Research Program in Finance
Graduate School of Business Administration

WORKING PAPER NO. 58

Security Appraisal and Unsystematic Risk in Institutional Investment

by

Barr Rosenberg
RESEARCH PROGRAM IN FINANCE AT THE
WALTER A. HAAS SCHOOL OF BUSINESS,
UNIVERSITY OF CALIFORNIA, BERKELEY

The Research Program in Finance in the Walter A. Haas School of Business at the University of California has as its purpose the conduct and encouragement of research in finance, investments, banking, securities markets, and financial institutions. The present reprint and working paper series were established in 1971 in conjunction with a grant from the Dean Witter Foundation.

INSTITUTE OF BUSINESS AND ECONOMIC RESEARCH
Carl Shapiro, Director

The Institute of Business and Economic Research is an organized research unit at the University of California, Berkeley, whose mission is to promote research by faculty and graduate students in the fields of business and economics. The Institute carries out its mission by organizing programs and activities that enrich the research environment, administering extramural research awards, publishing working papers, and making direct grants for research.
SECURITY APPRAISAL AND UNSYSTEMATIC RISK
IN INSTITUTIONAL INVESTMENT

by

Barr Rosenberg*

ABSTRACT

Stock selection within the equity portfolio reflects an attempt to exploit judgmental appraisals of common stocks to achieve superior return. Superior return (otherwise referred to as appraisal premium or positive alpha) is one of the services promised by an institutional investor who manages a client's portfolio, a promised benefit against which must be debited three forms of cost: management fee, transaction costs, and the unsystematic risk of the investment strategy. This paper addresses the question of how much unsystematic risk is warranted in the pursuit of superior return. In simple terms, how aggressively should the manager tilt the portfolio toward stocks that are believed to be undervalued and away from stocks that are believed to be overvalued? This problem is distinct from the choice of appropriate systematic risk (beta) but the solution is partially determined by that choice. The analysis attempts to resolve two interlocking questions: What investment goals should the client transmit to the manager, and how should the money manager determine the appropriate level of unsystematic risk? The recommended solution contradicts existing practice.

*Associate Professor, Schools of Business Administration, University of California, Berkeley. The insights in the paper have emerged in discussions with colleagues at Berkeley and Stanford, and with friends and clients among institutional investors. All those interchanges are gratefully acknowledged. Ellen McGibbon's assistance in preparing the manuscript was indispensable.
INTRODUCTION AND SUMMARY

Institutional investment is an important function within the capital markets. Personal savings are increasingly entrusted to banks, employee benefit funds, insurance associations, mutual funds, and endowment funds. Each pool of funds is entrusted to one or more "institutional investors" for management. The institutional investor thus assumes the role of "money manager" for his clients, and determines the investment strategy for that portion of the clients' money which he manages.

The client must somehow select one or more managers and apportion his funds among them. Also, he must instruct them so that they can invest optimally on his behalf, and he must monitor their performance. The investment goals transmitted to the manager must provide appropriate guidance without interfering with his special investment abilities. The difficulty of the problem is multiplied when the work of more than one manager must be coordinated. These complexities have given rise to consultants who advise on the selection of money managers, upon the appropriate investment goals to transmit, and upon the monitoring process.

The client's decisions with regard to the portfolio as a whole are considered below. The approach closely resembles the pioneering work of Treynor and Black (1973). A complete solution is provided for the simplest case, where the entire equity portfolio is managed by one manager. The sequel to this paper, "Institutional Investment with Multiple Portfolio Managers," covers the more common case. The multiple-manager problem includes a higher-level decision that is identical to the subject of this paper: the choice of overall optimal strategy, once the separate managers have been optimally coordinated.

The client's instructions to the single manager are expressed below as four parameters: one relating to systematic risk, the "normal portfolio beta"; and three relating to unsystematic risk—the "risk acceptance parameter," the "covenant information ratio," and the "active investment proportion." In the multiple-manager case, under
reasonable assumptions, slightly modified versions of these parameters, and one additional parameter, suffice to achieve the optimal decentralized policy.

The fundamental decision for the client is to select the appropriate exposure to the return on the "market portfolio" through the portfolio "beta." Greater exposure offers greater expected reward, but at the cost of greater risk. The client must determine the optimal beta: the one which provides the most satisfactory combination of risk and reward. The risk which is taken as a result of this decision is the "systematic risk" of the portfolio. Ordinarily it will amount to 90 percent or more of total risk.

Properly framed, the decision as to appropriate systematic risk is a long-term decision, and the forecast reward and risk for the market portfolio are those which characterize a long future period. Special forecasts of abnormal reward in the near future, to be exploited through "market timing," are distinct from this. Since the horizon is a long one, the prediction of reward and risk of the market are strongly determined by the century-long history of American capital markets, during which there has been little substantial change in real reward or risk. Hence the reward and risk opportunities perceived in the market portfolio are likely to approximate consensus values.

However the choice of beta is made, an essential parameter, the "risk acceptance parameter" (RAP) can be deduced from the chosen beta. To each forecast of the risk and reward opportunities of the market, corresponds a "reward to variability ratio" (RVR). The RVR, together with the RAP, determine the chosen beta. Conversely, for a given beta, the RAP can be deduced from the RVR. Since the RVR may safely be assumed to approximate a known consensus value, a reasonably accurate assessment of the RAP may be deduced from the chosen beta alone.

The RAP which applied to the choice of beta is informative concerning the client's other choices between risk and reward. In particular, it is relevant to the determination of exposure to unsystematic risk.
The decision with regard to unsystematic risk has been phrased in terms of a choice between active and passive management. At the extreme of passivity, the portfolio may be wholly invested in a proxy for the market portfolio, analogous to an index fund, with no unsystematic risk. As investment becomes more active, managers take positions that differ increasingly from the market portfolio, so as to profit from the information embodied in their appraisals. The difference between the portfolio holdings and a (risk-adjusted) market portfolio is the "active portfolio." Some active holdings are positive, and some are negative. All active holdings can be scaled up or down by a common "aggressiveness factor"; the result is an aggressive strategy, if that factor is large, or a cautious strategy, if that factor is small. The factor is thus a controllable determinant of the aggressiveness of active management.

The decision as to appropriate aggressiveness may be usefully framed in terms of fixed cost or overhead, and variable costs. The fixed cost is the management fee, paid to the manager, which is usually a fixed percentage of funds under management. The passive strategy possesses minimal fixed cost, while the fixed cost of active management is higher; the cost increase, or "active management fee," is presumably employed by the active manager to support the appraisal and investment process.

The active strategy possesses one benefit and one cost which vary in direct proportion to the aggressiveness factor. The benefit is the portfolio "alpha," the expected reward from the profitable exploitation of appraisals. (Of course, this benefit would be a debit if the manager were inferior to the average, but we are here presuming that the client has somehow identified a superior manager.) The cost is the transaction cost of carrying out the strategy. When transaction costs are subtracted from the gross portfolio alpha, the result is a "net alpha." The net alpha also changes in linear proportion to aggressiveness.

The final cost of an active strategy is the risk that the expected appraisal premium will not be obtained, "the unsystematic risk" of the portfolio. The standard deviation of this risk also increases linearly
with aggressiveness, so its variance increases as the square of aggressiveness. Unsystematic variance is the important construct, from the point of view of the total portfolio, because it adds to the systematic variance to determine total variance.

One central point of this paper concerns the evaluation of unsystematic risk. How severe a problem is it? How should the client evaluate it? The conclusion is quite simple: the risk acceptance parameter which applies to the evaluation of systematic risk should be a lower bound for the RAP applied to unsystematic risk. Once the RAP is known, it is a straightforward matter to convert the residual risk to an equivalent "risk premium" or "disutility of risk," defined as the amount of certain reward which is required to offset that risk. The RAP may also be applied to compute the optimal aggressiveness factor, that value which achieves an optimal trade-off between the net portfolio alpha and the unsystematic variance.

The funds are placed under active management because of the client's belief in the manager's appraisal ability. This belief must be expressed explicitly, if the manager is to know how to manage the client's money optimally. The natural construct by which to express such ability is the reward to variability ratio for the manager's appraisal ability, equal to the expected reward from his differential policy, divided by the standard deviation of the unsystematic risk to which that policy is exposed. Client and manager should reach a covenant concerning this ratio, which may be called the "covenant information ratio" (CIR). The higher this ratio, the greater is the expected reward to the client from active management.

The appropriate level of unsystematic risk for the active portfolio is determined by three parameters: the RAP, the CIR, and the proportion of total funds which are actively managed \( w_A \). The investment weight enters the equation because of the diversification effect. As funds are diverted to passive management, the exposure to unsystematic risk from active management declines more rapidly than the exposure to
appraisal reward; consequently it becomes beneficial to take a more aggressive strategy on behalf of the client within the part of the portfolio that is actively managed.\footnote{The "dependence adjustment factor" also enters the equation in a multiple-manager context. It takes account of the fact that the information processes of two different managers may or may not be completely distinct. If two or more managers are responding to similar information and reaching appraisals in a similar manner, the appraisals which they produce will be correlated, and each may be dependent on the others. The greater the interdependence among managers, the smaller will be the value of each manager's information to the client. This contraction in the value of information is accomplished by the DAF.}

When these three parameters are taken into account, the appropriate level of unsystematic risk for the portfolio can be computed. Through this approach the client can instruct the manager as to the appropriate aggressiveness of his differential policy. Figure 1 shows the relationship between client and money manager schematically.

Conversely, if we know any two of these parameters and also know the unsystematic risk level, we can deduce the third parameter. This reverse approach is apt and powerful, because it can be used to infer from publicly available data the underlying CIR. This strategem leads to conclusions which are somewhat revolutionary in their implications for investment practice. For this reason, it is important that the argument underlying the approach is straightforward, robust, and reasonable in its assumptions.

Once the argument is accepted, a simple procedure for the assessment of active portfolio strategies is available, as follows: (1) infer the client's RAP from the chosen beta of the total portfolio (or the equity portion thereof); (2) estimate the unsystematic risk of the active strategy, and employ the RAP to deduce the CIR which would exactly warrant
Forecast expected market excess return ($\mu_M$) and variance of market return ($\sigma_M^2$). These should be long-term forecasts of the "normal" environment.

Determine optimal exposure to market return ($\eta\beta$) for the total portfolio, and partition total funds among cash, bonds, and equity. Deduce "risk-acceptance parameter" ($\lambda$).

Evaluate equity manager's appraisal ability. Determine the "covenant information ratio" for the active policy, $z$.

Partition equity funds between active and passive management. Let $w_A$ be the proportion of the total portfolio actively managed.

Determine warranted aggressiveness in the active portfolio from $\lambda$, $w_A$, and $z$.

The expected appraisal premium ($\alpha_A$) and unsystematic standard deviation ($\omega_A$) should be:

$$\alpha_A = \frac{\lambda}{w_A} \cdot z^2$$
$$\omega_A = \frac{\lambda}{w_A} \cdot z$$
the existing level of unsystematic risk; (3) compute the implied portfolio alpha and estimate the net utility to the client from active management. Thus the expected appraisal premium which would be required to lead to the chosen level of aggressiveness is inferred. This may be termed the required portfolio alpha. In short, from a few basic characteristics of the portfolio, the client's assessment of his manager's appraisal ability has been inferred!

As one application, the disutility of unsystematic risk may be expressed in basis points, and compared with management fee and transaction costs. When this is done, it is found that the disutility of unsystematic risk is typically smaller than active management fee or transaction costs. Consequently, among the three reasons for abandoning active management and moving to a market fund, added diversification against unsystematic risk is less important than the avoidance of management fee and transaction costs.

A much more powerful conclusion is found by comparison of the inferred benefit from the net alpha of active management, with the increase in fixed fee entailed in such management. This comparison implements an elementary test of rationality: if the expected benefit from active management does not exceed the increased management fee, the active management strategy should be abandoned in favor of a market fund strategy.

Thus we should find that in all cases of rational behavior by the client, the expected benefit is at least as great as the management fee. Based upon representative characteristics of employee benefit funds, the opposite appears to be the case. The implied benefit appears to be frequently smaller than the management fee.

The key to this paradoxical situation seemingly lies in the client's irrationally exaggerated aversion to unsystematic variance. Unless he is forcefully reminded of the comparability of systematic and unsystematic variance, there are three plausible reasons, inherent in current institutional arrangements, which would tend to lead to an exaggerated aversion to unsystematic risk. Such exaggerated aversion, in turn, would lead to overly cautious policy in regard to active management. The effects of such
excessive caution are easily understood: the management fee is unchanged, so fixed costs remain the same; but the net benefit from the manager's appraisal ability is reduced, and may no longer be adequate to cover the fixed cost of the operation.

Recognition of the equivalence between systematic and unsystematic variance must therefore have one of three results: (1) the abandonment of active management; (2) the continuation of active management at a lower management fee; or (3) the continuation of active management at a more aggressive level, with greater unsystematic risk.

The following sections of this paper cover these topics in greater detail, as follows: section 1, the choice of systematic risk and the RAP; section 2, the appraisal process, the active strategy, the CIR, and the choice of unsystematic risk; section 3, the relationship between systematic and unsystematic risk; section 4, inferences concerning existing practice. The body of the paper is couched in verbal arguments, illustrated by geometric diagrams. The appendix derives the underlying relationships via calculus and matrix algebra.

1. The Appropriate Exposure to Equity Returns

It will be important to distinguish among three aspects of investment strategy, with each of which is associated a potential reward and an associated risk. These aspects, and the elements of risk and return with which they are associated, are respectively: (1) the normal systematic risk level or beta, associated with long-run expectations of return to the market portfolio; (2) fluctuations in beta for purposes of market timing; (3) active holdings, different from the market portfolio, established with the intention of profiting from forecasts of common factors in security returns and the specific returns of individual companies.

The normal beta (NB), is determined as the basic element of investment strategy. The client considers alternative mixes of stocks, bonds, and cash, and selects that mix which appears to best meet his needs. Often this choice is made with reference to the historical performance of
diversified capitalization-weighted stock and bond portfolios, akin to "market" portfolios for these asset categories. Because of the far greater risk and expected return of stocks, the ultimate choice is dominated by the readiness of the investor to bear the risk of the equity market.

To understand the relation between the asset mix and portfolio beta, some background is necessary. The client's total portfolio may include investments in cash, government bonds, corporate equity, corporate senior liabilities, real estate, foreign assets, and possibly other miscellaneous categories. For present purposes, it is best to focus upon domestic financial assets: cash, government bonds, and corporate liabilities. These may be grouped according to systematic risk into three categories: cash and cash equivalents (essentially zero beta); bonds (low beta); and equities (wide range of beta). All corporate liabilities with a conversion possibility into equity are grouped with equity, and all non-convertible preferred stocks are grouped with bonds. The three curves in figure 2 roughly outline the distribution of beta across different assets in each category, with each asset weighted by its market value (capitalization).

Cash and cash equivalents include short-term treasury notes, certificates of deposit and commercial paper, and government and corporate nonconvertible bonds with maturities of up to five years. Almost all of these assets have betas that are zero, although the beta rises toward about .15 as maturity increases. The exceptions are corporate liabilities, where there is a significant probability of default, so that the value of the debt instrument moves with the fortunes of the firm and hence acquires the characteristics of an equity holding.

The category of long-term bonds and nonconvertible corporate liabilities again offers a restricted range of betas, with the betas of individual assets ranging between .15 and perhaps .4. Again, the major exceptions are the few long-term liabilities where there is a significant probability of default.

Finally, for common equity and convertible senior liabilities, there is a much wider beta range. The great bulk of capitalization falls
FIGURE 2

THE APPROXIMATE DISTRIBUTION OF "CASH," "BONDS," AND "STOCKS" BY "BETA," WITH EACH ASSET WEIGHTED BY MARKET VALUE

Note: Throughout this paper, beta is (mis)defined as the familiar regression coefficient on the S&P 500 index, rather than as the regression coefficient on the true market portfolio.

Within the range of .5 to 2.0. Lower beta assets are typically convertible senior liabilities, or regulated companies such as utilities, which become almost bondlike; also some industries, such as gold, show a low beta because the related business risk is little correlated with risks elsewhere in the economy.

Portfolio beta is the investment-weighted average of the individual asset betas. Since the beta of cash and bond categories is so small, relative to the beta of equities, portfolio beta is largely determined by the amount invested in equities. Thus, in selecting the asset mix, the client effectively selects beta. Usually, the asset-allocation decision occurs under the hypothesis that the risk of the equity portfolio equals that of the equity market. When this is not the case, the second important determinant of portfolio beta becomes the chosen beta of the equity subportfolio.
The overall portfolio beta, $\beta_p$, approximately equals $w_E \beta_E$, where $w_E$ is the investment proportion in equity, and $\beta_E$ is the beta of the equity subportfolio.

Granted that the client, by whatever process, has chosen the appropriate beta, what can be deduced concerning his preferences with regard to risk and reward? To explain this relationship, a few terms are needed: Let $r_F$ be the rate of return on the riskfree asset, and let $\mu_M$ denote the expected excess return on the market portfolio (in excess of this riskfree rate). Let $\sigma_M^2$ denote the variance of the market portfolio. Similarly, let $\mu_{PS}$ and $\sigma_{PS}^2$ denote the expected excess return and variance of return for the portfolio (subscript P) from systematic return (subscript S). Figure 3, a graph in four quadrants, shows the "attainable frontier" of efficient reward-risk combinations which can be attained through systematic returns. To each value of beta, there corresponds one point on this frontier, that is, a particular combination of reward and risk. In choosing beta, the client is expressing a preference among the points on this attainable frontier.¹

The shape of the attainable set is determined by the parameters $\mu_M$ and $\sigma_M$, the mean and standard deviation of market return. Their ratio, $\mu_M / \sigma_M$, is termed the "reward to variability" ratio for systematic return. Since the choice of normal beta is made in the context of exceedingly long-term forecasts of the behavior of investment returns, forecasts of $\mu_M$ and $\sigma_M$ are likely to approximate consensus values.² For purposes

¹If the investor cannot obtain a higher-than-unity beta by borrowing and leveraging the market portfolio, then this must be accomplished by selecting an imperfectly diversified portfolio of equities, weighted in the direction of high beta assets. The imperfect diversification results in unsystematic risk. The portfolio standard deviation will then increase more than linearly with beta, as the beta increases beyond one, and the attainable frontier will flatten somewhat beyond the market variance, reflecting the greater variance required for an expected reward beyond the market reward. In practice, this effect is not substantial until beta increases beyond 1.3.

²An informal survey of about two dozen institutions resulted in values of between 4.5 percent and 7 percent for $\mu_M$, and 18 percent to 22 percent for $\sigma_M$. 
Each quadrant graphs a separate relationship, and each of the four axes plots a separate variable. Starting from any point on the left-hand horizontal axis (a beta value), a rectangle traced in the manner of the dashed line will locate the corresponding point on the risk/reward frontier.

Of illustration, 6 percent per annum for $\mu_M$, and 20 percent per annum for $\sigma_M$ will be used.

In figure 4, the attainable combinations of mean and variance of systematic return are repeated, this time superimposed on a hypothetical set of linear indifference curves. A higher indifference curve, which
gives greater expected return at each level of risk, is of course preferred.

The equation of an indifference curve may be written as:

\[ U = \mu_P - \lambda_Y \sigma_P^2, \]

where \( U \) is the level of utility provided by portfolio return distributions lying on the curve, and \( \lambda_Y \) is the slope of the indifference curve, equal to the cost, in terms of certain excess return, of one unit of variance.

The highest level of satisfaction for the client is reached at the point where an indifference curve is tangent to the attainable frontier. Now, we may not know the underlying preferences of the client, but we do know beta and therefore we can find the point on the attainable frontier that has been chosen. Since the indifference curve must be tangent to the attainable frontier at that point, it follows that we also know the slope.
of the indifference curve at that point. The slope of the indifference curve expresses the investor's tradeoff between variance and expected return, or between risk and reward. Hence, knowledge of the chosen beta allows the applicable risk-reward tradeoff to be inferred.

Further, since the indifference curves are linear, knowledge of the slope at any point allows us to find the intercept on the vertical axis. That intercept corresponds to the riskless portfolio which is indifferent to the chosen portfolio. The height of that intercept is the amount of return, which, if riskless, would be indifferent to the risky return offered by the client's chosen exposure to the equity market. The height above the riskfree rate is the certain excess return which is indifferent, or the "certainty equivalent excess return" (CEER). For linear indifference curves, the CEER is always equal to one-half of $\mu_{PS}$.

Thus the client's selected exposure to the systematic risk of the equity market results in an increase in expected return of $\mu_{PS} = \beta_P \mu_M$, at the cost of systematic variance $\sigma_{PS}^2 = \beta_P^2 \sigma_M^2$, with the same degree of satisfaction as would be provided by a CEER equal to $\frac{1}{2} \mu_{PS}$. In other words, of the total increase in expected return of $\mu_{PS}$, one-half is lost as the disutility of the systematic risk, and the other half survives as the CEER.

Linear indifference curves allow us to quickly deduce the CEER of a risky portfolio, and hence are a very convenient assumption to make. However, to determine the risk-reward tradeoff, linearity of the indifference curve is not essential. When the indifference curve is curvilinear, the slope of the indifference curve at the optimum equals the known slope of the frontier at the chosen beta. The slope of the indifference curve determines the risk-reward tradeoff that applies to a small change in the portfolio.

Moreover, unless the curvature of the indifference curve is very sharp at the optimum, the indifference curve in a small region around the point approximates a straight line. Hence, when a small change in portfolio variance is being studied, the approximate resulting CEER can be found by assuming the indifference curve is linear. This fact will be very useful in the following sections, where the CEER of unsystematic
risk is considered. Unsystematic variance is rarely as much as 10 percent of total variance, so that a large proportional change in unsystematic variance is a "small change" insofar as the total portfolio is concerned.

One further aspect of the solution is important. This is the "risk acceptance parameter" (RAP), defined as the ratio of variance to expected excess return, and denoted by \( \lambda \):

\[
\lambda = \frac{\sigma_{PS}^2}{\mu_{PS}} = \frac{\beta_p^2 \sigma_M^2}{\beta_p \mu_M} = \frac{\beta_p \sigma_M^2}{\mu_M}.
\]

\( \lambda \) is the reciprocal of the slope of the dashed line drawn in figure 4. The slope of that line is one-half of the slope of the linear indifference curve.\(^1\) Hence the risk-reward tradeoff may be summarized either by the slope of the indifference curve, \( \lambda_Y \), or by the RAP \( \lambda \), with \( \lambda = \frac{1}{2 \lambda_Y} \).

The RAP is much more useful, when we turn to the next topic, the analysis of unsystematic risk.

The expression for \( \lambda \) was obtained from the portfolio as a whole, including cash, bonds, and equity. This approach is justified in the appendix (A.4, A.5, and A.6). Alternatively, we may be concerned with an analysis of the equity component alone, with risk acceptance defined within this pool of funds ignoring the balance of the portfolio. It is shown in (A.6) that an analysis of the equity portfolio, using the formula

\[
\lambda \propto \frac{\beta_E \sigma_M}{\mu_M},
\]

and ignoring the nonequity components entirely, provides a good approximation of the correct procedure. In other words, the properties of the

\(^1\)Over the horizontal distance \( \sigma_{PS}^2 \), the indifference curve rises by \( \frac{1}{2} \mu_{PS} \), while the dashed line rises by \( \mu_{PS} \). Hence, the slope of the latter is twice as great. When the indifference curve is curvilinear, the slope of dashed line is still twice the slope of the indifference curve at the point of the optimum portfolio.
optimal equity portfolio, after it has been optimized as a component of the total portfolio, are as if it had been treated as a separate entity and optimized with RAP as given above.

Since equity portfolio betas do not vary widely, it follows that the RAP within the equity portfolio, \( \beta_{EM}^2 \), changes little from one client to another. Great differences between clients' RAPs do exist, but these are largely expressed in different proportional equity investments, not in different risk acceptances within the equity portfolio.

2. Active Portfolio Management and Unsystematic Risk

The client can maintain his equity portfolio as a "market portfolio" without unsystematic risk. Alternatively, some or all of the equity funds can be placed under active management. In active management, the client pays a higher management fee to support an appraisal process and an investment process.

The manager's appraisal process begins with valuations of common stocks, resulting in designations of undervaluation and overvaluation. Equally important, the manager establishes the estimated information content or validity of the valuation process. This may be summarized, for example, by the correlation between forecasts of subsequent returns and actual subsequent returns. The process of establishing information content is an exceedingly subtle one in which historical performance measurement can, at most, play an auxiliary role. Important insights come from a comparison of the process whereby information is generated and analyzed with procedures of other market participants, to determine where and to what degree the manager may have a comparative advantage.

Given this information, there are formulas to convert the forecasts of return to unbiased predictions of abnormal return, or "judgmental alphas." These were introduced by Treynor-Black (1973) and Ambechtsheer, among others, and are considered from a statistical perspective in sections 7 through 9 of the appendix. The outcome is a set of unbiased judgmental alphas, otherwise termed "appraisal premia," for the stocks in the universe.
The investment effects of transaction costs can be approximated by adjusting the alphas to reflect net returns, after deduction of the costs of creating and then eliminating a position in the asset. The result will be a forecast of net profit potential, rather than gross. In the balance of the paper, it will be assumed that the appraisal premia are defined net of transaction costs.¹

The investment process implements these judgmental alphas in a portfolio. The portfolio profit from appraisals is determined by its "differential holdings," the differences between portfolio holdings and those of a passive strategy. For any given portfolio beta, say, β_p, the passive strategy would be a levered market portfolio, with proportional investment of (1 − β_p) in the risk-free asset and β_p in the market portfolio. Thus, the passive holding in each asset, say, asset n, is β_p times the market proportion in that asset, or $h_{Mn}$, where $h_{Mn}$ is the market proportion. If $h_n$ is the portfolio holding, then $δ_n = h_n - h_{Mn}$ is the active portfolio holding.

Another aspect of active management is market timing. This may be treated as another element in the active strategy, with the active position being the difference between the current beta and normal beta. The appraisal premium is the expected difference between the near-term market return and the long-term forecast of return, and the associated unsystematic risk is the outcome of the abnormal exposure to market variance.

Let $k_A$ be a constant, the "aggressiveness factor," which will be considered later. All optimal active holdings are shown in (A.4) to vary

¹There is an implicit assumption that the unit transaction cost does not vary with the size of the transaction. This is certainly not exact, but the approximation is reasonable for present purposes.

Some transactions in active management would also occur in passive management; these are transactions which coincide with contributions or withdrawals. The passive management transaction costs are analogous to a fixed cost. Since this component is usually a small proportion of active transaction costs, it is included in variable costs to simplify the argument. The true "break-even point" for active management is thereby understandable.
proportionally to \( k_A \), so that \( \delta_n = k_A c_n \), for a set of constants \( c_n \).

When the alphas are properly constructed, the necessary additional data to determine the \( c \)'s are predictions of unexplained variance of returns. As section A.9 demonstrates, the variance is essentially identical to the total investment risk in the assets, and may be approximated thereby. The manager, by means of his alphas and a prediction of investment risk, computes the constants \( c_n \) for the assets in his universe.

As the factor \( k_A \) is increased, active holdings (whether positive or negative) grow in proportion. The expected extraordinary (risk-adjusted) portfolio return, or portfolio alpha \( \alpha_p \), increases proportionally. Also, the unsystematic standard deviation of the portfolio (to be denoted by \( \omega_p \)) increases proportionally.\(^1\) The ratio of these two \( \alpha_p / \omega_p \) is the reward-to-variability ratio of the unsystematic strategy. The RVR remains constant and is not affected by aggressiveness.\(^2\) Thus, the

---

\(^1\) The unsystematic standard deviation of the portfolio, as employed in this paper, is the standard deviation of the difference between the annual portfolio return and the annual return to the equal-beta levered market portfolio. From a statistical point of view, it is the "residual return" relative to the market portfolio, and may alternatively be called the "residual return." In a time-series regression of portfolio return on the market portfolio return, this would correspond to the annualized "standard error" of the regression. It is twice as great as the quarterly standard deviation, and 3.46 times as great as the monthly standard deviation.

\(^2\) As the active position grows larger, some holdings may run up against legal constraints: Maximum position sizes are sometimes established by law or convention, and since short-selling is prohibited, legal minimum position sizes are usually zero. (Notice that such constraints can be relaxed within the environment of a multiple-managed portfolio containing a "market-like" inventory pool; the individual manager can effectuate a large active holding by causing an active holding in that pool in addition to his managed portfolio.) When such constraints exist, the risk-reward trade-off becomes less favorable as the aggressiveness parameter forces positions against the constraints, and less satisfactory positions must be substituted. The result is that the RVR declines gradually with increased aggressiveness, so that the efficient frontier tails off more sharply than would otherwise be the case.
choice of aggressiveness is a choice of exposure to the fixed RVR of the unsystematic appraisal-generated return. The situation is exactly analogous to the case of systematic risk, with $k_A$ replacing $\beta_P$, and with the RVR of the appraisal-generated return replacing the RVR of systematic return. Figure 5 shows the development of the attainable frontier of appraisal-generated risk and reward.

The crucial determinant of the shape of the frontier is the RVR of the appraisal-generated return. This determines the opportunities to the investor. (It is meaningless to talk about the portfolio alpha in a vacuum, since a more extreme alpha can be achieved with greater aggressiveness.) The single most important product of the client's evaluation of the manager is a conclusion as to the RVR which the client believes the manager can achieve. This conclusion, which should be disclosed to the client, may be termed the "covenant information ratio, or CIR."\(^1\)

$$\text{CIR} = \frac{\text{expected appraisal-generated unsystematic reward}}{\text{standard deviation of resulting unsystematic return}}$$

The CIR expresses the client's and manager's agreement concerning the risk-reward trade-off attainable in a portfolio invested by that manager.

It is interesting to consider the changes that occur when part of the portfolio is actively invested and the balance is in an index fund. Let $w_A$ denote the portion that is actively managed, and let $\alpha_A$ and $\omega_A$ denote the appraisal premium and unsystematic standard deviation of the managed portfolio.\(^2\) Then, $\alpha_P = w_A \alpha_A$ and $\omega_P = w_A \omega_A$. Thus, the mean

---

\(^1\)The square of this ratio, a "Sharpe Ratio," has been called the "appraisal ratio" (Treynor-Black, 1973, and Ferguson, 1975). The information ratio, being the ratio of a mean to a standard deviation, is more meaningful. It is analogous to a standardized variate in statistics, and its reciprocal is called the "Coefficient of Variation." Hence, it is natural to call it a "Coefficient of Information" or "Information Ratio."

\(^2\)A related problem is that where part of the portfolio is managed by one manager whose strategy is under examination (manager i), and the balance is managed by others, with some other manager also being an active manager. The only difference between that situation and the one
When 100% of the portfolio is under active management, a rectangle traced from $k_1$ locates a point on the attainable frontier. Alternatively, when only 50% of the portfolio is under active management ($w_A = .5$), the broken lines give $\alpha_p$ and $\omega_p$ as functions of aggressiveness in the active portfolio. Then a rectangle traced in the manner of the dotted lines, starting at the point of doubled aggressiveness ($k_2 = 2k_1$), results in the same point on the frontier.

being considered here is that the balance of the portfolio is not a perfect market portfolio in that case, but instead is an active strategy driven by other managers' appraisals. Another dimension then enters the situation—namely, possible interdependence among managers' appraisals. The information content for manager $i$ may be adjusted for such dependence by means of the "dependence adjustment parameter" $b_i$. In the special case, where the manager's information is independent of other managers, the adjustment factor is unity and may be ignored. In that case, the analysis of the manager's risk-reward trade-off can be undertaken separately from the others. With linear indifference curves, it proceeds as if all other funds were passively invested: the proportion under management by the manager in question, $w_i$, replaces $w_A$ (Rosenberg, 1977a).
and standard deviation of portfolio unsystematic return are both reduced by the same factor, \( w_A \). For example, the broken lines in figure 5 represent the new relationships that would apply if \( w_A \) were one-half. When the functional relations in figure 5 are traced through, with this new value of \( w_A \), the efficient frontier is found to remain unchanged. In other words, a division of the portfolio between active and passive management does not change the attainable frontier at all. What is does accomplish is to provide the same risk-reward combination as could previously be obtained, at a lower proportion of funds under active management and at a higher level of aggressiveness within the managed portfolio. For example, if \( w_A \) is reduced to one-half, the same risk-reward combination is obtained by doubling the aggressiveness coefficient \( k_A \). Since the management fee for passive funds is lower, the end result is to reduce the total management fee.

Figure 6 illustrates the client's choice of optimal aggressiveness. As in the case of systematic return, the risk-reward trade-off may be expressed in terms of the slope of the indifference curve \( \lambda_V \).

**FIGURE 6**

THE CHOICE OF OPTIMAL APPRAISAL-GENERATED RETURN

![Diagram](image-url)
the RAP \( \lambda \). Also, the CEER due to unsystematic return, \( \text{CEER}_{PU} \) is one-half of the portfolio alpha. The equations characterizing the optimum portfolio are easily derived from the definition of the RAP:

\[
\lambda = \frac{\text{variance}}{\text{expected excess return}} = \frac{\omega_p^2}{\alpha_p} = \frac{\omega_p}{\alpha_p/\omega_p} = \frac{\omega_p}{\text{CIR}}.
\]

By rewriting this equation in various ways, the optimal strategy is found to be characterized by:

\[
\omega_p = \lambda \cdot \text{CIR} \quad \alpha_p = \lambda \cdot (\text{CIR})^2 \quad \text{CEER}_{PU} = \frac{\lambda}{2} \cdot (\text{CIR})^2
\]

When part of the portfolio is passive, substitution of \( \omega_p = w_A \omega_A \) and \( \alpha_p = w_A \alpha_A \) results in the following characteristics for the optimal active strategy:

\[
\omega_A = \frac{1}{w_A} \lambda \cdot \text{CIR} \quad \alpha_A = \frac{1}{w_A} \lambda \cdot (\text{CIR})^2 \quad \text{CEER}_{AU} = \frac{1}{w_A^2} \cdot \lambda \cdot (\text{CIR})^2
\]

\( \text{CEER}_{AU} \) is defined as the appraisal-generated CEER, per dollar invested in the active portfolio.

3. **Systematic and Unsystematic Return**

The return to the client is the return to the total portfolio, inclusive of systematic and unsystematic components. The distinction between these components is for analytic purposes. Suppose that the investor is concerned with the expected reward and variance of return of the portfolio. Expected excess reward is the sum of expected excess reward from systematic return and the appraisal premium:

\[
\mu_p = \mu_{PS} + \alpha_p.
\]

Variance of total return is the sum of variance of systematic return and unsystematic variance:

\[
\sigma_p^2 = \sigma_{PS}^2 + \omega_p^2.
\]

(The covariance between systematic and unsystematic risk is zero, by the definition of unsystematic risk.) Figure 7 illustrates the optimal portfolio, incorporating systematic and unsystematic return.
FIGURE 7
THE ATTAINABLE FRONTIER FOR TOTAL RETURN

NOTE: The figure shows the relationship between the overall optimum, with expected return $r_f + \mu_p$ and variance $\sigma_p^2$, and the optimum systematic and unsystematic risks and rewards. The systematic strategy has mean excess return $\mu_{PS}$ and variance $\sigma_{PS}^2$. The active strategy has mean $\alpha_p$ and variance $\omega_p^2$. The dotted line is tangent to the systematic frontier at the optimum, and is parallel to the tangent line to the optimal overall strategy, also tangent to the appraisal-generated frontier at its optimum. When indifference curves are linear, the CEER of the systematic, unsystematic, and total strategies are CEER$_{PS}$, CEER$_{PU}$, and CEER$_{PS} +$ CEER$_{PU}$. 
The overall efficient frontier is drawn as a heavy line. The efficient frontier for systematic return (from figure 4) is drawn as a lighter line. The overall return frontier lies above the systematic frontier, since the availability of appraisal premia improve opportunities for total return. Beginning at any point on the systematic frontier (a passive strategy), the additions to mean and variance attainable from increased aggressiveness can be shown by reproducing the appraisal-generated frontier (figure 6) with origin at the passive strategy. This is correct because mean and variance are additive.

The appraisal-generated frontier touches the overall frontier at one point only, where the two frontiers are tangent. That point is the optimal overall strategy, since the indifference curve is tangent to the attainable frontier at that point. The slope at that point is the same as the slope of the tangent line to the passive strategy on the systematic frontier.¹ Thus, the same risk-reward trade-off applies to the passive as to the active strategy.

Further, optimal systematic and unsystematic returns are characterized by the same RAP, since they both lie along the broken line with slope 1/RAP. In other words, if the client is concerned only with the mean and variance of total return, the RAP applying to systematic and unsystematic return must be the same. If a different RAP were applied, so that the chosen slope of the systematic frontier were different than

¹This fact is proved in Theorems A.4 and A.5 in the appendix. The reader can also prove it for himself by means of the graph with the following line of argument: If the passive strategy designated on the graph is the best, then the designated active strategy is obviously the best possible active strategy to put in conjunction with it, since it touches the highest indifference curve among all points on the appraisal frontier. Therefore, the statement could be false only if the indicated passive strategy is not the best one. But if we consider any other passive strategy, and draw the appraisal-generated frontier starting at that other point, the starting point will lie below the dotted line in figure 7, so that no point on the appraisal frontier will reach the indifference curve. Hence, all other passive strategies are ruled out, and the one that is optimum for this indifference curve is the one where the tangent line (the dotted line) has the same slope as the indifference curve at the efficient frontier.
the chosen slope of the appraisal frontier, the resulting utility would always be lower.

Thus, the RAPs applied to the two sources of return should be the same, unless there is some reason to regard systematic return as distinct from unsystematic return. As a statistical matter, expected return from any two sources is indistinguishable. Thus, there can be no reason to distinguish between the two different modes of expected reward.¹

The situation is otherwise with the risks of these returns. These returns are differently correlated with other eventualities in the client's environment and can therefore be differently valued. This point is particularly strong in the case of a corporate pension fund. Risk in such a fund would, from the point of view of stockholders in the company, contribute a possible drain on future earnings. The drain from unsystematic risk would be exactly correlated with the unsystematic return on the portfolio and hence would lead to added unsystematic risk in the common stock of the company. According to capital market theory, the shareholders could diversify away all such unsystematic risk in a larger portfolio of common stocks, so that it would possess no disutility at all. In contrast, systematic risk of the pension-fund portfolio would produce systematic risk in the common stock, which would be undiversifiable and hence a disutility to the shareholder. So, from the viewpoint of the shareholder, there is no disutility of unsystematic risk in the pension fund, but normal disutility to systematic risk. Therefore, the corporation should

¹One possible contrary argument would be that the philosophical justification for the expected appraisal premium is weaker than for the expected systematic reward, since the former contradicts efficient markets, while the latter does not. But the appraisal premium is, by definition, the expected reward from appraisal premia and should already reflect any "discounting" of promised returns due to lack of credibility. It is imperative that the CIR and resulting alpha fully incorporate the skepticism of the client for, otherwise, the expected benefit, expressed in terms of α, will not be comparable with the expected costs—management fee and disutility of unsystematic risk—and resulting decisions will be suboptimal and possibly foolish. In short, if the expected reward from appraisals is correctly formulated, no justification can exist for distinguishing between systematic and unsystematic expected reward.
have a much higher risk-acceptance parameter for unsystematic risk than for systematic risk. 1

This same argument applies to a lesser degree to all forms of institutional investment. To the extent that the invested funds are part of a larger portfolio of the ultimate investor, so that diversification is provided for the ultimate investor relative to unsystematic risk in the investment pool but not relative to systematic risk, the risk-acceptance parameter for unsystematic risk will be higher than for systematic risk. Thus, the RAP for appraisal-generated returns should, in general, be greater than the RAP for systematic returns. An assumption of equality between the RAPs provides a lower limit on the RAP for unsystematic risk.

4. Implications for Investment Practice

In section 1, it was shown that the RAP for systematic return in the equity portfolio (\( RAP_{ES} \)) could be inferred from the long-term forecasts of equity market reward and risk that the client considered in choosing the normal beta, and from the chosen value. There are certainly variations across clients in \( RAP_{ES} \), since normal equity betas range from .6 to 1.4, and long-term forecasts are not identical, so the lowest and highest RAP may differ by a factor of two. The client's exact \( RAP_{ES} \) would be known by the manager and employed in portfolio management. For present purposes, it will suffice to use a representative value that probably approximates the \( RAP_{ES} \) for the majority of clients. Taking as representative values, \( \mu = 1 \); \( \mu_M = 6\% \); \( \sigma_M = 20\% \), the result is:

---

1 The argument implies that the RAP for unsystematic risk should be infinite, since no risk disutility exists. In practice, institutional constraints would limit the degree of risk taken in the portfolio. The matter is complicated by pension fund insurance. This creates an implicit right on the part of the corporation to hand over the liabilities of the fund to a federally sponsored agency in case of disastrous loss. As Sharpe and Treynor, among others, have pointed out, this corresponds to a put option on pension assets in the hands of the corporation, the value of which increases as risk increases. Thus, an increase in risks in the pension fund, by increasing the value of the put, increases the net wealth of the shareholders. A corporate shareholder would actually have a positive utility of unsystematic risk, since a part of the downside risk is borne by the federal government.
\[ \hat{R}_{ES} = \frac{N \cdot \sigma^2}{\mu_M} = \frac{1 \cdot (20\%)^2}{6\%} = 66 \frac{2}{3}\% . \]

This is approximately correct for all clients' equity portfolios. (The major differences in risk aversion between clients will have been previously implemented by the choice of the proportion to invest in equity. The risk attitude toward the equity portfolio itself is much more nearly constant.)

As argued in section 3, the RAP for unsystematic risk in the equity portfolio, \( R_{EU} \), will be at least as large as this and probably should be much larger. Let us then consider the common case where the entire equity portfolio is under active management, possibly by multiple managers. As a first application, let us compute the disutility of unsystematic risk. The variance coefficient in the utility function \( \lambda_U \) is equal to \( 1/(2 \cdot R_{EU}) \). Since \( R_{EU} \geq \hat{R}_{ES} \), we have:

\[ \lambda_U \leq 1/(2 \cdot 66 \frac{2}{3}\% ) = 1/(133 \frac{1}{3}\% ) . \]

Therefore, the disutility of unsystematic standard deviation \( \omega_p \) is at most

\[ \lambda_U \omega_p^2 \leq \frac{\omega_p^2}{133 \frac{1}{3}\% } . \]

For example, with unsystematic standard deviation of 3% per annum, the disutility is \((3\%)(3\%)/133 \frac{1}{3}\% = .0675\% .\) This function is plotted in figure 8.

It is interesting to consider, in the light of this equation, the grounds for argument in favor of passive management. The excess of the management fee for an active manager, over that for passive management, is generally .12% per annum or more. The transaction costs entailed by

\[ ^1 \text{The computation of disutility actually depends on the entire shape of the indifference curve, not just its slope at the point of the optimal portfolio. If the curve is not linear, the accuracy of the approximation will worsen for large values of unsystematic variance, when the effect on total variance is relatively great. However, the approximation that the slope is constant should be a good one, since the removal of unsystematic risk involves a small change in the variance of the overall portfolio, e.g., a reduction from } (20^2 + 3^2 = 400 + 9 = 409) \text{ to } (20^2 = 400). \]
active management are probably at least .20% per annum of portfolio value. Hence, if the active strategy for the portfolio as a whole exhibits an unsystematic standard deviation of 3% per annum, the disutility of residual risk, which is at most .0675%, is the smallest of the three "costs" of active management. The value of 3% per annum for residual standard deviation approximates the median value for large pension and employee benefit funds.¹

Next, let us consider a more crucial question, the client's choice of appropriate aggressiveness. The active manager(s) offer

¹Exact data on employee benefit funds' unsystematic risk are not generally available, as distinct from the data on the separate components of those funds managed by individual managers. Recent surveys of individual subportfolios have shown median figures of about 6% per annum unsystematic standard deviation. Employee benefit funds in excess of $100 million typically have four or more equity managers—and sometimes fifteen or more. In addition, managers are often chosen so as to represent a variety of "styles," so that unsystematic covariance among managers is often small and sometimes negative. Hence, unsystematic variance for the portfolio as a whole is often of the order of 1/J times individual unsystematic variance, where J is the number of managers. Taking 4 as the typical number of managers, portfolio variance of $(6\%)^2/4 = (3\%)^2$ results.
market timing and active holdings, which provide a net information ratio (net of transaction costs). If we were privy to the client's decision process, information concerning his assessment of that ratio would be available, ideally in the form of a covenant between the client and manager. What can be said in the absence of that information? Actually, much can be done: given the RAP, the client's choice of aggressiveness is uniquely determined by the CIR. Conversely, if we know the unsystematic standard deviation of the portfolio and know or can approximate the RAP, it is possible to solve for all other aspects of the optimal unsystematic strategy. The equations at the end of section 2 can be rewritten as follows:

\[
\begin{align*}
\text{Entire Equity Portfolio} & \\
\text{Active Equity Component} & \\
\text{CIR} & = \frac{1}{\lambda} \omega_P \\
\alpha_p & = \frac{1}{\lambda} \omega_P^2 \\
\text{CEER}_{PU} & = \frac{1}{2\lambda} \omega_P^2 \\
\text{CIR} & = \frac{\omega_A}{\lambda} \omega_A \\
\alpha_A & = \frac{\omega_A^2}{\lambda} \omega_A \\
\text{CEER}_{AU} & = \frac{\omega_A^2}{2\lambda} \omega_A
\end{align*}
\]

Using the lower bound for the RAP, we obtain an upper bound for its reciprocal, and hence upper bounds for CIR, \( \alpha \), and CEER. Thus, considering first the entire portfolio,

\[
\begin{align*}
\text{CIR}_{req} & = \frac{\omega_P}{\text{RAP}_{EU}} \leq \frac{\omega_P}{\text{RAP}_{ES}} \sim \frac{\omega_P}{66 \frac{2}{3}} \\
\alpha^\text{req}_P & = \frac{\omega_P^2}{\text{RAP}_{EU}} \leq \frac{\omega_P^2}{\text{RAP}_{ES}} \sim \frac{\omega_P}{133 \frac{1}{3}} \\
\text{CEER}_{PU}^\text{req} & \sim \frac{\omega_P^2}{2 \cdot \text{RAP}_{EU}} \leq \frac{\omega_P^2}{2 \cdot \text{RAP}_{ES}} \sim \frac{\omega_P}{266 \frac{2}{3}}
\end{align*}
\]

The superscript "req" indicates that these are the levels of information content and utility that are required to make the chosen level of residual risk the optimal one. To put it another way, if the client's views as to
information content and utility were not equal to these values, the chosen level of unsystematic risk would be wrong for him.

In figure 9, the required appraisal premium and required CEER are plotted against $\omega_p$. To highlight the meaning of these figures, a representative active management cost of .12% is also shown. The CEER rises above the management fee at an unsystematic standard deviation of 4%. The client's actual benefit from active management fee is equal to the CEER less the management fee. For values of $\omega_p$ less than 4%, this is negative. Thus, if the portfolio is entirely under active management at this fee, then an unsystematic standard deviation less than 4% per annum is evidence of irrationality. Since many pension funds are currently operated in exactly this way, with added active management fees of 12 basis points or more, and unsystematic standard deviation of 4% or less, it follows that the current practice is incorrect in these cases.

To repeat the argument succinctly, we were able to show that appropriate risk acceptance toward unsystematic risk in equities, as expressed by the $R_{EU}$, is certainly no less than that for systematic risk $R_{ES}$; we found a procedure to compute this exactly from the portfolio beta chosen by the client and were therefore able to approximate the representative value for most clients. This supplied the risk preferences of the client. When the choice of optimal aggressiveness for the active management strategy was analyzed with this risk preference in mind, we were able to infer from the chosen risk level (the standard deviation of unsystematic risk) what underlying information content, net appraisal premium, and net contribution to client utility would be required to make the choice the correct one. If the information content is so small that optimal residual standard deviation is 4% or less, then that information is not enough to overcome the .12% added expense of active management. Hence, a client who finds that appraisal information does not warrant an unsystematic standard deviation of 4% or more should abandon active management—at least in part—because appraisal benefit is too small to offset the management fee.  

1It is interesting to consider the case where 50% of the portfolio is passively managed. In this case, $\omega_A$ is .5, and the required
CIR, $\alpha_A$, and $\text{CEER}_{AU}$ for each dollar in the active component are reduced by one-half. Consequently, the optimal active strategy at any CIR is more aggressive: the "break-even point," where $\text{CEER}_{AU}$ is equal to the active management fee of .12% now occurs with optimal unsystematic standard deviation of $\omega_A = 5.7\%$ per annum. This is consistent with the principle that, with a smaller proportion of the portfolio under active management, the active component should be more aggressive.
In the case of institutionally invested employee benefit funds, the client is the pension fund sponsor. Some sponsors—possibly the majority—currently employ active management at so low an unsystematic standard deviation that, if it were optimal, it would imply information appraisal benefit less than the management fee for active management. What explanation can be found for this apparent irrationality? There are two possible reasons, with sharply different implications. First, sponsors might be applying a correct trade-off between appraisal-generated risk and reward, fully aware that appraisal-generated CEER is less than the active management fee, but might not yet be ready to take the step to passive management. Second, sponsors might believe that the appraisal-generated CEER is greater than the current management fee, but might be less than optimally aggressive because of excessive prudence in regard to unsystematic risk.

I conjecture that the second explanation is the correct one in most cases. In other words, sponsors are convinced that the selection process has found above-average managers whose appraisal ability would obtain a net portfolio alpha of more than .24% per annum at \( \omega_p = 4\% \). Rationality would then lead the sponsor to undertake more aggressive differential strategies, resulting in higher \( \alpha_p \), higher \( \omega_p \), and higher CEER. However, for three natural reasons, the sponsor errs in the direction of irrational prudence and holds \( \omega_p \) to a lower suboptimal level.

First, the mandatory diversification requirement in ERISA does not explicitly state that unsystematic risk, even when deliberately undertaken from a justified belief in appraisal premia, is of comparable nature to systematic risk. The threat of a lawsuit under ERISA against the sponsor, justified by inadequate relative performance—with total return compared to a visible index such as the S&P 500—would place pressure on the sponsor to hold \( \beta_E \) near unity and \( \omega_p \) near zero. (Nonstandard exposure to systematic return and to unsystematic risk would figure indistinguishably in determining the downside risk that might lead to vulnerability to a suit based on total return.) Further, the legal precedent for suit on an individual asset basis—which might conceivably extend to suit on a manager-by-manager basis—results in special vulnerability from unsystematic risk.
due to positive active holdings. Thus anachronisms in legal precedent may be partially responsible for excessive prudence with regard to unsystematic risk.

Second, the typical employee benefit fund of $100,000,000 or more has four or more money managers, and the number of managers ranges upward to fifteen or more. The sponsor, in evaluating the individual managers, tends to concentrate separately on the performance of each. There is no fallacy in this, as long as the sponsor recalls that the total portfolio is diversified, by virtue of multiple managers (let J be the number of managers), so that total unsystematic variance is approximately of the order of \(1/J\) of the variance of the individual manager. The sponsor who appreciates this point will take care to apply a risk aversion to evaluating each manager, which is approximately \(J\) times as small as for the total portfolio. On the other hand, if the sponsor ignores this point and applies the same risk aversion to the individual managers as he has in mind for the total portfolio, the end result is portfolio aversion to unsystematic risk which is too great by a factor of \(J\), exactly the sort of excessive prudence which has been suggested.

Third, the present mode of performance evaluation relies on comparison of the performance of the managed portfolio with the performance of a representative index, usually the S&P 500. The sponsor will naturally experience risk in terms of the size of the difference between the S&P 500 outcome and the portfolio outcome. Natural impressions are such as "the portfolio underperformed the S&P 500 by 5% this quarter, when the S&P 500 rose by 12%," or, "in the last five years, the S&P 500 returned a compound annual rate 7% below the risk-free rate, and the portfolio further underperformed the S&P 500 by 2% per year." In these statements, one is comparing two random variables realized from different distributions—the market or systematic return and the unsystematic portfolio return. It is natural to assess the risk contributions of the two sources by their magnitudes. For instance, when the excess return on the market portfolio was -7% and the unsystematic portfolio return was -2%, for a total portfolio return of -9%, the sponsor would naturally conclude that roughly two-ninths of total risk was due to unsystematic risk.
This conclusion is incorrect and reflects a natural mistake. The risk of total portfolio return is determined by its standard deviation, which, in turn, is the square root of total variance. Total variance is the sum of systematic and unsystematic variance. Thus, the contributions to total risk are in proportion to variances, which are the squares of standard deviation. But an assessment of the relative importance of sources of risk by typical magnitude, or by the magnitudes of worst cases, is using a measure that is proportional to the standard deviation.

Consider a typical portfolio with unsystematic standard deviation of 6% per annum, and systematic standard deviation of 20%. The client might expect to experience a worst-case market return of \(-30\%\) \(-1.5\sigma_m\) and a worst-case portfolio unsystematic return of \(-9\%\) \(-1.5\sigma_p\). He might therefore conclude that the ratio of contributions to total risk was 9%/30%, equal to 6%/20%, the ratio of standard deviations. But, in fact, the ratio of contributions to total risk is the ratio of variances, \((6\%)^2/(20\%)^2\). Thus, the assessment of relative importance errs by a factor of 20/6 and overstates unsystematic risk by that amount. This confusion between variance and standard deviation leads to an exaggeration of unsystematic risk by a factor of three or more in relation to systematic risk.

The diversification effect, resulting in a natural tendency for the sponsor to overstate portfolio unsystematic risk by a factor of 4 or more, combined with this standard deviation confusion, results in a tendency to overstate the importance of unsystematic risk by a factor of 12 or more. In the absence of a careful quantitative treatment, it is therefore not at all surprising that excessive prudence, with regard to unsystematic risk, is the order of the day.

The cure for this error is to focus on the always available opportunity for the sponsor to balance the total portfolio by shifting some risk from systematic to unsystematic return. For example, suppose the current portfolio has a beta of 1, with systematic variance of \(400(\%)^2\) and unsystematic standard deviation of 3%, resulting in unsystematic variance of \(9(\%)^2\), for a total portfolio variance of \(409(\%)^2\). The following exchange could be made: reduce portfolio beta from 1.0 to .966,
resulting in a systematic variance of \(0.966^2 \cdot 400(\%^2) = 373(\%^2)\); double the aggressiveness of all differential holdings, so that unsystematic standard deviation doubles to 6%, resulting in unsystematic variance of 36(\%^2); resulting portfolio total variance, 373(\%^2) + 36(\%^2) = 409(\%^2), is unchanged. Thus, a reduction of beta of only .034 reduces systematic risk enough to allow the aggressiveness of the differential strategy to be doubled. The reduction in expected excess return from reducing beta is roughly .034(6%) = .204%. The exchange should be made if the increase in expected appraisal premium is .204% or more, which, in turn, will be the case if the existing appraisal premium is .204% or more, since doubling aggressiveness doubles the appraisal premium.\(^1\)

In conclusion, it appears that existing anomalies in management of employee benefit funds will be resolved when corporate sponsors correctly perceive the risk-reward trade-off for unsystematic return. At that time, funds under active management will exhibit greater unsystematic risk than is now the case. Annual unsystematic standard deviations of 8% and above will be common in the components of a multiple-managed portfolio.\(^2\) This will inevitably strain the relationship between money manager and client, since underperformance by one standard deviation or more, due purely to chance factors, can be expected about one year in six. Thus, in one-sixth of annual reviews, the manager will be reporting highly disappointing results, where the portfolio underperforms the expected appraisal premium by 8% or more. It is to be hoped that the client, having previously determined a covenant information ratio and, understanding unsystematic risk as necessary in the exploitation of appraisals, will be able to interpret these results from a proper perspective.

\(^1\) Notice that this argument shows that an exchange would be beneficial but does not show that the exchange in question is the best possible one. The optimum modification can be found by applying the optimality conditions.

\(^2\) Taking the break-even point at \(\omega_p = 4\%, a \) multiple-managed portfolio, equally divided among four managers with essentially uncorrelated information processes, would be at break-even when each component portfolio had an unsystematic standard deviation of 8%. With five managers, the break-even would be at approximately 9% (\(\sqrt{80}\%\)). If the funds are unequally apportioned among managers, break-even for the smaller pools of funds would be at higher risk levels (Rosenberg, 1977a).
This appendix rigorously presents the relationships underlying the statements in the text. Where similar material appears previously in the literature, proofs of theorems are only sketched, with references to earlier work for the interested reader.

A.1. Notational Conventions and Glossary

The important writings concerning the optimal use of judgment in portfolio management unfortunately employ several conflicting notational conventions. No single paper is sufficiently extensive in its coverage to serve as a unique reference. Hence, the notation in this appendix must be an amalgam of previous conventions, which attempts to maintain the best of each. Table A.1 shows the correspondence between the notation in this and prior articles.

The portfolio management problem is a multivariate problem. For all multivariate problems, in whatever branch of quantitative modeling, matrix algebra is a valuable tool. Earlier articles have employed the simplified "diagonal" model of portfolio optimization as a means to avoid matrix notation, but only at two costs: first, unrealistic representation of intercorrelation of asset returns; second, an oblique derivation that obscures the direct derivation of the general model. In this appendix, which is intended for application, the restrictive and unrealistic assumptions of the diagonal model are inadmissible. Hence, matrix notation is indispensable. Happily, the results in matrix notation are brief and transparent, once the original barrier of unfamiliarity with matrix notation is surmounted.

Following common convention in econometrics, vectors and matrices will be represented by bold-face characters, rendered in typescript by writing a tilde beneath the symbol. A lower-case letter with a tilde beneath is a column vector, e.g., \( \tilde{a} \). A row vector is written as the transpose of a column vector, with the transpose operator represented by an
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_n$</td>
<td>$E_{t-p} - \beta_{t} (E_{t-M} - p)$</td>
<td>$\alpha_t$</td>
<td>$\tilde{z}<em>{t} \mu</em>{t}$</td>
<td>$\tilde{r}_{t,\alpha}$</td>
</tr>
<tr>
<td>$\beta_n$</td>
<td>$b_i$</td>
<td>$\beta_{im}$</td>
<td>$b_i$</td>
<td>$\beta_i$</td>
</tr>
<tr>
<td>$\mu_n$</td>
<td>$E_{t-p}$</td>
<td>$E_{t-p}$</td>
<td>$E_{t-p}$</td>
<td>$E_{t-t}$</td>
</tr>
<tr>
<td>$\nu_{nn}$</td>
<td>$C_{ii}$</td>
<td>$\sigma_{2}$</td>
<td>$\sigma_{2}$</td>
<td>$\sigma_i$</td>
</tr>
<tr>
<td>$R_{nn}, \sigma^2_{nn}$</td>
<td>$\sigma^2_{ii}$</td>
<td>$\text{Var}({\tilde{c}_i})$</td>
<td>$\sigma^2_{ii}$</td>
<td>$\sigma_{y_i}$</td>
</tr>
<tr>
<td>$\gamma_n$</td>
<td>$R_{i-t} - p$</td>
<td>$\tilde{R}_{i-t} - \pi$</td>
<td>$y_i$</td>
<td>$y_{i}$</td>
</tr>
<tr>
<td>$h_n$</td>
<td>$X_i$</td>
<td>$X_i$</td>
<td>$h_i$ then $h_i + \beta_{h_{mi}}$</td>
<td>$h_i$ then $h_i + \beta_{h_{Mi}}$</td>
</tr>
<tr>
<td>$h_{Mn}$</td>
<td></td>
<td>$h_{ni}$</td>
<td>$h_{ni}$</td>
<td>$h_{ni}$</td>
</tr>
<tr>
<td>$\delta_n$</td>
<td></td>
<td>$h_{i-p_{mi}}$ then $h_i$</td>
<td>$h_{i-p_{Mi}}$ then $h_i$</td>
<td>$h_{i-p_{Mi}}$ then $h_i$</td>
</tr>
<tr>
<td>$V_{nn}$</td>
<td>$C_{ij}$</td>
<td>$C_{ij}$</td>
<td>$C_{ij}$</td>
<td>$C_{ij}$</td>
</tr>
<tr>
<td>$\gamma_F$</td>
<td>$p$</td>
<td>$\pi$</td>
<td>$\gamma_F$, then $\gamma$</td>
<td>$R_F$</td>
</tr>
<tr>
<td>$\gamma_M$</td>
<td>$R_{M-t} - p$</td>
<td>$\tilde{R}_{M-t} - \pi$</td>
<td>$y_{m}$</td>
<td>$y_{m}$</td>
</tr>
<tr>
<td>$\mu_M$</td>
<td>$E_{M-t} - p$</td>
<td>$E_{M-t} - \pi$</td>
<td>$y_{m1} \mu_{n+1}$</td>
<td>$y_{m1} \mu_{n+1}$</td>
</tr>
<tr>
<td>$\sigma^2_M$</td>
<td>$\sigma^2_{M}$</td>
<td>$\sigma^2_{M}$</td>
<td>$\sigma^2_{M}$, $\sigma_{n+1}^2$</td>
<td>$\sigma^2_{M}$, $\sigma_{n+1}^2$</td>
</tr>
<tr>
<td>$\gamma_P$</td>
<td>$E_{P-t} - p$</td>
<td>$E_{P-t} - \pi$</td>
<td>$\mu_{P}$</td>
<td>$\mu_{P}$</td>
</tr>
<tr>
<td>$\sigma^2_P$</td>
<td>$V_{P}, \sigma^2_P$</td>
<td>$V_{P}$</td>
<td>$\sigma^2_{P}$</td>
<td>$\sigma^2_{P}$</td>
</tr>
<tr>
<td>$\omega^2_P$</td>
<td>$V_{P}$</td>
<td>$\omega^2_{P}$</td>
<td>$\sigma^2_{P}$</td>
<td>$\sigma^2_{PA}$</td>
</tr>
<tr>
<td>$\alpha_P$</td>
<td>$A_{P}$</td>
<td>$A_{P}$</td>
<td>$\mu_{P}$</td>
<td>$\mu_{PA}$</td>
</tr>
</tbody>
</table>
### Table A.1—Continued

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda$</td>
<td>$\lambda/2$</td>
<td>$1/(2\lambda)$</td>
<td>$\lambda$</td>
<td></td>
</tr>
<tr>
<td>$\lambda_V$</td>
<td>$1/\lambda$</td>
<td>$\lambda$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$U$</td>
<td>$-\pi+\rho$</td>
<td>$\pi-$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$z_p$, $z_M$, $z_n$</td>
<td></td>
<td></td>
<td></td>
<td>$\rho$, $\lambda$, $\gamma_i$</td>
</tr>
<tr>
<td>$f$</td>
<td></td>
<td>$f_{i}$</td>
<td></td>
<td>$P_i$</td>
</tr>
<tr>
<td>$\rho$</td>
<td></td>
<td>$\rho$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Notes to Table A.1

The notation of Fisher (1975) is identical to Sharpe (1970), except that $E_i$ replaces $p$, $V_i$ replaces $C_{ii}$, and $s_i^2$ replaces $\sigma_{c_i}^2$. Where two symbols correspond to one, both notations were deliberately used in an article to represent the same construct. Where two symbols are separated by the work "then," the latter appearance resulted from a redefinition of the problem, in which the same construct acquired a different meaning.
apostrophe, e.g.,  \( a' \). A matrix will be written as an upper-case letter with a tilde beneath, e.g.,  \( A \). Ordinary numbers (scalars) will be represented as lower- or upper-case letters, without tildes beneath.

A generally applicable construct, such as variance, is written as a single symbol, e.g.,  \( \sigma^2 \), with the subscript denoting the item to which the construct is applied. For example,  \( \sigma_M^2 \) is the variance of the market portfolio return. Mnemonic subscripts which denote a particular item (such as "M" for "Market") are in upper case. Lower-case subscripts represent individuals in a class (e.g., the subscripts "m" and "n" for individual assets within the class of assets), and corresponding upper-case letters denote the number of such individuals in the class (e.g., \( N \) for the number of assets).

With this preamble, the notation used in the paper can be briefly summarized in the following glossary.

**TABLE A.2**

**GLOSSARY OF TERMS AND NOTATION**

1. **Upper-case subscripts** are used consistently to denote asset and return categories as follows:
   
   C  "Cash," including cash equivalents
   
   B  "Bonds," including medium and long-term nonconvertible corporate liabilities and governments
   
   E  "stocks" or "Equity," including common stocks, convertible senior liabilities, warrants
   
   F  the total Portfolio
   
   M  the "Market" portfolio
   
   F  the "Risk-Free" asset
   
   S  Systematic component of return
   
   U  Unsystematic (or residual) component of return

2. **Lower-case subscripts** for categories of items, and total numbers of items:

   \( n,m = 1,\ldots,N \) individual assets among \( N \) assets
\( j, i = 1, \ldots, J \) individual managers among \( J \) managers
\( t, s = 1, \ldots, T \) individual time periods among \( T \) periods

3. The following constructs are defined for portfolios and some asset categories:
   - \( \omega \) the investment proportion in that item, equal to dollars invested in that item divided by total dollars
   - \( 1+i \) total return over some holding period
   - \( r \) excess return over some designated holding period (a year)
   - \( \sigma^2 \) variance of return for that period
   - \( \mu \) expected excess return for that period
   - \( \alpha \) judgmental appraisal premium or expected abnormal return
   - \( \beta \) regression coefficient on the market portfolio (misdefined as the S&P 500 for expository purposes)
   - \( \mu_S, \sigma_S^2 \) mean and variance of excess systematic return
   - \( u \) unsystematic return
   - \( \omega^2 \) (omega squared) variance of unsystematic return
   - \( z = \mu / \sigma \) the ratio of expected excess return to standard deviation of return, or reward to variability ratio (RVR)
   - \( z^2 = \mu^2 / \sigma^2 \) the ratio of squared expected excess return to variance of return, or the Sharpe Ratio

4. The following constructs are defined for individual equity issues, in addition to \( \alpha, \mu, \beta, r, u, \) and \( z \) as defined above:
   - \( \hat{\omega}_n \) forecast residual return for asset \( n \)
   - \( h_n \) investment weight, equal to proportion of equity portfolio invested in that asset
   - \( h_{Mn} \) market weight, equal to proportion of market capitalization invested in that asset
   - \( \sigma_n^2 \) specific or independent variance of that asset
   - \( \delta_n \) the "active" holding of asset \( n \) in a portfolio, equal to \( \delta_n = h_n - 3p h_{Mn} \)
5. The following constructs are defined across different equity issues:

\( V_{mn} \) covariance between returns \( r_m \) and \( r_n \) (variance if \( m = n \))

\( R_{mn} \) covariance between unsystematic returns \( u_m \) and \( u_n \) (variance if \( m = n \), elsewhere defined as \( \sigma_n^2 \))

\( C_{mn} \) covariance between forecast unsystematic return \( \hat{u}_m \) for asset \( m \), and actual future unsystematic return for asset \( n \)

\( D_{mn} \) covariance between forecast unsystematic returns \( \hat{u}_m \) and \( \hat{u}_n \) (variance if \( m = n \))

\( M_{mn} \) covariance between forecast errors \( (u_m - \alpha_m) \) and \( (u_n - \alpha_n) \) (variance if \( m = n \))

\( A_{mn} \) covariance between \( \alpha_m \) and \( \alpha_n \) (variance if \( m = n \))

6. The following constructs apply to the advisor's management strategy:

\( z \) "covenant" information ratio

\( b \) dependence adjustment factor

\( \lambda \) risk acceptance parameter, equal to ratio of variance of return to mean excess return

\( U \) the certainty equivalent excess return (CEER) or utility of the portfolio

\( \lambda_U \) coefficient for variance in the utility function

\( \beta \) "normal" beta for long-term horizon

A.2. Representation of Investment Returns and Risk

For some fixed holding period, let \( i \) denote the rate of return on an asset or, equivalently, let \( 1+i \) denote the total return, so that:

\[
1 + i = \frac{\text{value at end of period (including reinvested cash payout, if any)}}{\text{value at beginning of period}}
\]

Then, for any asset, say, \( n \), the arithmetic excess return is defined as

\[
r_n = i_n - i_F = (1+i_n) - (1+i_F),
\]

where \( i_F \) is the rate of return on the risk-free asset. Alternatively, the logarithmic excess return is defined as:
\[ r_n = \ln(1+i_n) - \ln(1+i_F) . \]

In what follows, the "excess return" may be taken as either the arithmetic excess return or the logarithmic excess return. In the latter case, the statements on joint distributions of assets are only approximations, but the mean-variance representation of preferences is more realistic.

For \( N \) individual assets, let \( \tilde{r} \) be the column vector of excess returns, \( \tilde{r} = (r_1 : \ldots : r_N)' \) (where the column vector is written as a transposed row vector to save space.) Let \( \mu \equiv E(\tilde{r}) \) be the column vector of expected excess returns, and let \( \Sigma \equiv \text{Var}(\tilde{r}) \) be the \( N \times N \) matrix of variances and covariances of returns. We assume that there is no redundant asset that is exactly correlated with some other(s); then the matrix \( \Sigma \) will be positive definite and will possess an inverse.

Let \( h^{-}_M \) be the column vector of weights in the market portfolio, and let \( r^-_M \) be the excess return on the market portfolio, given by:

\[ r^-_M = \sum_{n=1}^{N} h^-_n r_n = h^-_\Sigma . \]

Let \( \mu^-_M \) and \( \sigma^-_M \) be the expected excess market return and variance of the market return. For each asset \( n \), let \( \beta^-_n = \text{Cov}(r^-_n, r^-_M) / \sigma^-_M \), and let the unsystematic or residual return be \( u^-_n \), where \( r^-_n = \beta^-_n r^-_M + u^-_n \). Let \( \tilde{u} \equiv E(\tilde{u}) \) be the column vector of expected unsystematic returns and \( \tilde{\Sigma} \) the \( N \times N \) matrix of variances and covariances of unsystematic return.

Let a portfolio \( P \) be constructed with investment weights \( h_1, \ldots, h_N \), such that if these weights do not sum to unity, the difference is invested in the risk-free asset. Thus, \( h_0 = 1 - \sum_{n=1}^{N} h_n \) is the investment weight in the risk-free asset. Then the expected excess return on the portfolio is easily shown to obey the following formulae.

---

**Theorem 1 (Expected Excess Portfolio Return)**

Expected excess return:
\begin{align}
\mu_p &= h'\mu = \sum_{n=1}^{N} h_n \mu_n . \\
\text{Decomposition into expected systematic and unsystematic return:} \\
\mu_p &= \mu_{PS} + \alpha_p ; \quad \mu_{PS} = \beta_p \mu_M \text{ and } \alpha_p = h'\alpha = \sum_{n=1}^{N} h_n \alpha_n . \\
\text{Portfolio beta:} \\
\beta_p &= h'\beta = \sum_{n=1}^{N} h_n \beta_n . \nonumber
\end{align}

These familiar formulae give the expected return, beta, and alpha of the portfolio as investment-weighted averages of the characteristics of the individual assets. The variance of the portfolio return is somewhat more complex. The important constructs is the joint distribution of asset returns, as derived in Rosenberg (1974), as are stated in Theorem 2.

\begin{align}
\text{Theorem 2 (Representation of Investment Risk)} \\
\text{Asset variance decomposes into systematic and unsystematic (residual) variance:} \\
\sigma_p^2 &= \sigma_{PS}^2 + \omega_p^2 ; \quad \sigma_{PS}^2 = \beta_p^2 \sigma_M^2 \text{ and } \omega_p^2 = h'Rh . \\
\text{Market variance and beta:} \\
\sigma_M^2 &= h_M' \Sigma h_M ; \quad \beta = \frac{1}{\sigma_M^2} h_M \Sigma h_M . \\
\text{Total portfolio variance:} \\
\sigma_p^2 &= h'\Sigma h . \nonumber
\end{align}

Portfolio variance decomposes into systematic and unsystematic variance:

\begin{align}
\sigma_p^2 &= \sigma_{PS}^2 + \omega_p^2 ; \quad \sigma_{PS}^2 = \beta_p^2 \sigma_M^2 , \quad \omega_p^2 = h'Rh . \nonumber
\end{align}
Portfolio unsystematic variance in terms of active holdings:

\[ \omega_p^2 = \delta^\prime \delta \]

Total asset variance may be substituted for unsystematic variance when active holdings are used:

\[ \omega_p^2 = \delta^\prime \psi \delta \]

Proof: The proof is contained in Rosenberg (1974). The only difficult step involves the observation that:

\[ \bar{R} = (I - \beta M^\prime)(I - \beta M) \quad \text{and} \quad \delta = (I - \beta M^\prime)h. \]

The easily verified fact that \((I - \beta M^\prime)(I - \beta M)\) is idempotent, i.e.,

\[ (I - \beta M^\prime)(I - \beta M) = (I - \beta M^\prime) \]

then leads to (T2.5) and (T2.6).

---

A.3. A Comparison with the Single-Index Model

The major contributions to the use of judgmental appraisals in portfolio management have been framed in terms of the "diagonal model" or "single-index model." This model was originally formulated by Sharpe (1963). The essential formula of the single-index model (SI) gives portfolio variance as:

\[ \sigma_p^2 = \beta_p^2 \sigma_M^2 + \sum_{n=1}^{N} h_n^2 \sigma_n^2. \]

Thus, unsystematic risk arises only from the specific or independent risk of individual companies \((\sigma_n^2)\), and portfolio exposure is the squared investment weight \(h_n^2\).

This formula conveniently removes all unsystematic covariances between assets. One might expect it to be a simplified case of the general model developed in section A.2. If this were so, comparison of present results with earlier articles would be straightforward. Unfortunately, the single-index formula is not a natural simplification of the more
general model. Moreover, this unnatural quality has caused some confusion, which it is important to remove. Consequently, in the balance of this section, the natural simplification is derived, the discrepancy between it and the single-index model is explained, and the previous literature is interpreted from this perspective.

To dispose of all unsystematic covariances among assets, one posits the "single-factor" model, in which covariance among assets arises from a single "market factor." All investment risk derives from this factor or from the specific returns of individuals securities, which are independent of one another. Let the variance of the market factor be \( \sigma_x^2 \), and for each asset \( n \), let \( \beta_n \) be the regression coefficient of return onto that factor, with \( \beta \) the column vector of such coefficients. The scale on the market factor is arbitrary: without loss of generality, it may be scaled so that the market-capitalization-weighted-average regression coefficient, 

\[
\begin{align*}
    h'x &= \sum_{n=1}^{N} h_n x_n, \\
    \frac{1}{N} h'x &= 1
\end{align*}
\]

is unity. Let \( \sigma_n^2 \) be the specific variance of security \( n \). Then the properties of investment risk are given in Theorem 3, below.

---

**Theorem 3 (The Single-Factor Model)**

Variance matrix of investment returns:

\[
(T3.1) \quad \Sigma = \sigma_x^2 \beta \beta' + \text{diag}(\sigma_n^2).
\]

Market variance:

\[
(T3.2) \quad \sigma_M^2 = \sigma_x^2 + \sum_{n=1}^{N} h_n^2 \sigma_n^2.
\]

Beta:

\[
(T3.3) \quad \beta_n = \frac{1}{\sigma_M^2} \left( \sigma_x^2 \beta_n + \sigma_n^2 h_n \right), \quad n = 1, \ldots, N.
\]

Simplest expression for unsystematic portfolio variance:

\[
(T3.4) \quad \omega_p^2 = \delta' \Sigma \delta = \sigma_x^2 \left( \sum_{n=1}^{N} h_n (\beta_n - x_n) \right)^2 + \sum_{n=1}^{N} (h_n - \beta_n h_n) \sigma_n^2.
\]
When the term involving the market factor is ignored, an approximation emerges as:

\[ \omega_p^2 = \sum_{n=1}^{N} \left( h_n - \beta_p h_n^M \right)^2 \sigma_n^2 = \sum_{n=1}^{N} \delta_n^2 \sigma_n^2. \]

**Proof:** (T3.1) follows from the definition of the single-factor model. (T3.2) and (T3.3) are found by application of (T2.2), using the fact that \( h_n^M x = 1 \) by construction. (T3.4) is found by application of (T2.6).

Although there is only a single factor, a small amount of unsystematic covariance exists. This occurs because the market portfolio return, by definition the weighted sum of individual security returns, includes the market factor plus a weighted sum of specific returns. Because the specific returns enter into the market return, their residuals, relative to the market return, are slightly correlated. The correlation is slight, so (T3.5) is a good approximation. Comparing (T3.5) with (SI), the discrepancy is evident: the correct approximation uses the active investment weight \( \delta_n^2 = (h_n - \beta_p h_n^M) \) as the measure of exposure to specific risk, while the single-index model uses the raw investment weight \( h_n^M \).

This error in the single-index model leads to misleading conclusions of two kinds.

The first error is that the specific variances of individual stocks, as they enter into the market portfolio, are double-counted in the single-index model. This is easily seen by applying formula (SI) to an index fund with investment weights equal to the market portfolio:

\[ \sigma_M^2 = \sigma_P^2 = \beta_M^2 \sigma_M^2 + \sum_{n=1}^{N} h_n^M \sigma_n^2 = \sigma_M^2 + \sum_{n=1}^{N} h_n^M \sigma_n^2. \]

Actually, the index fund perfectly matches the market and has zero unsystematic variance. Thus, the market portfolio improperly exhibits unsystematic variance relative to itself. This error is quite substantial: for the five-year period, 1972-1976, the formula gives the S&P 500 unsystematic risk relative to itself, with standard deviation of 2.2% per annum!
The second error is that the single-index model improperly locates the minimum risk holding of an asset, the holding where the portfolio is unexposed to its specific risk. The correct minimum risk position is $\beta_P h_{Mn}$, which simplifies to the market holding $h_{Mn}$, when $\beta_P = 1$. The single-index model asserts that the minimum-risk position is always zero. The authors who employed the single-index model were, I believe, fully aware that the minimum-risk position was at the market holding rather than at zero. If formula (T3.5) had been used as a simplifying assumption, the results would have naturally reflected this understanding. Because formula (SI) was used, there were some undesirable implications, incidental to the main thrust of the articles, which faced the authors.

Treynor and Black were convinced that the single-factor model was correct (see the concluding paragraphs of their appendix). To escape the undesirable implication of the single-index model, they created two different vehicles for purchase of the market portfolio: one through portfolio investment weights, which had the defect of double-counted specific risk as explained above; the second through a hypothetical "market asset," for which unsystematic risk was defined to be zero. Since the latter vehicle allows purchase of the market portfolio without spurious unsystematic risk, it is clearly the preferred vehicle. Hence, the holding of the market portfolio through the portfolio weights is always zero, and the error vanishes. While this device did lead to results that were correct for the single-factor model, it nevertheless led to some confusion: Treynor-Black and Ferguson (1975) both used the investment holding $h_{I}$ to mean $h_n$ at one point in their papers, and to mean the active holding $(h_n - \beta_P h_{Mn})$ at a later point.

---

1This stratagem appears in the transition from equation (4) to equation (6): in the former equation, the market portfolio can only be purchased with unsystematic risk, while in the latter equation, a purchase vehicle with zero unsystematic risk has "magically" appeared.
On the other hand, Sharpe (1974) and Fisher (1975) carefully retained the single-index model, so that their results reflect the original paradox: the market portfolio exhibits unsystematic risk relative to itself. As a consequence, the articles lead to slightly defective conclusions concerning the imputed appraisal premium or required alpha. For example, to justify holding the market portfolio, in their formulas, one would require that it exhibit an (impossible) appraisal premium relative to itself to offset its (spurious) unsystematic risk.

When formula (T3.5) is substituted for (SI), the previous results do become special cases of the present results. They were pioneering efforts, which anticipated much of what is contained in this appendix.

The single-index model was correctly viewed as a useful simplifying device, which would clearly show the relationship between investment judgment and investment risk. The validity of that model, abstracting from its heuristic usefulness, is an empirical matter. The model of investment risk estimated by Rosenberg and Marathe (1975, 1976), applied to managed portfolios, has shown that from 20 to 70 percent of unsystematic variance is due to common factors of return which are excluded from a single-factor model. Thus, from the point of view of application, the single-factor model must be rejected.

A.4. Portfolio Optimization with a Mean-Variance Utility Function

Suppose that the client summarizes his attitudes with regard to reward and risk by a utility function which is a weighted sum of expected portfolio excess return and portfolio variance. The coefficient for expected reward may be set to unity, so that the units of the utility function are "certain excess return." The coefficient for variance will presumably be negative, expressing risk aversion, and may therefore be written as $-\lambda_v$, where $\lambda_v$ is the penalty, in terms of certain excess return, that arises from one unit of variance. Thus the utility function

$$U = \mu_p - \lambda_v \sigma_p^2$$

gives the "certainty equivalent excess return" (CEER) of a portfolio.
Consider the problem of choosing the vector of portfolio investment weights, \( \mathbf{h} \), so as to maximize the utility of portfolio return. Assume that there are no constraints on these investment weights. Then the solution to this optimization problem, the "optimal portfolio," exhibits the properties stated in Theorem 4. The basic result was first expounded in a portfolio context by Markowitz (1959), and its further properties have been developed by several authors, notably Sharpe (1970) and Treynor-Black (1973) in the context of the single-index model, and Mossin (1973).

The properties of the solution can be characterized in a number of useful ways. We will be concerned with the following:

\[
\begin{align*}
    z_p &= \mu_p / \sigma_p & \text{the ratio of expected excess return to the standard deviation of return, or the reward to variability ratio (RVR) of the portfolio.} \\
    z_p^* &= \mu_p^2 / \sigma_p^2 & \text{the ratio of squared expected excess return to the variance of return, or the Sharpe Ratio.} \\
    \lambda &= \sigma_p^2 / \mu_p & \text{the ratio of portfolio variance to mean excess portfolio return, or the Risk Acceptance Parameter (RAP).}
\end{align*}
\]

---

**Theorem 4 (Optimal Portfolio with Quadratic Objective Function, Unlimited Borrowing and Lending at the Riskless Rate, and No Constraints on Investment Holdings)**

Risk Acceptance Parameter (RAP):

\[
\text{(T4.1)} \quad \lambda = \frac{1}{2\lambda_v}.
\]

Equations that relate the holdings to expected excess returns:

\[
\text{(T4.2)} \quad \frac{1}{\lambda} \mathbf{Vh} = \mathbf{\mu}.
\]

Optimum investment holdings:

\[
\text{(T4.3)} \quad \mathbf{h} = \lambda \mathbf{V}^{-1} \mathbf{\mu}.
\]
Sharpe ratio for the portfolio:

\[(T4.4)\]
\[z_P^2 = \mu^T \Sigma^{-1} \mu.\]

RVR for the portfolio:

\[(T4.5)\]
\[z_P = \sqrt{\mu^T \Sigma^{-1} \mu}.\]

Expected excess return and variance of return:

\[(T4.6)\]
\[\mu_P = \lambda z_P^2 ; \quad \sigma_P^2 = \lambda z_P^2.\]

Utility or CEER of optimum solution:

\[(T4.7)\]
\[U = \frac{1}{2} \mu_P = \frac{1}{2} \lambda z_P^2.\]

Proof: By the use of differential calculus in matrix notation (cf. Theil (1971), Ch. 1), the first-order conditions for an optimal solution are:

\[\frac{\partial U}{\partial \mu} = \mu - 2\lambda \nu \Sigma = 0.\]

Also, the matrix of second partial derivatives,

\[\frac{\partial^2 U}{\partial h \partial h} = -2\lambda \nu \nu.\]

is negative definite, since \(\nu\) is positive definite, so that the extremum given by the first-order conditions is indeed a maximum. The first-order conditions can be reexpressed by substitution of \(\lambda = 1/2\nu\), as (T4.2). Premultiplication by \(\nu^{-1}\) (which exists since \(\nu\) is positive definite) yields (T4.3). When (T4.3) is substituted into (T1.1) and (T2.3), (T4.6) results. (T4.1), (T4.4), (T4.5), and (T4.7) are then easily verified.

Next consider the decomposition of investment returns into systematic and unsystematic returns. To sensibly decompose the investments, the forecasts of extraordinary return or appraisal premia, \(\alpha\), are assumed to be "market neutral," as explained in section A.7. Allow a more
general utility function,

\[ U = \mu_P - \lambda v \sigma^2_{PS} - \kappa \lambda v \omega^2_P, \]

where \( \kappa = (\text{aversion to unsystematic risk}) / (\text{aversion to systematic risk}) \).

Then the formulas in Theorem 5 follow. These are mainly generalizations of previous results of Treynor-Black (1973).

---

**Theorem 5 (Portfolio Optimization in Terms of Systematic and Unsystematic Investments)**

Risk acceptance parameters for systematic and unsystematic risk:

\[ \text{RAP}_{ES} \equiv \lambda = \frac{1}{2\lambda v}; \quad \text{RAP}_{EU} \equiv \lambda^* = \frac{\lambda}{\kappa} = \frac{1}{2\lambda v}. \]

Equations that relate investments to expected excess returns:

\[ h = \delta + \beta_P h_M; \quad \frac{1}{\lambda} \sigma^2_M \beta_P = \mu_M; \quad \frac{1}{\lambda^*} \delta v = \alpha. \]

Optimum investment holdings:

\[ h = \delta + \beta_p h_M; \quad \beta_p = \frac{\lambda \mu_M}{\sigma^2_M}; \quad \delta = \lambda^* v^{-1} \alpha = \lambda^* R^\alpha. \]

Sharpe Ratio for the portfolio:

\[ z_p^2 = \mu_p^2 / \sigma_p^2; \quad z_{PS}^2 = z_M^2; \quad z_{PU}^2 = \alpha^* R^\alpha. \]

Simplification when \( \kappa = 1 \):

\[ z_f^2 = z_{PS}^2 + z_{PU}^2. \]

Expected excess return and variance of return:

\[ \mu_p = \mu_{PS} + \alpha_p; \quad \mu_{PS} = \lambda z_M^2; \quad \alpha_p = \lambda z_{PU}^2. \]

\[ \sigma_p^2 = \sigma^2_{PS} + \omega_p^2; \quad \sigma_{PS}^2 = \lambda^2 z_M^2; \quad \omega_p^2 = \lambda^2 z_{PU}^2. \]

Utility or CEER of optimum solution:

\[ U_p = U_{PS} + U_{PU}; \quad U_{PS} = \frac{1}{2} \mu_{PS}; \quad U_{PU} = \frac{1}{2} \alpha_p. \]
"Balancing" relationship among investment components:

\[ \frac{\sigma_{PS}}{z_{PS}} = \lambda ; \quad \frac{\omega_p}{z_{PU}} = \lambda \gamma ; \quad \text{when } \kappa = 1, \quad \frac{\sigma_{PS}}{z_{PS}} = \frac{\omega_p}{z_{PU}} = \frac{\sigma_p}{z_p} = \lambda . \]

**Proof:** From (T2.1) and (T2.4), the first-order conditions for an optimum, analogous to (T4.2), are \( \mu - 2\lambda \psi (\alpha M \tilde{\beta} \psi) h - 2k \lambda \psi h = 0 \). By substituting \( h = \beta_p h_p + \delta \), and \( \mu = \mu_M \beta + \alpha \), one obtains:

\[ \frac{1}{\lambda} (\sigma_M^2 \beta \psi (\psi_R)(\beta_p h_p + \delta)) = \mu_M \beta + \alpha . \]

The left-hand side may be simplified by observing that since \( h_M \beta = \beta_M = 1 \),

\[ Rh_M = R(I - h_M \beta')h_M = Rh_M - Rh_M = 0, \text{ and } \beta' \delta = \beta' (I - h_M \beta') h = 0 . \]

Therefore, multiplying out the terms on the left-hand side,

\[ \frac{1}{\lambda} (\sigma_M^2 \beta \psi (\psi_R \delta)) = \mu_M \beta + \alpha . \]

Next, we break this first-order condition into two parts: a matrix equation of rank \((N-1)\) relating to unsystematic return, and a matrix equation of rank 1 relating to systematic return. The matrix equation of rank \(N-1\) is obtained by premultiplying by \((I - \beta M \dot{x})\) on both sides, which, after simplification, yields

\[ \frac{K}{\lambda} R \dot{x} = \gamma . \]

This is the desired condition for residual return. Next, premultiplying by \( \beta M \dot{x} \) to complete the equation system (note that \( I - \beta M \dot{x} + \beta M \dot{x} = I \), so that the sum of the two new equations is the original system), and simplifying,

\[ \frac{1}{\lambda} \beta_p \sigma_M^2 \beta = \mu_M \beta . \]

(The simplification of the right-hand side required the "market-neutrality" condition that \( h_M \alpha = 0 \).) This equation can be satisfied if
only if \( \frac{1}{\lambda} \beta_p \sigma_p^2 = \mu_{M^-} \), which is the desired condition on systematic return. This completes the derivation of (T5.2).

The above condition is immediately solved to derive \( \beta_p \). To solve for \( \delta \), note that matrix \( \overset{\sim}{R} \) is an \( N \times N \) matrix of rank \( N - 1 \). As such, it does not possess an inverse. The condition \( \frac{K}{\lambda} \overset{\sim}{R} \delta = \alpha \) can be solved for \( \delta \) in two ways. First, by use of the idempotent matrix, we may show that the condition is equivalent to \( \frac{K}{\lambda} \overset{\sim}{V} \delta = \delta \), which may be solved by \( \overset{\sim}{V}^{-1} \) to obtain \( \delta = \overset{\sim}{\lambda} \overset{\sim}{V}^{-1} \alpha \). Alternatively, the pseudoinverse of \( \overset{\sim}{R} \) may be used, which is written as \( \overset{\sim}{R}^+ \), to obtain \( \delta = \overset{\sim}{\lambda} \overset{\sim}{R}^+ \alpha \).

These two solutions are identical, of course.

This completes the derivation of (T5.3) and ends the difficult part of the proof. When (T5.3) is substituted into (T1.2), (T2.4), and (T2.5), one finds (T5.1) and (T5.4)-(T5.8), after noting that \( \overset{\sim}{R} R^+ = R \).

The above derivation employed some matrix algebra which is unfamiliar to most social scientists—specifically, the pseudoinverse \( \overset{\sim}{R}^+ \) of the singular matrix \( \overset{\sim}{R} \). Because of the property of the active holdings that \( (I - h_M \overset{\sim}{M}') \delta = \delta \), \( \overset{\sim}{V} \) and \( \overset{\sim}{V}^{-1} \) may be used to replace \( \overset{\sim}{R} \) and \( \overset{\sim}{R}^+ \) in all expressions involving \( \delta \). Thus, \( \overset{\sim}{V} \) may be regarded as an invertible substitute for \( \overset{\sim}{R} \). In practice, we may also use an approximation to \( \overset{\sim}{R} \), in which the diagonal entries are increased slightly relative to the off-diagonal entries, and thereby obtain an approximation \( \overset{\sim}{R}^* \), which is positive definite and possesses an inverse. Since invertible substitutes for \( \overset{\sim}{R} \) are available, the notation \( \overset{\sim}{R}^{-1} \) will be used in the balance of this appendix: it may be understood as representing either the pseudoinverse of \( \overset{\sim}{R} \) itself, or the inverse of a substitute matrix.

A.5. Optimal Portfolios with a More General Utility Function

Thus far, it has been assumed that the investor possessed preferences that could be represented by the utility function \( U = \mu_p - \lambda \sigma_p^2 \).
or \( U = \mu_P - \lambda \sigma^2_P - \kappa \omega \omega^2 \). What if his preferences are more general in character?

For an arbitrary preference function, suppose that the optimal portfolio is somehow discovered. Now consider a small change in some aspect of the portfolio, represented by a continuous variable "x." When the preference function is of the form \( U = U(\mu, \sigma_P^2) \) [a general function of the mean and variance of portfolio return], then the effect of such a change upon utility is:

\[
\frac{\partial U}{\partial x} = \frac{\partial U}{\partial \mu_P} \frac{\partial \mu_P}{\partial x} + \frac{\partial U}{\partial \sigma P^2} \frac{\partial \sigma_P^2}{\partial x}.
\]

Alternatively, if the utility function is not in the mean-variance form, suppose that the only important considerations, with regard to small modifications of the portfolio, are the effects of such changes upon the mean and variance of portfolio return. This assumption amounts to assuming the utility function is of the mean-variance form, but only in the locality of the optimum. In either of these cases, the first-order condition for optimality of the portfolio is that:

\[
\frac{\partial U}{\partial \mu_P} \frac{\partial \mu_P}{\partial x} + \frac{\partial U}{\partial \sigma P^2} \frac{\partial \sigma_P^2}{\partial x} = 0.
\]

This, in turn, will hold, if and only if

\[
\frac{\partial \mu_P}{\partial x} - \lambda \frac{\partial \sigma_P^2}{\partial x} = 0, \quad \lambda = -\frac{\partial U}{\partial \sigma_P^2} \frac{\partial \sigma_P^2}{\partial \mu_P}.
\]

There is no requirement that the utility function, when viewed over the entire space of possible portfolio distributions, is of the mean-variance form; the only requirement is that it can be sufficiently well-approximated by a function of a portfolio mean and variance when small changes in portfolio composition are considered. This seems to be an entirely plausible assumption, as far as the client's attitude toward risk and return is concerned. However, it does rule out other aspects of the portfolio that might be important in some cases (cash yield as distinct from total return, tax effects on transactions, etc.).
Here, $\lambda_Y$ is the required risk premium per unit of variance that applies to the optimal solution. (The prior assumption that the utility function was linear placed the additional restriction that $\lambda_Y$ applied at all possible solutions, not just at the optimal solution.)

The above condition must apply for local changes in all aspects of the portfolio. In particular, suppose that the aspect to be changed is overall aggressiveness, that every holding is increased by a factor $x$. Then the new vector of holdings is given by $(1+x)h$. It is easily verified that:

$$\frac{\partial \mu_p}{\partial x} = \mu_p \quad \text{and} \quad \frac{\partial \sigma_p^2}{\partial x} = 2\sigma_p^2,$$

where the subscript $p$ denotes the optimal portfolio. When these are substituted into the previous formula, one finds that

$$\mu_p - \lambda_Y \left(2\sigma_p^2\right) = 0 \quad \text{or} \quad \lambda \equiv \frac{\sigma_p^2}{\mu_p} = \frac{1}{2\lambda_Y}.$$

Thus, the relationship, at the optimum portfolio, between $\lambda$ and $\lambda_Y$, holds just as it does for the linear utility function. The first-order conditions for the optimum take the same form as for a linear utility function, with the slope of the indifference curve at the optimum determining the coefficients $\lambda$ and $\lambda_Y$. Following this approach, Theorems 4 and 5 may be derived for a general utility function which is locally of the mean-variance form. All propositions of those theorems remain valid, except concerning the CEER, which cannot be computed because the intercept of the indifference curve is unknown.

**Theorem 6 (Generalization to a Nonlinear Mean-Variance Utility Function; Approximation for an Arbitrary Utility Function)**

Suppose that the utility function is of the form $U = U(\mu_p, \sigma_p^2, \omega_p^2)$, or, alternatively, that in the neighborhood of the optimal portfolio, the utility of slightly modified portfolios may be approximated by a function of this form. Define
\[ -\lambda_\psi = - \frac{\partial U}{\partial \psi^2} \left/ \frac{\partial U}{\partial \psi} \right. \]  
* \[ \kappa = \frac{\partial U}{\partial \psi^2} \left/ \frac{\partial U}{\partial \psi} \right. \]

where the partial derivatives are evaluated at the optimum point. Then all statements of Theorems 4 and 5 hold, excepting (T4.7) and (T5.7).

### A.6. The Equity Investment in the Context of a Portfolio of Equity, Bonds, and Cash

Let the portfolio be divided among cash, bonds, and equity, with investment proportions \( w_c, w_b, \) and \( w_e \). Let \( \beta_c, \beta_b, \) and \( \beta_e \) be the betas of these three component portfolios. Then the portfolio beta is

\[ \beta_p = w_c \beta_c + w_b \beta_b + w_e \beta_e. \]

Assume that the cash category is a risk-free asset, with return identically equal to the risk-free rate. Hence, the variance of the return to cash is zero, and its beta is therefore also zero. Thus, \( \beta_p = w_b \beta_b + w_e \beta_e. \)

Consider a bond "submarket portfolio" and a stock "submarket portfolio," each a capitalization-weighted portfolio of the assets in that category. Each will exhibit unsystematic returns relative to the true market portfolio comprising both. These unsystematic returns will have zero covariance with the overall market return, by construction. Each unsystematic return will have a variance (residual risk), denoted by \( T_b^2 \) and \( T_s^2 \). Also, the covariance of the residual return on the bond component with the residual return on the equity component is ordinarily so small that it can safely be neglected. (To elucidate this point,

---

1. The assumption is certainly plausible, for the variance of the return to the typical cash equivalent is very small compared to other variances in the economy. However, since the uncertainty in return arises primarily from unpredictable changes in medium-term rates and frashures of defaults brought about by economywide financial crises, the component of return is likely to be highly correlated with events in the other two categories and, hence, a nonnegligible contributor to overall portfolio risk. When we express the investors' desires in terms of real return, rather than nominal return, by adjusting for inflation, unpredictable inflation becomes another element of risk in cash equivalents, which is correlated with the other categories returns, strengthening this argument. Nevertheless, the risk of cash equivalents is relatively small, and it will simplify the exposition to assume that cash is risk-free.
note that the covariance would be large if the two portfolios consisted of the bonds and stock, respectively, of a company threatened with bankruptcy, such as Pan American Airways in the recent past. However, when the bond portfolio is well diversified, as it almost always is, the common exposure of bonds and stocks to companies' specific risk occurs with such small portfolio weights as to be safely neglected.)

Hence, the risk of a weighted combination of the two submarket portfolios is approximately

\[ \sigma_P^2 = (w_B \beta_B + w_E \beta_E)^2 \sigma_M^2 + (w_B - w_M)^2 \tau_B^2 + (w_E - w_M)^2 \tau_E^2. \]

Next, consider the possibility of active holdings in the two submarkets. If at least one portfolio is diversified, the covariance between submarkets can be neglected safely. Hence, the added risk will arise from the unsystematic variance of differential holdings within the two submarkets, denoted by the characters \( \omega_{PB}^2 \) and \( \omega_{PE}^2 \). Thus, total portfolio risk is given approximately by:

\[
\sigma_P^2 = (w_B \beta_B + w_S \beta_S)^2 \sigma_M^2 + \{ (w_B - w_M)^2 \tau_B^2 + \omega_{PB}^2 \}
+ \{ (w_E - w_M)^2 \tau_E^2 + \omega_{PE}^2 \}. 
\]

The next approximation is to take the equity market portfolio as a "pseudomarket" portfolio. This is permissible because the variance of the bond market and the capitalization weight of the bond market are both small relative to those characteristics of the stock market. Consequently, the overall market return is highly correlated with the equity market return, and there is only a small loss of validity when the equity market is misdefined as the "Market Portfolio." Since this is currently universal practice, it is convenient to do so here.

When the equity market becomes the market portfolio, \( \beta_E \) of the equity market portfolio falls to unity, \( \beta_B \) falls by a slightly greater proportion, and \( \sigma_M^2 \) rises in inverse proportion to preserve systematic risk at nearly the original level. With respect to unsystematic risk, \( \tau_E^2 \) falls to zero and \( \tau_B^2 \) rises slightly. The model for portfolio risk becomes:
\[ \sigma_p^2 = (\omega_B \beta_B + \omega_E \beta_E)^2 \sigma_M^2 + \omega_E^2 \omega_{PE}^2 + \{\omega_B^2 (\tau_B^2 + \omega_{PB}^2)\} \]

\[ \tau = \beta_P^2 \sigma_M^2 + \omega_E^2 \omega_{PE}^2 + \kappa \sigma \]

where \( \kappa \) does not involve equity investments in any way. Similarly, the mean portfolio return may be shown to be

\[ \mu_P = \beta_P \mu_M + \omega_E \alpha_E + \kappa \mu \]

where \( \kappa \) is again not dependent on equity investments.

These two equations have been obtained through a series of approximations, none of which importantly affect the results. Their form is identical to that for an all-equity portfolio, save for the appearance of \( \omega_E \) multiplying \( \alpha_E \) and \( \omega_{PE} \). When the derivations for Theorems 5 and 6 are repeated in the new format, the only consequence is that \( h \) and \( \delta \) are now defined as proportions of the total portfolio. To convert these to proportions in the equity portfolio, one would have to divide by \( \omega_E \).

Suppose that the bond contribution to portfolio beta is negligible, so that \( \beta_P \approx \omega_E \beta_E \). Then \( \beta_E \approx \beta_P / \omega_E \). Thus, all portfolio equity positions \( (\beta_E, h_E, \delta_E) \) are \( 1/\omega_E \) times the corresponding positions of the total portfolio. These positions would result if the theorem were applied to the equity portfolio alone, with modified RAP \( \lambda_E = \lambda / \omega_E \). Following this approach, the following theorem can easily be verified:

---

**Theorem 7 (Approximation for the Equity Component of a Larger Portfolio)**

Suppose that we are concerned with the equity component of a larger portfolio of cash, bonds, and equity. Let the market portfolio be defined as the equity market portfolio. Then two approximations to Theorems 5 and 6 are possible:

I. Let \( \beta_P \) be the beta of the total portfolio with respect to the equity market, let \( h \) and \( \delta \) be defined as investment proportions...
of the total portfolio in individual equity assets. Then all propositions of Theorems 5 and 6 remain valid except that the unsystematic risk \( \kappa_\sigma \), appraisal premium \( \kappa_u \), and utility contribution of the bond component are not included in portfolio properties.

II. Alternatively, ignoring the nonequity and components of the portfolio, let \( \beta_E, h_E, \) and \( \delta_E \) denote the beta and investment proportions of the equity subportfolio. Thus, \( \delta_E = \frac{1}{w_E} \delta \) and \( h_E = \frac{1}{w_E} h \). Assume that the balance of the portfolio makes no contribution to portfolio beta, so that \( \beta_E \sim \frac{1}{w_E} \beta_P \). Then all propositions of Theorems 5 and 6 hold approximately, with \( \beta_E, h_E, \) and \( \delta_E \) substituted for \( \beta_P, h, \) and \( \delta \). All constructs now refer to the equity component. Properties of the total portfolio may be computed as weighted sums of the properties of this equity component and the balance of nonequity assets.

Both approximations err due to the improper market portfolio. If \( \beta_E = \beta_P/w_E \), the approximations are identical. Otherwise, the second is inferior. When \( \beta_E < \beta_P/w_E \) (the usual case of a bond portfolio with slightly positive beta), approximation II understates \( \text{RAP}_{EU} \) relative to \( \text{RAP}_{ES} \). Thus, in the argument of section 4, the approximation causes an upward bias in \( \alpha_p^{\text{req}} \), and hence an upward bias in the break-even point.

A.7 Construction of Judgmental Alphas from Forecasts of Return

Thus far in the appendix, the mean vector for unsystematic returns, \( \alpha \), has been accepted as given. The remainder of the appendix is concerned with the process whereby the manager formulates \( \alpha \). Similar conclusions were reached by Treynor-Black (1973), following a less rigorous approach.

The basic process for returns forecasting will be taken as given. It may be viewed as providing a "window" into the future. Forecasts of unsystematic returns are generated, by means of this window. At any time, let \( \hat{u}_1, \ldots, \hat{u}_N \) be the \( N \) forecasts. The information process will sometimes work better and sometimes less well. The important property of the process is the typical or average information which it contains: in this sense, the actual values of the forecasts,
at any point in time, are a "realization" of the process. At different times there will be different realizations. All have the common property of a certain expected information content.

Let \( \hat{\mathbf{u}} \) be the column vector of \( N \) forecasts. The quality of these forecasts is summarized by two properties: the covariance of each forecast with actual future returns—a measure of information content—and the variance of the forecast, which is the sum of true information and noise. Let \( C_{mn} \) denote \( \text{Cov}[\hat{u}_m, u_n] \), the covariance between the forecast for stock \( m \) and the actual future residual return for stock \( n \). Also, let \( D_{mn} \) denote \( \text{Cov}[\hat{u}_m, \hat{u}_n] \), the covariance between the \( m \)th and \( n \)th forecasts. These terms can be collected in two square \( N \times N \) matrices: \( C \), the covariance matrix between the forecasts and the actual returns; and \( D \), the variance matrix of the forecasts. Thus,

\[
\tilde{C} = \text{Cov}[\hat{u}, u] \quad \text{and} \quad \tilde{D} = \text{Var}[\hat{u}].
\]

In the absence of the information provided by the manager's window on the future, the unsystematic returns are distributed with mean vector \( \bar{u} \), and with variance matrix \( R \).

Then the joint distribution of his forecast vector \( \hat{\mathbf{u}} \), and the vector of returns \( \mathbf{u} \), is given by the matrix equations:

\[
\begin{align*}
\mathbb{E} \left( \begin{bmatrix} \hat{u} \\ u \end{bmatrix} \right) &= \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\
\text{Var} \left( \begin{bmatrix} \hat{u} \\ u \end{bmatrix} \right) &= \begin{bmatrix} D & C \\ C' & R \end{bmatrix}.
\end{align*}
\]

Define \( \mathbf{a} \) as the best prediction for \( \mathbf{u} \), and \( \mathbf{M} \) as its mean square error matrix, \( \mathbf{M} = \mathbb{E}[(\mathbf{u} - \mathbf{a})(\mathbf{u} - \mathbf{a})'] \).

**Theorem 8 (Optimal Unbiased Prediction)**

The optimal unbiased predictor for \( \mathbf{u} \) is:

\[(T8.1) \quad \mathbf{a} = \tilde{C}^{-1} \tilde{u}.\]

The variance of \( \mathbf{a} \), denoted by \( \mathbf{A} \), equals its covariance with \( \mathbf{u} \):

\[(T8.2) \quad \mathbf{A} = \text{VAR}(\mathbf{a}) = \text{COV}(\mathbf{a}, \mathbf{u}) = \tilde{C} \tilde{D}^{-1} \tilde{C}.\]
The mean square prediction error matrix:

\[(T8.3) \quad \mathbb{M} = \mathbb{R} - \mathbb{C} \mathbb{D}^{-1} \mathbb{C} = \mathbb{R} - \mathbb{A} .\]

Notice that total variance \( \mathbb{R} \) therefore decomposes into Mean Square Prediction Error \( \mathbb{M} \) plus explained variance \( \mathbb{A} \).

\( \hat{\mathbb{C}} \) is the minimum-mean-square-error linear unbiased predictor for \( \mathbb{C} \). Furthermore, if \( \mathbb{C} \) and \( \hat{\mathbb{C}} \) are jointly normally distributed, \( \hat{\mathbb{C}} \) equals the posterior mean for \( \mathbb{C} \), conditional on the value of \( \hat{\mathbb{C}} \), and is therefore the minimum-mean-square error predictor among all possible unbiased predictors (not just linear ones).

**Proof:** The formula for \( \alpha \) is a well-known theorem in statistics (e.g., Theil (1971)). The equality of variance and covariance follows from:

\[
\text{Var}[\hat{\alpha}] = \mathbb{C} \mathbb{D}^{-1} \text{Var}[\hat{\mathbb{C}}] \mathbb{D}^{-1} \mathbb{C} = \mathbb{C} \mathbb{D}^{-1} \mathbb{A} \mathbb{D}^{-1} \mathbb{C} = \mathbb{C} \mathbb{D}^{-1} \mathbb{C} \\
\text{Cov}[\hat{\alpha}, \mathbb{C}] = \mathbb{C} \mathbb{D}^{-1} \text{Cov}[\hat{\mathbb{C}}, \mathbb{C}] = \mathbb{C} \mathbb{D}^{-1} \mathbb{C} .
\]

Thus, when the original forecasts of return are transformed so as to be judgmental alphas, the variance of the alphas is equal to their covariance with the returns. Elsewhere in the appendix it is assumed that the forecasts have been transformed in this way.

By definition, the unsystematic return of the market portfolio is zero. Therefore, an elementary and important property of the judgmental alphas is "market neutrality," the requirement that the alphas imply a zero market alpha (Treynor-Black, 1973). Mathematically, this requirement is:

\[ \sum_{n=1}^{N} h_{Mn} \alpha_n = h_M' \alpha = 0 . \]

---

1This property is defined as follows. Let any other linear predictor be denoted by \( \mathbb{K} \hat{\mathbb{C}} + \mathbb{k} \), for some matrix \( \mathbb{K} \) and vector \( \mathbb{k} \). For this predictor to be unbiased, \( E(\mathbb{K} \hat{\mathbb{C}} + \mathbb{k}) = E(\mathbb{C}) \). Subject to this condition, the mean square error matrix always exceeds \( \mathbb{M} \), in that

\[ E[(\mathbb{K} \hat{\mathbb{C}} + \mathbb{k} - \mathbb{C})(\mathbb{K} \hat{\mathbb{C}} + \mathbb{k} - \mathbb{C})'] = \mathbb{M} + \mathbb{Q} , \]

where \( \mathbb{Q} \) is a positive semi-definite matrix.
This property will automatically result if the matrices $\hat{C}$ and $\hat{D}$ are carefully constructed so as to imply zero unsystematic risk for the market portfolio ($\hat{C}_M = \hat{D}_M = 0$), in which case pseudoinverses must replace inverses for these matrices of rank $N-1$. Otherwise, a constant may have to be added to every alpha, so that the adjusted alphas are market neutral.

A.8. The Simplification When a Constant Proportion of Variance Is Explained

The matrix equations developed in the previous section become much more transparent when the forecasts account for the same proportion of variance for all aspects of future returns. This may be termed the "constant proportional explanation" (CPE) assumption. In this case, the covariance of forecasts with future returns simplifies to $\hat{C} = \hat{cR}$, for some constant $\hat{c}$ that measures the typical covariance or "information content." The variance matrix $\hat{D}$ simplifies to $\hat{D} = \hat{dR}$, for some constant $\hat{d}$ which captures the total variability of the forecasts. We may write $\hat{d} = \hat{c} + \hat{e}$, where $\hat{c}$ is the valid information content and $\hat{e}$ is the "noise" in the forecasts. Define $f = \hat{d}/\hat{c}$ as the ratio of valid covariance to total variability. Notice that the correlation coefficient between alphas and returns may now be defined as:

$$\rho = \frac{\hat{c}}{\sqrt{\hat{d}(1)}} = \frac{\hat{c}}{\sqrt{\hat{d}}}.$$

With these simplifications, the formulae of the previous section become:

$$\hat{\alpha} = \hat{c}^{-1}\hat{\beta} = \hat{cR}(\hat{dR})^{-1}\hat{\beta} = \left(\frac{\hat{c}}{\hat{d}}\right) \hat{R}^{-1}\hat{\beta} = \left(\frac{\hat{c}}{\hat{d}}\right) \hat{\beta} = f\hat{\beta}$$

$$\hat{A} = \hat{C}^{-1}\hat{C} = \hat{cR}(\hat{dR})^{-1}\hat{cR} = \left(\frac{\hat{c}}{\hat{d}}\right) \hat{cR} = \rho^2\hat{R}$$

$$\hat{M} = \hat{R} - \hat{C}^{-1}\hat{C} = \hat{R} - \hat{cR}(\hat{dR})^{-1}\hat{cR} = (1 - \hat{c}^2/\hat{d})\hat{R} = (1 - \rho^2)\hat{R}.$$  

All formulae involve the ratio $f$, which is the ratio of valid information in the forecasts (the legitimate covariance with future returns) to the total variability of the forecasts. The forecasts $\hat{\beta}$ are scaled down by this factor to obtain the judgmental alphas or, equivalently,
the covariance coefficient \( c \) is scaled down by this factor to obtain the explained variance \( \rho^2 \).

With the simplification of constant proportional explained variance, the results in this section become entirely comparable to previous results. For example, Treynor and Black's (1973) adjustment rule includes the terms \( f \) and \( \rho \), and their additive adjustment would correspond to achieving a market alpha of zero.

A.9. The Expected Sharpe Ratio and the Approximation That \( M \sim R \)

When the appraisal premia from a manager are available, uncertainty concerning residual returns is reduced. Thus, instead of using the unsystematic return variance matrix \( \tilde{R} \), the mean-square error matrix \( \tilde{M} \) should properly be used. In the CPE case, this amounts to substituting \( \tilde{M} = (1 - \rho^2) \tilde{R} \) for \( \tilde{R} \). One purpose of this section is to show that \( \tilde{M} \) is negligibly different from \( \tilde{R} \), so that \( \tilde{R} \) may be retained as the variance of unsystematic return.

From (A5.4), the optimal active holdings based upon the judgmental alphas have a Sharpe Ratio of

\[
\alpha^2 = \alpha^\top \tilde{M}^{-1} \alpha.
\]

This value depends upon the set of alphas which the judgmental process generates. The prior expectation for this value, which will give an average value for possible realizations of \( \alpha \), is easily evaluated:

\[
E(\alpha^2) = E(\alpha^\top (R - \alpha) \alpha) = E(\text{Trace}((R - \alpha) \alpha^\top)) = \text{Trace}(E(R - \alpha) \alpha^\top) = \text{Trace}(\tilde{R} - \tilde{A}^\top),
\]

where \( \text{Trace} \) denotes the trace of a matrix, defined as the sum of the diagonal elements. Under the CPE assumption, this simplifies to:

\[
E(\alpha^2) = \text{Trace}(R - \rho^2 \tilde{R}) = \text{Trace}\left(\frac{\rho^2}{1 - \rho^2} R \tilde{R}\right)
\]

\[
= \frac{\rho^2}{1 - \rho^2} \text{Trace}(R \tilde{R})
\]

\[
= (N - 1) \frac{\rho^2}{1 - \rho^2}.
\]
Thus, the expected Sharpe Ratio is equal to the ratio of explained to residual variance \( \rho^2 /(1-\rho^2) \) times \( N-1 \). The term \( N-1 \) appears rather than \( N \), because one dimension among the \( N \) assets corresponds to the market portfolio, with \( \alpha \) being identically zero.\(^1\) Conceptually, the expected Sharpe Ratio is just the sum of the expectations for each of the remaining \( N-1 \) dimensions. Compare Treynor and Black, where similar results are obtained for the diagonal model, and Ferguson's (1975) extension, where the possibility of differing \( \rho_n^2 \) for different assets is allowed, and the expected Sharpe Ratio is plotted as a function of the number, \( N \), of assets which are followed.

Suppose now that an optimal portfolio, based upon the judgmental alphas, is maintained for \( T \) years. For the moment, neglect transaction costs, so that the portfolio can be appropriately maintained at a continual optimal status. Then, the expected Sharpe Ratio at each moment of time, expressed in terms of annual return, is as given in the above formula. After \( T \) years, it will be possible to estimate the Sharpe Ratio, based upon the cumulative performance of the portfolio. As is shown in Rosenberg (1977b), the estimated cumulative annualized Sharpe Ratio for unsystematic return equals the squared "t-statistic" for the portfolio appraisal premium. The expected value for the squared t-statistic is approximately

\[
E(t^2) \approx T E(z_{Pu}^2) + 1 = T(N-1)(\rho^2/(1-\rho^2)) + 1.
\]

Conversely, the value of \( \rho^2 \) may be solved for as a function of \( N \), \( T \), and \( E(t^2) \):

\[
\rho^2 = \frac{E(t^2) - 1}{T(N-1) + E(t^2) - 1}.
\]

This formula allows us to obtain some insight into values of \( \rho^2 \) which may reasonably be expected in the institutional environment.

\(^1\)The product \( R^+R \), where \( R^+ \) is the pseudoinverse of \( R \), is the idempotent matrix corresponding to the perpendicular projection onto the \( N-1 \) dimensional subspace orthogonal to \( h_{N}^\prime \), and hence will have trace equal to \( N-1 \), the subspace dimension.
Historical studies of the performance of institutional portfolios over five-year periods have shown that the number of squared t-statistics greater than 4.0 is little greater than would be expected from chance variation with appraisal ability of zero, in which case \( E(t^2) = 1 \). Hence, cases of institutional investors having \( (E(t^2) = 5) \) must be rare, and institutions with expected values as high as \( (E(t^2) = 17) \) must be rare indeed. Thus, a very conservative upper bound for \( \rho^2 \) may be obtained by substituting \( E(t^2) = 17 \). Thus,

\[
\rho^2 \leq \frac{17 - 1}{5(N-1) + 17 - 1} = \frac{16}{5(N-1) + 16} = \frac{3.2}{(N-1) + 3.2}.
\]

Since the typical value for the list of followed assets \( N \) is at least 100, a conservative upper bound\(^1\) for \( \rho^2 \) is .03, corresponding to an explanation of 3% of total variance, and correlation of forecasts with returns of \( \rho = .175 \).

Since the value of \( \rho^2 \) is less than .03, and probably much less than that, very little difference is made by substituting \( 2 \) for \( M = (1-\rho^2)R \). Consequently, to simplify the notation in this appendix, the approximation \( M = R \) is used.

---

\(^1\) It is possible that measured values of \( t \) are downward biased as measures of true information content. The bias due to incurred transaction costs, and the management fee, if any, can be eliminated by the use of gross rather than net returns. A more subtle bias arises because portfolio weights may inadequately reflect underlying judgment. Such might be explained by suboptimal portfolio construction techniques, by externally imposed constraints on portfolios, or by limited portfolio adjustments to changing judgment (appropriate to avoid transaction costs). The upper bound for \( \rho^2 \) of .03 appears to be conservative enough to allow for all such biases.
REFERENCES


