A Practitioner's Guide to Arbitrage Pricing Theory

Edwin Burmeister
Duke University
Richard Roll
University of California, Los Angeles
Stephen A. Ross
Yale University

A fundamental principle of finance is the trade-off between risk and return. Unless a portfolio manager possesses special information, one portfolio can be expected to outperform another only if it is riskier in some appropriate sense. The crucial question is: "What is the appropriate measure of risk?"

Many attributes might be related to an asset's risk, including market capitalization (size), dividend yield, growth, price–earnings ratio (P/E), and so on. Use of these traditional descriptors, however, presents at least three problems:

1. Most are based on accounting data, and such data are generated by rules that may differ significantly across firms.
2. Even if all firms used the same accounting rules, reporting dates differ, so constructing time-synchronized interfirm comparisons is difficult.
3. Most importantly, no rigorous theory tells us how traditional accounting variables should be related to an appropriate measure of risk for computing the risk–return trade-off. Even if historical empirical relationships can be uncovered, without the foundation of a rigorous theory, one must be concerned that any historical correlation might be spurious and subject to sudden and material change.

Currently, only two theories provide a rigorous foundation for computing the trade-off between risk and return: the capital asset pricing model (CAPM) and the arbitrage pricing theory (APT).

The CAPM, for which William F. Sharpe shared the 1990 Nobel Memorial Prize in Economic Sciences, predicts that only one type of nondiversifiable risk influences expected security returns, and that single type of risk is "market risk."[1] In 1976, a little more than a decade after the CAPM was proposed, Stephen A. Ross invented the APT. The APT is more general than the CAPM in accepting a variety of different risk sources. This accords with the intuition that, for example, interest rates, inflation, and business activity have important impacts on stock return volatility.

Although some theoretical formulations of the APT can be more intellectually demanding than the CAPM, the intuitively appealing basics behind the APT are easy to understand. Moreover, the APT provides a portfolio manager with a variety of new and easily implemented tools to control risks and to enhance portfolio performance.

In the remainder of this paper, we will explain APT basics and the equations of the APT. We will also discuss macroeconomic forces that are the underlying sources of risk. We will then illustrate some risk exposure profiles and the resulting APT-based risk–return trade-offs, and we will show how these fundamental risks contribute to the expected and unexpected components of realized return. Finally, we will discuss several uses of the APT that every practitioner could easily apply.

The APT Basics

The CAPM and the APT agree that, although many different firm-specific forces can influence the return on any individual stock, these idiosyncratic effects tend to cancel out in large and well-diversified portfolios. This cancellation is called the *principle of diversification,* and it has a long history in the field of insurance. An insurance company has no way of knowing whether any particular individual will become sick or will be involved in an accident, but the company is able to predict its losses accurately on a large pool of such risks.

---

[1] More precisely, if $r_m(t)$ is the return (in time period t) on a market index, such as the S&P 500, the CAPM measure of the riskiness for asset i with return $r_i(t)$ is equal to that asset's CAPM beta defined by $\beta_i = \text{cov}[r_i(t), r_m(t)]/\text{var}[r_m(t)]$.

The CAPM is equivalent to the statement that the market index is itself mean–variance efficient in the sense of providing maximum average return for a given level of volatility. The index used to implement the CAPM is implicitly assumed to be an effective proxy for the entire market of assets.

An insurance company is not entirely free of risk, however, simply because it insures a large number of individuals. For example, natural disasters or changes in health care can have major influences on insurance losses by simultaneously affecting many claimants. Similarly, large, well-diversified portfolios are not risk free, because common economic forces pervasively influence all stock returns and are not eliminated by diversification. In the APT, these common forces are called *systematic* or *pervasive risks.*

According to the CAPM, systematic risk depends only upon exposure to the overall market, usually proxied by a broad stock market index, such as the S&P 500. This exposure is measured by the CAPM *beta,* as defined in Footnote 1. Other things equal, a beta greater (less) than 1.0 indicates greater (less) risk relative to swings in the market index.[2]

The APT takes the view that systematic risk need not be measured in only one way. Although the APT is completely general and does not specify exactly what the systematic risks are, or even how many such risks exist, academic and commercial research suggests that several primary sources of risk consistently impact stock returns. These risks arise from unanticipated changes in investor confidence, interest rates, inflation, real business activity, and a market index.

Every stock and portfolio has exposures (or betas) with respect to each of these systematic risks. The pattern of economic betas for a stock or portfolio is called its *risk exposure profile.* Risk exposures are rewarded in the market with additional expected return, and thus the risk exposure profile determines the volatility *and* performance of a well-diversified portfolio. The profile also indicates how a stock or portfolio will perform under different economic conditions. For example, if real business activity is greater than anticipated, stocks with a high exposure to business activity, such as retail stores, will do relatively better than those with low exposures to business activity, such as utility companies.

Most importantly, an investment manager can control the risk exposure profile of a managed portfolio. Managers with different traditional styles, such as small-capitalization growth managers and large-capitalization value managers, have differing inherent risk exposure profiles. For this reason, a traditional manager's risk exposure profile is congruent to a particular *APT style.*

Given any particular APT style (or risk exposure profile), the difference between a manager's expected return and his or her actual performance is attributable to the selection of individual stocks that perform better or worse

[2] Of course, "other things equal" can only be expected to hold on average over many time periods.

than a priori expectations. This extraordinary performance defines *ex post* APT selection.

APT Equations

The APT follows from two basic postulates:

Postulate 1. In every time period, the difference between the actual (realized) return and the expected return for any asset is equal to the sum, over all risk factors, of the risk exposure (the beta for that risk factor) multiplied by the realization (the actual end-of-period value) for that risk factor, plus an asset-specific (idiosyncratic) error term.

This postulate is expressed by equation (1):

$$r_i(t) - E[r_i(t)] = \beta_{i1} f_1(t) + \ldots + \beta_{iK} f_K(t) + \varepsilon_i(t), \qquad (1)$$

where

$r_i(t)$ = the total return on asset i (capital gains plus dividends) realized at the end of period t,
$E[r_i(t)]$ = the expected return, at the beginning of period t,
β_{ij} = the risk exposure or beta of asset i to risk factor j for $j = 1, \ldots, K$,
$f_j(t)$ = the value of the end-of-period realization for the jth risk factor, $j = 1, \ldots, K$, and
$\varepsilon_i(t)$ = the value of the end-of-period asset-specific (idiosyncratic) shock.

It is assumed that the expectations, at the beginning of the period, for all of the factor realizations and for the asset-specific shock are zero; that is,

$$E[f_1(t)] = \ldots = E[f_K(t)] = E[\varepsilon_i(t)] = 0.$$

It is also assumed that the asset-specific shock is uncorrelated with the factor realizations; that is,

$$\text{cov}[\varepsilon_i(t), f_j(t)] = 0 \text{ for all } j = 1, \ldots, K.$$

Finally, all of the factor realizations and the asset-specific shocks are assumed to be uncorrelated across time:

$$\text{cov}[f_j(t), f_j(t')] = \text{cov}[\varepsilon_i(t), \varepsilon_i(t')] = 0$$
$$\text{for all } j = 1, \ldots, K \text{ and for all } t \neq t'.$$

The above conditions are summarized by saying that asset returns are generated by a *linear factor model.* Note that the risk factors themselves may be correlated (inflation and interest rates, for example), as may the asset-specific shocks for different stocks (as would be the case, for example, if some unusual event influenced all of the firms in a particular industry).

Postulate 2. Pure arbitrage profits are impossible. Because of competition in financial markets, an investor cannot earn a positive expected rate of return on any combination of assets without undertaking some risk and without making some net investment of funds.

Postulate 2 is, in fact, an appealing equilibrium concept that has far-ranging implications for broad areas of financial economics well beyond the determination of asset prices. It is hard to imagine any model of financial behavior that fails to conclude that pure arbitrage profits tend to zero. This generality brings many advantages. The APT is free of restrictive assumptions on preferences or probability distributions, and it provides a rigorous logical foundation for the trade-off between expected returns and risks.

Given Postulates 1 and 2, the main APT theorem is that there exist $K + 1$ numbers $P_0, P_1, \ldots, P_K$, not all zero, such that the expected return on the ith asset is approximately equal to P_0 plus the sum over j of β_{ij} times P_j; that is,

$$E[r_i(t)] \approx P_0 + \beta_{i1}P_1 + \ldots + \beta_{iK}P_K. \tag{2}$$

Although equation (2) holds only approximately, with additional assumptions, it can be proved that it holds exactly (see, e.g., Chen and Ingersoll 1983). More importantly, even without any additional assumptions, it has been proved that the approximation in equation (2) is sufficiently accurate that any error can be ignored in practical applications (see, e.g., Dybvig 1983). Thus the approximation symbol, $\approx$, can be replaced by an equal sign:

$$E[r_i(t)] = P_0 + \beta_{i1}P_1 + \ldots + \beta_{iK}P_K. \tag{3}$$

Here, P_j is the price of risk, or the risk premium for the jth risk factor. Via equation (3), these P_j's determine the risk-return trade-off.[3]

Imagine a portfolio that is perfectly diversified (i.e., one for which $\varepsilon_p(t) = 0$) and with no factor exposures ($\beta_{pj} = 0$ for all $j = 1, \ldots, K$); such a portfolio has zero risk, and from equation (3) its expected return is P_0. Thus, P_0 must be the risk-free rate of return. Reasoning similarly, the risk premium for the jth risk

[3] An equivalent interpretation of equation (3) uses an analogy to the familiar relationship that "quantity × price = value." Thus, if we think of β_{ij} as the quantity of type-j risk in the ith asset and P_j as the price of type-j risk, then the product $\beta_{ij}P_j$ is the value of the contribution of type-j risk to the expected return of the ith asset. If we let V_{ij} denote this value, then it follows from equation (3) that the sum of all the values is equal to the expected excess return (the expected return in excess of the risk-free rate) for the ith asset; that is, $E[r_i(t)] - P_0 = V_{i1} + \ldots + V_{iK}$.

factor, P_j, is the return, in excess of the risk-free rate, earned on an asset that has one unit of risk exposure to the jth risk factor ($\beta_{ij} = 1$) and zero risk exposures to all of the other factors ($\beta_{ih} = 0$ for all $h \neq j$).

The full APT is obtained by substituting equation (3) into equation (1), which after rearranging terms yields:

$$r_i(t) - P_0 = \beta_{i1}[P_1 + f_1(t)] + \ldots + \beta_{iK}[P_K + f_K(t)] + \varepsilon_i(t). \tag{4}$$

It is at this level of the determination of expected returns that the CAPM and the APT differ. In the CAPM, the expected excess return for an asset is equal to that asset's CAPM beta times the expected excess return on a market index, even for multifactor versions of the standard CAPM. For such a multifactor CAPM to be true, the APT risk premiums—the P_j's—must satisfy certain restrictions. In statistical tests, these CAPM restrictions have repeatedly been rejected in favor of the APT.

A portfolio manager controls a portfolio's betas—the portfolio's risk exposure profile—by stock selection. Note that as the risk exposure to a particular factor is, for example, increased, the expected return for that portfolio is also increased (assuming that this risk factor commands a positive risk premium). Thus, risk exposures and hence the implied expected return for a portfolio are determined by a manager's stock selection.

In many applications, data are observed monthly, and the 30-day Treasury bill rate is taken as a proxy for risk-free rate; that is, P_0 in equation (4) is replaced by $TB(t)$, the 30-day Treasury bill rate known to investors at the beginning of month t. Then, for a model with N assets ($i = 1, \ldots, N$) and a sample period of T time periods ($t = 1, \ldots, T$), the data are the asset returns, $r_i(t)$, the Treasury bill rates, $TB(t)$, and the factor realizations, $f_j(t)$. From these data, the statistical estimation problem is to obtain numerical values for the N P_j's and the $(N \times K)\beta_{ij}$'s. Discussion of this econometric problem is beyond the scope of this paper, but the bibliography lists further readings that cover the topic in detail.[4]

Macroeconomic Forces Impacting Stock Returns

Taking the time period to be one month and using the 30-day Treasury bill rate as a proxy for the risk-free rate of return, the APT model, equation (4), becomes:

---

[4] See, for example, Brown and Weinstein (1983); McElroy, Burmeister, and Wall (1985); Chen, Roll, and Ross (1986); Burmeister and McElroy (1988); and McElroy and Burmeister (1988).

$$r_i(t) - TB(t) = \beta_{i1}[P_1 + f_1(t)] + \ldots + \beta_{iK}[P_K + f_K(t)] + \varepsilon_i(t). \qquad (5)$$

From this point, there are three alternative approaches to estimating an APT model:

1. The risk factors $f_1(t)$, $f_2(t)$, . . . , $f_K(t)$ can be computed using statistical techniques such as factor analysis or principal components.
2. K different well-diversified portfolios can substitute for the factors (see Appendix B).
3. Economic theory and knowledge of financial markets can be used to specify K risk factors that can be measured from available macroeconomic and financial data.

Each of these approaches has its merits and is appropriate for certain types of analysis. In particular, the first approach is useful for determining the number of relevant risk factors, or the numerical value of K. Many empirical studies have indicated that $K = 5$ is adequate for explaining stock returns.

The estimates extracted using factor analysis or principal components have an undesirable property, however, that renders them difficult to interpret; this problem arises because, by the nature of the technique, the estimated factors are nonunique linear combinations of more fundamental underlying economic forces. Even when these linear combinations can be given an economic interpretation, they change over time so that, for example, Factor 3 for one sample period is not necessarily the same combination—in fact, it is almost certainly different—as the combination that was Factor 3 in a different sample period.

The second approach can lead to insights, especially if the portfolios represent different strategies that are feasible for an investor to pursue at low cost. For example, if K were equal to 2, one might use small- and large-capitalization portfolios to substitute for the factors.

The advantage of the third approach is that it provides an intuitively appealing set of factors that admit economic interpretation of the risk exposures (the β_{ij}'s) and the risk premiums (the P_j's). From a purely statistical view, this approach also has the advantage of using economic information in addition to stock returns, whereas the first two approaches use "stock returns to explain stock returns." This additional information (about inflation, for example) will, in general, lead to statistical estimates with better properties, but of course, insofar as the economic variables are measured with errors, these advantages are diminished.

Selecting an appropriate set of macroeconomic factors involves almost as much art as it does science, and by now, it is a highly developed art. The practitioner requires factors that are easy to interpret, are robust over time, and explain as much as possible of the variation in stock returns. Extensive research work has established that one set of five factors meeting these criteria is the following:

- $f_1(t)$: *Confidence risk.* Confidence risk is the unanticipated changes in investors' willingness to undertake relatively risky investments. It is measured as the difference between the rate of return on relatively risky corporate bonds and the rate of return on government bonds, both with 20-year maturities, adjusted so that the mean of the difference is zero over a long historical sample period. In any month when the return on corporate bonds exceeds the return on government bonds by more than the long-run average, this measure of confidence risk is positive ($f_1 > 0$). The intuition is that a positive return difference reflects increased investor confidence because the required yield on risky corporate bonds has fallen relative to safe government bonds. Stocks that are positively exposed to this risk ($\beta_{i1} > 0$) then will rise in price. Most equities *do* have a positive exposure to confidence risk, and small stocks generally have greater exposure than large stocks.
- $f_2(t)$: *Time horizon risk.* Time horizon risk is the unanticipated changes in investors' desired time to payouts. It is measured as the difference between the return on 20-year government bonds and 30-day Treasury bills, again adjusted to be mean-zero over a long historical sample period. A positive realization of time horizon risk ($f_2 > 0$) means that the price of long-term bonds has risen relative to the 30-day Treasury bill price. This is a signal that investors require less compensation for holding investments with relatively longer times to payouts. The price of stocks that are positively exposed to time horizon risk ($\beta_{i2} > 0$) will rise to appropriately decrease their yields. Growth stocks benefit more than income stocks when this occurs.
- $f_3(t)$: *Inflation risk.* Inflation risk is a combination of the unexpected components of short- and long-run inflation rates. Expected future inflation rates are computed at the beginning of each period from available information: historical inflation rates, interest rates, and other economic variables that influence inflation. For any month, inflation risk is the unexpected surprise that is computed at the end of the month—the difference between the actual inflation for that month and what had been expected at the beginning of the month. Because most stocks have negative exposures to inflation risk ($\beta_{i3} < 0$), a positive inflation surprise ($f_3 > 0$) causes a negative contribution to return, whereas a negative inflation surprise ($f_3 < 0$), a deflation shock, contributes positively toward return.

Luxury-product industries are most sensitive to inflation risk. Consumer demand for luxury goods plummets when real income is eroded through inflation, thus depressing profits for industries such as retailing, services, eating places, hotels and motels, and toys. In contrast, industries least sensitive to inflation risk tend to sell necessities, the demands for which are relatively insensitive to declines in real income. Examples include foods, cosmetics, tires and rubber goods, and shoes. Also, companies that have large asset holdings such as real estate or oil reserves may benefit from increased inflation.

- $f_4(t)$: *Business cycle risk.* Business cycle risk represents unanticipated changes in the level of real business activity. The expected values of a business activity index are computed both at the beginning and end of the month, using only information available at those times. Then, business cycle risk is calculated as the difference between the end-of-month value and the beginning-of-month value. A positive realization of business cycle risk ($f_4 > 0$) indicates that the expected growth rate of the economy, measured in constant dollars, has increased. Under such circumstances, firms that are more positively exposed to business cycle risk—for example, firms such as retail stores, which do well when business activity increases as the economy recovers from a recession—will outperform those such as utility companies that respond only weakly to increased levels in business activity.
- $f_5(t)$: *Market-timing risk.* Market-timing risk is computed as that part of the S&P 500 total return that is not explained by the first four macroeconomic risks and an intercept term. Many people find it useful to think of the APT as a generalization of the CAPM, and by including this market-timing factor, the CAPM becomes a special case. If the risk exposures to all of the first four macroeconomic factors were exactly zero (if $\beta_{i1} = \ldots = \beta_{i4} = 0$), then market-timing risk would be proportional to the S&P 500 total return. Under those extremely unlikely conditions, a stock's exposure to market-timing risk would be equal to its CAPM beta. Almost all stocks have a positive exposure to market timing risk ($\beta_{i5} > 0$), and hence positive market-timing surprises ($f_5 > 0$) increase returns, and vice versa.[5]

A natural question, then, is whether confidence risk, time horizon risk, inflation risk, and business cycle risk help to explain stock returns better than the S&P 500 alone. This question has been answered using rigorous statistical tests, and the answer is very clearly that they do.[6]

---

[5] Market-timing risk is not required in an APT model that includes *all* the relevant macroeconomic factors. As a practical matter, some relevant macroeconomic factor may be difficult to measure or may not even be observable. Market-timing risk will capture the effects of any such unobserved macroeconomic factor.

[6] The probability that the first four macroeconomic factors do not add any information that is useful for explaining stock returns is less than the probability that a standard normal variable (a

Risk Exposure Profiles and Risks–Return Trade-off

The risk exposure profile for the S&P 500 and the corresponding prices of risk (the risk premiums) are shown in Table 1.[7] For each risk factor, the contribution to expected return is the product of the risk exposure (Column 1) and the corresponding price of risk (Column 2), and the sum of these products is equal to the expected return in excess of the 30-day Treasury bill rate (Column 3). Thus, if the 30-day Treasury bill rate were, say, 5.00 percent, the forecasted return for the S&P 500 would be 5.00 + 8.09 = 13.09 percent a year.

TABLE 1. Calculation of Expected Excess Return for the S&P 500

Risk Factor	Exposure	×	Price of Risk (%/year)	=	Contribution of Risk Factor to Expected Return (%/year)
Confidence risk	0.27		2.59%		0.70%
Time horizon risk	0.56		−0.66		−0.37
Inflation risk	−0.37		−4.32		1.60
Business cycle risk	1.71		1.49		2.55
Market-timing risk	1.00		3.61		3.61
Expected excess return					8.09%

In general, then, for any asset, i, the APT risks–return trade-off defined by equation (3) is:

random variable that is normally distributed with a mean of zero and standard deviation of 1) exceeds 20 in value; that is, it is virtually zero. See McElroy and Burmeister (1988).

[7] The model presented in this section uses parameters estimated by the *BIRR® Risks and Returns Analyzer®* ("BIRR" is an acronym for Burmeister, Ibbotson, Roll, and Ross). The model is re-estimated every month, and the examples here and in the next sections use numbers taken from the April 1992 release, which is based on monthly data through the end of March 1992.

The *Risks and Returns Analyzer®* is a PC-based software package for doing APT-based risk analysis with a model of the sort described in this paper. Although econometric estimation of APT parameters (the risk exposures, β_{ij}'s, and the risk premiums or prices, P's) is beyond the scope of this paper, complete discussions of the more technical statistical issues involved in parameter estimation can be found in Brown and Weinstein (1983); McElroy, Burmeister, and Wall (1985); Chen, Roll, and Ross (1986); Burmeister and McElroy (1988); and McElroy and Burmeister (1988).

$$E(r_i) - TB = \beta_{i1}(2.59) + \beta_{i2}(-0.66)$$

$$+ \beta_{i3}(-4.32) + \beta_{i4}(1.49) + \beta_{i5}(3.61),$$

where TB is the 30-day Treasury bill rate. The following four observations will help clarify this risks–return trade-off:

1. The price of each risk factor determines how much expected return will change because of an increase or decrease in the portfolio's exposure to that type of risk. Suppose, for example, a well-diversified portfolio, p, has a risk exposure profile identical to that of the S&P 500, except that it has an exposure to confidence risk, β_{p1}, of 1.27 instead of 0.27 (= $\beta_{S\&P,1}$). Because the price of confidence risk (from Table 1) is 2.59 percent a year, the reward for undertaking this additional risk is 1.00 times 2.59—that is, the portfolio will have an expected return that is 2.59 percent a year higher than the expected return for the S&P 500.

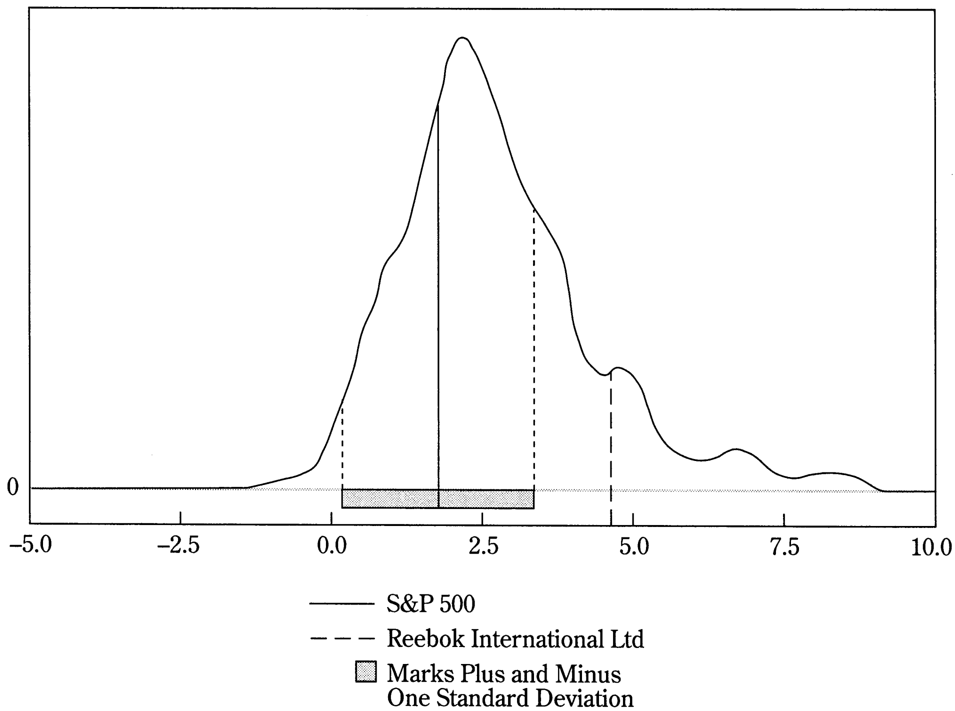
2. APT risk prices can be negative, and they are for both time horizon risk and inflation risk ($P_2 < 0$ and $P_3 < 0$). Consider first inflation risk. Almost all stocks have negative exposures to inflation risk because their returns decrease with unanticipated increases in inflation. Thus, the inflation risk contribution to expected return is usually positive (the negative risk exposure times the negative price for inflation risk equals a positive contribution to expected return). That is, for most i, $\beta_{i3} < 0$, and because $P_3 < 0$, $\beta_{i3} \times P_3 > 0$ for most i.

3. Many stocks have a positive exposure to time horizon risk ($\beta_2 > 0$), however, and thus, when the price of long-term government bonds rises relative to the price of 30-day Treasury bills, their return increases. Because the reward for time horizon risk is negative ($P_2 < 0$), time horizon's contribution to the expected return for such stocks is negative; for stocks with a negative exposure to time horizon risk, its contribution is positive.

Why should this be the case? The answer is that, just as you pay for an insurance policy that pays off when your house burns down, investors desire to hold stocks with returns that increase when the relative price of long-term government bonds rises. The fact that investors want to hold stocks having this characteristic means that the prices of those stocks have been driven higher than they otherwise would have been, and therefore, their expected returns are lower. Thus, the negative price for time horizon risk produces the desired result: stocks with larger (positive) exposures to time horizon risk also have lower expected returns.

4. Table 1 is based on raw values that were not standardized.[8] A good approach for judging whether or not a particular value is significantly different from another is to plot the actual empirical distribution function across stocks and make a visual assessment. Figures 1 and 2, computed from more than 3,200 stocks in the BIRR database, illustrate this empirical distribution function for business cycle risk and the P/E, respectively.

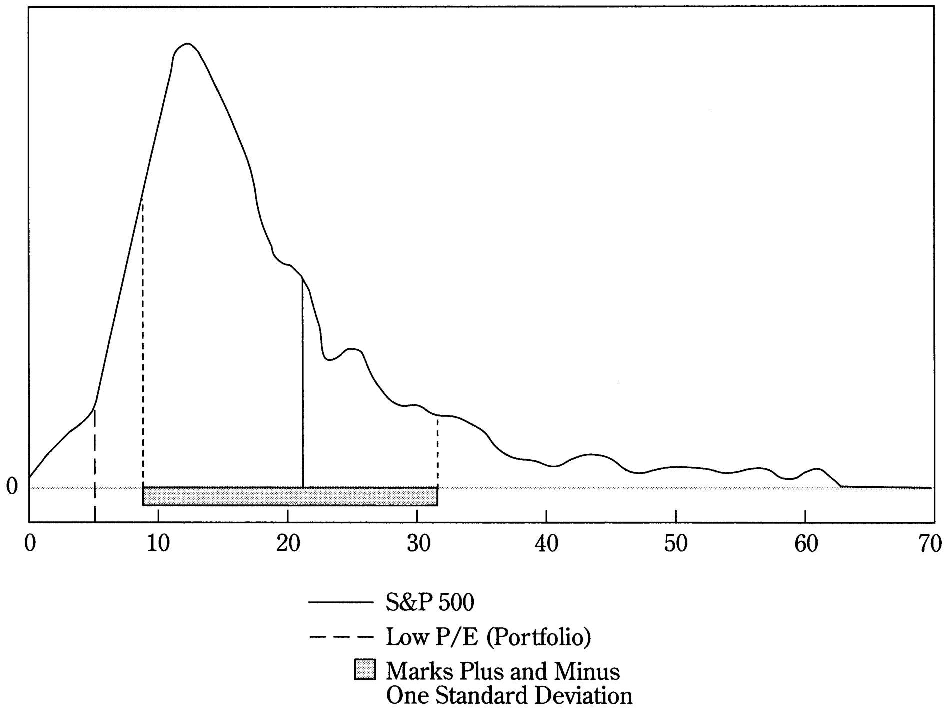
FIGURE 1. Empirical Distribution for Business Cycle Risk (β_{i4}), Reebok and S&P 500, March 1992



Source: BIRR Portfolio Analysis, Inc.

[8] It is not uncommon for the numerical values of financial attributes (such as P/E's) to be reported in units of standard deviation. A standardized value is computed by first transforming the variable so that it has a mean of zero and unit variance; a standardized value of 1.0 (−1.0) means that the value lies one standard deviation above (below) the mean. Provided the attribute is distributed normally, 68.26 percent of the observations lie between the standardized values −1.0 and 1.0, 95.44 percent lie between −2.0 and 2.0, and so forth. Such standardized values can be

FIGURE 2. Empirical Distribution for Price—Earnings Ratios, Low-P/E Stocks and S&P 500, March 1992



Source: BIRR Portfolio Analysis, Inc.
Note: Low-P/E firms are represented by a value-weighted portfolio of the 50 lowest P/E stocks listed on the NYSE.

In Figure 1, the empirical distribution for business cycle risk, β_{i4}, for i = the S&P 500 is indicated by the solid vertical line and the β_{i4} for i = Reebok International Ltd. is indicated by the vertical dashed line. The box on the horizontal axis centered on the S&P 500 line indicates plus and minus one standard deviation of business cycle risk for all the stocks in the database—that is, the width of the box is two standard deviations for the across-firm distribution. Note that the distribution does not appear normal.

In Figure 2, the P/E for the S&P 500 is indicated by the solid line and the P/E for a market-value-weighted portfolio of the 50 lowest P/E stocks listed on the

misleading, however, and even dangerous if the underlying financial attribute is not distributed normally.

New York Stock Exchange is indicated by the vertical dashed line. Again, note that the distribution is not normal and appears to be skewed to the right.

As is evident from Figure 1, the business cycle risk for Reebok is much larger than for the S&P 500. These risk exposure profiles are shown below.

	Exposure for Reebok	*Exposure for S&P 500*
Confidence Risk	0.73	0.27
Time Horizon Risk	0.77	0.56
Inflation Risk	−0.48	−0.37
Business Cycle Risk	4.59	1.71
Market-Timing Risk	1.50	1.00

These exposures give rise to an expected excess rate of return for Reebok equal to 15.71 percent a year, compared with the 8.09 percent a year computed for the S&P 500. Figure 3 compares the risk exposure profiles for Reebok and the S&P 500.[9]

In general, the risk exposure profiles of individual stocks and of portfolios can differ significantly. For example, Figures 4, 5, 6, and 7 compare the respective risk exposure profiles for portfolios of low-capitalization versus high-capitalization stocks, growth stocks versus the S&P 500, a value portfolio versus the BIRR stock database, and a growth versus high-yield portfolio. These risk exposure profiles define APT styles, and they enable us to view traditional portfolio management styles from a new perspective that reveals their inherent macroeconomic risks.

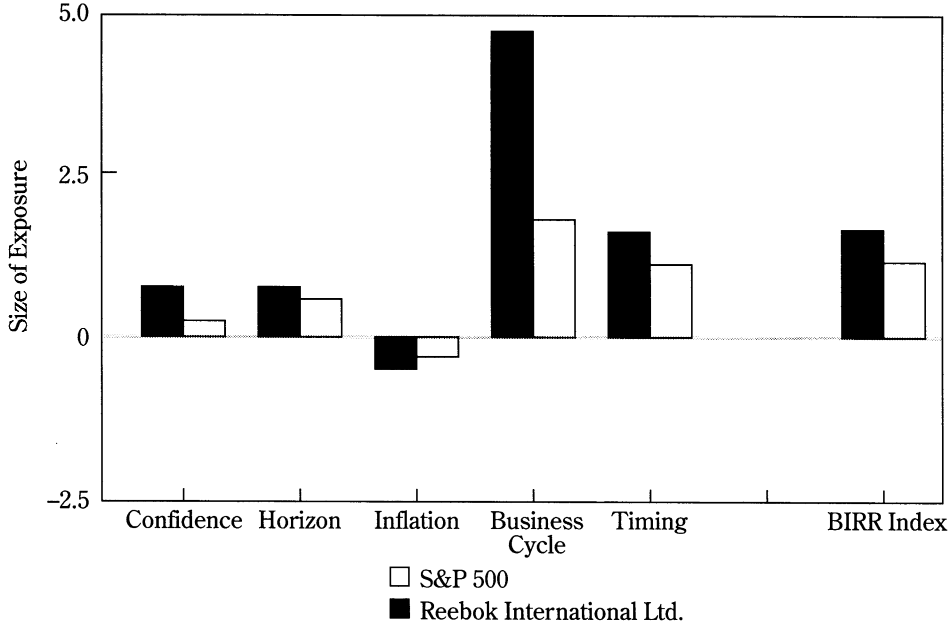
The usefulness to practitioners of risk exposure profiles and the risk–return trade-off is an empirical issue. Abundant evidence shows that market indexes are not mean–variance efficient; if so, the usual implementations of the CAPM using some market index as a proxy are invalid. More importantly, recent empirical evidence demonstrates that CAPM betas do not accurately explain returns.

The multifactor APT approach has far greater explanatory power than the CAPM. Many econometric studies have verified the superior performance of models that include multiple factors (Postulate 1 of the APT) to explain returns and that use multiple factor premiums (Postulate 2 of the APT) to explain expected returns. These results are discussed in some of the papers listed in

---

[9] The BIRR Risk Index plotted in this and the following graphs is a single number that gives an approximate answer to the question, "Does A have more systematic risk, relative to the market, than B?"

FIGURE 3. Risk Exposure Profiles for Reebok and the S&P 500, March 1992



Source: BIRR Portfolio Analysis, Inc.

the bibliography, especially Roll and Ross (1980), Burmeister and McElroy (1988), McElroy and Burmeister (1988), and Fama and French (1992).

Contribution to Return from Macroeconomic Surprises

No matter how precise the model of expected return, surprises always occur, and expected returns differ from actual returns. Taking expectations of equation (5), it follows that the expected return for the *i*th asset in period *t* is given by

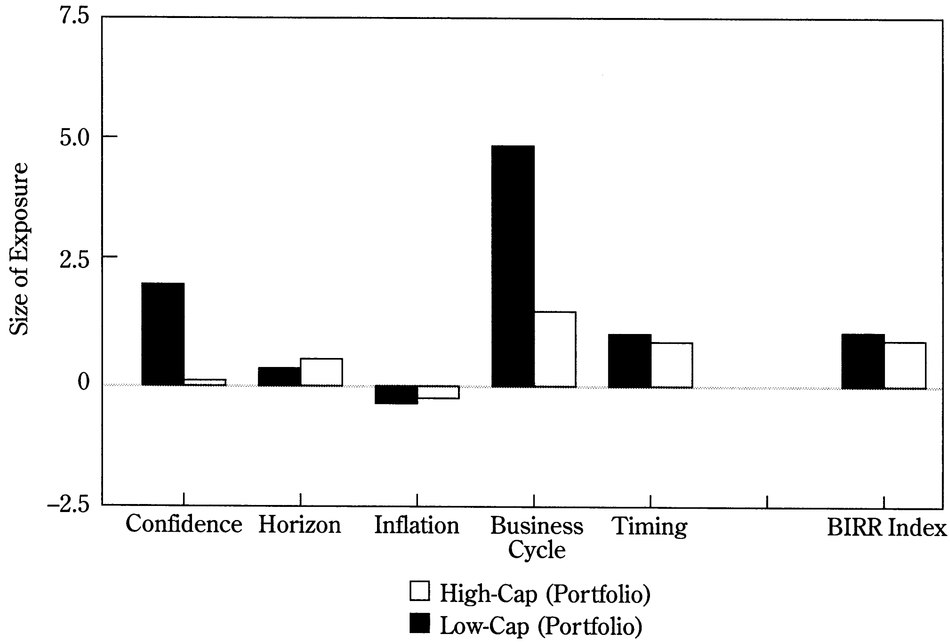
$$E[r_i(t)] = TB(t) + \beta_{i1}P_1 + \ldots + \beta_{iK}P_K. \tag{6}$$

The expected return given by equation (6) is rarely equal to the actual return. Because factors seldom do exactly what is forecast for them and because the idiosyncratic portion of return, $\varepsilon_i(t)$, is almost never zero, the actual return for the *i*th asset is

$$r_i(t) = E[r_i(t)] + U[r_i(t)], \tag{7}$$

where $U[r_i(t)]$ is the unexpected return given by

FIGURE 4. Risk Exposure Profiles for Market-Value-Weighted Portfolios of the 50 Lowest and Highest Capitalization Stocks listed on the NYSE, March 1992



Source: BIRR Portfolio Analysis, Inc.

$$U[r_i(t)] = \beta_{i1}f_1(t) + \ldots + \beta_{iK}f_K(t) + \varepsilon_i(t). \tag{8}$$

Suppose now we consider a historical sample period $t = 1, \ldots, T$ and let bars denote sample period means. The mean *ex post* actual return for the ith asset is

$$\begin{aligned} \bar{r}_i &= \bar{E}(r_i) + \bar{U}(r_i) \\ &= \bar{E}(r_i) + \beta_{i1}\bar{f}_1 + \ldots + \beta_{iK}\bar{f}_K + \bar{\varepsilon}_i. \end{aligned} \tag{9}$$

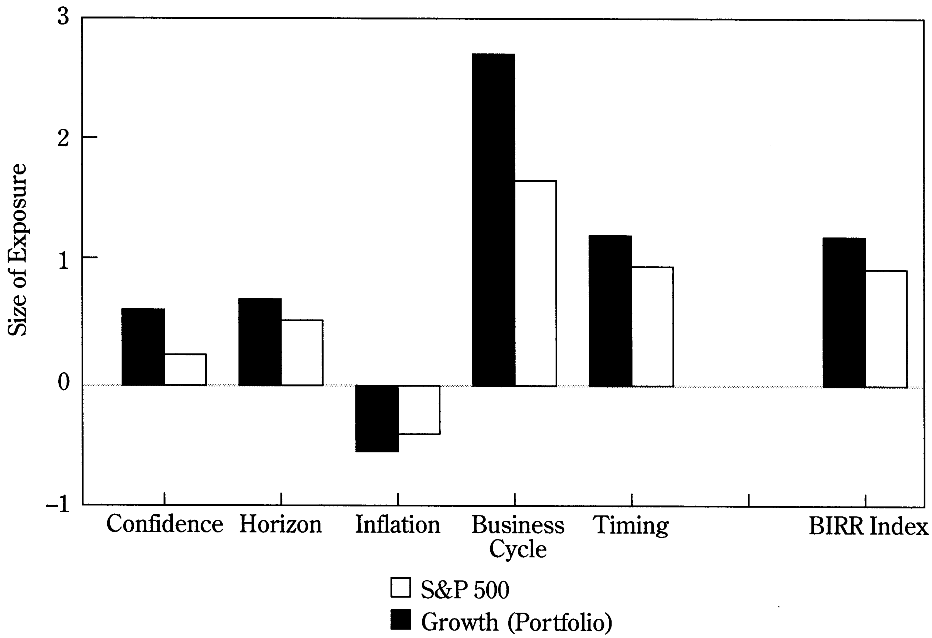
That is, the historical mean return for the ith asset is equal to the sum of the mean *ex post* expected return and the mean of the surprise components of return.

The mean *ex post* unexpected macroeconomic factor return is

$$\beta_{i1}\bar{f}_1 + \ldots + \beta_{iK}\bar{f}_K,$$

and the *ex post* sample period alpha for the ith asset is

FIGURE 5. Risk Exposure Profiles for a Market-Value-Weighted Portfolio of the 50 Highest Growth Stocks Listed on the NYSE and for the S&P 500, March 1992

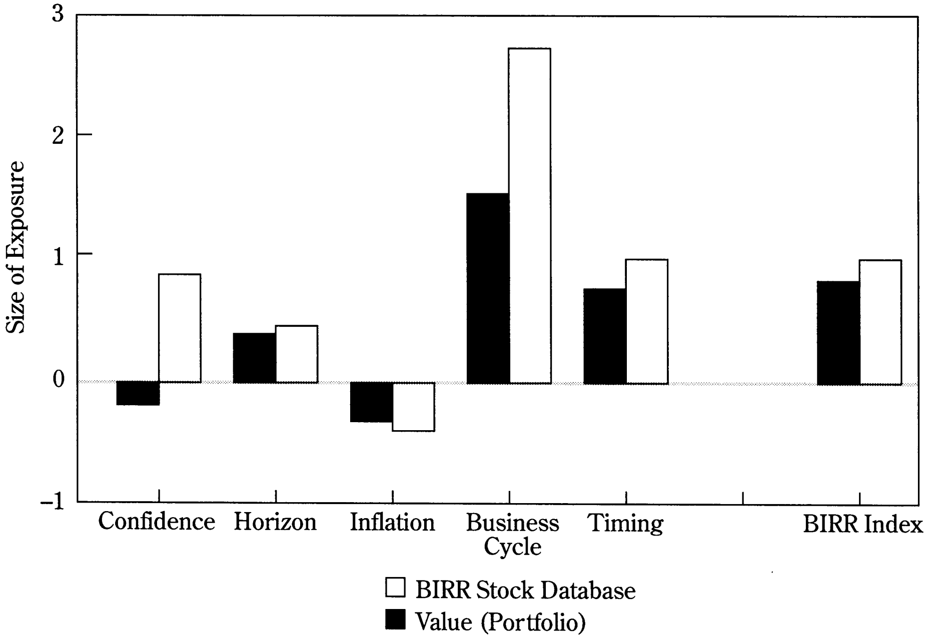


Source: BIRR Portfolio Analysis, Inc.

$$\alpha_i \equiv \bar{\varepsilon}_i.$$

Putting all this together, for the *i*th asset, the mean *ex post* actual return is equal to the mean *ex post* expected return plus the mean *ex post* unexpected macroeconomic factor return plus α_i. The first term on the right side of this equality measures the rewards for risks; it is the reward a manager receives that is attributable to the risk exposure profile for the portfolio. The second term has two possible interpretations: (1) If a manager has taken intentional macroeconomic bets (e.g., a "bet" on an economic expansion through an unusually large exposure to business cycle risk), the unexpected macroeconomic factor return measures the success or failure of those bets; but (2) if a manager is not intentionally making factor bets, the unexpected return can be interpreted simply as a measure of good or bad luck in this sample period. The last term, α_i, is a measure of a manager's selection of individual stocks that perform better or worse than a priori expectations and is the measure of APT selection.

FIGURE 6. Risk Exposure Profiles for a Market-Value-Weighted Portfolio of the 50 Lowest P/E Stocks among the 500 Largest NYSE Firms and for an Equal-Weighted Portfolio of All Stocks in the BIRR Data Base, March 1992

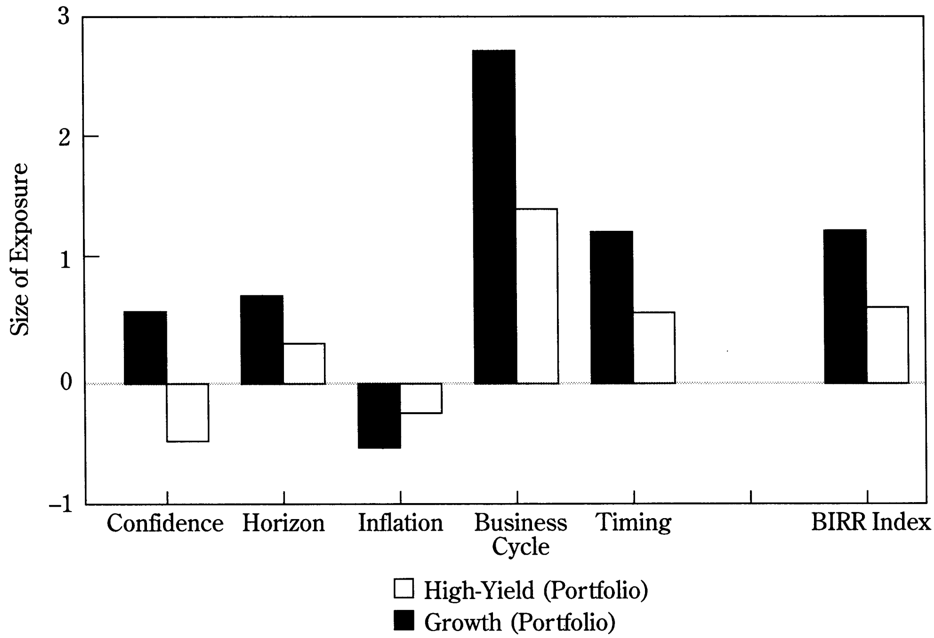


Source: BIRR Portfolio Analysis, Inc.

By construction, all of the macroeconomic factors have zero population means (they have zero-mean probability distributions), so over long historical periods, their sample means will be approximately zero. Thus, over long historical sample periods, the contribution to return from macroeconomic surprises will be approximately zero.[10] Over long time periods, then, almost all of the mean realized return will be rewards for risks and, possibly, APT selection.

[10] This statement is literally true for a portfolio with constant betas. It is possible, however, that "timing" managers can successfully alter betas from period to period so that the average contribution of the factor surprises to portfolio returns is *not* zero. For instance, if managers can predict changes in real business activity (as measured, say, by industrial production) better than the market as a whole, they could structure their portfolios to have high (low) business cycle exposure when they predict an increase (decrease) in business activity.

FIGURE 7. Risk Exposure Profiles for a Growth Portfolio and for a Market-Value-Weighted Portfolio of the 50 Highest Dividend Yield NYSE Stocks, March 1992



Source: BIRR Portfolio Analysis, Inc.
Note: The growth portfolio is as described in Figure 5.

Over short time periods, this will not be the case, even for managers with timing skills; the surprises arising from the macroeconomic factors can have significant impacts on realized returns, as Table 2 shows for Reebok and the S&P 500.

TABLE 2. Annual Mean *Ex Post* Unexpected Macroeconomic Factor Return, Reebok and S&P 500

Sample Period	Reebok	S&P 500
4/91–3/92 (12 months)	−2.03%	−1.58%
10/90–3/92 (18 months)	14.24	9.31
4/90–3/92 (24 months)	−0.95	−0.86
4/87–3/92 (60 months)	−4.01	−2.95
4/86–3/92 (72 months)	−0.26	−0.56

How to Use the APT: Some Examples

A primary concern for practitioners is not only to acquire an understanding of the APT but also to learn how to use it to enhance their investment performance. So far, we have concentrated on explaining the APT; now, we will briefly discuss several uses of the APT that every practitioner could easily apply. The following list is chosen to be exemplary of some widely used APT techniques, but it is by no means exhaustive.

Evaluation of Macroeconomic Risk Exposures and Attribution of Return. Risk exposure profiles can vary widely for stocks and portfolios. They are determined by the risks a manager undertakes through stock selection, and in turn, determine a manager's APT style. A basic first task, then, is to identify the risk exposure profiles for portfolios. Usually, managers will want to compare their risk exposure profiles with those for appropriate benchmarks. A small-cap manager, for example, should know whether his or her portfolio differs in its exposure to macroeconomic risks from an appropriate index of small-cap firms. Any differences will account for performance differentials from the index. Only if the risk exposures are the same as the index can *ex post* superior performance be attributed to APT selection—selection of individual stocks that returned more than would be expected on the basis of the risks undertaken.

Whatever the manager's risk exposure profile, the APT should be used to divide the mean *ex post* actual return into: (1) expected return, which is the reward for the risks taken, (2) unexpected macroeconomic factor return, which arises from factor bets and factor surprises, and (3) α, which arises from stock selection. Moreover, expected and unexpected factor return can be attributed to the manager's risk exposure profile. Thus, APT analysis will provide a better understanding of the true sources of actual portfolio performance.

Index Portfolios. A closely related use of the APT is in the formation of index portfolios designed to track particular well-diversified benchmarks. The APT provides powerful tools for tracking any such benchmark portfolio. A tracking portfolio can be constructed simply by forming a portfolio with a matching risk exposure profile. The *ex post* APT α can be made small by making the tracking portfolio well diversified so that the portfolio-specific return, call it ε_p, is near zero.

Tracking a benchmark that itself is not well diversified, in the sense that its *ex post* α usually is not near zero, is more difficult. In this case, not only the risk exposure profiles but also the benchmark's α must be matched. One way to do this is to form the tracking portfolio by random sampling from the stocks that constitute the benchmark.

Tilting, or Making a Factor Bet. Good managers may possess superior knowledge about the economy. Suppose, for example, a manager believes that the economy is going to recover from a recession faster than most market participants do. If the manager is correct in this belief, the realizations of business cycle risk will be positive ($f_4 > 0$), and stocks that have greater risk exposures to business cycle risk (stocks for which β_{i4} is larger) will, *ceteris paribus*, outperform.

To take advantage of this superior knowledge, the manager will want to make a factor bet on (or tilt toward) business cycle risk—alter the existing portfolio to increase its business cycle risk exposure without changing any other macroeconomic risks. Conversely, if a manager has special knowledge that the economy is going to slide into a recession, he or she will want to lower the portfolio's exposure to business cycle risk.

Multimanager Fund Performance. Most sponsors employ more than one manager. Even though each may perform well when compared with a particular style benchmark, that is not the issue of most importance to a sponsor. A sponsor wants to evaluate the risks and performance of the overall fund.

The sponsor should combine the portfolios of individual managers into one overall fund portfolio and then use the APT to examine the risk exposure profile and performance of the fund portfolio. Often, the combination of managers leads to risk exposures that the sponsor finds uncomfortable. If so, funds should be reallocated among the managers to achieve the desired fund risk exposure profile.

The sponsor must also examine whether or not the overall fund return exceeds the benchmark and determine the sources of differences.

Optimized Risk Control with Manager-Supplied Rankings. Many managers have their own proprietary methods for evaluating stock return performance, yet lack adequate methods for estimating their accompanying risks. The APT, or more accurately, part of the APT, is a perfect tool for such managers.

To keep matters simple, suppose a manager has a personal ranking system that scores every stock on a scale from 1 to 10, where 10 is the score given to the stocks in the best expected return category. The objective is to emulate the volatility of the S&P 500 but achieve a higher return. How could the manager use the APT?

Let s_i be the score from 1 to 10 assigned to the ith stock ($i = 1, \ldots, N$). The formal problem is to find portfolio weights, $w_1, w_2, \ldots, w_N$, for the N stocks in the selection universe such that the portfolio score is maximized but the risk exposure profile is similar to that of the S&P 500. More formally, the weights should result in the highest possible value for

$$w_1 \times s_1 + w_2 \times s_2 + \ldots + w_N \times s_N,$$

subject to the constraint that the portfolio betas,

$$\beta_{pj} = w_1 \times \beta_{1j} + w_2 \times \beta_{2j} + \ldots + w_N \times \beta_{Nj}$$

for $j = 1, \ldots, K$, are close to the betas for the S&P 500. That is, the weights should make the risk exposure profile for the portfolio close to the risk exposure profile for the S&P 500 while maximizing the value of the portfolio's ranking score. If the ranking system works, the return will be superior to the S&P 500. If the resulting portfolio is well-diversified, it and the S&P 500 will have approximately equal volatilities. The proper diversification can be achieved by making N sufficiently large and by imposing a maximum value for the weights so that the portfolio contains a large number of stocks. This optimization problem is easily solved using linear programming.

Long–Short Investment Strategies. Long–short, or market-neutral, investment strategies are receiving increased attention. The pure APT view of such strategies will be discussed first; then, it will be shown how managers with superior knowledge can use the APT to implement those strategies effectively.

Suppose a manager holds a long portfolio with return $r_L(t)$ and a short portfolio with return $r_S(t)$; both have equal dollar values. Let the risk exposures for these portfolios be denoted by β_{Lj} and β_{Sj}, $j = 1, \ldots, K$. Assuming that the short position earns the 30-day Treasury bill rate, the manager's total return is

$$r_L(t) - r_S(t) + TB(t).$$

Now, let the risk exposure profile on the long portfolio exactly match the risk exposure profile on the short position. Then, using equation (3), the *expected* returns on the long and short portfolios are equal, the expected return to the long–short strategy is simply $TB(t)$, and the variance of the realized return is

$$var[\varepsilon_L(t) - \varepsilon_S(t) + TB(t)].$$

Because no stock is held in both the long and short portfolios, this variance is approximately

$$\text{var}[\varepsilon_L(t)] + \text{var}[\varepsilon_S(t)] + \text{var}[TB(t)].$$

The position has greater volatility than 30-day Treasury bills but no greater mean return. Therefore, it is not a very attractive strategy, particularly after trading costs.

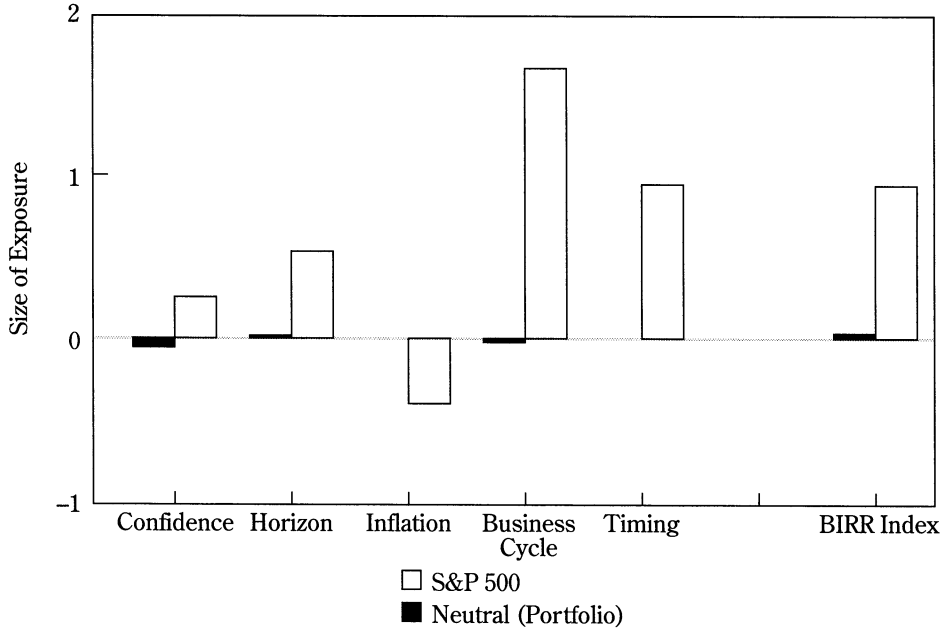
This strategy could become attractive if the APT alphas on the long position were significantly larger than the APT alphas on the short position; that is, it is an attractive strategy for a manager with superior APT selection. Consider an exceptional manager who can pick two well-diversified portfolios of stocks, with no stocks in common, such that $\alpha_L > 0$ for the long portfolio and $\alpha_S < 0$ for the short portfolio. If the manager also can match the risk exposure profiles of the long and short positions, the return would be $\alpha_L - \alpha_S + TB(t)$ with a volatility approximately equal to that of 30-day Treasury bills.

The APT can play a crucial role for such a manager: It provides an easy and quick way to match the risk exposure profiles of the long and short positions. As an example of this role of the APT, we constructed a long portfolio consisting of approximately 50 NYSE-listed stocks with the largest *ex post* alphas over a sample period of 72 months (April 1986 to March 1992). We then computed the risk exposure profile for this long portfolio. A short portfolio of approximately 50 NYSE-listed stocks, not in the long portfolio, was also selected. An optimization problem was solved to find portfolio weights for the short position that matched its risk exposure profile to that of the long position. The resulting risk exposure profile for the overall long–short strategy is illustrated in Figure 8; it has essentially zero systematic risk. The sole source of volatility (beyond the volatility of 30-day Treasury bills) for this long–short strategy comes from the ε's for the long and short positions. By having portfolios of 50 stocks or more, this volatility can be kept small.

The performance of this long–short, or market-neutral, strategy for the most recent 12 months of the sample period (April 1991 to March 1992) is illustrated in Figure 9. The mean realized return was 30.04 percent a year, compared with 11.57 percent for the S&P 500, and the standard deviation of this realized return was only 6.26 percent a year, compared with 18.08 percent for the S&P 500.

Mean–Variance Efficiency. The standard optimization problem of finding the portfolio with the highest expected rate of return for a given variance is easily solved within an APT framework. For this problem, the expected return could either be given by the APT equation, equation (3), or it could come

FIGURE 8. Risk Exposure Profile for the Market-Neutral Strategy and for the S&P 500



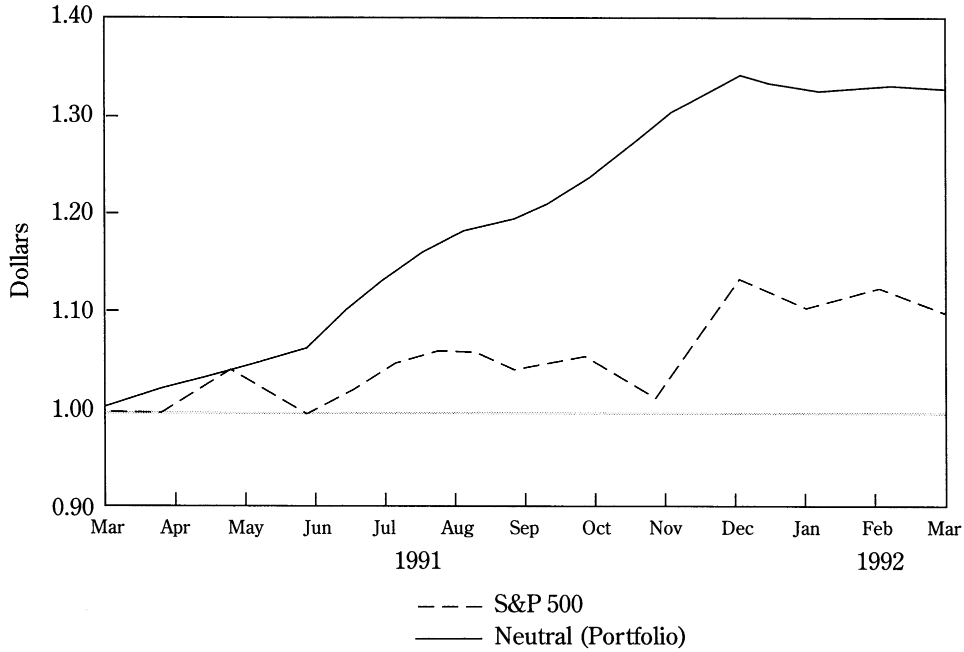
Source: BIRR Portfolio Analysis, Inc.
Note: The risk profile values for the market-neutral strategy are as follows: confidence risk = −0.02, time horizon risk = 0.02, inflation risk = 0.00, business cycle risk = −0.02, and market-timing risk = 0.00; the BIRR risk index = 0.01.

from manager-supplied rankings. In either case, a variety of computational methods can be used to calculate the optimal portfolio weights. In such problems, one often takes the APT's systematic variance (rather than total variance) as given and then imposes constraints to assure that the resulting portfolio is well diversified. This procedure often produces superior results because estimates of stock return variances and, especially, covariances tend to have large out-of-sample errors.

Conclusion

What we have described in this paper is the foundation for the more sophisticated portfolio management techniques that the APT makes possible. Our hope is that a careful reading will enable practitioners to apply the APT to

FIGURE 9. Cumulative Wealth, Market-Neutral Strategy and the S&P 500



Source: BIRR Portfolio Analysis, Inc.

the construction of superior portfolios and will help provide an understanding of the true sources of actual risks and returns. In contrast to other common return measurement approaches, the APT offers fewer explanations for return differences. This simplicity is a great virtue. Understanding the true sources of stock returns is much easier with an APT model than with models having dozens and dozens of parameters that supposedly influence returns.

The basic APT model described here can be enhanced in many ways. Some of the generalizations now in use include the following:

- Allowing the risk prices, P_j's, to vary over time.
- Allowing the risk exposures, the β_{ij}'s, to vary over time.
- Using Bayesian methods to produce optimal out-of-sample forecasts for the risk exposures and hence for the expected returns.
- Introducing additional factors with zero-risk prices, which are typically used to capture industry and sector effects. Although such nonpriced

factors do not contribute to expected return, they do help to explain volatility, and they provide managers with a tool to evaluate the diversification of their portfolios.

Other enhancements are being invented every day. As more and more tools become available and as understanding of the APT spreads, so does its application to portfolio management problems.

Appendix A

To derive the restrictions that a multifactor CAPM must obey, suppose that the CAPM were true for some market index of N assets. This index has a return denoted by $r_m(t)$ and has weights $w_{m1}, w_{m2}, \ldots, w_{mN}$ summing to 1. Suppose also that Postulate 1 of the APT holds, that is, that the N asset returns are generated by the linear factor model (LFM) given in equation (1). We will then show that the APT is valid and find the CAPM restrictions that the APT risk prices must satisfy.

This problem is solved by recognizing that the CAPM beta for any asset can be computed as a linear function of the LFM risk exposures; that is, the CAPM beta is equal to a linear function of the APT β_{ij}'s.

First note that the return on the market index is

$$r_m(t) = w_{m1} \times r_1(t) + \ldots + w_{mN} \times r_N(t)$$

and hence is generated by a LFM with

$$\beta_{mj} = w_{m1} \times \beta_{1j} + \ldots + w_{mN} \times \beta_{Nj} \text{ for } j = 1, \ldots, K.$$

Using Footnote 1, the CAPM beta for the ith asset is

$$\beta_i = \frac{\text{cov}[r_i(t),\ r_m(t)]}{\text{var}[r_m(t)]}.$$

The latter can be computed from the LFM generating the return for the ith asset:

$$\beta_i = \frac{\beta_{i1} \times \text{cov}[f_1(t),\ r_m(t)]}{\text{var}[r_m(t)]} + \ldots$$

$$+ \frac{\beta_{iK} \times \text{cov}[f_K(t),\ r_m(t)]}{\text{var}[r_m(t)]}$$

$$+ \frac{\text{cov}[\varepsilon_i(t),\ r_m(t)]}{\text{var}[r_m(t)]}.$$

Because by Postulate 1 $\text{cov}[\varepsilon_i(t), f_j(t)] = 0$, it follows that $\text{cov}[\varepsilon_i(t), r_m(t)] = \text{cov}[\varepsilon_i(t), \varepsilon_m(t)]$. Thus, under the usual assumption that the market index is well diversified and $\varepsilon_m(t)$ is approximately zero, we may set the last covariance term in the above expression for β_i equal to zero.

Under the CAPM, $E[r_i(t) - TB(t)] = \beta_i \times E[r_m(t) - TB(t)]$. The APT is true when there exist numbers $P_1, \ldots, P_K$ such that

$$E[r_i(t) - TB(t)] = \beta_{i1} \times P_1 + \ldots + \beta_{iK} \times P_K.$$

It then follows immediately that the APT holds provided that

$$\frac{P_j = \text{cov}[f_j(t), r_m(t)] \times E[r_m(t) - TB(t)]}{\text{var}[r_m(t)]}$$

for all $j = 1, \ldots, K$. Conversely, if the APT is true and the above K CAPM restrictions on the P_j's hold, then the CAPM is also true. Given an LFM for asset returns, these are the CAPM restrictions that are rejected in favor of the APT in statistical tests.

Appendix B

We will show that K well-diversified portfolios can substitute for the factors in an APT model. To simplify the computations, we assume that $K = 2$; the general case is easily handled using matrix algebra. Thus, suppose that two different well-diversified portfolios have returns given by

$$R_1(t) = TB(t) + \beta_{11}[P_1 + f_1(t)] + \beta_{12}[P_2 + f_2(t)] + \varepsilon_1(t)$$

and

$$R_2(t) = TB(t) + \beta_{21}[P_1 + f_1(t)] + \beta_{22}[P_2 + f_2(t)] + \varepsilon_2(t).$$

Also assume that the risk exposure profiles for the two portfolios are not proportional. We will show below that

(a) The APT equation for the return on the ith asset, $r_i(t)$, given by equation (5), can be rewritten in terms of the portfolios with returns $R_1(t)$ and $R_2(t)$.

(b) Given the answer to (a), $E[r_i(t)]$ can be expressed in terms of the expected returns for the two portfolios.

To prove (a) and (b), we introduce the following simplifying notation:

$$y_1(t) \equiv R_1(t) - TB(t) - \varepsilon_1(t),$$
$$y_2(t) \equiv R_2(t) - TB(t) - \varepsilon_2(t),$$
$$z_1(t) \equiv [P_1 + f_1(t)], \text{ and}$$
$$z_2(t) \equiv [P_2 + f_2(t)].$$

In this notation, the APT equations for the two portfolios are

$$y_1(t) = \beta_{11}z_1(t) + \beta_{12}z_2(t)$$

and

$$y_2(t) = \beta_{21}z_1(t) + \beta_{22}z_2(t).$$

Taking $y_1(t)$ and $y_2(t)$ as given, the latter are two equations in two unknown z's, and they may be solved for

$$z_1(t) = b_{11}y_1(t) + b_{12}y_2(t)$$

and

$$z_2(t) = b_{21}y_1(t) + b_{22}y_2(t),$$

where

$$\begin{aligned} b_{11} &= \beta_{22}/\delta, \\ b_{12} &= -\beta_{12}/\delta, \\ b_{21} &= -\beta_{21}/\delta, \\ b_{22} &= \beta_{11}/\delta, \end{aligned}$$

and

$$\delta = (\beta_{11}\beta_{22} - \beta_{12}\beta_{21}).$$

Note that as long as the risk exposure profiles for the two portfolios are not proportional, $\delta \neq 0$ and the solution given above exists.

Given these results, with straightforward algebraic manipulation, equation (5) may be rewritten as

$$r_i(t) - TB(t) = c_{i1}[R_1(t) - TB(t)] + c_{i2}[R_2(t) - TB(t)] + e_i(t),$$

where

$$\begin{aligned} c_{i1} &= \beta_{i1}b_{11} + \beta_{i2}b_{21}, \\ c_{i2} &= \beta_{i1}b_{12} + \beta_{i2}b_{22}, \end{aligned}$$

and

$$e_i(t) = \varepsilon_i(t) - c_{i1}\varepsilon_1(t) - c_{i2}\varepsilon_2(t).$$

This exercise establishes (a) above.

Finally, taking expectations of the latter equation gives

$$E[r_i(t) - TB(t)] = c_{i1}E[R_1(t) - TB(t)] + c_{i2}E[R_2(t) - TB(t)] + 0.$$

This formulation proves (b) above.