In recent years, risk management has been of growing interest to institutional investors, including pension funds, insurance companies, endowments, and foundations as well as the asset management firms that manage funds on their behalf. Traditionally, institutional investors, and particularly pension funds, have emphasized measuring and rewarding investment performance by their portfolio managers. In the past decade, however, many U.S. pension funds have significantly increased the complexity of their portfolios by broadening the menu of acceptable investments. These investments can include foreign securities, commodities, futures, swaps, options, and collateralized mortgage obligations. At the same time, well-publicized losses among pension funds, hedge funds, and municipalities have underlined the importance of risk management and measuring performance on a risk-adjusted basis.

One approach to risk management, known as Value at Risk (or VaR), has gained increasing acceptance in the last five years. However, institutional investors’ quest for a VaR-based risk-management system has been hampered by several factors. One is a lack of generally accepted standards that would apply to them. Most work in the area of VaR-based risk measurement and standard-setting has been done at commercial and investment banks in conjunction with managing market risk. VaR originated on derivatives trading desks and then spread to other trading operations. The implementations of VaR developed at these institutions naturally reflected the needs and characteristics of their trading operations, such as very short time horizons, generally liquid securities, and market-neutral positions. In contrast, investment managers generally stay invested in the market, can have illiquid securities in their portfolios, and hold positions for a long time.
Moreover, many risk-management systems developed for trading operations are expensive to implement and beyond the budget and manpower of smaller pension funds. Nevertheless, recent developments in web-based technologies, which application service providers use to make risk measurement available to clients over the Internet, hold promise of bringing affordable risk management to the cross-section of smaller institutional investors. This makes it important to explore the practical issues institutional investors have to consider while implementing a VaR-based risk management system.

VaR is a measure of risk based on a probability of loss and a specific time horizon in which this loss can be expected to occur. Bank regulators use VaR to set capital requirements for bank trading accounts because VaR models can be used to estimate the loss of capital due to market risk. Pension plans are generally concerned not with the loss of capital, but with underperforming their benchmarks. Pension plans distinguish between a long-term or strategic asset allocation, also known as the “policy portfolio,” and a short-term or tactical asset allocation. The policy portfolio is typically aimed to match the plan’s liabilities. The actual portfolio, which represents the tactical asset allocation, can differ from the policy portfolio because fund managers implement market views with the goal of outperforming the policy portfolio. Thus, the policy portfolio represents the benchmark against which the actual portfolio performance is measured. Because performance is measured against the benchmark, the risk should be measured the same way. At the same time, for defined-benefit plans, VaR can represent the risk that assets fall below a certain target, in particular the risk that assets would be insufficient to fund the benefits due employees.

VaR has advantages as a risk measure for institutional investors. Specifically, it is based on the current portfolio composition rather than the historical return on the portfolio, and it can be aggregated across many asset classes. The more traditional risk measures used in investment management have one of these characteristics, but not both. For example, tracking error is a measure of the deviation of the portfolio’s historical return from the return on the benchmark index. It may not be useful if the current composition of the portfolio differs from the one that produced these historical returns. On the other hand, two traditional asset-specific measures, beta for stocks and duration for bonds, are based on the current portfolio composition. Beta measures the portfolio’s systematic risk, that is, the degree to which its return is correlated with the return on the market as a whole. Duration measures the sensitivity of a bond portfolio to changes in interest rates. The higher the duration, the more sensitive it is to changes in interest rates. These measures, while useful, cannot be combined to provide an overall measure of risk.

Thus, VaR is particularly useful to a pension plan sponsor that has a multi-asset-class portfolio and needs to measure its exposure to a variety of risk factors. VaR can measure the risk of stocks and bonds, commodities, foreign exchange, and structured products such as asset-backed securities and collateralized mortgage obligations (CMOs), as well as off-balance-sheet derivatives such as futures, forwards, swaps, and options. VaR is useful to plan sponsors who have their portfolios managed by a variety of external asset managers and need to compare their performance on a risk-adjusted basis.

A survey of major pension fund sponsors and several asset management firms by one consulting firm (Kerrigan 1999) found that the demand for portfolio managers to produce VaR reports comes both from the senior management of their firms and from clients. Sometimes clients specify the confidence interval and time horizons used to calculate VaR. Portfolio managers report both absolute and relative VaR measures. VaR does not replace tracking error but is used along with it. The survey reports that institutional investors mostly use parametric VaR, unless they have options in the portfolio, in which case they also use simulations.

This article is organized as follows: Section I describes one proposed set of risk management standards. Section II introduces the concept of Value at Risk and describes the parametric VaR, the most common method of its calculation. Section III compares VaR to tracking error, a common measure of risk employed by institutional investors, showing that tracking error can
be seen as a special case of Value at Risk. Section IV discusses the issues surrounding measures of risk-adjusted performance. Section V describes the major difficulties institutional investors may encounter when implementing VaR analysis. Section VI discusses some public policy implications of widespread VaR adoption.

---

VaR is particularly useful to a pension plan sponsor that has a multi-asset-class portfolio and needs to measure its exposure to a variety of risk factors.

---

I. Elements of Risk Management Standards for Institutional Investors

While no generally accepted standards exist for risk management and measurement for institutional investors, one major study, by the Risk Standards Working Group (1996), addressed many of the issues in general terms. The study formulated 20 risk standards, grouped into three categories of management, measurement, and oversight. A summary of the Working Group’s guidelines that are related specifically to the measurement of risk follows:

**Risk Measurement Guidelines of the Risk Standards Working Group**

1. Valuation procedures

   Readily priced instruments such as publicly traded securities, exchange-listed futures and options, and many over-the-counter derivatives should be priced daily.

   Less readily priced instruments such as complex CMOs, exotic derivatives, and private placement notes should be priced as often as possible and at least weekly. For such instruments, the model and price mechanism must be made explicit so that they can be independently verified.

   Non-readily priced assets such as real estate and private equity stakes should be valued as frequently as is feasible and whenever a material event occurs. For such instruments, the valuation method (such as theoretical model, appraisal, committee estimate, or single-dealer quote) should be made explicit to facilitate independent evaluation.

2. Valuation reconciliation, bid-offer adjustments, and overrides

   Material discrepancies from different sources, such as managers and custodians, should be reconciled following established procedures at least monthly, or more frequently if material difference occurs.

3. Risk measurement and risk/return attribution analysis

   Risk should be measured in the overall portfolios, individual portfolios, and each instrument. Return attribution analysis should be performed to determine the key historical drivers of returns on the portfolios. Risk attribution analysis should also be performed to determine the key sources of volatility of returns in the current or anticipated portfolio. For example, a risk attribution analysis of a U.S. bond portfolio might quantify duration, yield curve, convexity, and sector risk in absolute terms or relative to a benchmark. A risk attribution analysis of a U.S. equity portfolio might use a risk-factor model to quantify the various sources of absolute and benchmark-relative risk.

4. Risk-adjusted return measures

   Investors should compare all managers on a risk-adjusted basis. By taking into account both risk and return, they will be able to better evaluate performance of two managers. Risk-adjusted measures also highlight instances in which a manager’s outperformance is the result of incurring misunderstood, mispriced, unintended, or undisclosed risks.

5. Stress testing

   Stress testing should be performed to ascertain how the aggregate portfolio and individual portfolios would behave under various market conditions. These include changes in key risk factors, correlations, and large market moves. Stress testing should be per-
formed at least quarterly or whenever significant changes occur in market conditions or in the composition of the portfolio.

Stress tests should take into account all types of leverage and related cash flows, including such items as repurchase agreements, options, structured notes, and high-beta stocks, as well as instruments requiring initial and valuation margin requirements.

6. Back testing

Investors should back test all models and forecasts of expected risk, return, and correlations for instruments, asset classes, and strategies. Back tests evaluate how a model actually performed for a given period versus what was predicted.

7. Assessing model risk

Dependence on models and assumptions for valuation, risk measurement, and risk management should be evaluated and monitored. Important dimensions of model risk to analyze include the following:

- Data integrity (for example, curve construction, differing sources of data, representativeness and statistical significance of samples, the time of day data are extracted, data availability, and errors)
- Definition and certainty of future cash flows (formula-driven cash flows or flows that depend on an option)
- Formula or algorithm (Black–Scholes versus Hull and White for options valuation)
- Liquidity assumptions (length of time to liquidate and bid-ask spreads)
- Model parameter selection (choice of spreads, discount rates, scenario and stress-test parameters, probability intervals, time horizon, correlation assumptions).

II. Calculating Value at Risk

VaR answers the question, “Over a given period of time with a given probability, how much could the value of the portfolio decline?” If VaR equals a thousand dollars, and the probability is 1 percent, then one can say that the chance of losing one thousand dollars over the holding period is 1 in 100. One advantage of VaR is that it is an intuitively appealing measure of risk that can be easily conveyed to the firm’s senior management.

The three main methods of calculating VaR are the parametric (or analytic, or variance-covariance) method, the historical method, and the Monte Carlo simulation. Detailed descriptions of these methods can be found in RiskMetrics Technical Document (Longerstaey and Zangari 1996), Simons (1997), and Duffie and Pan (1997). Briefly, parametric VaR assumes that the returns on the portfolio can be approximated by a normal distribution, and it draws on the properties of that distribution to calculate the probability of loss. Thus, it conveys the same information as the standard deviation, but on a different scale. Among the relevant properties of the normal distribution is that 67 percent of returns will fall within one standard deviation around the mean, while 33 percent will lie outside it. Since normal distributions are symmetric and we are concerned only with the loss (the left tail of the distribution), losses in excess of one standard deviation will occur 16.5 percent of the time. One minus the probability is referred to as the confidence level. Table 1 summarizes some common confidence levels that can be used for calculating parametric VaR.

| Time Horizon |

The time horizon used to calculate VaR should depend on the liquidity of the securities in the portfolio and how frequently they are traded. Less liquid securities call for a longer time horizon. The most common time horizons used by commercial and investment banks to calculate VaRs of their trading rooms are one day, one week, and two weeks. The Basle Committee on Banking Supervision mandates that banks using VaR models to set aside capital for market risk of their trading operations use a holding period of two weeks and a confidence level of 99 percent. In con-
trast, institutional investors have long holding periods for investments, ranging from one month to as long as five years.

Long time horizons complicate VaR modeling, because the use of the daily data to estimate volatilities and correlations among assets may not be valid over these long time horizons. Moreover, a VaR estimate for a given time interval implies that the investor cannot or will not trade out of the position during this time. If “mid-course” corrections are possible, the VaR can overstate the probable losses when the investor takes conservative action. Also, using derivatives to hedge the portfolio, such as purchasing put options and other “portfolio insurance” techniques, complicates VaR calculations.

III. Comparing Value at Risk and Tracking Error

Traditionally, portfolio managers and institutional investors measure both risk and return relative to a benchmark. The commonly used benchmarks for measuring stock returns are the S&P 500 for stocks in general and large capitalization funds, the Wilshire 5000 and the Russell 3000 for the U.S. market in general, the Russell 2000 for small stocks, and the Morgan Stanley EAFE for international portfolios.

Tracking error is a measure of risk based on the standard deviation of portfolio returns relative to the chosen benchmark return. It is defined as the standard deviation of the excess return, that is, the difference between the return on a portfolio and the return on its benchmark. Unlike VaR, which is usually measured for shorter periods, tracking error is typically measured in terms of monthly returns. However, returns can be measured over a period of any length.

\[
ER_t = R_p - R_b
\]

\[
TE = \sqrt{\frac{1}{T^2} \sum (ER_t - \overline{ER})^2}
\]

In equation (1), \(ER_t\) is the excess return of the portfolio over the benchmark return in period \(t\), \(\overline{ER}\) is the average excess return, \(TE\) is the tracking error, and \(T\) is the number of periods over which the tracking error is being calculated.

Unlike tracking error, which is measured in percent relative to the benchmark, VaR is usually measured as a dollar amount of loss that can occur with a given probability. However, it is possible to calculate “tracking VaR,” which is also measured relative to the benchmark. One can think of tracking VaR as measuring a loss in a hypothetical portfolio consisting of a long position in the actual portfolio being measured and a short position in its benchmark. Tracking VaR is usually expressed in terms of return, rather than an absolute amount of money the portfolio may lose.

**Tracking VaR is usually expressed in terms of return, rather than an absolute amount of money the portfolio may lose.**

Thus, the tracking error can be seen as a special case of tracking VaR where the confidence level and holding period are fixed—at 83.5 percent and one month, respectively. So, a tracking error of \(X\) percent means that a monthly underperformance greater than \(X\) percent relative to the benchmark can be expected to occur 16.5 percent of the time, or once every six months.

**Instrument Mapping**

For large investors with portfolios containing many securities, VaR calculations require vast quantities of data to construct the variance-covariance matrix of their returns. Thus, most users need some way of mapping of instruments into a smaller number of standard equivalents (commonly known as risk factors) for which data are available. For example, a very popular data set of volatilities and correlations is J. P. Morgan’s RiskMetrics, which is free to the public, downloadable from the Internet, and updated daily. The data include a number of major currencies, interest rates, commodities, and equity indexes for major markets and countries.

**An Example of VaR and Tracking Error Calculation**

We will use an example of a U.S. equity mutual fund and find its parametric and historical VaRs and its tracking error. We first calculate the fund’s daily return for 360 trading days as well as its daily excess return over the S&P 500, which is the customary benchmark for U.S. equity funds. The daily returns
used were for the period between 5/11/1999 and 10/10/2000. Table 2 shows the fund’s actual daily returns for selected days in column 2, the daily returns on the S&P 500 in column 3, and the excess return, which is the difference between them, in column 4. Average daily returns and standard deviations over the entire period are shown at the bottom of the table.

We can see from the last row in column 4 that the fund’s daily tracking error, that is, the standard deviation of its excess return, is 0.98 percent. The fund’s daily parametric VaR at the 99th percent confidence level is the standard deviation of its returns (which is equal to 1.24) multiplied by 2.33 (see Table 1), or 2.89 percent. This means, roughly speaking, that a negative daily return of 2.89 percent is expected to occur about 1 out of a 100 trading days.

One often wishes to calculate VaR for periods longer than one day, since it may not be possible to close a position in one day, especially if it is illiquid. If, in addition to normality, we assume that returns are serially independent, then the standard deviation of longer-period returns increases with the square root of time. A one-month (24 trading days) VaR is the daily VaR times \( \sqrt{24} \) (= 4.9). Thus, if the returns can be approximated by the normal distribution, then VaR is simply a linear function of the standard deviation.

Selected values of absolute and tracking VaR (defined at the beginning of Section III) for one day and one month are shown in Table 3. They are calculated by multiplying the standard deviation of actual daily returns (1.24 for absolute VaR) or the standard deviation of excess return (0.98 for tracking VaR) by the appropriate numbers of standard deviations for the given confidence interval from Table 1, and then multiplying by \( \sqrt{24} \) in case of the monthly holding period. Note that the shaded cell, representing tracking VaR at the 83.5 percent confidence level, is also the conventional tracking error, or the monthly standard deviation of the fund’s excess returns over its benchmark, the S&P 500.

### IV. Risk-Adjusted Performance and Tracking Error

Both tracking error and tracking VaR show only how closely the returns on a given portfolio track the benchmark; they say nothing about performance. In fact, it is possible to underperform the benchmark quite dramatically while having a low tracking error or tracking VaR. This can be a serious weakness of the tracking error as a risk measure, since most portfolio managers would consider underperforming the benchmark to be perhaps their most significant risk.

Figure 1 illustrates how very different performance results can be associated with a similar tracking error. The figure shows daily values of $1 invested in three hypothetical portfolios: Fund A, Fund B, and a benchmark portfolio. These are simulated results that were produced by adding a random component drawn from a normal distribution to a different predetermined growth trend. The trend was 0.001 per day for Fund A, 0 for Fund B, and 0.0005 for the benchmark portfolio. As can be seen from the graph, over the course of one year (250 trading days), Fund A outperformed the benchmark by 15 percentage points, while Fund B underperformed it by 11. These large differences in performance occurred despite the fact that
both funds managed to have an annualized tracking error of around 3 percent. This example shows that systematic trends in returns can have a powerful cumulative effect over the long-term investment time scales, even if the period-by-period tracking error is low.

This example also illustrates a serious difficulty with using simple measures based on standard deviations over long time horizons. Standard VaR methods usually assume that the expected return on the portfolio is zero, or, at most, the risk-free rate. This is because trading portfolios are assumed to be market-neutral or held for such a short time that the expected return can be ignored. For measuring a long-term absolute VaR of an investment portfolio, it can make sense to incorporate an estimate of expected return on the asset. However, doing so is problematic for tracking error or tracking VaR that measures underperformance relative to the benchmark, because there really is no such thing as “expected” underperformance. If the manager of our hypothetical Fund B had known that the portfolio would underperform the benchmark by 11 percent, he would not have put on these positions in the first place! On the other hand, it can be quite tempting for a portfolio manager to incorporate an expected outperformance relative to the benchmark into VaR. However, the majority of active managers regularly underperform their benchmarks, so doing so may be an unwarranted underestimation of risk.

Despite the fact that similar tracking errors or tracking VaRs can accompany large differences in returns, they can provide important information for adjusting performance for risk. This can be useful to a pension plan sponsor that is choosing between two or more funds representing the same asset class for inclusion in the plan. The plan sponsor could simply choose the fund that had the highest return relative to the benchmark. However, the fund may have achieved its high return by taking higher risk, not through any particular skill of its manager. Therefore, the plan sponsor may wish to make the selection on the basis of a risk-adjusted measure. Tracking error can be used for this purpose, to calculate the measure of risk-adjusted performance known as the Sharpe ratio (Sharpe 1966). Using the same notation as in equation 1, the Sharpe ratio can be expressed as follows:

\[
\text{Sharpe Ratio} = \frac{\text{ER}}{\text{TE}}.
\]
The investment with the higher Sharpe ratio is preferable because it provides a higher return per unit of risk. Several points should be made about the Sharpe ratio. The same benchmark must be used for all the portfolios being compared; otherwise the comparison will be misleading. In comparing funds that would normally use different benchmarks, such as a bond fund and a stock fund, the Sharpe ratio would typically use a sort of “universal benchmark,” namely a return on a risk-free investment such as Treasury bills. In this case, the Sharpe ratio is calculated as the expected excess return over the risk-free rate divided by the standard deviation of the excess return. Lastly, unless the returns of the funds are perfectly correlated, one can usually achieve a higher risk-adjusted return through a combination of available funds. However, a plan sponsor can only have a limited number of funds in the plan, so in practice the sponsor will often have to forgo making such a combination. In this case, however, the optimal combination will still have a higher Sharpe ratio than any of the individual funds, so the principle of choosing the higher Sharpe ratio still holds.

V. Difficulties with VAR Implementations

One of the most serious and well-known shortcomings of parametric VaR is that it underestimates the frequency of “extreme events,” such as outcomes several standard deviations away from the mean. This is because asset return distributions exhibit “fat tails,” meaning that more of the outcomes are located in the tails rather than toward the center of the distribution.

Using the historical returns for the fund in the example above, we can compare the probable losses implied by the normal distribution to the actual size and frequency of losses that did, in fact, occur. The frequency distribution of the fund’s returns is pictured in Figure 2 with the normal curve superimposed on it. We see that the 95th percentile loss was 1.72 percent, better than the 2.05 percent daily VaR implied by the normal distribution. For the 99th percentile, the actual loss was 2.61 percent, also better than the 99th confidence level VaR of 2.89 percent. So far, parametric VaR seems to hold up well in this example, being more conservative than the actual outcomes. However, this impression is misleading. The normal distribution assigns virtually zero probability to events that are greater than 3 standard deviations. In fact, three events in our time series represented losses that were 4.5 standard deviations away from the mean, including a 12 percent loss which is a 9-standard-deviations event. Since the main objective of risk management models is to measure losses in the tails, this is a serious shortcoming.

The degree of “fat-tailness” of a distribution can be measured by kurtosis, which is defined as the fourth moment of the distribution (that is, the mean to the power of four) divided by the square of the variance. Thus, if \( r_i \) is the return on day \( i \) and \( \sigma^2 \) is the variance, kurtosis is defined as follows:

\[
\kappa = \frac{\sum r_i^4}{n\sigma^4}.
\]

The normal distribution has a kurtosis of 3. Any distribution that has a kurtosis greater than 3 is said to be leptokurtotic, that is, it has a lower central “hump” and fatter “tails” than the normal distribution. Our sample of mutual fund returns has a kurtosis of 25, indicating a rather high degree of leptokurtosis.

The fat tails of asset return distributions have elicited alternative approaches to parametric VaR. One is to simply use the percentiles of the actual historical returns on the portfolio to calculate VaR, the way we have done in this example. While this approach completely avoids the issue of choosing a distribution of asset returns, its applicability in practice is limited. In our example, we have used the returns on the actual portfolio for simplicity. In practice, one is generally interested in predicting the variance of the portfolio on the basis of its current composition, not its own historical returns, which may not reflect its current composition. Thus, one is interested in calculating VaR on the basis of the prospective variances and covariances of the instruments that are currently included in the portfolio, if these are available. (Even then, the variances and covariances of the risk factors may not stay constant for long because of structural shifts in the market, changes in fiscal or monetary policy, tax treatment...
of various assets, and other changes.) Unfortunately, this information often can be difficult or even impossible to assess. This is the reason why parametric VaR is often used instead of historical VaR.

One of the most serious and well-known shortcomings of parametric VaR is that it underestimates the frequency of “extreme events,” such as outcomes several standard deviations away from the mean.

To correct for the weaknesses of parametric VaR, or to test the consequences of changing composition of the portfolio, one can employ stress tests and scenario simulations. Such simulations can be useful to test the portfolio for hypothetical future events, such as increases in oil prices or an inflation surprise, and for the extreme effects of financial crises that occurred in the past, such as the culmination of the Asian crisis and Russian crisis in August 1998. It is not always clear, however, which scenarios should be tested for particular portfolios and how to interpret the results, since the probabilities that these scenarios will actually occur are unknown. It is also possible to model fat-tailed distributions explicitly. Approaches range from using a mixture of normals approach to stochastic volatility (see Simons 1997).

It should be noted that coping with fat-tail distributions is not unique to institutional investors. The problem of fat-tailed returns is a major issue of VaR modeling that banks’ trading desks have been struggling with for years. It can even be argued that the longer-term perspective of institutional investors makes these short-term market swings less important to them than to investment and commercial banks, to the extent that asset prices exhibit mean reversion.
VI. Policy Implications

VaR has become an accepted standard in the banking industry and it forms the basis of bank capital requirements for market risk. VaR adoption has been slower in the investment management industry, but as demand grows and consensus about the standards emerges, its use can be expected to accelerate. This will be a mixed blessing, as VaR has a number of serious limitations. It is based on volatilities and correlations that can work in normal market conditions but break down in times of market crises. Factors that exhibit low levels of correlation during normal market conditions can become highly correlated at times of high volatility. In such cases, the value of diversification across markets can be greatly reduced. Thus, VaR can understate potential losses during market turbulence and instill a false sense of security. Nevertheless, VaR can be useful for those organizations that understand its limitations and use stress testing to gauge their vulnerabilities to “tail events.”

There is another, more subtle, risk to the widespread adoption of VaR. During periods of turmoil, inefficient or illiquid markets could be destabilized if many market participants have rigid rules about exceeding VaR limits. That is because the VaR of a portfolio can change drastically as a result of changing market conditions, even if portfolio compositions do not change. If managers are given a mandate to keep VaR below a certain level, they will have no choice but to sell instruments causing high VaR at the moment. Used in this way, VaR can be seen as a type of dynamic hedging. As it gains worldwide acceptance, this type of VaR management has the potential to drive down asset prices and increase volatilities in thin or illiquid markets if enough market participants act together because they have similar positions and use similar models.

References
