Prediction of Beta from Investment Fundamentals

Barr Rosenberg and James Guy

PREDICTION CRITERIA

Systematic risk, as measured by beta, captures that aspect of investment risk that cannot be eliminated by diversification. Consequently, it plays the crucial role in evaluating *ex post* the degree of risk undertaken in a diversified investment program, hence in judging the ability of that investment program to achieve a desirable risk–return posture. Again, the prediction of beta essentially predicts the future risk of a diversified portfolio, hence its influence on portfolio beta is one of the key considerations in any investment decision. Therefore, among many possible risk measures beta deserves particular attention and will be the central topic of this article. Beta will be defined, and then, in our discussions of the applications of beta, criteria for optimal prediction and estimation of beta will emerge.

BETA

If the investment return on the market portfolio in any time period assumes any certain value, what return can be expected, on the average, for a security in the same time period? For example, if the market return in that period will be 10 percent, can the security return be expected, on the average, to be 20 percent, or five percent?

Notice that this question refers to the value of the security return to be expected "on the average," although it applies to a single security in a single period. The expectation is to be taken in the following sense. Suppose that, in view of everything we now know about the economy and the specific firm n, we imagine repeating many times the uncertain events that may occur in the time period with each repetition having the nature of an experiment. Each experiment yields some market return r_M and some security return r_n.

Each pair of returns (r_M, r_n) may be graphed, with the security return r_n on the vertical axis and the market return r_M on the horizontal axis. The slope of a regression line fitted through these points, which measures the degree to which higher market return leads to an expectation of greater security return, is the beta of the security (see Figure 1). When the stock market index rises or falls, the security price will tend to rise or fall also, and the rise will tend to be more or less than one. Typically, the slope (i.e., beta) will be greater than zero but less than three. Many securities have betas around one, and they tend to rise and fall in price roughly by the same percentage that the market index rises or falls. A security with a negative beta would tend to move against the market, but such securities are rare.

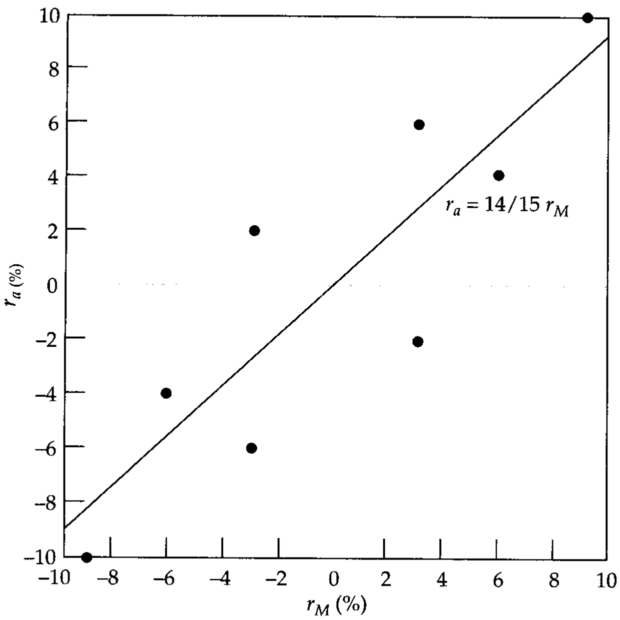
When each repetition is viewed in hindsight, a unique pair of returns (r_M, r_n) will have occurred, but we are concerned with the expectation that held looking forward in time, before the actual returns have occurred. The values actually realized will not ordinarily correspond to expectations: *Ex post* (i.e., hindsight) observation that $r_M = 10$ percent and $r_n = 20$ percent does not imply that the security's beta was two. The true beta could have been one with the additional 10 percent in security return being caused by random factors unique to that security. Beta gives an expected value just as a probabilistic prediction for the profit in a gamble does: *Ex post*, the gamble will have either succeeded or failed, but the result need not be equal to the expected value.

Beta is often explained by plotting a time series of pairs of returns. This corresponds to repeating the above experiment at a sequence of dates. In this way, we are able to observe more than one outcome and, therefore, to illustrate the relationship. Repetition is somewhat misleading, however, since it suggests that beta is unchanging over the sequence. Actually, as we will discuss below, there are reasons to expect that beta changes. The sequence is actually that of repeating similar but changing experiments. The essential meaning of beta applies distinctly at each point in time.

Note that, from an economic viewpoint, the market return does not cause the security return.

Reprinted from Financial Analysts Journal *(May/June 1976):60–72.*

Figure 1. Possible Security Returns Plotted against Corresponding Market Return for the Hypothetical Security "A"



Instead, both are caused by economic events. This point has created some confusion among analysts who interpret beta, which is a regression coefficient, as necessarily stating the causal relationship of market returns upon the security returns: That is, if beta is two, a market return of 10 percent *causes* a security return of 20 percent. The correct wording of this statement is that, as a consequence of the dependence of both market return and security return upon economic events, if a market return of 10 percent is observed, then the most likely value for the associated security return is 20 percent. The words "most likely" include the following pattern of inference: If the market return is 10 percent, then the associated economic events must be of certain types; if for each set of events that could induce a market return of 10 percent we compute the security return that would result, then on average the return, weighted by the probability of the events, is 20 percent.

BETA AS THE CONSEQUENCE OF UNDERLYING ECONOMIC EVENTS

It is instructive to reach a judgment about beta by carrying out an imaginary experiment as follows. One can imagine all the various events in the economy that may occur, and attempt to answer in each case the two questions: (1) What would be the security return as a result of that event? and (2) What would be the market return as a result of that event? Looking forward in time we can see that the market will be significantly affected by changes in the expected rate of inflation, interest rates, institutional regulations of alternative investment media, growth rate of real GNP, and many other factors. Further, there are a number of less broad events that also deserve attention: movements in international oil and other raw material prices, developments in alternative domestic energy supplies, changes in public attitudes toward pollution and consumer durables, and possible changes in tax law, among others. Each of these events is important in contributing to the uncertainty of future market returns. And for each we can anticipate the effect upon any particular security. Consider, for example, a domestic oil stock. "Energy crisis"-related events will have a proportionally greater effect upon such a stock, inflation-related events probably a relatively smaller effect, than for the market as a whole. As a result, if we foresee that the major source of uncertainty in future returns is from developments in the energy picture, we will anticipate an unusually high beta, but if we foresee that the major source of uncertainty lies in inflation-related events, we will anticipate an unusually low beta.

One could easily devote as much time to predicting beta as is usually devoted to predicting security returns in conventional security analysis. This parallel is, in fact, a valuable one to draw on in thinking about beta. In security analysis, it is customary to distinguish between the component of return resulting from events specific to the firm in question, and the component of return stemming from events affecting the economy or the market as a whole. When the sum of these two is expected to be positive, then the security is considered to be a good buy. Now, in enumerating the events specific to the firm in question, the analyst will formulate a prediction of the expected impact on return and also a forecast of the uncertainty of realizing that expectation. The former determines the expected specific return and the latter the magnitude of specific risk. Thus the tasks of predicting expected return and risk of return are clearly related in this case.

Similarly, in predicting the component of security return arising from economywide events rather than from events specific to that particular firm, the analyst estimates the probabilities of the various possible outcomes of the event, and the magnitude of the response of the security return to that event. The product of these two is the expected effect of the event upon the security return. These effects are then summed over all economywide events that may impact the stock to obtain the expected security return due to economywide factors. Here again, all that is needed is a judgment as to the uncertainty attaching to the economywide events, and we find a prediction of the uncertainty of the security return due to economic events. The return on the market portfolio is the weighted average of the individual security returns, so this same approach yields a prediction of the uncertainty of the market return due to economywide events. Since the events specific to individual firms will tend to average out and contribute little to the market return, the economywide events will account for the great bulk of market risk.

Thus the risk of market return is largely accounted for by economic events that impact many stocks. For each stock, we find that these events also have an effect that can be predicted by security analysis. As an illustration, consider Table 1, where we give two imaginary future events with equal probability of good, bad, and no-change outcomes, and describe the resulting percentage returns on the market, stock A and stock B. Relative to the market, stock A responds two-thirds as much to the energy event and two times as much to the inflation event. Relative to the market, stock B responds four-thirds as much to energy and responds nil to inflation. (These are later referred to as relative response coefficients.)

Table 1.

		% Contribution to Return		
Event	Outcome	Market	Stock A	Stock B
Energy	Good	+6	+4	+8
	No change	0	0	0
	Bad	–6	–4	–8
Inflation	Good	+3	+6	0
	No change	0	0	0
	Bad	–3	–6	0

Because the effects of the two events are independent, the information given in this table can be represented by the tree diagram given in Figure 2. Using this diagram, it is easy to derive the expected value and variance of returns on the market, r_M, as a result of the two events:

$$E(r_M) = 0$$

$$\mathrm{VAR}(r_M) = 1/9\ (9^2 + 6^2 + 3^2 + 3^2 + 3^2 + 3^2 + 6^2 + 9^2)$$
$$= 30.$$

This variance of future market returns can be decomposed into the variances induced by the two independent events. The variance in market returns caused by energy uncertainty alone is equal to $1/3[6^2 + 0^2 + (-6)^2] = 24$, while that caused by inflation uncertainty alone is equal to $1/3[3^2 + 0^2 + (-3)^2] = 6$. Because these two events are independent, the sum of these two subvariances should equal the total variance of market returns, and indeed we have

$$24 + 6 = 30.$$

The variance of the future market return stems from uncertainty in energy and inflation. Energy is the greater source of future variance (actually four-fifths of the total in this example). Stock B is more responsive to the energy factor than the market, and it will show a high volatility if the energy situation changes. Stock A will show the higher volatility if the inflation situation changes, since its response coefficient to inflation is higher. Since energy is the greater source of uncertainty, it turns out that stock B has the higher beta.

The betas of companies A and B can be easily calculated using this tree diagram. Consider, for example, the beta of company A, which is defined as[1]

$$\beta_a \equiv \frac{\mathrm{COV}(r_a, r_M)}{\mathrm{VAR}(r_M)} = \frac{E[(r_a - E[r_a])(r_M - E[r_M])]}{E[(r_M - E[r_M])^2]}.$$

We know that $\mathrm{VAR}(r_M) = 30$, so that all that remains is to calculate $\mathrm{COV}(r_a, r_M)$. Remembering that $E(r_a) = 0$, we have

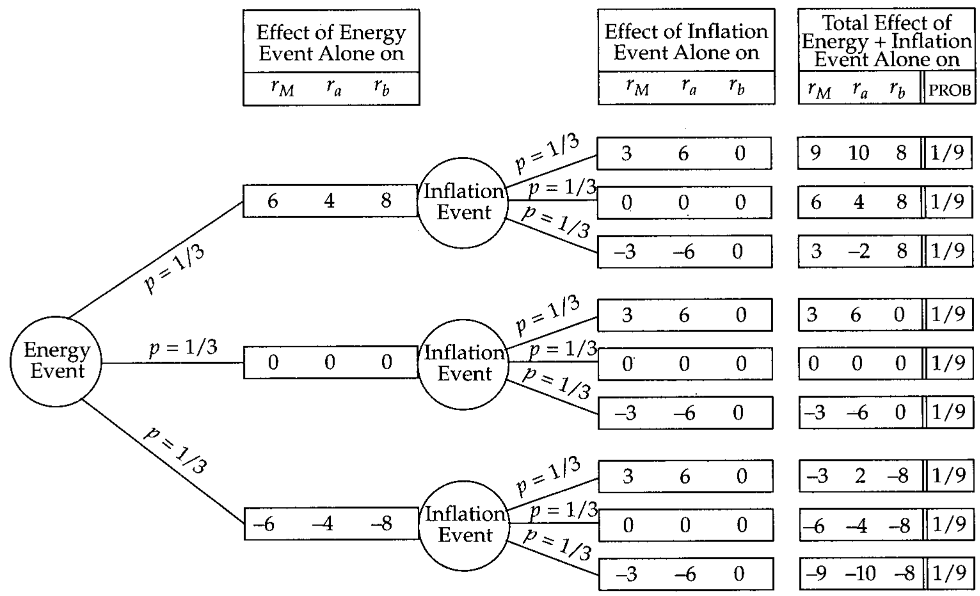
$$\mathrm{COV}(r_a, r_M) = 1/9[9.10 + 6.4 + 3(-2) + 3.6 + 0.0$$
$$+ (-3)(-6) + (-3).2$$
$$+ (-6)(-4) + (-9)(-10)]$$

= 28, substituting this result in the formula for β_a, we have,

$$\beta_a = 28/30 = 14/15.$$

This beta for company A can be decomposed into the component betas due to the two events. Let us define r_M^e, r_a^e, and r_b^e as the returns on the market, stock A and stock B due to the energy event alone,

Figure 2.



and r^i_M, r^i_a, and r^i_b as the corresponding returns due to the inflation event alone. Then,[2]

$$\mathrm{COV}(r_a, r_M) = \mathrm{COV}(r^e_a, r^e_M) + \mathrm{COV}(r^i_a, r^i_M)$$

$$= 1/3[4.6 + 0.0 + (-4)(-6)]$$

$$+ 1/3[6.3 + 0.0 + (-6)(-3)]$$

$$= 2/3\{1/3[6.6 + 0.0 + (-6)(-6)]\}$$

$$+ 2\{1/3[3.3 + 0.0 + (-3)(-3)]\}$$

$$= 2/3\ \mathrm{VAR}(r^e_M) + 2\ \mathrm{VAR}(r^i_M)$$

$$\beta_a = 2/3\,\frac{\mathrm{VAR}(r^e_M)}{\mathrm{VAR}(r_M)} + 2\,\frac{\mathrm{VAR}(r^i_M)}{\mathrm{VAR}(r_M)}.$$

Substituting in the values for $\mathrm{VAR}(r^e_M)$, $\mathrm{VAR}(r^i_M)$, and $\mathrm{VAR}(r_M)$, we obtain

$$\beta_a = 2/3 \cdot 24/30 + 2 \cdot 6/30 = 2/3 \cdot 4/5$$

$$+ 2 \cdot 1/5 = 14/15.$$

The first component of β_a reflects the behavior of the security relative to energy, and the second considers the effect of inflation. As indicated in the derivation, 4/5 and 1/5 are the proportional contributions of the energy and inflation events to market variance, and 2/3 and 2 measure the relative response coefficients of stock A to these events.

Similarly, it is possible to show that, for security B we have

$$\beta_b = 4/5 \cdot 4/3 + 1/5 \cdot 0 = 16/15.$$

The foregoing discussion illustrates the proposition that the level of beta is determined by two kinds of parameters: (1) The degree of uncertainty attached to various categories of economic events (the proportional contributions of the events to market variance), and (2) the response of the security returns to these events (relative response coefficients).

In general, if we assume, for expository purposes, that economic events are independent of each other, then the beta of the security n will be

$$\beta_n = \frac{\sum_{j=1}^{J} \gamma_{jn} V_j}{\sum_{j=1}^{J} V_j},$$

where V_j is the contribution of economywide event j to market variance in any period, and where γ_{jn} is the ratio of the responses of the nth security and the market to the jth event or the "relative response coefficient."[3] This expression can be rewritten as

$$\beta_n = \sum_{j=1}^{J} \left(\frac{V_j}{\sum_{j=1}^{J} V_j} \right) \gamma_{jn},$$

which clearly shows that the beta for any one security is the weighted average of its relative response coefficients, each weighted by the proportion of total variance in market return due to the event.

This insight into the fundamental determinants of beta will be exploited at many points in this article. For the moment it provides a grasp on the behavior of a security's beta over time. Is beta likely to be constant over time, to drift randomly, or to change in some predictable or understandable way? The answer is that beta will change when either the relative response coefficients or the relative variances of economic events change. To the degree that these changes can be predicted or explained, changes in beta can be predicted or explained. For example, the monthly dates on which the Bureau of Labor Statistics announces inflation rates will be dates upon which the inflation-oriented events will explain a larger proportion of market variance, and will therefore be dates when firms with high relative response to inflation will have higher than usual betas. For another example, if a firm changes its capital structure, thereby increasing its leverage, its relative response coefficient to virtually all economic events will increase, and so as a result will its beta.

Because beta need not be constant over time, it follows that estimating the average value of beta for a security in some past period is not the same problem as predicting the value of beta in some future period. This is the first distinction between historical estimation and future prediction. A second equally important distinction arises from the use of beta.

USES OF BETA

It is important to examine the uses of beta, not only as an aid in understanding it, but also because the criteria for prediction and estimation probably arise from the requirements of usage. In other words, in each application, that estimator or predictor should be used that will function best in that application. If different applications impose different requirements, then different estimators should be used. Recall that we never observe the "true" beta but rather outcomes that are randomly distributed about an expected value that is equal to beta. As a consequence, we must estimate from the observed outcomes the underlying value of beta that generated them. Similarly, we must predict from this same data the value of beta to be expected in the future, as distinct from the true value of beta in the past.

Performance Evaluation

The most widely recognized use of beta, at this writing, is in the evaluation of past investment performance. For reasons repeatedly discussed in the literature, this use of beta is strongly suggested by the theory of capital markets; the wisdom of this course has been confirmed by the extraordinary increase in the clarity with which investment performance is now being assessed and perceived.

For this purpose, the portfolio as a whole is the appropriate entity: One is interested in the degree of portfolio risk (the beta of the portfolio). There is only a derivative interest in the risks of the individual securities, to the degree that knowledge of these can be helpful in assessing risk for the overall portfolio.

Investment Strategy

We now turn to the use of beta in the selection of an investment policy, that is, to decision making as opposed to *ex post* evaluation.

Because the value of beta measures the expected response to market returns and because the vast majority of returns in diversified portfolios can be explained by their response to the market, an accurate prediction of beta is the most important single element in predicting the future behavior of a portfolio. To the degree that one believes that one can forecast the future direction of market movement, a forecast of beta, by predicting the degree of response to that movement, provides a prediction of the resultant portfolio return. To the degree that one is uncertain about the future movement of the market, the forecast of beta, by determining one's exposure to that uncertainty, provides a prediction of portfolio risk. For a less well diversified portfolio, the residual returns associated with the component investments assume greater proportional importance, but the influence of the overall market factor remains important even in a portfolio containing only one security.

Thus there is little doubt that, if one could make an accurate prediction of future beta for the portfolio, it would be an important ingredient in his investment decision making. And equally, if he could make accurate predictions of the betas for individual securities, these would be important ingredients of his portfolio revision decisions. For instance, if the manager decides to increase the portfolio beta, then he will seek to exchange current holdings with low beta for new purchases with high beta, and the success of this exchange will depend on his ability to forecast the difference in beta.[4]

In this same context it must also be noted that the decision to revise the portfolio cannot be separated from an implicit time horizon. If the asset is to be held for a four-year period, perhaps the average duration in large portfolios, then the appropriate horizon for the forecast of beta will be four years. However, if the asset is purchased with a view to exploiting an anticipated market movement in the short term, say the next half year, then the beta forecast should be made with a horizon of six months.

Thus far, two kinds of uses of beta in the decision-making aspects of portfolio management have been delineated: (a) By forecasting the response to market movement, it allows a forecast of security return when a forecast of market movement is made; and (b) to the degree that the market movement is uncertain, beta, in determining the response, determines the expected uncertainty of security or portfolio return. To develop criteria for predictors of beta, it is convenient to refer to a typical investment decision strategy (in the spirit of Treynor and Black) that relies, in part, on beta. This will be referred to as a "typical control strategy."[5] We assume that the strategy includes a target for the portfolio beta, which changes over time in response to (a) changing forecasts of the direction of market movement, or (b) changing assessments of the permissible level of systematic risk to be assumed. Transactions are motivated in part by considerations of security analysis, in the sense that securities regarded as overvalued are sold and securities regarded as undervalued are purchased. Transactions are also influenced by a desire to maintain an appropriate level of diversification. Also, each time that the beta target is changed, a set of transactions is undertaken with the intention of reaching the new target. To reach the new target with a minimum of transactions (hence a minimum of transaction costs), there is a preference for the purchase of securities with values of beta that differ from the existing portfolio in the direction of the new target, and for the sale of securities that differ in the opposite direction. Thus transactions are undertaken with the multiple goals of (1) reaching an appropriate portfolio beta with a minimal number of transactions; (2) increasing expected return; and (3) retaining an appropriate degree of portfolio diversification.

During periods when the target beta for the portfolio is not changing, there will be transactions motivated by the desire to increase expected return and to control diversification. Beta will remain an important consideration in these transactions, because the need to keep the portfolio beta near the target will serve as an indirect constraint on purchases and sales. Transactions involving stocks with betas differing from the target will require offsetting adjustments in other transactions. And, recalling that the beta of the portfolio, just as the beta of a security, may change over time, transactions may sometimes be required simply to adjust for an undesirable drift in the portfolio beta.

Thus a typical control strategy will involve a constraint on the portfolio beta that induces a preference for the purchase (or sale) of stocks with particular kinds of individual betas; in other words, the beta of each individual stock assumes importance as a means to achieve a portfolio target value. The portfolio beta being the average value of the individual betas, weighted by investment proportions, the importance of the individual betas will be determined by the investment proportions. Since the typical portfolio will by definition involve investments in securities that are proportional to their market capitalizations, it follows that the typical weight of an individual beta, as an ingredient in the control strategy, will be in proportion to the capitalization of the firm.

Valuation

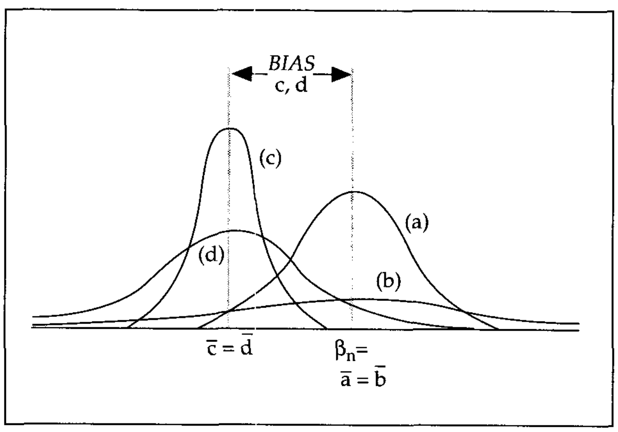
Finally, a third class of uses applies to the valuation of convertible assets. Consider any asset, such as convertible bonds, convertible preferred stock, warrants and options, that provides the opportunity to exercise a conversion into the underlying security. An important determinant of the value of any such asset is the total risk of the underlying security, for the simple reason that such assets provide one-sided claims on the underlying security. The higher the underlying risk, the more likely that the security price will change significantly. Since one profits (loses) if the security price goes in one direction and is unaffected if the security goes in the other, the greater the expected risk, the greater the expected profit (loss). Knowledge of the value of beta permits prediction of one important element of risk. Notice that this use of beta arises because its usefulness as a measure of risk of the underlying common implies an estimate of the value of the convertible asset.

CRITERIA FOR PREDICTION

For each use of beta described above, one should ask what properties an appropriate measure of beta should have. It is beyond the scope of this article to discuss criteria for the estimator of beta to be used in historical performance evaluation. We may note in passing that the appropriate measure relates to an average level of risk assumed in the portfolio during the evaluation period, so that it is an estimator of a past risk level. The problem of choosing among alternative estimators of the average value in the past provides a good vehicle for introducing the concepts of bias, variance, and mean square error as employed in the context of estimation problems.

How are we to choose among several alternative estimates of the average value of the portfolio beta over the historical period? (Recall that beta is no more than an underlying tendency and that the actual results observed *ex post* do not tell us what the exact underlying tendency was.) The distributions of estimated values for four imaginary estimators are plotted in Figure 3.

Figure 3.



Suppose that the true average for a portfolio or security beta was β_n, and that $\hat{\beta}_n$ is an estimator of this and has an expected value $\bar{\hat{\beta}}_n$. The quality of this estimator can be judged by three criteria: bias, variance, and mean square error.[6] If the estimator is unbiased, its expected value equals the true underlying average, and the bias, $\bar{\hat{\beta}}_n - \beta_n$, is zero. Estimators (a) and (b) are unbiased in Figure 3. Freedom from bias is obviously desirable.

Of a group of unbiased estimators, the most desirable is the one that is the most accurate. Accuracy may be defined by the smallness of the variance of estimation error. Thus the best unbiased estimator is the unbiased estimator with the smallest variance, i.e., minimum $E[\hat{\beta}_n - \bar{\hat{\beta}}_n]^2$. In Figure 3, a is the most desirable unbiased estimator. A criterion for comparing biased and unbiased estimators when it is not important whether the error in $\hat{\beta}_n$ is derived from the bias or estimation error is the mean square error, MSE. Whereas the variance of the estimator is the expected squared deviation of the estimated beta from its mean, the mean square error is the mean of the squared deviation of the estimated beta from the true value, i.e., $E[\hat{\beta}_n - \beta_n]^2$. Of course, when the estimator is unbiased these two measures are equivalent. For any estimator $\hat{\beta}_n$, the formal relationship between bias, $\text{BIAS}(\hat{\beta}_n)$, variance, $\text{VAR}(\hat{\beta}_n)$, and mean square error, $\text{MSE}(\hat{\beta}_n)$, is given by[7]

$$\text{MSE}(\hat{\beta}_n) = \text{VAR}(\hat{\beta}_n) + [\text{BIAS}(\hat{\beta}_n)]^2.$$

As can be seen, by minimizing the MSE of the estimate, we are in fact minimizing the sum of the variance and the squared bias of that estimator. As such, minimizing the MSE imposes an arbitrary judgment as to the relative importance of the bias and variance. If it is thought critical to have an unbiased estimator, then minimizing the MSE would not automatically provide one. It is quite possible that a biased estimator with low variance would be chosen in preference to an unbiased estimator with high variance. This point is amplified graphically in Figure 3. Estimates (c) and (d) are both biased to the same extent, but (c) is superior to (d) because it has a lower variance. Can (c) be superior to either (a) or (b), even though (c) is biased and (a) and (b) are not? Using the MSE criterion, it is quite possible that (c) is superior to (b) as long as

$$\text{VAR(c)} + [\text{BIAS(c)}]^2 < \text{VAR(b)}.$$

Let us now turn to the main topic of this article, namely, the prediction of beta and the criteria for good prediction. Consider the case where the criteria are concerned with the management of a portfolio of stocks and other nonconvertible assets, as distinct from convertible assets. Clearly, the first requirement is a prediction of the beta of the existing portfolio. This will provide an indication of the portfolio's response to anticipated market movements as well as a prediction of the portfolio's exposure to market risk. Naturally, the prediction should relate to the planning horizon.

That is, we are concerned with an estimate of beta for the future period for which plans are being made.

The portfolio beta in the future is the weighted average of the individual security betas, each weighted by the proportionate investment in that security,[8]

$$\beta_P = \sum_n W_{Pn}\beta_n,$$

where W_{Pn} is the proportion of the total investment now in stock n, with $\sum_n W_{Pn} = 1$. The predicted portfolio beta is[9]

$$\hat{\beta}_P = \sum_n W_{Pn}\hat{\beta}_n.$$

The prediction error will therefore be $\sum_n W_{Pn}(\hat{\beta}_n - \beta_n)$. Thus the prediction error for the portfolio beta is the weighted average of the prediction errors for the individual securities, each weighted in proportion to the value of the investment in that security. In order for the prediction error to be small, it is necessary that the prediction errors for the individual stocks be small and average out to zero. If the estimation errors are independent and are expected to equal zero (which will be the case if the estimators are unbiased) then the estimation error will tend to average out to zero.

The quality of the forecast beta for any one stock can be judged using the same criteria as was suggested in the evaluation of estimates of the historical average beta. If the true future beta is β_n, and the forecast beta is $\hat{\beta}_n$ and has an expected value of $\bar{\hat{\beta}}_n$, then the forecast is unbiased if $\bar{\hat{\beta}}_n - \beta_n = 0$. From a group of such unbiased forecasts, the optimal estimate is that with the minimum forecast variance. If, on the other hand, we are considering biased and unbiased forecasts of beta we should choose that one with the minimum mean square forecast error, MSE. Notice that it is the true future value of β_n, not the present value, that is to be predicted.

If we were concerned with estimating the beta for a single stock n, β_n, the preceding considerations would suffice. But since we are estimating beta for a number of securities, $n = 1, \ldots, N$, we must consider criteria for a collection of estimates $\hat{\beta}_n, \ldots n = 1, \ldots N$ such that the collection will perform optimally in use. Suppose that a prediction rule is defined that produces, for each n, a prediction $\hat{\beta}_n$. Then a criterion for this prediction rule might take the form of a condition applying to a weighted average of the properties of the estimator for the individual securities.

Consider, for example, the question of *unbiasedness*. The strongest requirement of unbiasedness would be that the expected value of the estimator for each and every individual security should equal the value of beta for that security. A weaker requirement would be that the average estimated beta for each industry should equal the true average beta for that industry. Comparing the requirement with the previous one, the difference here is that some estimators within the industry could be upward biased and others downward biased as long as the average bias were zero. A still weaker statement would be that the average predicted beta for all stocks should equal the true average value. This last statement is equivalent to asserting that the expected value for a predicted beta of a stock selected at random from the stock exchange should equal the expected true value for a security selected at random. This condition requires only that the average bias, averaged over all securities, is zero.

Each of these prediction criteria involve an average over many securities. Over what group of securities should this average be taken? How should the securities be weighted? These two questions can be collapsed into a single question of weighting within the universe of securities, because those securities not included in the group over which the average is taken would automatically have a weight of zero.

The answer to the weighting problem follows directly from the criterion that the errors in the predicted betas should average out when weighted by the future proportionate investments in the portfolio. What is desired is unbiasedness, when weighted by the future investment proportions.[10] Thus, ideally, a slightly different set of weights must be used to evaluate unbiasedness for each future investment portfolio. In practice, it is simpler and probably sufficient to achieve unbiasedness relative to the average investment weights to be expected for the user of the prediction rule. Since the sum of the investment weights, summed across all potential institutional users of the prediction rule, approximates the aggregate market values, a natural criterion is to define unbiasedness relative to a capitalization-weighted average.

Having settled the question of weighting, the next issue is that of the strictness of the unbiasedness condition: Must the prediction be unbiased for every security, for groups of securities such as industries, or only for the entire sample? The answer is again that the average expected prediction error for the group of securities in any portfolio should be zero. If all portfolios were identical to the market portfolio, then the absence of bias for the capitalization-weighted market would suffice. But in fact individual portfolios differ. Some emphasize one industry group, some emphasize another. Some concentrate on stocks with a particular fundamental characteristic, some on stocks with a particular technical characteristic. It follows that, if the average expected prediction error is to be zero for all portfolios, it is desirable that the predictor be unbiased for each industry group and for each fundamental or technical characteristic that may serve as a basis for portfolio selection.

The question of the appropriate criterion for *accuracy* of the estimators may be approached in a similar fashion. From the point of view of predicting portfolio risk, it is the size of the error in predicting the portfolio beta that is important, as distinct from the betas of individual stocks in the portfolio. Moreover, it is the error itself that matters, not the source from which it derives. Thus it is immaterial whether an error results from bias or from variance in the estimator. It follows that the appropriate criterion for accuracy in the prediction of portfolio risk is a minimum mean square error predictor. We are not only concerned with predictions of beta for the prediction of portfolio risk, but also for making decisions with regard to possible portfolio revisions. The respective criteria for prediction of individual security betas and of the present risk of the portfolio must be such as to yield a good control of risk for the eventual portfolio constructed using these predictions. Thus the form of these criteria must be derived from the decision procedure. If, for example, the manager follows a typical control strategy with a desired portfolio beta of 1.3, then a good beta predictor is one such that by relying on the predictor he will indeed tend to achieve a portfolio beta of 1.3. Because the portfolio revision decision entails the sale of specific securities within the portfolio and the purchase of others, it becomes necessary to predict the betas of individual securities—highlighting another essential distinction between future prediction and historical evaluation: In prediction, the risk levels of individual securities assume primary importance. Again, any error in the prediction of risk for the existing portfolio, regardless of its source or nature, will be equally serious as long as we accept the predicted value as the basis for subsequent portfolio revision.

However, in modifying the portfolio, we will consider alternative combinations of sales and purchases, following the "typical control strategy" outlined previously. Our decision will depend in some form on the predictions of the betas for the individual securities. Presumably, we will select a group of sales and purchases that move in the direction of the desired beta, while also achieving an increase in expected return. It is likely that certain "characteristics" of the stocks will influence the choice. Thus we consider currently "popular" stocks for purchase, and currently "unpopular" stocks for sale. Or we consider currently high P/E stocks for purchase, and currently low P/E stocks for sale. Any one of an infinite number of decision rules may be used in which the major ingredient is a forecast of excess return on the individual security. But if this forecast of excess return shows any dependence at all across different stocks, it is probable that the dependence will take the form of a belief on the part of the manager that stocks with more of some characteristics or groups of characteristics are desirable. Another form of this approach would be based on the belief that stocks in some sectors will outperform others.

Obviously, we want the prediction of beta to be as accurate as possible for each stock, so that its contribution to the expected change in beta is as accurately measured as possible. But it is also important that the law of averages will operate to reduce toward zero over a number of decisions the average value of the errors in the individual stocks selected. In other words, we want the prediction rule to be unbiased relative to the decision rule being used.

The importance of this point can be indicated by an illustration that we shall develop in some detail. Consider a portfolio manager who constructs his portfolio using stocks currently experiencing trading volume above their historical average. Then, when revising his portfolio, that portfolio manager might sell from the existing portfolio those stocks with below average volume, and might buy stocks with currently high trading volume. Now, suppose that at the same time the portfolio manager attempts to control the portfolio risk and limit beta to, for example, 1.2. If the predicted beta value on his current portfolio is 1.3, he might reasonably select for sale those stocks from the portfolio that were high in predicted beta, and replace these with stocks from among the actively traded list that were low in beta, while also meeting his other criteria for higher expected return.

Having set up this illustration, consider now the effects of a prediction rule that is negatively biased relative to changes in share trading volume in comparison to historical averages. In other words, if the stock is currently popular, the predicted beta will be too low, and, if the stock is currently unpopular, the predicted beta will be too high. It should be apparent that the portfolio manager would not achieve his goal of controlling risk by using such a rule. The average predicted beta for the stocks that he sold would be too high so that the sale would reduce the beta of his portfolio less than he expected, and the average predicted beta for the stocks he bought would be too low, resulting in a greater increase in beta from the purchase than he expected. These two effects combine to result in the transactions reducing beta less than expected. In fact, if the bias is large enough, the transactions might actually increase beta despite the fact that a reduction is predicted.

This example was developed at some length because the conventional methods now being used to predict beta *do show this kind of bias* and, as a result, this kind of error is being made on an everyday basis. It is quite conceivable for a portfolio manager, with the best intentions, to continue to produce a beta of 1.3 on a regular basis, although continually revising his portfolio to achieve an apparent beta of 1.2, simply because the prediction rule, by being biased relative to one of the characteristics employed for stock selection, asserts that beta will be reduced, when in fact it will not.

Thus we see that in selecting stocks it is desirable that the prediction rule for individual security betas again be unbiased relative to the characteristics employed in the decision rule.[11] Subject to this requirement, the prediction rule should be as accurate as possible—i.e., should exhibit minimum mean square error.

Finally, let us turn to the third use of predicted beta, namely, the valuation of convertible assets. Consider an investor in convertible assets who will repeatedly use the prediction rule to value a convertible asset prior to making a buy or sell decision. For this purpose the important point is that he make profitable decisions on average. So in this case our criterion for the choice of a prediction rule for beta is derived from the requirement that "good" valuations of convertible assets result, where "goodness" is measured by the profitability of an investment strategy based upon the valuations. Any error in the predicted beta feeds through to a consequent error in the valuation of the convertible asset, and the relationship between the former and the latter is a complicated one. It follows that a simple criterion applied to the valuation rule for convertible assets will result in a complicated criterion for the underlying prediction of risk. In particular the desire for a minimum-variance unbiased predictor of convertible asset value (not a bad criterion for a valuation rule), yields a highly complex criterion for the nature of the predictor of risk on the underlying common, that, among other things, does not require that the underlying predictor be unbiased.[12] Thus the criteria for beta predictions to be used for asset valuation are crucially dependent on the exact context.

FOOTNOTES

1. For an explanation of subscript notation, see J.L. Valentine and E.A. Mennis, *Quantitative Techniques for Financial Analysis*, 1st ed. (Charlottesville, Va.: CFA Research Foundation, 1971).
2. Note that $r_M = r_M^i + r_M^e$ and $r_a = r_a^i + r_a^e$. Because the events are independent, $E(r_M^i, r_M^e) = 0 = E(r_a^i, r_a^e)$. Further, because there is no reason to expect any dependence between r_a^e and r_M^i or r_a^i and r_M^e, $E(r_M^i, r_a^e) = 0 = E(r_M^e, r_a^i)$. Consequently,

$$\begin{aligned}\mathrm{COV}(r_a, r_M) &= E[r_a^i + r_a^e]\,[r_M^i + r_M^e] \\ &= E[r_a^i, r_M^i] + E[r_a^i, r_M^e] + E[r_a^e, r_M^i] \\ &\quad + E[r_a^e, r_M^e] \\ &= E[r_a^i, r_M^i] + E[r_a^e, r_M^e] \\ &= \mathrm{COV}(r_a^i, r_M^i) + \mathrm{COV}(r_a^e, r_M^e).\end{aligned}$$

3. A formal proof of this equation is given as follows: Let the market return generated by the j^{th} factor be denoted by f_j, with $r_M = \Sigma_j f_j$, and the market variance resulting from the j^{th} factor = $\mathrm{VAR}(f_j) = V_j$. For expository convenience, let us assume that the factors are independent, so $\mathrm{COV}(f_j, f_i) = 0$ for $i \neq j$. Without loss of generality, the factors are standardized so that the market response coefficient is 1.

Then $r_n = \Sigma_j \gamma_{jn} f_j + u_n$, where γ_{jn} is the security return caused by the j^{th} factor divided by the market return caused by the same factor, and u_n is the specific component of return for security n, independent of the factors. Therefore,

$$\begin{aligned}\beta_n &= \mathrm{COV}(r_n, r_M)/\mathrm{VAR}(r_M) \\ &= \frac{\mathrm{COV}(\sum_j \gamma_{jn} f_j + u_n, \sum_j f_j)}{\mathrm{VAR}(\sum_j f_j)} \\ &= \frac{\sum_j \gamma_{jn} V_j + \sum_i \sum_{j \neq i} 0}{\sum_j V_j + \sum_i \sum_{j \neq i} 0}.\end{aligned}$$

This equation can be viewed in another light. γ_{jn} can be considered to be that component of beta arising from a specific economic event. Consequently, to derive the overall beta, we should weight each one of these components by the importance of that specific event to overall market variance.

4. The discussion in the text indicates that an investor will make use of his predictions about the future and his attitude toward risk to derive a portfolio with a particular beta value. In this process, the investor is choosing between many portfolios with different beta values. When confronted with such a decision process, some market participants simplify the portfolio problem by advocating that an investor has to choose between just two extreme portfolios. If he expects the stock market to rise, he should be fully invested in common stocks with as high a beta value as is possible. If he expects the stock market to decline, however, he should hold no common stocks and should be fully invested in some fixed-interest assets whose value does not depend on movements in the stock market. Such an approach is based on the naive belief that we know with certainty whether the market will rise or fall. We can never be so certain. To reduce the exposure to this uncertainty, it is prudent to select an intermediate portfolio that balances the risks of an exposed position against the benefits from the expected movement. Consequently, at any point in time, the optimal portfolio will be some mixture of fixed-interest and equity securities, and, depending on the uncertainty of our predictions and our risk attitude, the portfolio could have one of many different beta levels.
5. See J.L. Treynor and F. Black, "How to Use Security Analysis to Improve Portfolio Selection," *The Journal of Business* (January 1973):66–86.
6. In principle, none of these criteria is really appropriate. One should first consider the investment strategy and evaluate the cost of making an error. Once this is decided, the error is measured in such a way as to maximize the present value of the contemplated investment strategy.
7. The derivation of this formula is simple:

$$\begin{aligned}\mathrm{MSE}[\hat{\beta}_n] &= E[\hat{\beta}_n - \beta_n]^2 \\ &= E[\hat{\beta}_n - E(\hat{\beta}_n) + E(\hat{\beta}_n) - \beta_n]^2 \\ &= E[\hat{\beta}_n - \bar{\hat{\beta}}_n + \bar{\hat{\beta}}_n - \beta_n]^2 \\ &= E[\hat{\beta}_n - \bar{\hat{\beta}}_n]^2 + E[\bar{\hat{\beta}}_n - \beta_n]^2 \\ &\quad + 2E[\hat{\beta}_n - \bar{\hat{\beta}}_n]\,[\bar{\hat{\beta}}_n - \beta_n].\end{aligned}$$

Now, $E(\hat{\beta}_n - \bar{\hat{\beta}}_n)^2 = \mathrm{VAR}(\hat{\beta}_n)$, by definition

$E(\bar{\hat{\beta}}_n - \beta_n)^2 = [\mathrm{BIAS}\ (\hat{\beta}_n)]^2$, since $\bar{\hat{\beta}}_n$ and β_n are both parameters, whose difference is equal to BIAS $(\hat{\beta}_n)$, the expectation of the BIAS $(\hat{\beta}_n)^2$ is equal to BIAS $(\hat{\beta}_n)^2$.

And $$\begin{aligned}E[\hat{\beta}_n - \bar{\hat{\beta}}_n]\,[\bar{\hat{\beta}}_n - \beta_n] &= [\bar{\hat{\beta}}_n - \beta_n]E[\hat{\beta}_n - \bar{\hat{\beta}}_n] \\ &= [\bar{\hat{\beta}}_n - \beta_n]\,[\bar{\hat{\beta}}_n - \bar{\hat{\beta}}_n] \\ &= 0.\end{aligned}$$

8. The variance of returns on an individual security, n, is related to its beta, and the variance of returns on the market by the following expression:

$$\mathrm{VAR}(r_n) = \beta_n^2\mathrm{VAR}(r_M) + \mathrm{VAR}(u_n),$$

where $\mathrm{VAR}(u_n)$ is the unsystematic risk of the security n. If we combine N securities in a portfolio with each security weighted by W_n, the expected return and variance of returns for the portfolio are

$$E(r_p) = \sum_{n=1}^{N} E[W_{Pn}(\alpha_n + \beta_n r_M + u_n)]$$

and

$$\mathrm{VAR}(r_p) = \sum_{n=1}^{N} W_{Pn}^2\beta_n^2\mathrm{VAR}(r_M) + \sum_{n=1}^{N} W_{Pn}^2\mathrm{VAR}(u_n).$$

In a diversified portfolio, the last term is close to zero and

$$\mathrm{VAR}(r_p) = \mathrm{VAR}(r_M)\sum_{n=1}^{N} W_n^2\beta_n^2 \approx \beta_P^2\mathrm{VAR}(r_M).$$

Also,

$$\beta_P = \frac{\mathrm{COV}(r_P, r_M)}{\mathrm{VAR}(r_M)} = \frac{\mathrm{COV}\left(\sum_{n=1}^{N} W_n r_n, r_M\right)}{\mathrm{VAR}(r_M)}.$$

$$= \frac{\sum_{n=1}^{N} W_n\mathrm{COV}(r_n, r_M)}{\mathrm{VAR}(r_M)} = \sum_{n=1}^{N} W_n\beta_n.$$

9. In future periods, the investment proportions will change as a consequence of stock price changes, and the portfolio beta will therefore also change. Nevertheless, the expected weights in the future will be close to the existing investment proportions, so that the predicted portfolio beta using current investment proportions is appropriate even when the uncertain future changes in investment proportions are taken into account.
10. There exists a problem of circularity here. The estimates of beta are used to determine the investment properties in any future portfolio, but yet these future investment proportions are needed in order to choose between the various estimates of $\hat{\beta}$. The choice of "typical investment proportions" suggested in the text sidesteps this problem.
11. As in the prediction of portfolio beta, there is the question of appropriate weights for the definition of unbiasedness. Paralleling the previous discussion, a natural criterion is to define the unbiasedness relative to a capitalization weighted average. For purposes of portfolio revision, however, this weighting is less clearly indicated. The problem is that the entire set of beta predictors for securities being considered for purchase and sale influences the transaction decision, although only a fraction of the securities under consideration may actually be traded. For instance, among eight securities regarded as candidates for above-average appreciation, the one with the highest predicted beta may be chosen for purchase. Whether this is also the stock with the highest true beta depends on the errors in estimating all eight statistics, regardless of the capitalization of those securities. Nevertheless, it is a reasonable approximation to assert that the expected influence of an error in estimating $\hat{\beta}_n$ is proportional to the capitalization of that asset.
12. To see this, note that the typical valuation rule for the estimated value $\hat{V}$ of a convertible asset, as a function of the

estimated mean $\hat{\bar{r}}$ and variance $\hat{s}$ of the return to the underlying common stock, has the properties of the integral

$$V \propto \int_{X_0}^{\infty} X \exp\{-1/2\,(X - \hat{\bar{r}})^2/\hat{s}\}\, dX.$$

The integral is a nonlinear function of $\hat{\bar{r}}$ and $\hat{s}$, so that a linear or quadratic criterion on $\hat{V}$ (e.g., $E[\hat{V}] = E[V]$, or MINIMIZE VAR[$\hat{V}$]) implies a nonquadratic criterion on $\hat{s}$. Indeed, the criterion can only be written in the form of an integral equation.