MARK GRINBLATT *

How to Evaluate a Portfolio Manager

I. Introduction

Many financial institutions and individual investors delegate the management of a large part of their assets to professionals, presumably for the benefit of expertise and/or diversification. It is commonly believed that there are differences in the abilities of these professional portfolio managers, which can be ascertained from past track records. This is evident from the multi-million dollar fees collected by the managers of the larger funds, who were often selected after successfully managing smaller amounts of money.

With billions of dollars at stake in some of the funds, it would be difficult to believe that anything but the most sophisticated evaluation methods would be used to promote and hire these managers. Nevertheless, while many financial institutions spend large sums of money to evaluate their portfolio managers, often employing outside consulting firms to conduct the evaluation, the procedures used are generally archaic and flawed.

For instance, the most common method of evaluating the performance of equity managers compares their portfolio returns (with all dividends reinvested) to that of the Standard and Poors 500 index, which is a weighted average of the prices of 500 equities on the New York Stock Exchange. However, the S&P 500 index tends to rise faster than the interest rate at which one borrows. Thus, a leveraged position in the S&P 500's stocks would beat the index, on average. Moreover, for a fair comparison of returns, it is necessary to reinvest the dividends of the stocks in the S&P 500. However, the correction to the S&P 500 for dividend reinvestment is not always done correctly, even in the official Standard and Poors computations. In these instances, a manager who bought the portfolio of 500 stocks corresponding to the index would generally appear to have done better than the index. In 1985, for example, a manager who invested in the stocks corresponding to the S&P 500 index would have earned over a 32% return, while the official S&P 500 return was only slightly over 31½%. If the same manager had borrowed $ 1000 at 10% interest at the beginning of 1985 and invested $ 2000 of the fund's capital in the S&P 500's stocks, he would have earned a return of almost 55%, beating the index by over 23%.

The second most common method of evaluation is a ranking based on raw returns within some class of portfolio managers (e.g. all mutual fund managers, all pension fund managers, all bank trust managers). At first glance, this is plausible, since at the most basic level, investors prefer managers who offer the highest return. Unfortunately, returns are random variables – the manager who obtained the highest return for his clients this year is unlikely to offer the highest return next year. Moreover, comparisons based solely on returns will, on average, favor those managers with the riskiest investment strategies. It is well known, for instance, that long term bonds have historically had higher average rates of return than short term notes and bills and that equities have had higher average rates of return than bonds. Differences in average returns also exist within these classes of securities. A fund manager who primarily invests in the equities of large well-established companies will generally earn a lower rate of return than a manager who primarily invests in the equities of newly-established companies with investment projects that are more speculative.

Table 1 presents excess return data (annualized) for 12 equally-weighted portfolios, each

* The author is grateful to WALTER WASSERFALLEN and GATAM DHINGRA for comments on an earlier draft.

with about two hundred securities, formed on the basis of firm size. All securities on the New York and American Stock Exchanges were first sorted monthly by market value and then grouped into one of twelve portfolios on the basis of their ranking. Portfolio 1 represents the portfolio of securities with the smallest market values, Portfolio 12 represents securities with the largest market values. The returns of these twelve portfolios were then tabulated in the month subsequent to the sorting, averaged over the 120 months from January 1975 to December 1984, and annualized by multiplying the average monthly return by twelve. The percentages reported in the table are more precisely referred to as 'excess returns' because they subtract the annualized returns of a proxy for the risk-free rate, the return of one month U.S. Treasury Bills. Table 1 indicates that average excess returns are inversely related to firm size. The average excess return of the portfolio of smallest firms was 33.4% per year, while the largest firms earned only 5.5% more than the Treasury Bill return.

Table 1: Annualized Excess Returns for Twelve Equally-Weighted Portfolios Based on Firm Size from 1975–1984

Portfolio	Excess Return (%)	Portfolio	Excess Return (%)
1	33.37%	7	19.24%
2	22.41%	8	16.97%
3	23.33%	9	15.30%
4	21.18%	10	13.36%
5	21.32%	11	11.00%
6	18.52%	12	5.50%

It would be wrong to conclude from this evidence that the securities of small firms are better buys than those of large firms or that equities are better investments than Treasury Bills. It has long been established that investors are averse to risk and when risk is taken into account, one cannot infer that one class of securities is superior to another. Indeed, according to various theories of capital market equilibrium, one expects to find substantial differences in average returns across securities. Such differences are needed to induce investors to hold the riskier securities.

All of this suggests that it is important to determine the risk of an investment strategy before comparing investment managers or judging whether a particular manager is superior to the market. Once risk is established, the risk premium of the investment strategy can be subtracted from the time-weighted return to yield a more basic measure of performance. Such a measure would not be subject to the biases described above.

Research in finance over the last twenty years has sought a measure that would properly adjust for risk. Beginning with the SHARPE (1966) Ratio, the TREYNOR (1965) Ratio, and the JENSEN (1968, 1969) Measure, and continuing to the 'state-of-the-art' measures proposed by CORNELL (1979), COPELAND and MAYERS (1982), and GRINBLATT and TITMAN (1986a), all measures of portfolio performance have attempted to adjust raw returns for risk in some manner. The evolution of the academic thinking on this topic is the focus of this paper. Section II discusses the relation between risk, return, and measured performance using a variety of evaluation techniques. Section III describes two types of performance evaluation procedures: measures that employ data on portfolio holdings and those that do not. Section IV briefly concludes the paper.

II. Risk and Performance Evaluation

a) Equilibrium Asset Pricing Theories and Risk

Altough the evaluation of portfolio performance has been an important area of research in finance over the last twenty years, it has also been an area of great controversy. There are all sorts of pitfalls associated with the measurement of risk. The importance of these has been the focus of an intense debate in the finance literature over the past ten years.

For one, the financial theories that describe the risk and risk premia of individual securities have had only limited success when measured against stock market data. Theories like the Capital Asset Pricing Model (CAPM) and the Arbitrage Pricing Theory (APT), for instance, exhibit strong firm size biases. Table 2 describes the difference between the actual average returns of the twelve portfolios in Table 1 and the average return predicted by the CAPM and the APT. As can be seen, both the CAPM and the APT tend to underpredict the expected returns of the equity securities of firms with very low market values and overpredict the expected returns of firms with very high market values.

Table 2: Prediction Error of the CAPM and APT for Twelve Equally-Weighted Portfolios Based on Firm Size from 1975–1984

Portfolio	Annualized Prediction Error (%) CAPM	APT	Portfolio	Annualized Prediction Error (%) CAPM	APT
1	7.20%	10.80%	7	0.72%	–1.20%
2	–0.36%	1.08%	8	0.60%	–0.96%
3	0.48%	0.12%	9	–0.48%	–0.84%
4	0.60%	–0.24%	10	–1.68%	–2.04%
5	2.04%	1.92%	11	–2.64%	–1.80%
6	–1.08%	–2.16%	12	–5.28%	–3.96%

Measures of portfolio performance that are inspired by these theories of asset pricing display similar biases. According to these measures, portfolio managers who invest primarily in the equity issues of small firms are generally superior to those who invest in the equity of large firms. One would suspect that this is not true, however, and that the difference in performance arises from some error in the measurement of risk with an imperfect theory. Similar risk measurement biases may exist for other security characteristics like dividend yield or price/earnings ratio. As a result, academic researchers cannot agree on the best way to measure risk.

b) Mean-Variance Analysis and Asset Pricing

Financial theorists have a geometric framework for interpreting their theories. This familiar framework, known as mean-variance analysis, plots the means and standard deviations of all feasible portfolio returns, as in Diagram 1. The feasible points of lowest variance for a given mean (i.e. furthest to the left in Diagram 1) are referred to as 'the mean-variance efficient frontier'. The diagram traces its origins to early work in finance by MARKOWITZ (1959) and TOBIN (1958), who proposed the standard deviation of return as the basic measure of total risk. (There was even a measure of performance developed from this diagram, known as the SHARPE Ratio. This measure, which has only limited applicability, is not discussed in the paper.) This should not be confused with 'beta', which is a measure of non-diversifiable risk that we will focus on shortly.

Recent work by ROLL (1977) and by GRINBLATT and TITMAN (1987) suggests that financial theories of the risk-return relation, like the CAPM and the APT, can alternatively be view-

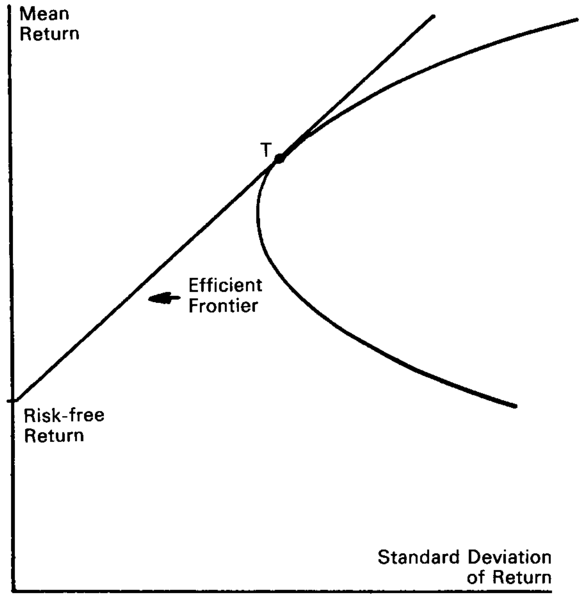


Diagram 1: The mean-standard deviation diagram.

ed as statements about the identity of portfolios on the mean-variance efficient frontier. The CAPM, for instance, states that the tangency portfolio at T is the market portfolio. Such a portfolio has a positive weight (between zero and one) on all assets in the world. The weight on the equity of a company like Nestlé would be the market value of all Nestlé stock divided by the market value of all risky assets. Similarly, in a K-factor model APT, there are K portfolios that are perfectly correlated with the common generating factors that determine the correlations between securities returns. According to the APT, some weighted average of these K portfolios is on the line connecting the risk-free return with the tangency portfolio at T.

These insights can be used to interpret the empirical evidence discussed above. The inability of the CAPM to explain expected returns across portfolios grouped by size indicates that the proxies used to represent the market portfolio in CAPM tests, like the Standard and Poors 500 index portfolio, lie inside the efficient frontier (as in Diagram 2). Similarly, the efficient frontier generated only by the factor portfolios never touches the efficient frontier of all assets if the APT cannot explain the expected returns of various portfolios grouped by firm size.

c) Benchmark Portfolio Efficiency and Performance

In the finance literature, there has generally been a link between asset pricing theories and

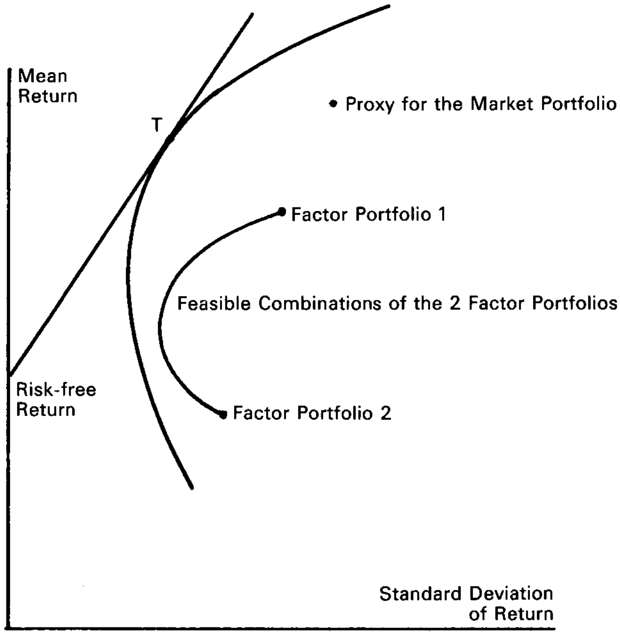


Diagram 2: The position of the Market Proxy Portfolio and the 2-Factor APT Factor Portfolios when the risk premia of assets are not explained by risk measures computed with these portfolios.

performance measurement. This linkage is not as necessary as some researchers have led us to believe. As can be seen in Diagram 2, even when the CAPM and APT are flawed theories, there is still a mean-standard deviation diagram that can be analyzed. In particular, the diagram still has a mean-variance efficient frontier. As we shall shortly see, if we can identify portfolios that are on (or are close to) that frontier, performance evaluation, properly done, will yield appropriate inferences about the identities of superior portfolio managers.

The market portfolio for the CAPM and the factor portfolios for the APT are referred to as benchmark portfolios or index portfolios. They are the portfolios that are used to compute non-diversifiable risk in the respective theory. The risk of a security is found by running a time series regression of the historical returns of a security against the historical returns of the market portfolio or the factor portfolios. On average, the slope coefficient in the CAPM regression, called 'beta', is the covariance between the return of a security and the return of the market portfolio divided by the variance of the return of the market portfolio. This is a traditional measure of (non-diversifiable) risk. In the case of the APT, one has a multiple regression and a number of slope coefficients, called 'factor loadings'. If the factor portfolio returns are uncorrelated, security i's factor loading on the jth factor will, on average, be the covariance between the return of security i and the return of the jth factor portfolio divided by the variance of the return of the jth factor portfolio. These factor loadings are multi-dimensional measures of risk – in a K-factor APT, there are K measures of risk and K district risk premia, one for each measure of risk.

As was hinted above, one can generalize this pattern to arrive at a generic view of risk. Risk may be defined relative to any benchmark portfolio or any group of benchmark portfolios. The risk measure(s) of a security can be found by regressing its historical returns against the historical returns of the benchmark portfolio(s). The slope coefficient(s) represents the measure(s) of risk, and if there is a single benchmark portfolio, risk (which henceforth will be called 'beta') is the covariance between the return of the investment being evaluated and the return of the benchmark portfolio, divided by the variance of the return of the benchmark portfolio.

In practice, when implementing this procedure, one has to deal with the problem of anticipated inflation. Since anticipated inflation tends to increase the expected nominal returns of individual securities and benchmark portfolios, the slope coefficients in these regressions are likely to be overstated. One way around this difficulty is to subtract the risk-free return (as measured by the return on a short term government bill) from all of the return variables in the regression. This excess return regression has one added benefit: the intercept, often called 'alpha', is a measure of performance. Diagram 3 illustrates the alpha of a security when there is a single benchmark portfolio used to compute risk. A line connecting the benchmark portfolio and the risk-free asset in mean-beta space, called the 'securities market line', represents the predicted return for an investment given its beta risk. The height of the actual return above or below this, which is the alpha in the excess return regression, represents the deviation of the actual return from the predicted return.

One complication that we have glossed over here is which benchmark portfolio to use, especially since popular financial theories may make contradictory statements about the identity of 'appropriate' benchmark portfolios. In a

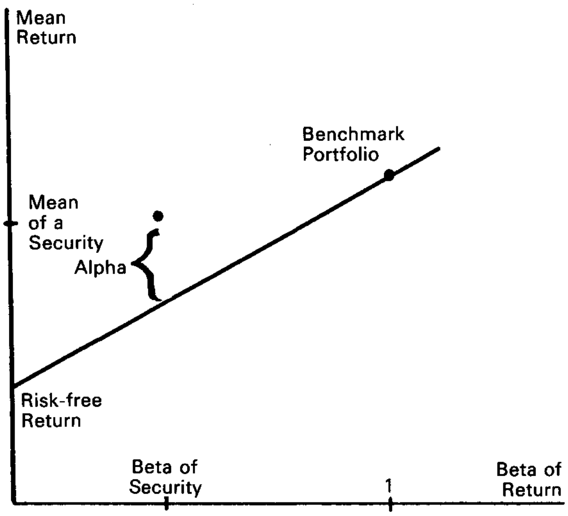


Diagram 3: Illustration of the alpha of a security when there is a single benchmark portfolio used to compute risk.

seminal article, ROLL (1978) examined this problem and concluded that performance evaluation with this risk adjustment procedure, known as 'securities market line analysis', is very sensitive to the choice of the benchmark portfolio and leads to ambiguous conclusions. The ambiguity stems from mathematical properties that relate the position of the benchmark portfolio in the mean-standard deviation diagram to the positions of securities and portfolios in the mean-beta diagram.

If, as in Diagram 4A, the benchmark portfolio lies on the mean-variance efficient frontier, all securities and portfolios of those securities have alphas of zero and lie on the securities market line. As illustrated in the diagram, there is no ability to discriminate between investments in this case. On the other hand, if the benchmark portfolio lies to the right of the efficient frontier, as in Diagram 4B, some securities and some portfolios lie above the securities market line (i.e. have positive alphas) and others lie below. One can now rank investments, but is this ranking an indication of true performance or does it merely reflect the position of the benchmark portfolio in the mean-standard deviation diagram? ROLL pointed out that an alternative benchmark portfolio, also not on the frontier, could virtually reverse the rankings of the investments. Hence, performance evaluation with the securities market line is more an artifact of the evaluator's choice of a benchmark portfolio than a valid tool for measuring investment skill.

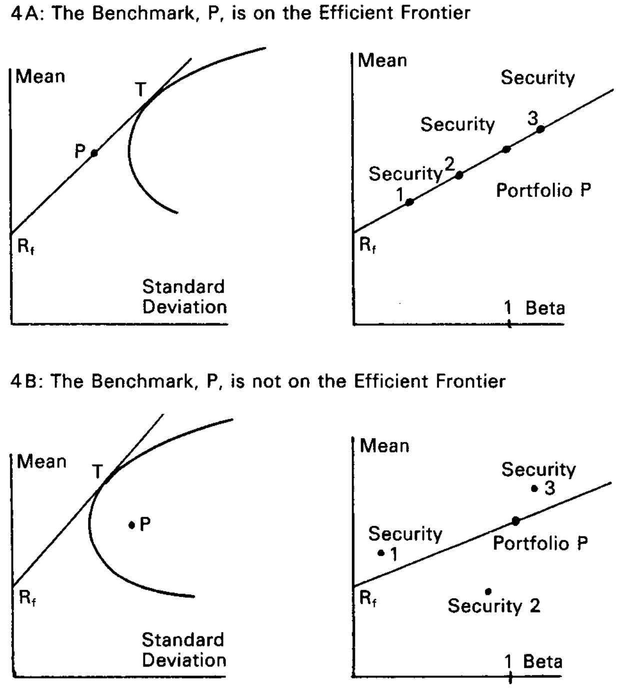


Diagram 4: The relation between benchmark portfolio efficiency and alpha.

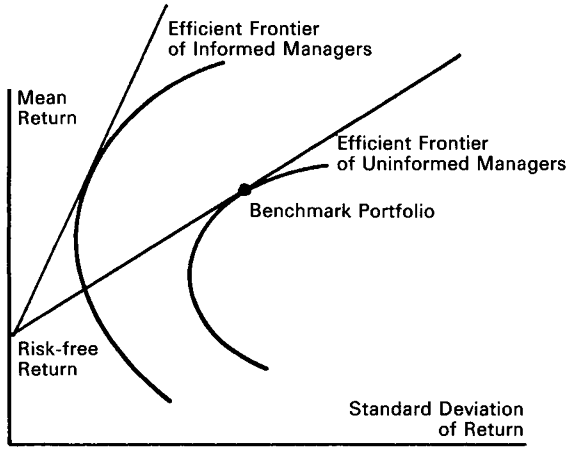
Similar statements apply to securities market line analysis with multiple benchmark portfolios. If the efficient frontier that is constructed from the benchmark portfolios alone touches the efficient frontier of all assets, one cannot discriminate between investments. If the former does not touch the efficient frontier of all assets, there are differences in performance, but the direction and magnitude of the differences stem entirely from the choice of the benchmark portfolios and are very sensitive to alternative choices.

Prior to ROLL's paper, there was a substantial body of research on the performance of mutual fund managers. The benchmark portfolios used in most of these studies were inspired by the Capital Asset Pricing Model and usually consisted of either the Standard and Poors 500 index portfolio or an equally-weighted portfolio of all stocks on the New York Stock Exchange. Since we now know that these benchmark portfolios have size-related biases and lie well inside the mean-variance efficient frontier, the findings of many of these mutual fund studies are questionable.

It is not as obvious, however, that the benchmark portfolios on the frontier are inappropriate. MAYERS and RICE (1979), DYBVIG and ROSS (1985b), and GRINBLATT and TITMAN (1986b) have explored conditions under which an active portfolio manager with superior information will, on average, display positive alphas while managers who lack superior information, on average, will display alphas of zero. The most general version of the major conditions is that

(i) The informed manager is a mean-variance optimizer.
(ii) The informed manager cannot forecast the return of the benchmark portfolio that is used to compute beta risk.
(iii) The benchmark portfolio is on the mean-variance efficient frontier from the perspective of managers who lack superior information.

This result can be reconciled with ROLL's analysis of the relation between benchmark portfolio inefficiency and security market line analysis as follows: The efficient frontiers of informed and uninformed portfolio managers, who disagree about the means, variances, and covariances of asset returns, are different. In particular, the efficient frontier of informed portfolio managers always lies outside the efficient frontier of uninformed portfolio managers. Hence, as in Diagram 5, the slope of a line connecting the risk-free return to a benchmark portfolio satisfying condition (iii) will always be less than the slope of the efficient frontier of the informed managers. DYBVIG and ROSS (1985a) have shown that the sign of alpha will, on average, be positive whenever the slope associated with the evaluated portfolio is steeper than the slope associated with the benchmark portfolio. With the exception of some classes of cases analyzed by GRINBLATT and TITMAN (1986b) (where the alpha is positive, anyway), informed managers essentially choose portfolios with steeper slopes than the benchmark portfolio. Thus, ROLL's dilemma is resolved by the benchmark portfolio being both on and off the efficient frontier at the same time. Since it is on the efficient frontier of uninformed managers, it cannot (and should not) distinguish between passive 'buy and hold' investment strategies or between active strategies based on information that is readily available in the market. Since the benchmark portfolio is off the efficient frontier of informed managers, active strategies based on this superior private information plot off the securities market line. Even more importantly, they plot above the securities market line.



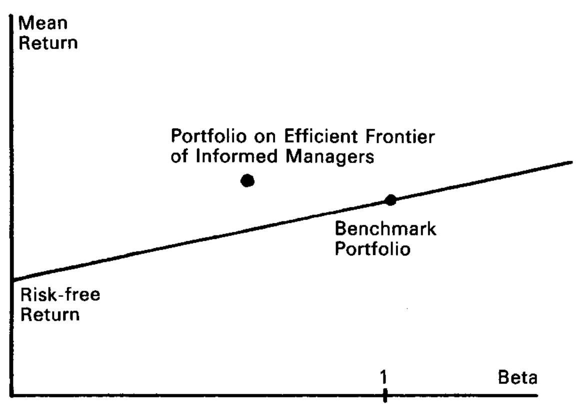


Diagram 5: Efficient frontiers of informed and uninformed portfolio managers and performance of informed managers with the JENSEN measure.

In practice, the sign of alpha is critical. Brokerage commissions, management fees, and embezzled funds reduce mean returns and may result in positions significantly below the securities market line for uninformed portfolio managers. If informed portfolio managers could plot below the securities market line, we would have difficulty distinguishing them from uninformed managers who incur these 'transaction costs' or from informed managers who invest in a manner contrary to that suggested by their information. In fact, there are classes of cases where informed investors can plot below the securities market line, even in the absence of transaction costs or contrary investment behavior. Because of this, some researchers have labeled the assumptions behind the MAYERS/RICE, DYBVIG/ROSS and GRINBLATT/TITMAN results as too restrictive to offer much hope to performance evaluators who wish to use securities market line analysis.

III. Performance Measures

ROLL's paper presented a tremendous challenge to financial researchers. Two lines of research sprang from it. The first attempted to identify alternative measures of performance that were not based on security market line analysis and were not sensitive to the choice of a benchmark portfolio. This line of research led to measures that employ data on the asset holdings of the portfolio. The second, which led to measures that did not use portfolio holdings data, attempted to better understand the relation between the choice of a benchmark portfolio, the nature of superior information, and performance measurement with the securities market line. There was the hope that under a plausible set of conditions, a benchmark portfolio might be found that would allow securities market line analysis to discriminate between superior and ordinary portfolio managers. As we saw above, this research was successful, but in a limited way, since the conditions required seemed too stringent. As a result, further research was undertaken on other measures that would be appropriate under more general conditions.

a) Performance Measure that Use Portfolio Holdings Data

CORNELL (1979) developed a measure that did not require the computation of beta or the selection of a benchmark portfolio. It did, however, require the observation of the historical sequence of portfolio weights (i.e. the investment choices) for the investment strategy being evaluated. CORNELL'S idea was very simple. A successful portfolio manager holds securities only during periods when their returns are higher than usual. Thus, observing the returns of a security when it is held in the portfolio of a manager and comparing it to the return on the same security when it is not held by the manager indicates whether the manager is purchasing the security at the right time.

Various measures can be developed from this idea. Suppose, for instance, that you want to evaluate a portfolio manager in 1983. At the beginning of each quarter, let's assume that the manager analyzes his investment choices and revises his portfolio weights. A benchmark return can be established for each of the four quarters in 1983 by applying the portfolio weights in that quarter to some future benchmark period. (Note that this concept is different from a 'benchmark portfolio'. To avoid confusion, this section refers to the latter as an 'index portfolio'.) The difference between the actual return in the quarter and the benchmark return, when averaged, is the measure of performance.

To make this example very simple, let's suppose that the manager restricts his investment to three stocks: IBM, Exxon, and General Motors. In the first quarter of 1983, the manager divides his portfolio equally between the three stocks. In the second quarter, the manager, believing that IBM will do the best, puts all of the portfolio's wealth into IBM; in the third quarter invests all of it in Exxon, and in the fourth quarter, invests all in General Motors. Table 3 will now illustrate the calculation.

Table 3: Computation of the CORNELL Measure

Time Weighted Quarterly Return in Excess of the Risk-Free return

	1983				1984			
Quarter:	1	2	3	4	1	2	3	4
Column:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
IBM	0.063	0.187	0.061	–0.030	–0.058	–0.063	0.176	–0.002
Exxon	0.062	0.119	0.103	0.044	0.053	0.079	0.123	0.023
General Motors	–0.041	0.227	0.020	0.024	–0.120	0.026	0.185	0.056

	Portfolio Manager's Weights (1983)				**Benchmark Return for securities** = Average of Col's
Quarter:	1	2	3	4	(5) – (8)
Column:	(9)	(10)	(11)	(12)	(13)
IBM	1/3	1	0	0	0.013
Exxon	1/3	0	1	0	0.070
General Motors	1/3	0	0	1	0.037

Actual and Benchmark Portfolio Returns (1983)

Quarter:	1		2		3		4	
	Actual	Benchmark	Actual	Benchmark	Actual	Benchmark	Actual	Benchmark
	(9)×(1)	(9)×(13)	(10)×(2)	(10)×(13)	(11)×(3)	(11)×(13)	(12)×(4)	(12)×(13)
Column:	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)
	0.028	0.040	0.187	0.013	0.103	0.070	0.024	0.037

CORNELL Measure

$$\{[(14)-(15)]+[(16)-(17)]+[(18)-(19)]+[(20)-(21)]\}\div 4$$

(22)

0.04550

Multiply by 400 to compute annualized CORNELL Measure (%) = 18.2%

The benchmark excess return for each of the three securities is found by averaging the four 1984 quarterly time-weighted rates of return for these securities (reinvesting any dividends and adjusting for splits) and subtracting the average quarterly risk-free return. The benchmark excess return for a portfolio is then found by multiplying the portfolio weights in that quarter by the benchmark excess returns of the respective securities and summing. This benchmark excess return is then subtracted from the actual excess return for the quarter to yield an abnormal return for the quarter. The four quarterly abnormal returns are then averaged.

Obviously, other benchmark periods could be used here. The benchmark return for a security in the first quarter of 1983 could be its return in the second quarter of 1983; the benchmark return for a security in the second quarter of 1983 could be its return in the third quarter of 1983, etc. CORNELL, in first proposing this measure, suggested that the benchmark returns for securities be computed in prior periods. In this example, this might consist of averaging quarterly returns for the three securities in 1982. This method is not advisable, however, because one can easily design an investment strategy that will beat it. For instance, by purchasing in 1983 the worst performing stocks of 1982, one is almost guaranteed to display a highly positive measure – not because your returns in 1983 are abnormally high, but because your benchmark returns from 1982 are abnormally low. Since one cannot, in general, forecast 1984 returns in 1983, no similar strategy can be designed for benchmarks based on future periods.

In a similar vein, GRINBLATT and TITMAN (1986b) developed two measures called the Selectivity Measure and the Timing Measure. In contrast to the CORNELL Measure, which does not require an index portfolio, the Selectivity and Timing Measure may be sensitive to the choice of the index portfolio. They are not as sensitive, however, to problems associated with the risk and expected returns of assets changing over time, which is technically referred to as 'nonstationarity'. The measure is computed as follows: First, the betas of individual securities are computed by running the usual time series regressions (one for each security) against a mean-variance efficient index portfolio. The portfolio-weighted average of these betas yields the portfolio's beta at any instant in time. In the example of Table 3, there would be four different betas for the investment strategy of 1983. In each holding period, subtract the product of the beta of the portfolio being evaluated and the excess return of the index portfolio from the actual excess return of the portfolio being evaluated. Average this difference over all holding periods to get the Selectivity Measure. Examples of the calculations are provided in Table 4, which uses the data from Table 3's calculation of the CORNELL Measure.

The Timing Measure also uses the portfolio betas that were described above for the Selectivity Measure. It computes the covariance between the time series of portfolio betas and the time series of excess returns of the index portfolio. The calculation is also illustrated in Table 4.

Table 4: Computation of the Selectivity and Timing Measures

Step 1) Compute the beta of each asset (assume this was done and the following results were obtained):

Firm	Beta (22)
IBM	0.9
Exxon	1.1
General Motors	1.0

Step 2) Compute the portfolio beta in each quarter of 1983 (this uses information from Table 3):

1983 Quarter:	1	2	3	4
	(9) × (22)	(10) × (22)	(11) × (22)	(12) × (22)
Column:	(23)	(24)	(25)	(26)
Portfolio Beta:	1.0	0.9	1.1	1.0

Step 3) Compute the quarterly excess returns of the index portfolio for 1983:

1983 Quarter:	1	2	3	4
Column:	(27)	(28)	(29)	(30)
(Assumed) Excess Return for Index Portfolio:	0.05	0.20	0.05	−0.02

Selectivity Measure

$$\{[(14)-(23)\times(27)]+[(16)-(24)\times(28)]+[(18)-(25)\times(29)] + [(20)-(26)\times(30)]\} \div 4$$

(31)

0.01925

Multiply by 400 to compute annualized Selectivity Measure (%) = 7.7%

Timing Measure

$$\{[(23)\times(27)]+[(24)\times(28)]+[(25)\times(29)]+[(26)\times(30)]\}\div 4 - \begin{bmatrix}\text{Avg. of}\\ \text{Cols. (23)–(26)}\end{bmatrix}\times\begin{bmatrix}\text{Avg. of}\\ \text{Cols. (27)–(30)}\end{bmatrix}$$

(32)

−0.00375

Multiply by 400 to compute annualized Timing Measure (%) = −1.5%

An interesting variant of the CORNELL Measure, developed by COPELAND and MAYERS (1982), was used to study the recommendations of the Value Line Investment Survey. This measure, illustrated in Table 5, subtracts a benchmark period's Selectivity Measure from the actual Selectivity Measure. In contrast to the Selectivity Measure, for which an approximately mean-variance efficient index portfolio is required, the COPELAND-MAYERS Measure is valid for any index portfolio. (This is because errors in actual and benchmark alphas that stem from the index portfolio lying inside the efficient frontier cancel out in the subtraction of the two Selectivity Measures.)

Table 5: Computation of COPELAND-MAYERS Measure

Step 1) Compute the quarterly excess returns of the index portfolio for 1984:

1984 Quarter:	1	2	3	4
Column:	(33)	(34)	(35)	(36)
(Assumed) Excess Return for Index Portfolio	0.10	−0.01	0.05	0.01

Step 2) Compute the Selectivity Measure for the 1983 portfolio in 1984:

$$\{[(15)-(23)\times(33)]+[(17)-(24)\times(34)]+[(19)-(25)\times(35)]+[(21)-(26)\times(36)]\}\div 4$$

(37)

0.001

Step 3) Subtract the benchmark period Selectivity Measure from the Actual Selectivity Measure:

COPELAND-MAYERS Measure

(31)–(37)

0.01825

Multiply by 400 to get annualized COPELAND-MAYERS Measure (%) = 7.3%

Under relatively mild assumptions, these measures have the following desirable properties: on average, they are positive for active investment strategies based on superior information and zero, on average, for passive investment strategies or strategies based on public information. This has been shown in a recent paper by GRINBLATT and TITMAN (1986b).

b) *Measures of Performance that Do Not Use Portfolio Weights*

Most academic experts in the area of performance evaluation would conclude that, where possible, measures that make use of portfolio weight information are superior to those that do not make use of this information. The four measures presented above and variations of them represent 'the state of the art' in simple, universally applicable performance measures. Unfortunately, data on portfolio holdings is very costly and time-consuming to acquire and often cannot be acquired. For this reason, researchers and practitioners are also interested in measures that do not require data on portfolio holdings.

Security market line analysis, which was discussed earlier, is also known as the JENSEN Measure; see JENSEN (1968, 1969). To refresh your memory, it is the intercept (or alpha) in a regression of the time series of excess returns of the portfolio being evaluated against the time series of excess returns of a benchmark portfolio. In contrast to a Selectivity Measure that uses the same benchmark portfolio to compute betas, where the observation of portfolio weights allows one to calculate a different beta as portfolio holdings change, security market line analysis computes only a single beta over all holding periods. If this single beta were the time weighted average of the betas in the Selectivity Measure, and if the sequence of betas in the Selectivity Measure were uncorrelated with the return of the index portfolio, then the JENSEN Measure, on average, would yield the same performance conclusions as the Selectivity Measure. In this case, the single beta in the JENSEN Measure represents an average of risk over time and the sequence of betas in the Selectivity Measure represents risk as a function of time. However, if the informed manager can forecast the returns of the index portfolio, his sequence of portfolio betas will be correlated with the return of that portfolio. This type of investment behavior, known as 'timing ability' in the mutual fund literature, implies that the single beta computed in the JENSEN Measure tends to over-state the average risk of the investment strategy. This bias may be so severe that the JENSEN Measure will erroneously designate superior portfolio managers as inferior.

To avoid this bias, the results of MAYERS/RICE, DYBVIG/ROSS, and GRINBLATT/TITMAN, discussed above, preclude superior information about the direction of the business cycle, interest rates, exchange rates, or other macroeconomic variables. For the JENSEN Measure to be valid, only information that is relevant to picking individual securities, like information about managerial talent, takeover candidates, or which of two firms will win a large contract, would be permissible.

Fortunately, recent research has found a way out of this state of affairs. GRINBLATT and TITMAN (1986b) have developed a new measure, called the Positive Period Weighting Measure, that, on average, is zero for uninformed managers and positive for informed managers, even when there is timing ability. This measure finds a time series of weights that sum to one and have two properties:

(i) They are non-negative.
(ii) They make the weighted average of the time series of excess returns of a portfolio on the mean-variance efficient frontier zero.

Once having found such weights, apply them to the excess returns of the portfolio being evaluated. That is, the Positive Period Weighting Measure is

$$w_1 (r_{p1} - r_{f1}) + w_2(r_{p2} - r_{f2}) + \ldots + w_T(r_{pT} - r_{fT})$$

where
w_t = weight for period t,
r_{pt} = period t return of the portfolio being evaluated,
r_{ft} = period t risk-free return.

An example is provided in Table 6.

Table 6: Computation of the Positive Period Weighting Measure

Step 1: Find nonnegative weights that make the weighted excess returns of the index portfolio equal zero in the evaluation period. In this case, the excess returns come from columns (27) to (30) in table 4. The following two equations must be solved for nonnegative weights w_1, w_2, w_3 and w_4.

eq. (1): $0.05\, w_1 + 20\, w_2 + 0.05\, w_3 - 0.02\, w_4 = 0$
eq. (2): $w_1 + w_2 + w_3 + w_4 = 1$

There are many solutions to this equation. One is

Weight:	w_1	w_2	w_3	w_4
Column:	(38)	(39)	(40)	(41)
	0.072	0.045	0.072	0.81

Step 2: Apply the weights to the excess returns of the portfolio being evaluated.

Positive Period Weighting Measure
(38) × (14) + (39) × (16) + (40) × (18) + (41) × (20)
(42)
0.03729

Multiply by 400 to get annualized measure (%) = 14.9%

V. Conclusion

Academic research on performance evaluation has focused on the analysis of risk and its effect on returns. Fair comparisons can be made only when risk premia are subtracted from returns. This presents a fundamental problem, in that financial theories are deficient at explaining the risk premia of securities. Moreover, traditional methods of obtaining risk, even if they could explain the risk premia of individual securities, exhibit important biases when attempting to measure the risk of dynamic investment strategies. In particular, the ability to 'time the market' generally introduces an upward bias in the traditional beta estimate for the risk of that strategy.

There are two ways out of this dilemma. One is to obtain data on portfolio weights. If the risk premia of securities are stationary, some performance measures allow us to track the risk premia of the investment strategy through time as the portfolio-weighted average of the risk premia of individual securities. The latter measures can be estimated from differences in sample mean returns (or variants) across securities and require no elaborate theory of asset pricing.

The second alternative has two components. First, by recognizing that the ex-ante mean-variance efficient portfolio from the perspective of uninformed portfolio managers places all of these managers on the securities market line (i.e. yields no abnormal performance for them), we go a long way toward liberating performance evaluation from asset pricing theory, which has many deficiencies. Second, by distinguishing between selectivity ability and timing ability, we can identify the class of cases in which superior performers plot above the securities market line. In the cases where they do not plot above, an alternative measure, the Positive Period Weighting Measure, has been developed that is positive, on average, for managers with superior timing and/or selectivity information.

The major challenge to researchers who use the latter approach is the identification of an ex-ante mean-variance efficient portfolio (or a group of portfolios that, with some weighting, add up to a point on the mean-variance efficient frontier). Procedures for deriving such a portfolio are not routine and take a considerable amount of theoretical intuition and empirical skill. Less complex is the procedure for testing the efficiency of the portfolio, which has been well documented in the literature. See, for example, GRINBLATT and TITMAN (1986a), GIBBONS (1982), SHANKEN (1985), or JOBSON and KORKIE (1982).

Another challenge to both approaches is the degree of randomness in securities prices. Although research has focused on the large sample properties of various measures of performance, these methods may not be powerful enough to yield meaningful inferences about performance for the typical sample sizes used.

Despite these reservations, financial research on performance evaluation holds lots of promise for practitioners. GRINBLATT and TITMAN (1986a), for instance, recently completed a large study of mutual fund investment performance and are in the process of completing another. The former study showed that the Positive Period Weighting Measure for 279 funds is highly correlated with the JENSEN Measure for those funds. However, where differences exist, they are significantly explained by measures designed to detect timing ability. The paper also showed that new tests with the JENSEN and Positive Period Weighting Measures can potentially surmount the problem of noisy data. For instance, it was possible to show that past performance is significantly related to future performance.

The benchmark portfolio selection procedure in GRINBLATT and TITMAN (1986a) also yielded some promising results. A benchmark consisting of eight portfolios, referred to as 'Port8', could not be rejected as mean-variance efficient. Most importantly, it seemed to properly indicate no performance for a variety of tests on 109 passive portfolios, constructed on the basis of public information. The same could not be said for more traditional benchmarks, like the S&P 500. While the details of how to select these benchmarks are beyond the scope of this paper, the discussion in GRINBLATT and TITMAN (1986a) should be accessible to anyone with a research orientation in finance.

Until the second study is complete, it is difficult to recommend a single evaluation technique that is best. Obviously, the optimal measure depends on the data that is available and on what type of performance is being measured. If there is data on portfolio holdings and the portfolio manager was instructed to be a 'stock picker', rather than a 'market timer', then you might attempt the Selectivity Measure or the COPELAND-MAYERS Measure. If you want to capture timing ability as well, then you would prefer the sum of the Selectivity and Timing Measures, the CORNELL Measure, or the Positive Period Weighting Measure. If you have poor (or no) data on portfolio holdings and are fairly certain that the manager has no timing ability, then the JENSEN Measure, which is fairly simple to compute, might be best.

On the other hand, there are other issues, like the sensitivity of the measures to various benchmarks, nonstationarities, and sampling error, that might also be important in choosing the right measure. These have not yet been thoroughly investigated in the finance literature.

To the extent that a financial institution has the resources to carry out a variety of evaluation techniques, the dilemma of which technique to use becomes less important. Despite our incomplete knowledge, however, it seems that even a rudimentary implementation of these techniques is better than the crude and ad hoc evaluation procedures currently in practice. Portfolio managers, being more technically sophisticated than most of their clients, find it easy to 'game' the evaluation techniques currently used. For example, we saw earlier that managers who buy the stocks of small firms can dramatically beat any evaluation technique based on raw returns or a CAPM-inspired benchmark portfolio (like the S&P 500). At the very least, the sophisticated evaluation techniques discussed in this paper are more difficult to deliberately 'game'. Thus, they provide an incentive for the security analysis and portfolio management of these managers to search for real rewards, rather than clues on how to beat an artificial measuring stick.

References

COPELAND, THOMAS, and MAYERS, DAVID (1982): 'The Value Line Enigma (1965–1978): A Case Study of Performance Evaluation Issues'. *Journal of Financial Economics,* 10, pp. 289–322.

CORNELL, BRADFORD (1979): 'Asymmetric Information and Portfolio Performance Measurement'. *Journal of Financial Economics,* 7, pp. 381–391.

DYBVIG, PHILIP, and ROSS, STEPHEN (1985a): 'The Analytics of Performance Measurement Using a Security Market Line'. *Journal of Finance,* 40, pp. 401–416.

DYBVIG, PHILIP, and ROSS, STEPHEN (1985b): 'Differential Information and Performance Measurement Using a Security Market Line'. *Journal of Finance,* 40, pp. 383–399.

GIBBONS, MICHAEL (1982): 'Multivariate Tests of Financial Models: A New Approach'. *Journal of Financial Economics,* 10, pp. 3–27.

GRINBLATT, MARK, and TITMAN, SHERIDAN (1986a): 'A Comparison of Measures of Abnormal Performance on a Sample of Monthly Mutual Fund Returns'. UCLA working paper.

GRINBLATT, MARK, and TITMAN, SHERIDAN (1986b): 'Portfolio Performance Evaluation: Old Issues and New Insights'. UCLA working paper.

GRINBLATT, MARK, and TITMAN, SHERIDAN (1987): 'The Relation Between Mean-Variance Efficiency and Arbitrage Pricing'. *Journal of Business,* 60, pp. 97–112.

JENSEN, MICHAEL (1968): 'The Performance of Mutual Funds in the Period 1945–1964'. *Journal of Finance,* 23, pp. 389–416.

JENSEN, MICHAEL (1969): 'Risk, the Pricing of Capital Assets, and the Evaluation of Investment Portfolios'. *Journal of Business,* 42, pp. 167–247.

JOBSON, J. D., and KORKIE, BOB: 'Potential Performance and Tests of Portfolio Efficiency'. *Journal of Financial Economics,* 10, pp. 433–466.

MARKOWITZ, HARRY (1959): *Portfolio Selection: Efficient Diversification of Investments.* Wiley.

MAYERS, DAVID, and RICE, EDWARD (1979): 'Measuring Portfolio Performance and the Empirical Content of Asset Pricing Models'. *Journal of Financial Economics,* 7, pp. 3–29.

ROLL, RICHARD (1978): 'Ambiguity When Performance is Measured by the Securities Market Line'. *Journal of Finance,* 33, pp. 1051–1069.

SHARPE, WILLIAM (1966): 'Mutual Fund Performance'. *Journal of Business,* 39, pp. 119–138.

TOBIN, JAMES (1958): 'Liquidity Preference as Behavior Towards Risk'. *Review of Economic Studies,* 25, pp. 65–85.

TREYNOR, JACK (1965): 'How to Rate Management of Investment Funds'. *Harvard Business Review,* 43, pp. 63–75.