tail events. We calculate a rolling measure of the absolute percentage error. Specifically, for any given period, we look forward 100 periods and ask how many tail events were realized. If the indicator variable is both unbiased and independent, this number is supposed to be the VaR's percentage level, namely x. We calculate the average absolute value of the difference between the actual number of tail events and the expected number across all 100-period windows within the sample. Smaller deviations from the expected value indicate better VaR measures.

The data we use include a number of series, chosen as a representative set of "interesting" economic series. These series are interesting since we *a priori* believe that their high order moments (skewness and kurtosis) and, in particular, their tail behavior, pose different degrees of challenge to VaR estimation. The data span the period from January 1, 1991 to May 12, 1997, and include data on the following:

- DEM the dollar/DM exchange rate;
- OIL the spot price for Brent crude oil;
- S&P the S&P 500 Index;
- BRD a general Brady bond index (JP Morgan Brady Broad Index).

We have 1,663 daily continuously compounded returns for each series.

In the tables, in addition to reporting summary statistics for the four series, we also analyze results for:

- EQW an equally weighted portfolio of the four return series
- AVG statistics for tail events averaged across the four series.

The EQW results will give us an idea of how the methods perform when tail events are somewhat diversified (via aggregation). The AVG portfolio simply helps us increase the effective size of our sample. That is, correlation aside, the AVG statistics may be viewed as using four times more data. Its statistics are therefore more reliable, and provide a more complete picture for general risk management purposes. Therefore, in what follows, we shall refer primarily to AVG statistics, which include 6,656 observations.

In the tables we use a 250-trading day window throughout. This is, of course, an arbitrary choice, which we make in order to keep the tables short and informative. The statistics for each of the series include 1,413 returns, since 250 observations are used as back data. The AVG statistics consist of 5,652 data points, with 282 tail events expected in the 5 percent tail, and 56.5 in the 1 percent tail.

In table 2.5 we document the percentage of tail events for the 5 percent and the 1 percent VaR. There is no apparent strong preference among the models for the 5 percent VaR. The realized average

Table 2.5 Comparison of methods – results for empirical tail probabilities

	Historical STD	Historical simulation	EXP		Hybrid	
			0.97	0.99	0.97	0.99
5% Tail						
DEM	5.18	5.32	5.74	5.18	5.25	5.04
OIL	5.18	4.96	5.60	5.39	5.18	5.18
S&P	4.26	5.46	4.68	4.18	6.17	5.46
BRD	4.11	5.32	4.47	4.40	5.96	5.46
EQW	4.40	4.96	5.04	4.26	5.67	5.39
AVG	**4.62**	**5.21**	**5.11**	**4.68**	**5.65**	**5.30**
1% Tail						
DEM	1.84	1.06	2.20	1.63	1.84	1.28
OIL	1.84	1.13	1.77	1.77	1.70	1.35
S&P	2.06	1.28	2.20	2.13	1.84	1.42
BRD	2.48	1.35	2.70	2.41	1.63	1.35
EQW	1.63	1.49	1.42	1.42	1.63	1.21
AVG	**1.97**	**1.26**	**2.06**	**1.87**	**1.73**	**1.32**

Table 2.6 Rolling mean absolute percentage error of VaR

	Historical STD	Historical simulation	EXP		Hybrid	
			0.97	0.99	0.97	0.99
5% Tail						
DEM	2.42	2.42	1.58	2.11	1.08	1.77
OIL	2.84	2.62	2.36	2.67	1.93	2.44
S&P	1.95	1.91	1.52	1.85	1.72	1.68
BRD	3.41	3.53	3.01	3.34	2.54	2.97
EQW	2.43	2.36	2.48	2.33	1.50	2.20
AVG	**2.61**	**2.57**	**2.19**	**2.46**	**1.76**	**2.21**
1% Tail						
DEM	1.29	0.87	1.50	1.12	1.02	0.88
OIL	1.71	0.96	1.07	1.39	0.84	0.80
S&P	1.45	1.14	1.40	1.42	0.99	0.82
BRD	2.15	1.32	1.98	2.06	1.03	1.12
EQW	1.57	1.52	1.25	1.25	0.72	0.87
AVG	**1.63**	**1.16**	**1.44**	**1.45**	**0.92**	**0.90**

varies across methods, between 4.62 percent and 5.65 percent.[19] A preference is observed, however, when examining the empirical performance for the 1 percent VaR across methods. That is, HS and Hybrid ($\lambda = 0.99$) appear to yield results that are closer to 1 percent than the other methods. Thus, the nonparametric methods, namely HS and Hybrid, appear to outperform the parametric methods for these data series, perhaps because nonparametric methods, by design, are better suited to addressing the well known tendency of financial return series to be fat tailed. Since the estimation of the 1 percent tail requires a lot of data, there seems to be an expected advantage to high smoothers ($\lambda = 0.99$) within the hybrid method.

In table 2.6 we document the mean absolute error (MAE) of the VaR series. The MAE is a conditional version of the previous statistic (percentage in the tail from table 2.4). The MAE uses a rolling 100-period window. Here again, we find an advantage in favor of the nonparametric methods, HS and Hybrid, with the hybrid method performing best for high λ ($\lambda = 0.99$) (note, though, that this is not always true: $\lambda = 0.97$ outperforms for the 5 percent for both the hybrid and the EXP). Since a statistical error is inherent in this statistic, we

Table 2.7 First-order autocorrelation of the tail events

	Historical STD	Historical simulation	EXP		Hybrid	
			0.97	0.99	0.97	0.99
5% Tail						
DEM	0.39	0.09	−2.11	−1.06	−2.63	−2.28
OIL	1.76	2.29	2.11	1.25	3.20	0.31
S&P	0.77	1.09	−0.15	0.94	0.77	2.46
BRD	11.89	12.69	13.60	12.27	10.12	12.08
EQW	5.52	2.29	3.59	4.26	−2.04	−0.14
AVG	**4.07**	**3.69**	**3.41**	**3.53**	**1.88**	**2.49**
1% Tail						
DEM	2.04	−1.08	1.05	2.76	−1.88	−1.29
OIL	−1.88	−1.15	2.27	2.27	−1.73	−1.37
S&P	4.94	9.96	7.65	8.04	2.04	8.70
BRD	15.03	9.30	10.75	12.60	−1.66	3.97
EQW	2.76	3.32	3.63	3.63	2.76	4.73
AVG	**4.58**	**4.07**	**5.07**	**5.86**	**−0.09**	**2.95**

cannot possibly expect a mean absolute error of zero. As such, the 38 percent improvement of the hybrid method with λ of 0.99 (with MAE of 0.90 percent for the AVG series' 1 percent tail) relative to the EXP method with the same λ (with MAE of 1.45), is an understatement of the level of improvement. A more detailed simulation exercise would be needed in order to determine how large this improvement is. It is worthwhile to note that this improvement is achieved very persistently across the different data series.

The adaptability of a VaR method is one of the most critical elements in determining the best way to measure VaR. When a large return is observed, the VaR level should increase. It should increase, however, in a way that will make the next tail event's probability precisely x percent. We can therefore expect these tail event realizations to be i.i.d. (independent) events with x percent probability. This independence can be examined using the autocorrelation of tail events, with the null being that autocorrelation is zero. As we see in table 2.7, the hybrid method's autocorrelation for the AVG series is closest to zero. Interestingly, this is especially true for the more fat-tailed series, such as BRD and OIL. As such, the hybrid method is very well suited for fat tailed, possibly skewed series.

Table 2.8(a) Test statistic for independence (autocorrelations 1–5)

	Historical STD	Historical simulation	EXP		Hybrid	
			0.97	0.99	0.97	0.99
5% Tail						
DEM	7.49	10.26	3.80	8.82	3.73	6.69
OIL	9.58	12.69	5.82	4.90	4.71	3.94
S&P	8.09	8.32	0.88	4.31	0.81	3.87
BRD	66.96	87.80	88.30	78.00	46.79	69.29
EQW	16.80	6.30	11.66	14.75	4.87	12.10
AVG	**21.78**	**25.07**	**22.09**	**22.16**	**12.18**	**19.18**
1% Tail						
DEM	3.34	5.33	4.56	4.39	7.58	3.83
OIL	33.98	8.29	3.82	18.89	8.53	3.54
S&P	14.67	36.15	22.68	25.18	3.26	24.10
BRD	88.09	29.37	41.60	82.77	11.26	11.36
EQW	41.55	14.69	16.85	16.85	5.08	13.05
AVG	**36.32**	**18.77**	**17.90**	**29.61**	**7.14**	**11.18**

Table 2.8(b) p-value for independence (autocorrelations 1–5)

	Historical STD	Historical simulation	EXP		Hybrid	
			0.97	0.99	0.97	0.99
5% Tail						
DEM	0.19	0.07	0.58	0.12	0.59	0.24
OIL	0.09	0.03	0.32	0.43	0.45	0.56
S&P	0.15	0.14	0.97	0.51	0.98	0.57
BRD	0.00	0.00	0.00	0.00	0.00	0.00
EQW	0.00	0.28	0.04	0.01	0.43	0.03
AVG	**0.09**	**0.10**	**0.38**	**0.21**	**0.49**	**0.28**
1% Tail						
DEM	0.65	0.38	0.47	0.49	0.18	0.57
OIL	0.00	0.14	0.58	0.00	0.13	0.62
S&P	0.01	0.00	0.00	0.00	0.66	0.00
BRD	0.00	0.00	0.00	0.00	0.05	0.04
EQW	0.00	0.01	0.00	0.00	0.41	0.02
AVG	**0.13**	**0.11**	**0.21**	**0.10**	**0.28**	**0.25**

In tables 2.8a and b we test the statistical significance of the autocorrelations in table 2.7. Specifically, we examine the first through fifth autocorrelations of the tail event series, with the null being that all of these autocorrelations should be zero. The test statistic is simply the sum of the squared autocorrelations, appropriately adjusted to the sample size. Under the null this statistic is distributed as $\chi^2(5)$. These test statistics are generally lower for the hybrid method relative to the EXP. For the specific series four rejections out of a possible eight are obtained with the hybrid method, relative to seven out of eight for the EXP method.

It is illegal to reproduce this material in any format without prior written approval of Blackwell Publishing Ltd.