21 Measuring Trading Performance

For every complex problem there is a solution that is simple, neat and wrong.

—H. L. Mencken

THE NEED TO NORMALIZE GAIN[1]

Too many investors make the mistake of focusing solely on return when they evaluate money managers.[2] It is critical to also incorporate some measure of risk as part of the evaluation process.

Consider the equity streams of the accounts of Manager A and Manager B in Figure 21.1.[3] Although Manager A produces the larger return for the period as a whole, he can hardly be considered the superior performer—note the many sharp retracements in equity.

This is not a negative feature merely because investors with Manager A will have to ride out many distressing periods. Even more critical is the consideration that investors who start with Manager A at the wrong time—and that is not hard to do—will actually have significant losses. In fact, assuming that accounts will be closed once 25–50 percent of the initial equity is lost, there is a significant chance investors with Manager A will be knocked out of the game before the next rebound in performance.

It seems reasonable to assume that most investors would prefer Manager B to Manager A because the modestly lower return of Manager B is more than compensated by the apparent much lower risk. Moreover, if Manager B had used a modestly higher margin-equity ratio, she could have exceeded

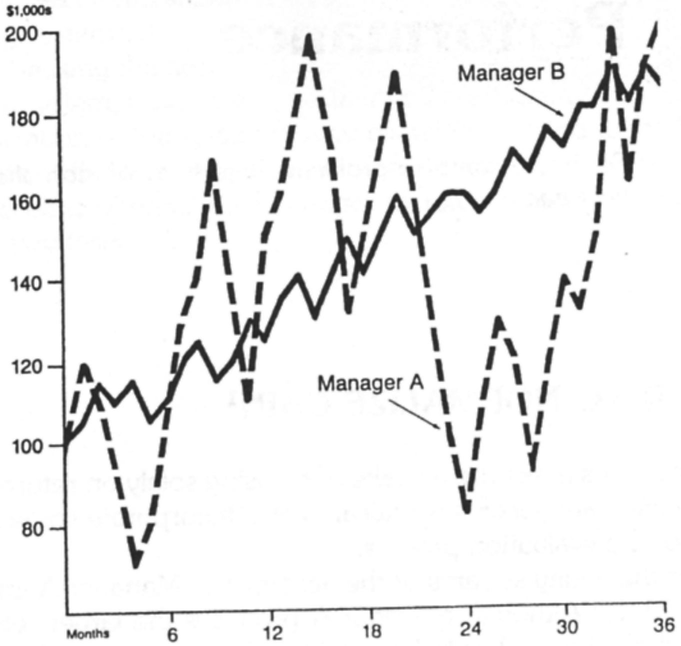
---

[1]The following section is adapted from J. Schwager, "Alternative to Sharpe Ratio Better Measure of Performance," *Futures,* pp. 56–57, March 1985.

[2]In the futures industry, most money managers (i.e., those registered with the Commodity Futures Trading Commission) are called "commodity trading advisors" (CTAs), an unfortunately inappropriate choice of names. In this chapter, the more generic term "money managers" is used and can be read as interchangeable with CTAs.

[3]Although the examples of this chapter are based on evaluating money manager performance, similar examples would also apply to trading systems. Distinctions between money manager and system performance evaluation are noted where appropriate.

Figure 21.1
THE NEED TO NORMALIZE GAIN



Source: J. Schwager, "Alternative to Sharpe Ratio Better Measure of Performance," *Futures,* p. 56, March 1985.

Manager A's return while still having much smaller retracements. (For money management reasons, all managers will limit the number of positions so that total margin requirements are well below total available equity; typically, the margin–equity ratio will be approximately 0.15–0.35.)

Clearly, Manager B has the better performance record. As illustrated by this example, any performance evaluation method must incorporate a risk measure to be meaningful.

THE SHARPE RATIO

The need to incorporate risk in evaluating performance has long been recognized. The classic return–risk measure is the Sharpe Ratio, which can be expressed as follows:

$$SR = \frac{E - I}{\text{sd}}$$

$SR = \frac{E - I}{sd}$

where E = the expected return
I = risk-free interest rate
sd = standard deviation of returns

E is typically stated in terms of percent return. Normally, the expected return is assumed to equal the average past return. In view of this fact, although E always refers to the expected return (i.e., applies to a future period), we will use it synonymously with the average past return.

The incorporation of I in the Sharpe Ratio recognizes that an investor could always earn a certain return *risk free*—for example, by investing in T bills. Thus, the return in excess of this risk-free return is more meaningful than the absolute level of the return.

The standard deviation is a statistic that is intended to measure the degree of dispersion in the data. The formula for the standard deviation is:

$$\text{sd} = \sqrt{\frac{\sum_{i=1}^{N}(X_i - \overline{X})^2}{N-1}}$$

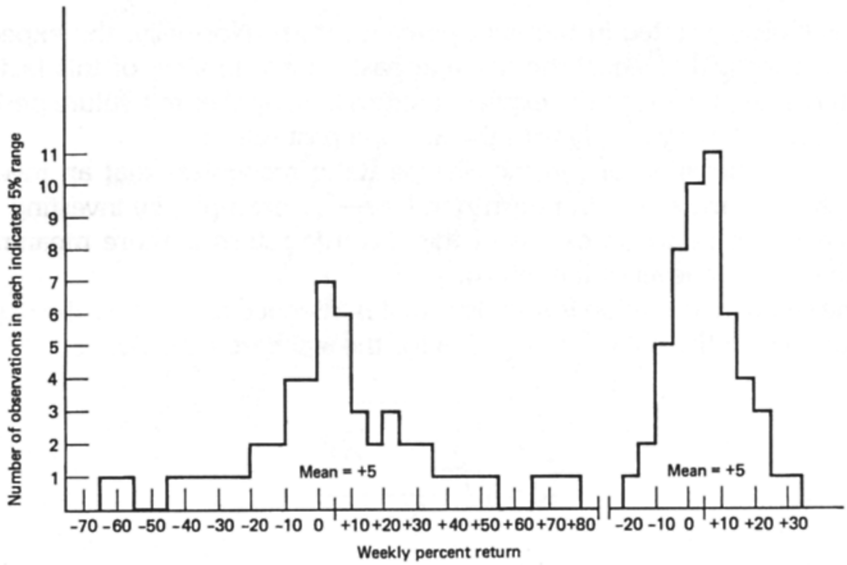
where $\overline{X}$ = mean
X_i = individual data values
N = the number of data values

In the Sharpe Ratio application, N is equal to the number of time intervals. For example, if monthly time intervals are used for a three-year survey period, $N = 36$.

In calculating the standard deviation, it is always necessary to choose a time interval for segmenting the total period equity data (e.g., weekly, monthly). If, for example, the percent return data for a given year were broken down into weekly figures, the standard deviation would be very high if the return of many of the individual weeks deviated sharply from the average for the period. Conversely, the standard deviation would be low if the individual weeks tended to cluster around the average. Figure 21.2 illustrates two sets of data with the same average weekly return but substantially different standard deviations.

The basic premise of the Sharpe Ratio is that the standard deviation is a measure of risk. That is, the more widespread the individual returns from the average return, the riskier the investment. In essence, the standard deviation measures the ambiguity of the return. It should be intuitively clear that if the standard deviation is low, it is reasonable to assume that the actual return will be close to the expected return (assuming, of course, that the expected return is a good indicator of actual return). On the other hand, if the standard deviation is high, it suggests that there is a good chance that the actual return may vary substantially from the expected return.

Figure 21.2
COMPARISON OF TWO MANAGERS WITH EQUAL AVERAGE RETURNS BUT DIFFERENT STANDARD DEVIATIONS



The Sharpe Ratio can be calculated rather directly for a money manager because we know the amount of funds on which percent return is based. This is not the case for a trading system. In applying the Sharpe Ratio to a trading system, we have one of two options:

1. Estimate the funds required to trade the system and use this figure to calculate a percent return.
2. Simplify the Sharpe Ratio by deleting the risk-free return *I*. (As is explained below, if this form of the Sharpe Ratio is used, it is not necessary to estimate the funds required to trade the system.) Thus, the Sharpe Ratio would reduce to

$$SR = \frac{E}{\text{sd}}$$

The second approach can be justified on the basis that, except for small accounts, the bulk of commodity margin requirements can be met by T-bill deposits. Thus, in contrast to the buyer of securities, the commodity trader does not sacrifice the risk-free return in order to participate in the alternative investment. The reduced form of the Sharpe Ratio also has a theoretical justification in the case of money managers: Whereas the Sharpe Ratio will increase if a manager increases her leverage—an undesirable feature—the reduced form will be unaffected by changes in leverage.

In the form E/sd, the Sharpe Ratio would be the same whether E were expressed in terms of dollar gain or in percent return. The reason for this is that the same unit of measurement would be used for the standard deviation. Thus, the funds requirement figure would appear in both the numerator and denominator and would cancel out.[4] To help clarify the exposition, the examples provided in the remainder of this chapter assume the reduced form of the Sharpe Ratio. This simplifying assumption does not meaningfully alter any of the theoretical or practical points discussed.

THREE PROBLEMS WITH THE SHARPE RATIO

Although the Sharpe Ratio is a useful measurement, it does have a number of potential drawbacks:[5]

1. *The Gain Measure of the Sharpe Ratio.* This measure—the annualized average monthly (or other interval) return is more attuned to assessing the probable performance for the next interval than the performance for an extended period. For example, assume that a fund manager has six months of 40 percent gains and six months of 30 percent losses in a given year. The annualized average monthly return would be 60 percent ($12 \times 5\%$). However, if position size is adjusted to existing equity, as is done by most managers, the actual return for the year would be −11 percent. That is because, for each dollar of equity at the start of the period, only \$0.8858 would remain at the end of the period—$(1.40)^6 \times (0.70)^6 = 0.8858$.

As this example illustrates, if you are concerned about measuring the potential performance for an extended period rather than just the following month or other interval, then the gain measure used in the Sharpe Ratio can lead to extreme distortions. This problem can be circumvented, however, by using an annualized geometric (as opposed to arithmetic) mean rate of return for the numerator of the Sharpe Ratio, as is frequently done. The annualized geometric return is precisely equivalent to the average annual compounded return, which is discussed in the section on the return retracement ratio later in this chapter.

---

[4]The implicit assumption here is that trading funds are constant (i.e., profits are withdrawn and losses replenished). In other words, there is no compounding (i.e., reinvestment of gains, reduction of investment in the event of losses). Generally speaking, although a compounded return calculation is preferable, this consideration is more than offset by the critical advantage of not having to estimate fund requirements for a trading system. Furthermore, in comparing two systems, the system with the higher noncompounded return will often exhibit the higher compounded return.

[5]This section is adapted from J. Schwager, "Alternative to Sharpe Ratio Better Measure of Performance," *Futures,* pp. 57–58, March 1985.

2. The Sharpe Ratio Does Not Distinguish between Upside and Downside Fluctuations. The Sharpe Ratio is a measure of volatility, not risk. The two are not necessarily synonymous.

In terms of the risk calculation employed in the Sharpe Ratio—that is, the standard deviation of return—upside and downside fluctuations are considered equally bad. Thus, the Sharpe Ratio would penalize a manager who had sporadic sharp increases in equity, even if the equity retracements were small.

Figure 21.3 compares the hypothetical equity streams of Manager C, who has intermittent surges in equity and no equity retracements, and Manager D, who experiences several equity retracements. Although both managers realize equal gains for the period as a whole and Manager D goes through several retracements while Manager C doesn't have any, the Sharpe Ratio would rate Manager D higher (see Table 21.1). This outcome is a direct consequence of the fact that the Sharpe Ratio penalizes upside volatility exactly the same as downside volatility.

3. The Sharpe Ratio Does Not Distinguish between Intermittent Losses and Consecutive Losses. The risk measure in the Sharpe Ratio (the standard deviation) is independent of the order of various data points.

Figure 21.4 depicts the hypothetical equity streams of $100,000 accounts handled by Manager E and Manager F. Each earns a total of $48,000 or $24,000 per year. However, Manager E alternates $8,000 monthly gains with

Figure 21.3
COMPARISON OF MANAGER WITH LARGE UPSIDE VOLATILITY AND NO RETRACEMENTS TO A MANAGER WITH RETRACEMENTS

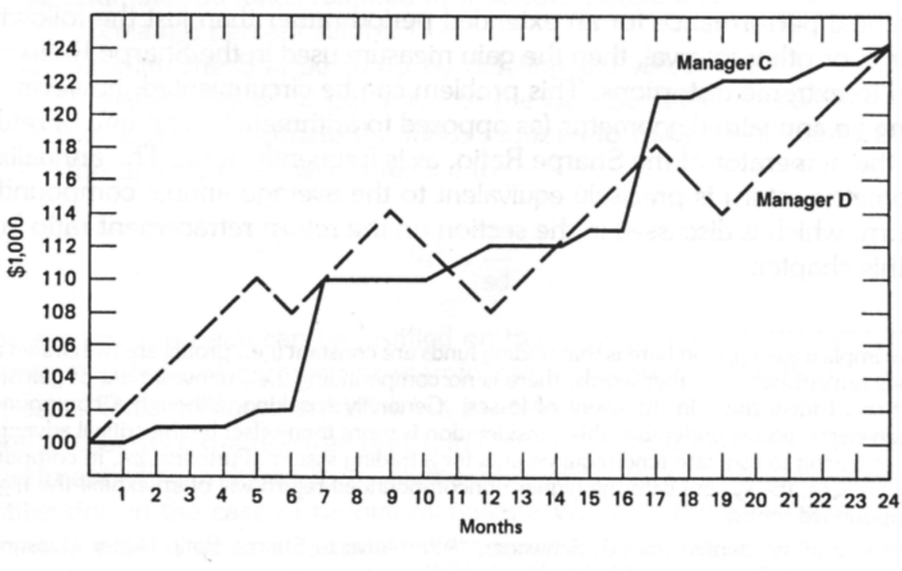


Table 21.1
COMPARISON OF MONTHLY RETURNS FOR TWO MANAGERS

	Manager C		Manager D	
Month	Equity Change	Cumulative Equity Change	Equity Change	Cumulative Equity Change
1	0	0	2,000	2,000
2	1,000	1,000	2,000	4,000
3	0	1,000	2,000	6,000
4	0	1,000	2,000	8,000
5	1,000	2,000	2,000	10,000
6	0	2,000	–2,000	8,000
7	8,000	10,000	2,000	10,000
8	0	10,000	2,000	12,000
9	0	10,000	2,000	14,000
10	0	10,000	–2,000	12,000
11	1,000	11,000	–2,000	10,000
12	1,000	12,000	–2,000	8,000
13	0	12,000	2,000	10,000
14	0	12,000	2,000	12,000
15	1,000	13,000	2,000	14,000
16	0	13,000	2,000	16,000
17	8,000	21,000	2,000	18,000
18	0	21,000	–2,000	16,000
19	1,000	22,000	–2,000	14,000
20	0	22,000	2,000	16,000
21	0	22,000	2,000	18,000
22	1,000	23,000	2,000	20,000
23	0	23,000	2,000	22,000
24	1,000	24,000	2,000	24,000
	Average monthly return = 1,000		Average monthly return = 1,000	

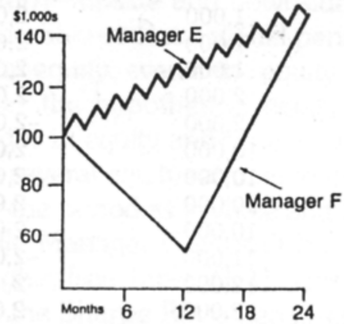
$$SR_C = \frac{E}{\text{sd}} = \frac{\frac{24,000}{2}}{\sqrt{12}\cdot\sqrt{\frac{14(1000-0)^2+8(1000-1000)^2+2(1000-8000)^2}{23}}} = 1.57$$

$$SR_D = \frac{\frac{24,000}{2}}{\sqrt{12}\cdot\sqrt{\frac{18(1000-2000)^2+6(1000+2000)^2}{23}}} = 1.96$$

The expected return, E, is equal to total equity gain for the period divided by the number of years, or equivalently, the average monthly return multiplied by 12. The *annualized* standard deviation is equal to the standard deviation of the monthly returns multiplied by $\sqrt{12}$.[a]

[a]To annualize an interval expected (average) return, it is necessary to multiply by the number of intervals in a year (12 for monthly data). To annualize an interval-based standard deviation, it is necessary to multiply by the square root of the number of intervals in a year ($\sqrt{12}$ for monthly data). This standard deviation conversion is a consequence of the fact that if the intervals are independent, the *variance* of return for longer interval data (e.g., year) would be equal to the variance of return for a shorter interval (e.g., month) times the number of shorter intervals in each longer interval (e.g., 12). Thus, the *standard deviation* of return for the longer interval would be equal to the standard deviation of return for the shorter interval times the *square root* of the number of shorter intervals in a longer interval (since the standard deviation is defined as the square root of the variance).

Figure 21.4
COMPARISON OF TWO MANAGERS WITH EQUAL RETURNS AND STANDARD DEVIATIONS BUT DIFFERENT SEQUENCE OF MONTHLY GAINS



Source: J. Schwager, "Alternative to Sharpe Ratio Better Measure of Performance," *Futures*, p. 56, March 1985.

$4,000 monthly losses, while Manager F first loses $48,000 in the initial 12 months and subsequently gains $96,000 during the remainder of the period.

Both managers would have identical Sharpe Ratios. Despite this fact, few traders would consider the two performance records equivalent in risk. Virtually all traders would agree that Manager F's performance implies a much greater risk level.

RETURN RETRACEMENT RATIO

The return retracement ratio (RRR) provides a return/risk measure that avoids the drawbacks of the Sharpe Ratio detailed in the previous section and also comes closer to defining risk in a manner consistent with the way most traders actually perceive risk. The RRR represents the average annualized compounded return (R) divided by an average maximum retracement (AMR) measure:

$$\text{RRR} = \frac{R}{\text{AMR}}$$

R is that return that when compounded annually for a period coinciding with a manager's (or system's) equity stream will yield the same ending equity, given the starting equity. The AMR is equal to the average of the maximum retracement (MR) for each data point (e.g., month), where the MR is equal to the greater of the following two measures:

1. Maximum retracement from a prior equity peak (MRPP); or
2. Maximum retracement to a subsequent low (MRSL).

As the name implies, the MRPP measures the percent decline in equity from the prior high point in equity. In effect, for each data point (e.g., month end) the MRPP reflects the worst retracement that theoretically could have been experienced *by any investor* with an account at that time. The MRPP would be equal to the cumulative loss that would have been realized by an investor starting at the worst possible prior time (i.e., the prior equity peak). Note that if a new equity peak is set in a given month, the MRPP for that point will be equal to 0. One problem with the MRPP is that for early data points the drawdown measure may be understated because there are few prior points. In other words, if more prior data were available, the MRPP for those points would very likely be larger.

As the name implies, the MRSL measures the percent decline in equity to the subsequent lowest equity point. In effect, for each data point (e.g., month end), the MRSL measures the worst retracement that would be experienced *at any time* by investors starting in that month—that is, the cumulative loss that would be realized by such investors at the subsequent low point in equity. Note that if equity never decreases below a level for a given month, the MRSL for that point will be equal to 0. One problem with the MRSL is that for latter data points this drawdown measure is likely to be understated. In other words, if more data were available, there is a good chance the MRSL would be greater—that is, the subsequent equity low may not yet have been realized.

The MRPP and MRSL complement each other. Note that each measure is most likely to be understated when the other measure is least likely to be understated. For this reason the MR for each point is defined as the greater of the MRPP and MRSL. In this sense, the MR provides a true worst-case scenario for each point in time (e.g., month end). The AMR simply averages these worst-case values. This approach is far more meaningful than methods that employ only the single worst case—the maximum drawdown.

The mathematical derivation of the RRR is summarized below:

$$\text{RRR} = \frac{\text{R}}{\text{AMR}}$$

where R = average annual compounded return (see below for derivation),

$$\text{AMR} = \frac{1}{n}\sum_{i=1}^{n} \text{MR}_i$$

where n = number of months in survey period,

$$MR_i = \max(MRPP_i, MRSL_i)$$

where

$$MRPP_i = \frac{PE_i - E_i}{PE_i}$$

$$MRSL_i = \frac{E_i - ME_i}{E_i}$$

where E_i = equity at end of month i
PE_i = peak equity on or prior to month i
ME_i = minimum equity on or subsequent to month i

Note that $MRPP_i$ will be equal to 0 for first month and $MRSL_i$ will be equal to 0 for last month.

R, the average annual compounded return, is derived as follows:[6]

$$S(1 + R)^N = E$$

where S = starting equity
E = ending equity
N = number of years
R = annualized compounded return (in decimal terms)

This equation can be reexpressed in terms of the annualized return (R):

$$R = \sqrt[N]{\frac{E}{S}} - 1$$

To facilitate solving this equation for R, it is necessary to reexpress it in terms of logarithms:

$$R = \text{antilog}\left[\frac{1}{N}(\log E - \log S)\right] - 1$$

For example, if a $100,000 account grows to $285,610 in four years, the annualized compounded return would be 0.30 or 30%:

$$R = \text{antilog}\,[\tfrac{1}{4}(\log 285{,}610 - \log 100{,}000)] - 1$$
$$R = \text{antilog}\,[\tfrac{1}{4}(5.4557734 - 5)] - 1$$
$$R = \text{antilog}[0.11394335] - 1 = 0.30$$

---

[6]The following derivation of R through the example where R = 0.30 is from J. Schwager, "Alternative to Sharpe Ratio Better Measure of Performance," *Futures,* p. 58, March 1985.

The calculation for the RRR can be applied directly in evaluating a money manager's performance, because the equity size of the account is known for each data point. However, a moment's reflection will reveal that in the case of trading systems, the equity size is not known; only the dollar gain/loss in each interval is available. How can percent return and retracements be calculated if we don't know the amount of funds needed to trade the system? The answer is that since the RRR value will be independent of the size of the funds assumed to be needed to trade the system,[7] any number can be used. Although it won't affect the calculation, as a means of selecting a plausible number, the trader can assume that the funds needed to trade the system are equal to four times the maximum dollar loss. For example, if the system's worst loss is $50,000, the funds needed to trade the system could be assumed to be $200,000.

Once the figure for the funds needed to trade the system (i.e., the assumed account size) is selected, monthly equity figures for the trading system can be generated as follows:

1. Divide all the monthly profit/loss figures by the *same* account size to generate monthly percent return figures.[8]
2. Use a chain multiple of the assumed account size and the monthly percent return numbers to generate monthly equity levels. For example, if the assumed account size is $200,000 and the percent returns for the first four months are +4%, −2%, −3%, and +6%, then the corresponding equity levels would be calculated as follows:

Start = $200,000
End of month 1 = ($200,000) (1.04) = $208,000
End of month 2 = ($200,000) (1.04) (0.98) = $203,840
End of month 3 = ($200,000) (1.04) (0.98) (0.97) = $197,725
End of month 4 = ($200,000) (1.04) (0.98) (0.97) (1.06) = $209,588

Once the monthly equity levels are obtained, the derivation of the R and AMR values in the RRR calculation would be exactly analogous to the money manager case.

---

[7]Since the assumed account size is used as a divisor in both the numerator of the RRR (to divide dollar gain/loss in the return calculation) and the denominator of the RRR (to divide dollar retracements), the figure will cancel out. For example, doubling the size of the assumed account size would halve both the average annual compounded return and the average maximum retracement, leaving the RRR value unchanged.

[8]Note that the implicit assumption is that the system trading results were based on a fixed portfolio. In other words, the test of the system doesn't increase the number of contracts traded when the system makes money and decrease the number when the system loses. (In actual trading, of course, such adjustments would be made.) Hence, using a *constant* account size as the divisor to transform monthly profit/loss figures into percent return figures is the appropriate procedure.

It should be noted that in actual trading, the individual would adjust the funds used for trading based on personal risk preferences. The actual level used could be greater or smaller than the four times maximum loss figure used as a starting assumption in calculating the RRR for a system. The RRR value of the system, however, would be unaffected by the specific choice of the account size assumed needed to trade the system.

ANNUAL GAIN-TO-PAIN RATIO

The annual gain-to-pain ratio (AGPR) represents a simplified type of return/retracement measure. The AGPR is defined as follows:

$$AGPR = AAR \div AAMR$$

where AAR = arithmetic average of annual returns

AAMR = average annual maximum retracement, where the maximum retracement for each year is defined as the percent retracement from a prior equity high (even if it occurred in a previous year) to that year's equity low

The RRR is a better return/retracement measure than the AGPR insofar as each data point is incorporated in the risk calculation and the measure does not artificially restrict the data (e.g., calendar year intervals). However, some traders may prefer the AGPR because it requires less computation and the resulting number has an easy-to-grasp intuitive meaning. For example, an AGPR of 3 would mean that the average annual return is three times as large as the average annual worst retracement (measured from a primary peak).

MAXIMUM LOSS AS A RISK MEASURE

One number of particular interest is the worst-case possibility in a given system. In other words, the largest retracement that would have been experienced during the entire survey period if trading was initiated on the worst possible start date. The maximum loss (ML) is merely the largest $MRSL_i$ (or largest $MRPH_i$, the two would be equivalent) and can be expressed as

$$ML = \max(MRSL_i)$$

See the section, "Return Retracement Ratio" for derivation of $MRSL_i$.

The ML is not recommended as a sole risk measure or the risk component in a return/retracement ratio because it depends on only a single event and hence may be very unrepresentative of the overall performance of a system. Furthermore, because of this characteristic, the value of the ML may be highly contingent on the choice of the survey period. As a related consideration, the use of ML introduces a negative bias for managers with longer track records. However, the ML does provide important information and should be consulted in conjunction with the RRR.

TRADE-BASED PERFORMANCE MEASURES

In addition to the performance measures just discussed, the following measures may also merit supplemental attention:

1. ***The Expected Net Profit per Trade.*** The expected net profit per trade (ENPPT) can be expressed as:

$$\text{ENPPT} = (\%\ \text{P})(\text{AP}) - (\%\ \text{L})(\text{AL})$$

err

where % P = percent of total trades that are profitable
% L = percent of total trades that result in net losses
AP = average net profit of profitable trade
AL = average net loss of losing trade

The usefulness of this indicator is that a low ENPPT figure will highlight systems that are vulnerable to a serious deterioration of profits given poor executions, increased commissions, or any other form of increased transaction cost. For example, if a system had an ENPPT of $50, its validity would be highly suspect, no matter how favorable the other performance measures. The critical disadvantage of the ENPPT is that it does not incorporate a risk measure. In addition, the ENPPT has the intrinsic drawback that it may unfairly penalize active systems. For example, a system that generated one trade with a net gain of $2,000 would rate better than a system that, during the same period, generated 100 trades with an ENPPT of $1,000 (with similar equity fluctuations).

2. ***Trade-Based Profit/Loss Ratio.*** The trade-based profit/loss ratio (TBPLR) can be expressed as follows:

$$\text{TBPLR} = \frac{(\%\text{P})(\text{AP})}{(\%\text{L})(\text{AL})}$$

This measure indicates the ratio of dollars gained to dollars lost in all trades. The appeal of the TBPLR is that it deflates profits by a measurement of total pain suffered. There are three drawbacks to the TBPLR: (1) Similar to the ENPPT, it is severely biased against systems with a higher frequency of trades. For example, consider the following two systems:

System	Average Profit ($)	Average Loss ($)	Percent Profitable Trades	Percent Losing Trades	TBPLR
A	400	200	75	25	6
B	200	100	50	50	2

Superficially, it might appear that System A is better (three times better to be exact). However, suppose you are now provided with the following additional information: system B generated 100 trades per year and system A only 10, while both systems had similar risk levels (e.g., AMR), and hence required equivalent funds to trade. In this case, system B's percent return would actually be double that of system A.[9] (2) The TBPLR gives no weight to open position losses. Thus, a trade that witnesses a huge loss before it is finally closed at a slight profit would have the same effect on the TBPLR as a trade that is immediately profitable and closed at the same slight profit. The two trades, however, would hardly be equivalent in the eyes of the trader. (3) The TBPLR does not distinguish between intermittent and consecutive losses—a potentially serious flaw there is if a tendency for losing trades to be clustered.

WHICH PERFORMANCE MEASURE SHOULD BE USED?

By using drawdowns (the worst at each given point in time) to measure risk, the risk component of the RRR (the AMR) comes closer to describing most people's intuitive sense of risk than does the standard deviation in the Sharpe Ratio, which makes no distinction between sudden large gains and sudden sharp losses—two events that are perceived very differently by traders (and investors). The RRR also avoids the Sharpe Ratio's failure to distinguish between intermittent and consecutive losses. For these reasons, the RRR is probably a superior return/risk measure to the Sharpe Ratio.

Even so, the RRR is being proposed as an additional rather than replace-

[9]Percent return = (ENPPT × N)/F, where N = number of trades and F = funds traded (assumed equal for each system). System A's percent return = (250 × 10)/F, while system B's percent return = (50 × 100)/F.

ment return/risk measure to the Sharpe Ratio. Reason: the Sharpe Ratio is a very widely used return/risk measure, whereas, at this writing, the RRR is not used at all. Hence, the trader or the system designer would still need to calculate a Sharpe Ratio for the purpose of comparing his results to CTA track records, industry indexes, or alternative investments. Together the Sharpe Ratio and the RRR provide a very good description of a system's or trader's relative performance.

In addition to these return/risk measures, the ENPPT should be calculated to make sure that the validity of the system would not be threatened by a moderate increase in transaction costs or a small deterioration in the average profit per trade. The maximum loss (ML) figure should be checked to make sure there was no catastrophic losing streak. Finally, the AGPR might be calculated as a supplemental measure, which yields a figure that is intuitively meaningful.

THE INADEQUACY OF A RETURN/RISK RATIO FOR EVALUATING MONEY MANAGER TRADING PERFORMANCE

In the case of evaluating trading systems, the selected return/risk measure would yield the same ranking order of systems as the *estimated* percent return. This observation, which is a consequence of the fact that fund requirements for trading a system can only be estimated basis risk, can be proved as follows:

$$\text{Selected return/risk measure} = \frac{G}{R}$$

$$\text{Estimated percent return for a system in a given market} = \frac{G}{F}$$

where G = average annual gain per contract
R = chosen risk measure (e.g., sd, AMR, ML)
F = total funds allocated for trading

The only practical way to estimate F is as a function of risk. Most directly, F might be estimated as some multiple of the chosen risk measure. That is,

$$F = kR$$

where k = multiple of risk measure (determined subjectively). Thus, the estimated percent return for a system could be expressed as

$$\frac{G}{F} = \frac{G}{kR} = \frac{1}{k}\left(\frac{G}{R}\right)$$

Note that G/R is the selected return/risk measure. Consequently, the percent return for a system will merely be equal to some constant times the return/risk measure. Although different traders will select different risk measures and values for k, once these items are specified, the return/risk measure and the estimated percent return would yield the same ranking order of systems. Also note that in the case of evaluating systems, the *percent risk*, which we define as the risk measure divided by fund requirements, is a constant (percent risk = $R/F = R/kR = 1/k$).

Whereas in the case of evaluating trading systems a higher return/risk ratio *always* implies higher percent return, this is not true for the evaluation of money managers. Also, the percent risk is no longer a constant, but instead can vary from manager to manager. Thus, it is entirely possible for a money manager to have a higher return/risk ratio than another manager, but to also have a lower percent return or a higher percent risk. (The reason for this is that in the money manager case, the link between fund requirements and risk is broken—that is, different money managers will differ in the level of risk they will assume for any given level of funds.) Consequently, a return/risk ratio is no longer a sufficient performance measure for choosing between alternative investments. We illustrate this point by using the Sharpe Ratio, but similar conclusions would apply to other return/risk measures. (In the following discussion, we assume that management fees are based entirely on profits and that interest income is not included in money manager return figures, but is received by investors. Consequently, the simplified form of the Sharpe Ratio, which deletes the riskless interest rate, is appropriate.)

Assume we are given the following set of *annualized* statistics for two money managers:

	Manager A	Manager B
Expected gain	$ 10,000	$ 50,000
Standard deviation of gain	$ 20,000	$ 80,000
Initial investment	$100,000	$100,000
Sharpe Ratio	.50	.625

Although Manager B has the higher Sharpe Ratio, not all traders would prefer Manager B, because he also has a higher risk measure (i.e., higher standard deviation). Thus, a risk-averse investor might prefer Manager A, gladly willing to sacrifice the potential for greater gain in order to avoid the substantially greater risk. For example, if annual trading results are normally distributed, for any given year, there would be a 10 percent probability of the return falling more than 1.3 standard deviations below the expected rate. In such an event, an investor would lose $54,000 with Manager B [$50,000 – (1.3 × $80,000)], but only $16,000 with Manager A. For a risk-averse investor, minimizing a loss under negative assumptions may be more important than maximizing gain under favorable conditions.[10]

Next, consider the following set of statistics for two other money managers:

	Manager C	Manager D
Expected gain	$ 20,000	$ 5,000
Standard deviation of gain	$ 20,000	$ 4,000
Initial investment	$100,000	$100,000
Sharpe Ratio	1.0	1.25

Although Manager D has a higher Sharpe Ratio, Manager C has a substantially higher percent return. Investors who are not particularly risk-averse might prefer Manager C even though he has a lower Sharpe Ratio. The reason for this is that for the major portion of probable outcomes, an investor would be better off with Manager C. Specifically, in this example, the investor will be better off as long as return does not fall more than .93 standard deviations below the expected rate—a condition that would be met 82 percent of the time (assuming trading results are normally distributed).[11]

Even more striking is the consideration that there are circumstances in which virtually all investors would prefer the money manager with the lower Sharpe Ratio. Consider the following two money managers:[12]

---

[10]Implicit assumption in this example: The investor can't place a fraction of the stated initial investment with Manager B. In other words, the minimum unit size of investment is $100,000. Otherwise, it would always be possible to devise a strategy in which the investor would be better off with the manager with the higher Sharpe Ratio. For example, placing $25,000 with Manager B would imply the same standard deviation as is the case for a $100,000 investment with Manager A but a higher expected gain ($12,500).

[11]Implicit assumption in this example: Borrowing costs for the investor are significantly greater than the interest income return realized by placing funds with a money manager. This assumption prohibits the alternative strategy of borrowing funds and placing a multiple of the initial $100,000 investment with the manager with the higher Sharpe Ratio. If borrowing costs and interest income were equal (an assumption not likely to be valid in the real world), it would always be possible to devise a strategy in which the investor would be better off with the manager with the higher Sharpe Ratio. For example, the strategy of borrowing an additional $400,000 and placing $500,000 with Manager D would imply the same standard deviation as is the case for a $100,000 investment with Manager C, but a higher expected gain ($25,000).

[12]The Sharpe Ratios used in this example are considerably higher than the levels likely to be found in the real world. However, the assumption of such higher Sharpe Ratios ellucidates the intended theoretical point.

	Manager E	Manager F
Expected gain	$ 10,000	$ 50,000
Standard deviation of gain	$ 2,000	$ 12,500
Initial investment	$100,000	$100,000
Sharpe Ratio	5.0	4.0

In this example, virtually all investors (even those that are risk-averse) would prefer Manager F, despite the fact that he has a lower Sharpe Ratio. The reason is that the percent return is so large relative to the ambiguity of that return (standard deviation), that even under extreme adverse circumstances, investors would almost certainly be better off with Manager F. For example, once again assuming that trading results are normally distributed, the probability of a gain more than three standard deviations below the expected gain is only 0.139 percent. Yet even under these extreme circumstances, an investor would still be better off with Manager F: Gain = $12,500/year (12.5 percent) compared with $4,000/year (4 percent) for Manager E. This example illustrates, even more dramatically the fact that, by itself, a return/risk ratio does not provide sufficient information for evaluating a money manager.[13] (This conclusion applies to all return/risk measures, not just the Sharpe Ratio.)

The key point is that in evaluating money managers, it is also important to consider the percent return and risk figures independently rather than merely as a ratio.

GRAPHIC EVALUATION OF TRADING PERFORMANCE

Graphic depictions can be particularly helpful in comparing the performance of different money managers. Below we consider two types of charts:

1. ***Net Asset Value.*** The net asset value (NAV) indicates the equity at each point in time (typically, month-end) based on an assumed beginning equity of $1,000. For example, an NAV of 2,000 implies that the original investment was doubled as of the indicated point in time. By definition, the NAV at the start of the survey period is equal to 1,000. Subsequent values would be derived as follows:

[13]Comments analogous to footnote 11 also apply here.

End of Month	Monthly Dollar Return Divided by Equity at Start of Month	NAV
1	r_1	$(1{,}000)(1 + r_1)$
2	r_2	$(1{,}000)(1 + r_1)(1 + r_2)$
3	r_3	$(1{,}000)(1 + r_1)(1 + r_2)(1 + r_3)$
.	.	.
.	.	.
.	.	.
n	r_n	$(1{,}000)(1 + r_1)(1 + r_2)(1 + r_3) \cdots (1 + r_n)$

For example, if a money manager witnesses a + 10 percent return in the first month, a −10 percent return in the second month, and a +20 percent return in the third month, the NAV at the end of the third month would be:

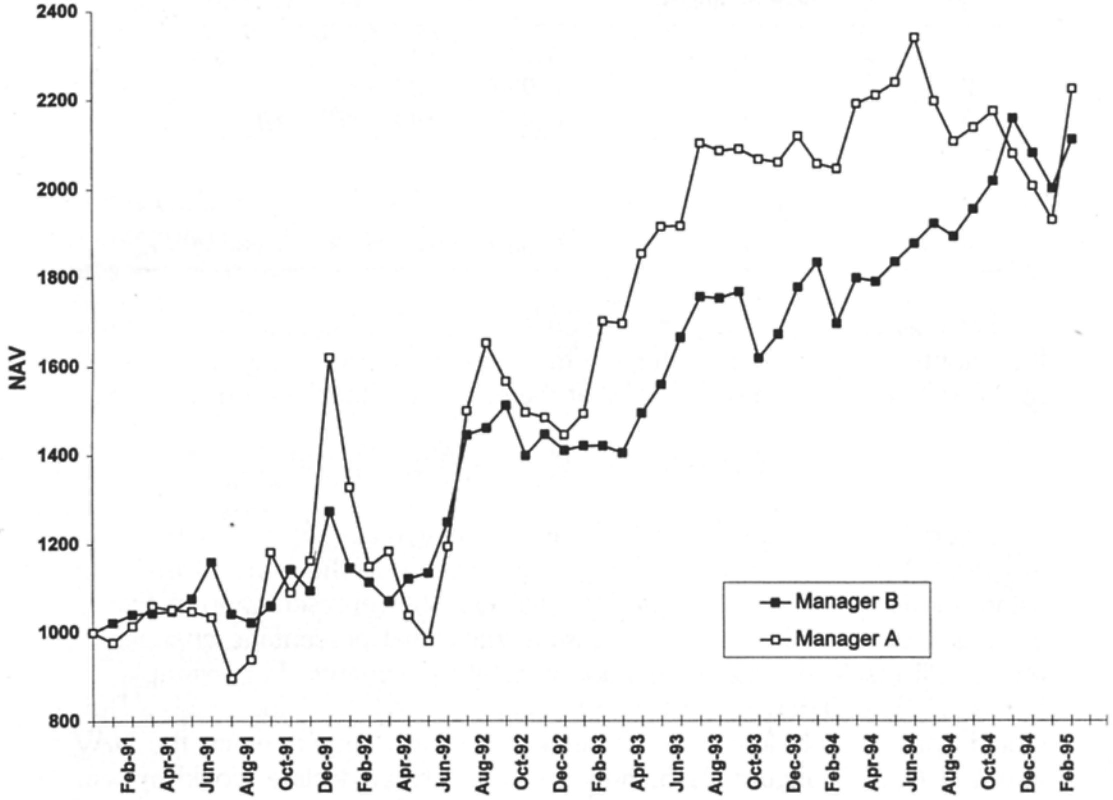
$$(1{,}000)(1 + 0.1)(1 - 0.1)(1 + 0.2) = 1{,}188$$

Figure 21.5 illustrates the NAV for two money managers during the January 1991–February 1995 period. Figure 21.6 presents the same information using a logarithmic scale for the NAV values. The representation in Figure 21.6 is preferable because it will assure that equal percentage changes in equity will result in equal magnitude vertical movements. For example, in Figure 21.6, a 10 percent decline in equity when the NAV value equals 2,000 would appear equivalent to a 10 percent decline in equity when the NAV equals 1,000. In Figure 21.5, however, the former decline would appear twice as large. In any case, regardless of the type of scale used to depict NAV curves, it should be stressed that only comparisons based on exactly the same survey period are meaningful.

Although the NAV is primarily a return measure, it also reflects risk. All else being equal, the more volatile a money manager's performance, the lower the NAV. For example, consider the five money managers below who, during a given year, witness the following monthly gains and losses:

Manager	Six Months of Percentage Gains Equal to:	Six Months of Percentage Losses Equal to:	NAV at End of Year
1	+11%	−1%	$(1{,}000)(1.11)^6(.99)^6 = 1{,}760$
2	+21%	−11%	$(1{,}000)(1.21)^6(.89)^6 = 1{,}560$
3	+31%	−21%	$(1{,}000)(1.31)^6(.79)^6 = 1{,}230$
4	+41%	−31%	$(1{,}000)(1.41)^6(.69)^6 = 850$
5	+51%	−41%	$(1{,}000)(1.51)^6(.59)^6 = 500$

Figure 21.5
NAVS FOR TWO MANAGERS



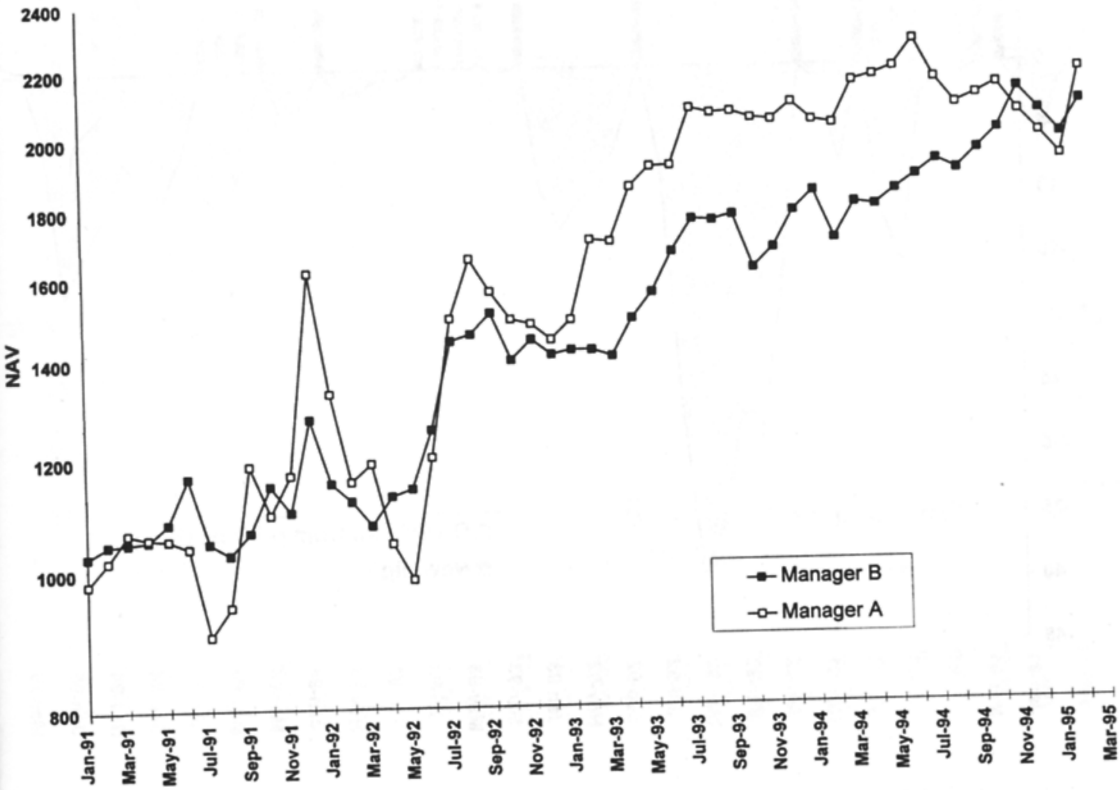
Note the dramatic differences between the ending NAV values despite the equal absolute differences between the percentage gains in winning months and percentage declines in losing months.

The degree to which the NAV incorporates risk may not be sufficient for risk-averse investors. For example, although Manager A witnesses a greater ending NAV than Manager B (see Figure 21.6), many investors might still prefer Manager B because her performance is less volatile. Clearly a more explicit depiction of risk, such as the "underwater" chart described below, would be helpful as a supplement to the NAV chart.

2. *Underwater Curve.*[14] The underwater curve depicts the percent drawdown as of the end of each month, measured from the previous equity peak. In other words, assuming beginning-of-month trading start dates, the underwater curve reflects the largest percentage loss as of the end of each

[14]The term "underwater curve" was first coined by Norman D. Strahm.

Figure 21.6
NAVS FOR TWO MANAGERS (LOG SCALE)

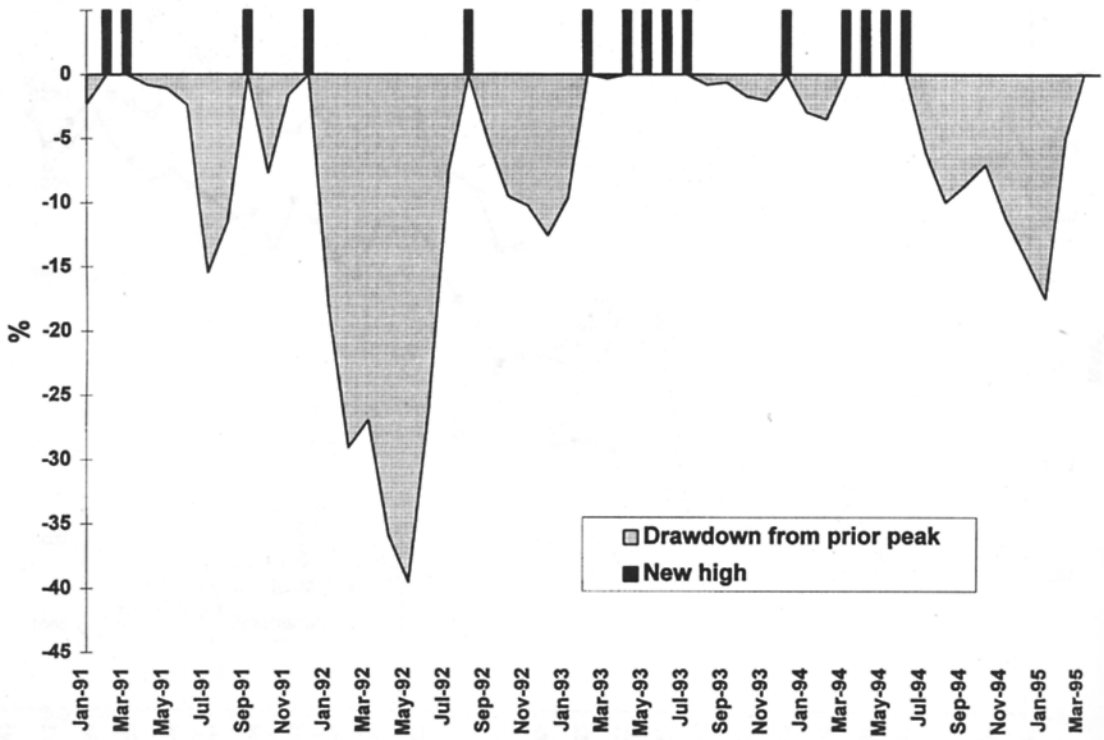


month, assuming an account had been initiated at the worst possible prior entry point (i.e., prior equity peak). Insofar as it reflects the maximum possible equity retracement at each point, the underwater curve is conceptually similar to the previously described MRPP in the RRR calculation. Figures 21.7 and 21.8 illustrate the underwater curves for the two money managers depicted in Figures 21.5 and 21.6. (The vertical bars above the zero line indicate that the given month witnessed a new equity high.) These charts clearly demonstrate the greater degree of risk implied by Manager A's performance.

Which manager (A or B) has the better performance? The answer must unavoidably be subjective because Manager A achieves the higher end of period NAV value, but also exhibits more extreme drawdowns.[15] However, the key point is that by using both the NAV and underwater charts, each

[15]Although this statement is theoretically true, for the example given, it is likely that the vast majority of investors would prefer Manager B because Manager A's marginally higher return hardly seems worth the substantial increase in risk.

Figure 21.7
UNDERWATER CURVE: MANAGER A



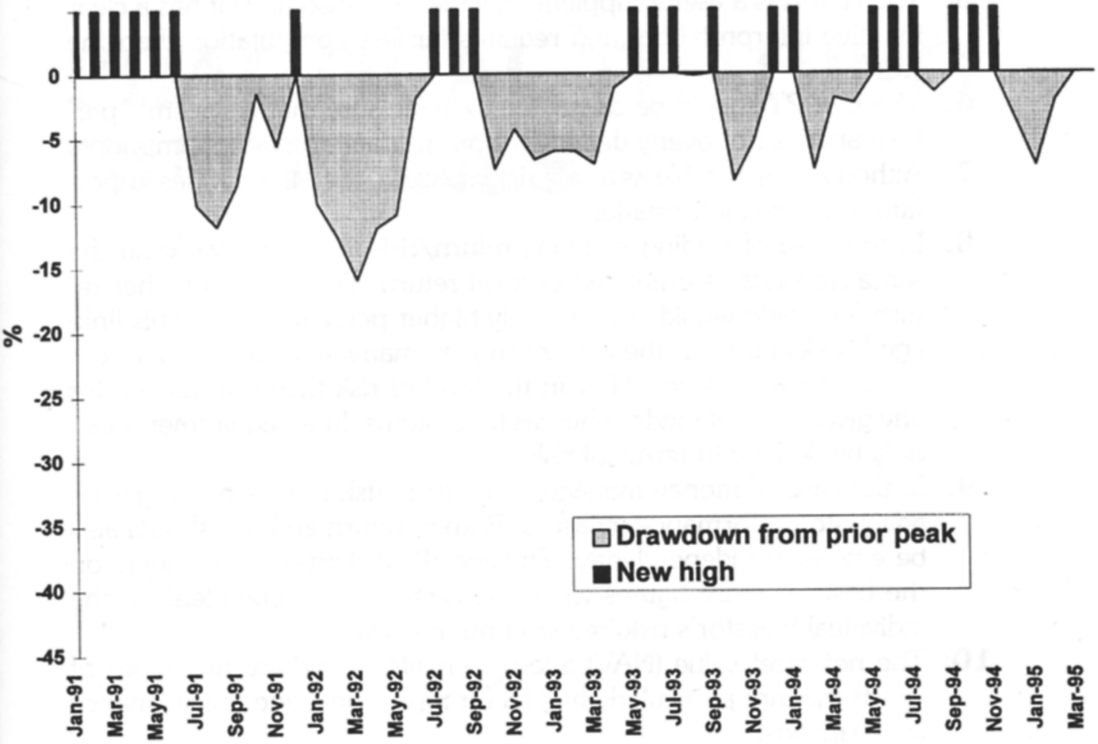
investor should have sufficient information to choose the money manager he prefers, given his personal return/risk preferences. In fact, given the relative ease with which the NAV and underwater charts can be derived, and the depth of the information they provide, for many investors, the combination of these charts may offer the ideal methodology for money manager performance comparisons.

Although this section was described in terms of depicting money manager performance, the same types of charts could be generated for trading systems. The trader would merely have to transform the system's dollar profit/loss figures into percent return figures based on the account size the trader deems necessary to trade the system. The NAV for the system could then be derived by creating a chain multiple of 1,000 and these percent return numbers.

CONCLUSIONS

1. By itself, dollar gain per unit time is an insufficient measure for evaluating the performance of a trading system or money manager.

Figure 21.8
UNDERWATER CURVE: MANAGER B



2. In evaluating the performance of a system, a return/risk measure serves a dual role:

   a. It incorporates risk;
   b. It provides a proxy percent return measure.

3. The Sharpe Ratio has several potential drawbacks as a trading performance measure:

   a. Failure to distinguish between upside and downside fluctuations;
   b. Failure to distinguish between intermittent and consecutive losses;
   c. Potential distortions in the gain measure in assessing performance for an extended period.

4. The RRR is an alternative performance measure that seems to be preferable to the Sharpe Ratio in that it appears to reflect more closely the behavioral preferences of the trader (that is, traders are generally concerned about *downside* volatility in equity rather than volatility in equity). However, the Sharpe Ratio should still be considered as a supplemental measure because it is the most used return/

risk measure and hence is essential for comparing one's own track record or system to industry money managers.

5. The AGPR is a useful supplemental measure insofar as it has a clear intuitive interpretation, and requires far less computation than the RRR.
6. The ENPPT should be calculated to make sure that a system's performance is not overly dependent on transaction cost assumptions.
7. Although not suitable as a sole risk measure, the ML provides important additional information.
8. In the case of trading systems, return/risk measures will yield the same rankings as estimated percent return. Therefore, a higher return/risk ratio would always imply higher percent return. This linkage breaks down in the case of money managers because different money managers will differ in the level of risk they will assume for any given level of funds, whereas for systems, fund requirements can only be defined in terms of risk.
9. In the case of money managers, a return/risk ratio is no longer an adequate performance measure. Rather, return and risk should also be evaluated independently. The specific ordering of managers on the basis of these figures will be subjective (i.e., dependent on the individual investor's risk/reward preferences).
10. The net asset value (NAV) and underwater curve are two types of charts that are particularly helpful for money manager performance comparisons.