RESEARCH PROGRAM IN FINANCE

WORKING PAPER NO. 33

THE PREDICTION OF SYSTEMATIC RISK

by

Barr Rosenberg
and
James Guy

February 1975

Associate Professor and Assistant Professor, respectively, Graduate School of Business Administration, University of California, Berkeley. The research for this paper was supported by NSF Grant GS 3306. Preparation of this manuscript was supported in part by the Dean Witter Foundation.
THE PREDICTION OF SYSTEMATIC RISK

Barr Rosenberg and James Guy*

ABSTRACT

The systematic risk of securities and portfolios, or "beta," plays a central role in the evaluation of past investment performance and in the prediction and control of future investment risk. There has been a pronounced tendency to overlook the distinction between these two functions: historical evaluation and prediction. The purpose of this paper is to point out that estimates of systematic risk that are ideal for historical evaluation are far from ideal for predictive purposes, to explore criteria for optimal risk prediction, and to develop the methodology for risk prediction based upon fundamental concepts.

Introduction

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, and 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 and, as such, determines a major part of the context within which investment decisions are made. Therefore, among many possible risk measures, beta deserves particular attention, and will be the central topic of this paper. In the first section, beta will be defined, and then, in discussing the applications of beta, criteria for optimal prediction and estimation of beta emerge. In the second section, alternative methods of estimating and predicting beta are developed from a fundamental perspective, in which beta is seen as the consequence of economic variables. The third section becomes more technical, and develops further properties of alternative predictors of beta for securities and portfolios.

*Schools of Business Administration, University of California, Berkeley. This research was supported by NSF Grant GS 3306.
I. BETA: ITS USES AND CRITERIA FOR ITS PREDICTION

1.a. Beta

For clarity, let there be \( N \) securities, \( n = 1, \ldots, N \), and \( T \) time periods, indexed by the subscript \( t \) with the subscript for the present period being taken as \( t = 0 \). For each security \( n \), and for each time period \( t \), the following question may be posed: If the investment return on the market portfolio in that time period, \( r_{M_t} \), assumes any certain value, what return can be expected, on the average, for security \( n \), \( r_{nt} \)? For example, if the market return in that period will be +10 percent, can the security return be expected to be +20 percent, or +5 percent? The value of the ratio of the expected return on the security to the return on the market is the beta of security \( n \) in period \( t \), or \( \beta_{nt} \). Thus if a market return of +10 percent is associated with an expected security return of 20 percent, and a market return of -10 percent with an expected security return of -20 percent, the beta coefficient is \( 20%/10% = -20%/10% = 2 \).

Notice that the definition 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. This 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 were to imagine repeating many times the uncertain events which may occur in period \( t \), with each repetition having the nature of an experiment. Each experiment would yield some market return \( r_{Mt} \) and some security return \( r_{nt} \), and therefore some ratio of the two \( r_{nt}/r_{Mt} \). Beta is the expected value, or the average across these many experiments, of this ratio. Ex post, with hindsight, a unique value of this ratio will have occurred, but we are concerned with the expectation that held ex ante, looking forward in time. The value actually realized will not ordinarily equal beta. If we observed ex post that \( r_{Mt} = 20 \) percent and \( r_{nt} = 40 \) percent, the security's beta may not have been two. The true beta could have been one with the additional 20 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 either succeed or fail, but the result need not be equal to the expected value.

Note that from an economic viewpoint, the market return does not cause the security return. 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 from these sets of events, then the weighted average of these returns weighted by the probability of the events is 20 percent.

1.b. Beta as the Consequence of Underlying Economic Events

It is an instructive exercise to reach a judgment about beta by carrying out an imaginary experiment as follows: One can imagine all of the various events in the economy that may occur, and attempt to answer in each case the two questions: What would be the security return as a result of that event, and 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, related to these, that deserve greater 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 laws, among others. Now, 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 proportionally 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 in predicting beta as is usually devoted to predicting security returns in conventional security analysis. This parallel is, in fact, a valuable one to draw in thinking about beta. In security analysis, it is customary to distinguish between the component of return resulting from events that are specific to the firm in question, and the component of return stemming from events that affect 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 that are 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 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 anticipated value 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 which may impact the stock to obtain the expected security return due to economywide factors. Now the important point is that 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 two imaginary future events are given, with equal probability of good, bad, and no-change outcomes.

<table>
<thead>
<tr>
<th>Event</th>
<th>Outcome</th>
<th>Contribution to Return</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Market</td>
</tr>
<tr>
<td>Energy</td>
<td>good</td>
<td>+6</td>
</tr>
<tr>
<td></td>
<td>no change</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>bad</td>
<td>-6</td>
</tr>
<tr>
<td>Inflation</td>
<td>good</td>
<td>+3</td>
</tr>
<tr>
<td></td>
<td>no change</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>bad</td>
<td>-3</td>
</tr>
</tbody>
</table>

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.
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.

Notice that the level of beta is determined by two kinds of parameters: the degree of uncertainty attached to various categories of economic events and the response of security returns to these events. Although the economic events are not independent of one another, it is useful for expository purposes to assume that they are. Then the beta for any security n, will be

\[ \beta_{nt} = \frac{\sum_{j=1}^{J} \gamma_{jnt} V_{jt}}{\sum_{j=1}^{J} V_{jt}} \]

where \( V_{jt} \) is the contribution of economywide event j to market variance in period t, and where \( \gamma_{jnt} \) is the ratio of the responses of the nth security and the market to the jth event or the "relative response coefficient." This expression could be rewritten as,

\[ \beta_{nt} = \frac{\sum_{j=1}^{J} \left( \frac{V_{jt}}{\sum_{j=1}^{J} V_{j}} \right) \gamma_{jnt}}{\sum_{j=1}^{J} \left( \frac{V_{jt}}{\sum_{j=1}^{J} V_{j}} \right)} \]

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 variance in market return due to the event.

Returning to the illustration in table 1, if the energy and inflation events are independent of one another, the betas of the two stocks are

\[ \beta_A = \frac{4}{5} \cdot \frac{2}{3} + \frac{1}{5} \cdot \frac{2}{2} = \frac{14}{15} \]

\[ \beta_B = \frac{4}{5} \cdot \frac{4}{3} + \frac{1}{5} \cdot 0 = \frac{16}{15} \]

where (4/5) and (1/5) are the proportional contributions of the two events to market variance, and the other terms are the relative response coefficients of the stocks to these events.
This insight into the fundamental determinants of beta will be exploited at many points in the paper. 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 the relative response coefficients or the proportional variances of economic events change. To the degree that these changes can be predicted or explained, changes in beta can be predicted or explained. As one example, the monthly dates that 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. And, 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.

I.c. Uses of Beta

It is important to examine the uses of beta, not only as an aid in understanding it, but because the criteria for prediction and estimation should 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 that are 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. It is also clear that the beta of interest is the one that was assumed during the period of performance evaluation. In fact, it is absolutely imperative that the measure of beta apply to exactly the same period as the measure of cumulative return, for any mismatch can only worsen the efficiency of the performance measure. What if the beta during the period was changing, as is the usual case in the light of nonconstant management policy? Then, it turns out that the obvious answer, namely, to measure a beta that is in some sense the average over the historical period, is also a correct answer from the point of view of measuring investment performance.

**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 return in diversified portfolios can be explained by the 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 an accurate prediction of future beta for the portfolio were made, it would be an important ingredient in investment decision making. And equally, if accurate predictions were made of the betas for individual securities, these would be important ingredients of the portfolio revision decision. 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.

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. (b) To the degree that the market movement is uncertain, beta, in determining the response, determines the expected uncertainty of security or portfolio return.

In order to develop criteria for predictors of beta, it is convenient to refer to a typical investment decision strategy that relies, in part, on beta. This will be referred to as a "typical control strategy." We assume that the strategy consists in setting a target for the portfolio beta, which changes over time in response to (a) changing forecasts of the direction or market movement, or (b) changing assessments of the permissible level of systematic risk to be assumed.

Each time that the 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 (and 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. At the same time, these 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. Finally, these transactions are influenced by a desire to maintain an appropriate level of diversification. Thus, a set of transactions is undertaken with the multiple goals of (1) reaching a new 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 by 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 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). 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.

I.d. 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 paper 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.

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 1.
Suppose that the true average for a portfolio or security beta was \( \beta_n \), and that \( \hat{\beta}_n \) is an estimator of this and has an expected value \( \overline{\beta}_n \). The quality of this estimator can be judged by three criteria: bias, variance, and mean square error. If the estimator is unbiased, its expected value equals the true underlying average, or the bias, \( \overline{\beta}_n - \beta_n \), is zero. Estimators (a) and (b) are unbiased in figure 1. Such a characteristic is obviously desirable, and confronted by a group of estimators satisfying this criteria the most desirable one is that which 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 \( E[\hat{\beta}_n - \beta_n]^2 \). In figure 1, a is the most desirable unbiased estimator.

If we want to compare biased and unbiased estimators, then the appropriate criterion to evaluate them 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, BIAS(\( \hat{\beta}_n \)), variance, VAR(\( \hat{\beta}_n \)), and mean square error, MSE(\( \hat{\beta}_n \)), is given by,

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

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 estimate with low variance would be chosen in preference to an unbiased estimate with
high variance. This point is amplified graphically in figure 1. 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) so long as,

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

Let us now turn to the main topic of this paper, namely, the prediction of beta and 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:

\[
\hat{\beta}_p = \sum_n W_n \hat{\beta}_n,
\]

where \( W_n \) is the proportion of the total investment now in stock \( n \), with \( \sum_n W_n = 1 \). The predicted portfolio beta \( \hat{\beta}_p \) is

\[
\hat{\beta}_p = \sum_n W_n \hat{\beta}_n.
\]

The prediction error will therefore be \( \sum_n W_n (\hat{\beta}_n - \beta_n) \). 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.

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. Namely, if the true future beta is \( \beta_n \), and the forecasted beta is \( \hat{\beta}_n \) and has an expected value of \( \overline{\beta}_n \), then the forecast is unbiased if \( \overline{\beta}_n = \beta_n \). 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 $\beta_n$, not the present value, that is to be predicted.

If we were concerned with estimating the beta for a single stock $n$, $\beta_n$, the preceding considerations would suffice. But, as we are estimating beta for a number of securities, $n=1,\ldots,N$, it is necessary to consider criteria for a collection of estimates $\hat{\beta}_n, \ldots, n=1,\ldots,N$ so that the collection will perform optimally in use. Suppose that a prediction rule is defined which 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 and 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 this problem of weighting follows directly from the criterion that the errors in the predicted betas should average out when weighted by the proportionate investments in the portfolio. What is desired is unbiasedness, when weighted by the investment proportions. Thus, ideally, a slightly different set of weights must be used to evaluate unbiasedness for each 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 arises 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 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 that is 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 that predictor he will indeed tend to achieve a portfolio beta of 1.3. Now, the portfolio revision decision involves the sale of specific securities within the portfolio and the purchase of others. Consequently, it becomes necessary to predict the betas of individual securities. This highlights another essential distinction between future prediction and historical evaluation; in prediction, the risk levels of individual securities assume primary importance. Notice again that 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 values 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. The decision will depend in some form on the predictions of the betas for the individual securities. We presume that a group of sales and purchases will be selected that move in the direction of the desired betas, while also achieving an increase in expected return. It is likely that certain "characteristics" of the stocks will influence the choice. Thus, currently "popular" stocks may be considered for purchase, and currently "unpopular" stocks may be considered for sale. Or, currently high "P/E" stocks may be considered for purchase, and currently low "P/E" stocks considered 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 or 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 is developed in some detail. Consider a portfolio manager who constructs his portfolio using currently "popular" stocks, that is stocks that are 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 among portfolio managers of the hypothesized kind. 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. Subject to this requirement, the prediction rule should be as accurate as possible, so that it 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 will 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. Now 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
Thus the criteria for beta predictions to be used for asset valuation are crucially dependent on the exact context and will not be examined further in this paper.

II. ALTERNATIVE PREDICTION METHODS

Forecasting involves, of necessity, an extrapolation of some past relationships into the future. It further involves the use of currently available information to which the relationship is applied to predict a future value. Of course, one wishes to use as much information as possible and to extrapolate those relationships that are most likely to remain stable.

II.a. The Historical or Technical Beta: Its Limitations

At this writing, the most popular method of predicting beta is to take an estimate of the historical value of beta, \( \beta_h \), and to use this in an unadjusted fashion as a predictor. The estimate of the past average beta can be obtained by regressing historical security returns on market returns. This "historical beta" is an unbiased estimator of the true historical beta averaged over past periods and the historical beta is the minimum variance unbiased estimator of the true historical value if certain common assumptions hold; namely, the regression model is correctly specified, and the specific returns have zero mean, constant variance, and no correlation with each other.

One of the moving forces behind this approach has been to use the same estimator of beta for prediction purposes as for evaluation of historical performance, since the historical beta is also somewhat suitable for performance measurement. However, does this estimator meet our criteria of minimum variance subject to unbiasedness relative to characteristics for individual securities and of minimum mean square error for portfolios?

Whether the historical beta is an unbiased predictor of the future value of beta depends entirely on whether the change from the past average to the future value is expected to be zero. If the expected change is zero, then the average prediction error will be the sum of the average estimation error in the historical period (or zero) plus the expected change from the historical to the future value (or zero). However, if the average change from past to future value is non-zero, then a bias in this amount will occur. Therefore, if any information, presently available to implement a decision rule, contains implications about the difference between the future beta and the historical average, then the expected value of the change in beta, conditional on that information, will be non-zero, and the prediction rule will be biased.
Figure 2 illustrates several circumstances in which the expected change from the historical average to the future value is non-zero. In figure 2a, the beta of an imaginary security is plotted over a fairly long period, perhaps fifteen years. Two pronounced drops in beta occurred in the past, one perhaps arising from the acquisition of a less risky firm, another from a secondary offering of equity that reduced leverage. One of these occurred in the recent past, so that the average value of beta in the historical period indicated by the dashed line is substantially higher than the current value. The best forecast of beta is indicated by the heavy line extending out to the right, and the probable range of future values of beta, taking into account the possible future changes in the fundamentals of the firm, is shown by dotted lines that spread wider as time elapses and the likelihood of cumulative change grows.

Notice that the historical average is a biased predictor of future beta, because the fundamentals of the firm have changed within the historical interval. This indicates the first defect of the historical beta. Because it ignores possible changes in the fundamentals of the firm, it is potentially out of date and inaccurate. Moreover, it is biased with respect to any change in the fundamentals that took place within the historical period. It is necessary to incorporate the current fundamental position of the firm in risk prediction in order to remove this defect.

Figure 2b graphs the beta of another stock that tends to drift about an industry norm, assumed equal to 1.2. Sometimes the firm's beta is substantially above the norm, sometimes it is below, but the economic conditions of the industry, together with the standards of management in that industry, cause a tendency to revert toward the norm. The reversion may be the result of management's overt decision to achieve a more normal risk posture, or it may result from the disappearance of transitory abnormalities in the position of the firm within the industry. In either case, the best prediction for future beta includes a tendency for reversion to the industry norm, shown by the declining heavy line of best forecast. Again the historical average provides a biased forecast, this time because of the intrinsic tendency for the beta to return toward the norm from any abnormal historical average. The only way to remove this defect in the historical beta is to employ a weighted average including the industry norm and the historical beta.

Figure 2c graphs the beta of a company whose intrinsic exposure to economic events is changing over time as the result of unknown causes. This might be a consulting firm in pollution control, whose specialties are changing over time so that its fortunes are linked to different industries. Recently, the beta of the firm has been much higher than the historical average. The best forecast for the future, in the absence of information about future changes in the situation of the firm, is
that beta will persist at its present value. Thus, the historical average is out of date and provides a biased estimator with respect to any decision rule that responds to recent changes in beta. The only way to remove this defect is to obtain a more timely estimator of beta, either by changing the estimation procedure for historical beta itself or by obtaining other indicators of the current beta of the security.

Figure 2d graphs the beta of a security that has had a relatively constant level of risk in the past and for which the exposure to economic events is unchanging. However, this security is heavily exposed to one kind of economic event—the energy crisis—and the variance of that event is known to be increasing in the future. As a consequence, the exposure of the firm to that event will contribute to a progressively higher beta in the future, so that the best forecast of beta is shown by the rising heavy line. The historical average is a biased predictor because the variance of economic events is changing. The only way to remove this defect is to predict the relative response coefficient of the security to economic events and use the forecast of increasing variance for energy-related events to deduce a higher future beta.

There is one further defect of the historical beta as a predictor, which is that it inevitably involves estimation error. Specific events occurring to the firm in the past, which really had nothing to do with the economic events moving the market, will have matched the direction of market movement purely by chance, thereby causing the estimated value to differ from the underlying tendency. In figure 2e, a firm is graphed for which there have been no changes in the fundamentals and hence in the underlying beta. Nevertheless, the estimated historical beta will not be exactly equal to the underlying tendency, but instead will exhibit a random error as shown in the superimposed distribution. When an estimator exhibits error and when other information on the underlying quantity is available, the use of an appropriate weighted average of the two sources of information always reduces the mean square error of the predictor. Thus, even if beta is unchanging, we will want to include both fundamental and technical information in the predictor in order to achieve improved accuracy.

Now there are a number of possible improvements that lie strictly in the realm of greater efficiency in deducing from historical stock behavior the most likely current value of the beta. All of these can be viewed as improvements in the historical beta per se, and all amount to different forms of regression. All maintain the advantage of being unbiased estimators of some historical average of past beta for the stock. Some increase the timeliness of the estimate by weighting more recent observations more heavily than past observations, thus moving the weighting forward toward the present. Others attempt to deduce the underlying true value of beta more accurately by differential weights
for observations at different time intervals when more or less noise (specific risk) was present to obscure the underlying beta. Still others use different time intervals for observations. But the crucial point is that none escape the limitations discussed and exemplified above.

An implied predictor for the portfolio beta is found by summing the predictors for the individual stocks in the portfolio, with each weighted by the investment proportion in that stock. It turns out that the properties of this predictor as a predictor of future beta are similar to those for individual securities, so that an amelioration of the problems considered above would improve portfolio risk prediction as well.

II.b. The Use of Fundamental Information in Prediction

It is clearly an intriguing possibility to consider using other information to predict beta. At this point, an analogy may be helpful. Suppose that it is desired to estimate the volume of fluid in a human being, with the purpose of determining the dilution that will be experienced by a drug administered to him, in order to determine the correct drug dosage. Direct measurement of this figure is impossible, because it would entail invading the patient. A good indirect measure, based on the "fundamental" of physiology, is provided by the patient's weight, but this is imperfect, since other of his characteristics obscure the relationship between weight and volume. Other good "fundamental" measures can be based on his height, age, the amount of fat in his body as measured by skinfold thickness, and on knowledge of the manner in which the amount of fluid in the body is being self-regulated. To be even more precise, measurements of his body may be taken to estimate volume constructively. Now, each one of these measurements can add some information as to the prevailing volume. Alternatively, the volume concept can be measured directly by observing how much dilution actually occurs when the drug is administered. This is achieved by observing the concentration in the blood after administering some dosage. This latter approach is akin to a "technical" measure, for it measures the system's response to the drug. This measure may be unbiased, but it will have the associated problems of error (if concentration is perturbed by random factors) and of untimeliness (if the volume has changed since the time of the last measurement).

There is no theory as to which of these two procedures should be used. Both have a natural appeal. The former gets down to reliance on an intuitively appealing fundamental view of the patient, one that can be related to our understanding of physiology. The latter is a direct measurement of the construct that we are trying to estimate.

Precisely the same distinction arises in contrasting fundamental and technical approaches of predicting beta. The fundamental approach allows us to rely on our understanding of the firm in its
economic setting and on our understanding of the market's view of the firm. The technical approach is a direct measurement of the market's treatment of the firm that can neither be improved nor polluted by "knowledge" about the economy and the market.

In fundamental prediction, the fundamental characteristics of the stock and the market are analyzed in order to predict the responsiveness of the stock to the market.

In order to highlight the advantages of fundamental prediction and consequently the disadvantages of technical prediction, it is instructive to discuss the "conditional" and "unconditional" prediction of beta.

In the unconditional prediction of beta, the response coefficients to economic events are predicted, but the variances of the underlying events are considered unchanged. Consequently, beta is determined solely by the predicted response coefficients. This approach deals with the problem illustrated in figure 2a. Consider, as another illustration, a small undiversified automobile parts supplier, under the assumption that the future variance of economic events will be the same as in the past. Then, if the company's response to economic events is the same in the future as in the past, the unconditional forecast of beta will be the same as the best estimate of its value in the past. If this company decides to increase its long-term debt liability, however, its response coefficient will change, and there will be a different unconditional forecast of beta.

In the conditional forecast of beta, the variance of economic events and the response coefficients are both predicted, and the future beta is predicted conditional on the macroeconomic forecast. Although the energy crisis affects the stock market adversely, it has a more severe effect on the automobile industry. If the severity of the energy crisis increases, the prices of automobile stocks will fall more than the market. Consequently, if we believe the future variance of the macroeconomic event called "energy crisis" will increase, then the predictions of future beta for automobile companies, including the parts supplier introduced above, will be higher. Thus, the conditional approach deals with the problem illustrated in figure 2d.

Fundamental predictions specifically attempt to measure the effects on beta of changes in the response coefficients and the variances of the underlying events. Nonfundamental predictions could never predict a change in beta due to a change in the firm's response coefficient arising from a change in such characteristics as leverage and diversification. Unconditional fundamental forecasting achieves this goal. Further, nonfundamental predictions could never predict a change in beta due to a change in the variance of underlying macroeconomic events. Conditional fundamental forecasting performs this function.
II.C. Estimation of an Unconditional Fundamental Prediction Rule

To estimate an unconditional prediction rule, it is only necessary to observe the pattern of systematic risk that was realized in the past because the unconditional forecast presumes that the future economic variance is similar to the observed historical variance and extrapolates that pattern directly into the future. For example, if utilities have shown lower systematic risk in the recent past than airlines, then an unconditional forecast asserts that they will do so in the future.

This point is illustrated by some empirical results presented in a study by Rosenberg, Marathe, Houglet, and McKibben (1973). In column 1 of Table 2, the average betas for a number of industry groups are given as differences from the overall average for the market. These average betas were generated from more than 100,000 data points referring to the monthly behavior of returns for a large number of firms over a fifteen-year period. Although it is beyond the scope of this paper to discuss these results in detail, it is interesting to note that the standard errors are small and that highly significant differences in average beta conform, in the main, to the analysts' intuitions concerning industry risk. Because industry betas maintained these differences over the fifteen-year period, it is appealing to incorporate in an unconditional prediction of beta the assertion that the future beta for stocks in each industry will tend to be close to the historical average for that industry. Thus, the predicted beta for a stock will include some weight on the average historical beta for the industry.

If the future industry average beta for the utilities, for example, will not change relative to this historical average, this prediction rule will be unbiased across utilities, in the sense of yielding a predicted beta that is equal, on the average across all utilities, to the true beta for the utilities. This rule will be biased upward, however, for any one utility which, for unknown reasons, differs from the usual utility and has an inordinately low value of beta. This error arises because the utility is assumed to be similar to the other utilities, when in fact it is not.

Besides considering the industry effect, it is necessary to capture those other characteristics of a firm which influence its risk—for example, financial structure, aggressiveness of management or diversification. Using the various theories of the firm, it is possible to deduce the variables that are likely to affect risk and the direction of the influence. Increases in the firm's financial leverage or variability of earnings are likely to increase the firm's risk. But increases in the payout ratio or dividend yield are likely to lower risk as the management reduces its attempts to provide returns to the investors from uncertain capital gains based upon new projects.

The problem is to quantify these and similar concepts into a usable prediction rule. As one solution, the relationship between future risk and the fundamental characteristics of the firm may be
<table>
<thead>
<tr>
<th>Industry</th>
<th>Gross Differences from the Average</th>
<th>Remaining Unexplained Differences After Correction for Income Statement, Balance Sheet, and Market-Related Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>Standard Error</td>
</tr>
<tr>
<td>Air transportation</td>
<td>.47</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Electronics</td>
<td>.42</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Leisure time industry</td>
<td>.26</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Instruments</td>
<td>.26</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Office machines</td>
<td>.23</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Household white and brown goods and furniture</td>
<td>.21</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Newspapers, publishing, communications broadcasters</td>
<td>.20</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Nonelectrical machinery</td>
<td>.15</td>
<td>(0.02)</td>
</tr>
<tr>
<td>Electrical equipment</td>
<td>.13</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Aerospace</td>
<td>.13</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Motor vehicles and equipment</td>
<td>.07</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Textiles and apparel</td>
<td>.07</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Nonferrous metals</td>
<td>.07</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Construction materials</td>
<td>.05</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Miscellaneous consumer products</td>
<td>.05</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Chemicals</td>
<td>.03</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Iron and steel</td>
<td>.01</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Miscellaneous business services</td>
<td>.01</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Rubber and plastics</td>
<td>-.02</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Paper and associated products</td>
<td>-.02</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Finance companies</td>
<td>-.03</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Mining</td>
<td>-.06</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Retail merchandise</td>
<td>-.07</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Container</td>
<td>-.08</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Domestic oils and petroleum services</td>
<td>-.08</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Drugs and cosmetics</td>
<td>-.08</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Beverages</td>
<td>-.17</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Foods and associated products</td>
<td>-.17</td>
<td>(0.02)</td>
</tr>
<tr>
<td>Food stores</td>
<td>-.17</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Tobacco</td>
<td>-.18</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Utilities</td>
<td>-.38</td>
<td>(0.02)</td>
</tr>
<tr>
<td>International Oils</td>
<td>-.39</td>
<td>(0.05)</td>
</tr>
</tbody>
</table>

These results are taken from Barr Rosenberg et al. (1973), where the sample, estimation method, and industry groups are defined in detail.
estimated by regression methods. For this to be possible, quantitative variables describing a firm's characteristics must be defined. These descriptors of the current condition of the firm are then used to predict future risk, and the best prediction rule is fitted by linear regression. The better the descriptors, the more successful the prediction of risk will be.

This approach was used in Rosenberg and McKibben (1973) and in the study mentioned above. It is not the purpose of this paper to review those results in detail but, as an illustration of this approach, a few selected results are given in table 3.

TABLE 3
OPTIMAL ADJUSTMENTS TO SHORT-TERM BETA FORECASTS, \( \hat{\beta} \),
FOR SELECTED FUNDAMENTAL CHARACTERISTICS

<table>
<thead>
<tr>
<th>Descriptor</th>
<th>Adjustment in ( \hat{\beta} ) for a Difference of One Standard Deviation from the Mean of All Firms</th>
<th>t-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shares traded/shares outstanding</td>
<td>.095</td>
<td>10.</td>
</tr>
<tr>
<td>Book value/market value</td>
<td>.052</td>
<td>7.</td>
</tr>
<tr>
<td>Leverage</td>
<td>.038</td>
<td>5.</td>
</tr>
<tr>
<td>Dividend yield</td>
<td>-.037</td>
<td>-4.</td>
</tr>
<tr>
<td>Variability of earnings</td>
<td>.014</td>
<td>2.</td>
</tr>
</tbody>
</table>

These figures are standardized so as to be comparable with one another. For example, the coefficient of .095 for the share turnover ratio (STO) indicates that if the STO for a firm lies one standard deviation from the mean STO of all firms, the predicted beta for that firm is increased by .095. Roughly 17 percent of all firms will lie more than one standard deviation above the mean, and these will experience upward adjustments of .095 or more. Similarly, roughly 3 percent of all firms will lie two standard deviations or more above the mean and will experience upward adjustments of .19 or more as a result of this descriptor. And, of course, equally as many firms will lie below the mean and receive negative adjustments of the same amount. The coefficient of .014 for earnings' variability implies a smaller upward adjustment of .014 for a firm that is one standard deviation above the mean in earnings' variability. Roughly 17 percent of all firms will lie in this category and be adjusted by this amount or more.
This table gives the adjustments to predicted beta for a few descriptors. These adjustments are additive: in other words, the value of each descriptor for the firm is computed, and the adjustment for that descriptor is added to the adjustments for all other descriptors.

The final prediction is a weighted sum of many descriptors. A good prediction could be obtained by using only the variability of earnings: by adding leverage, a better prediction is obtained, and the importance of earnings' variability diminishes because it was, in part, serving as a means to predict the effects of financial leverage. Similarly, as dividend yield is added, the importance of both earnings' variability and leverage diminish, because they were both serving, in part, as measures of management conservatism, which is also measured by yield. If one characteristic of the firm cannot be measured accurately, the measurement, once obtained, may be useless, because the accurate measurements of other characteristics serve as better predictors of this one characteristic than does the doubtful measurement itself. This explains the quite low weight given to the variability of past earnings in the final prediction rule. The measure used in this study, the variability of annual historical earnings, apparently has less content than other fundamental variables, not all of which are included in this table, that serve as predictors of the intrinsic earnings' variability.

Taken as a group, the fundamental variables provided excellent predictors of risk. In fact, they were substantially better predictors than the historical beta in the sense that they achieved a smaller measurement error. The prediction rule combining the fundamental variables and the industry differences in an optimal way performed still better. In this combined prediction rule, industry differences need account for only those aspects of the firms in the industry that are not otherwise measured by fundamental characteristics. These remaining differences, after the contribution of the fundamental characteristics, are given in column 2 of table 2. The differences are now much smaller but remain highly significant statistically. These remaining differences may be interpreted as true "industry characteristics" as distinct from fundamental characteristics that are typical of firms in the industry.

Will this prediction rule, as fitted by regression, be unbiased? It may be shown that if any characteristic is included within this regression, the resulting prediction rule will necessarily be unbiased relative to that characteristic. However, if unbiasedness is taken with the items weighted by market capitalization, the data in the regression must be weighted in the same way.

This approach provides a very powerful method of constructing a prediction rule for beta using fundamental information. To clarify this point, suppose that we have initially a fundamental prediction rule that does not employ the distinction between utilities and other firms as one of its sources of information. Suppose, further, that this characteristic has a profound effect upon beta, over and
above all other observable fundamental characteristics that might have already been employed in the prediction rule. Consequently, because this item is correlated with the difference between the previous fundamental prediction and the actual beta, the previous fundamental prediction rule will be biased relative to this characteristic. In fact, it will be biased upward relative to this characteristic, because utilities have lower betas that could otherwise be accounted for by balance-sheet and income-statement characteristics.

The foregoing prediction rule would have been fitted by a regression in which the information about utilities was not included. The bias and inefficiency of that prediction rule can be overcome if an additional descriptor capturing the "utility" characteristic is added to the right-hand side of the equation. Since this variable carries significant new information about beta, the modified regression fits significantly better. On the average, the predictions of beta for utilities will drop by the desired amount, so that the resulting predictions will no longer be biased and will show an average error of zero across the group. Thus, by adding this additional item, the efficiency of the overall predictive performance has improved, and the bias of the predictor relative to the utility industry has been removed.

There is a clearcut moral here. Whenever an item of information will be helpful in predicting beta, it should be employed in the prediction rule. This improves efficiency in prediction and, whenever that characteristic might be used as the basis for a selection rule, removes bias with respect to that selection rule.

II.d. Combining Fundamental and Technical Approaches

In making an estimate, it is natural to combine information from several sources. In the fitted fundamental prediction rule, a variety of descriptors of fundamental characteristics and industry groups were combined into a prediction rule by least squares regression. It is not widely realized that these same sources of information may be further combined with technical predictors for beta by the simple strategy of including these technical measures as additional descriptors in the regression. In the augmented regression, descriptors may be any measures of risk-relevant aspects of the firm that are based upon published information—fundamental, technical, or otherwise. The dependent variable is the future beta, and the fitted prediction rule provides an optimal predictor of future beta from currently published descriptors of all sorts, in the sense of providing a minimum mean square error prediction rule that is also unbiased with respect to any descriptors included in the fitted rule, relative to the weights on firms that are used in the regression.

Intuitively, the regression places weight on a descriptor whenever accurate information about future risk can be obtained. If individual security risk is not changing over time, and if past
risk can be accurately estimated by a historical technical measure, then that technical measure will be given heavy weight and fundamental information will be given relatively little weight. If, on the other hand, the individual security betas are changing over time and/or the historical risk cannot be accurately estimated by a technical measure, then the technical measure will be downweighted and heavy weight will be placed on fundamental predictions.

The fundamental information turns out to carry the lion's share of the weight in the fitted prediction rule. The historical beta added only 4 percent to the total predictable variance, and the weight given to historical beta was only 0.2. That is, if the estimated historical beta were 2.5--1.5 above the average value of 1.0--then the predicted beta would be increased by $0.2 \times 1.5 = 0.3$.

Fundamental betas have certain advantages and disadvantages in comparison with technical betas. The technical beta uses the historical data of the stock to derive a direct measurement of the characteristic of interest, beta. Its weakness lies in measurement error and in inadequate response to changes in the company's characteristics and in the variances of the underlying economic events. In contrast, the weakness of the fundamental prediction rule, or of the most naive rule that all betas are akin to the norm, is that it applies a general pattern to an individual stock. Its strength lies in the ability to respond to changes in characteristics and in less measurement error, since the pattern is observed over many cases. The combined rule attains some of the strengths of both approaches.

II.e. Combining Predictions for Individual Security Betas into a Prediction for a Portfolio Beta

So far, the question of combining predictors of individual security betas into a predictor for the portfolio beta has been sidestepped. At first sight, the answer appears obvious. Since the portfolio beta is the sum of the individual security betas, each weighted by the appropriate investment proportion, the predictor of the portfolio beta should just be the weighted average of the predictors for the individual betas. The problem, however, is that the individual predictors were constructed so as to be optimal predictors for individual securities, and it is not necessarily true that they are also optimal as ingredients in a weighted sum being used to predict a portfolio beta.

Thus, we need to begin with the criterion of minimum mean square error for portfolio beta prediction and to deduce from this new criteria that can be applied to individual stocks. When this is done, the following fact emerges. If the individual predictors are optimal relative to some criterion, then their weighted sum will be optimal for the portfolio relative to that criterion, if and only if the prediction rule exhibits errors that are uncorrelated with the weights in the portfolio (i.e., the prediction rule is unbiased relative to the method of constructing the portfolio) and the prediction rule is optimal with respect to the weights used in the portfolio.
These two conditions must be satisfied. The second condition can be approximately satisfied by a single overall prediction rule, provided that it is unbiased with respect to capitalization weights. These weights will tend to reflect the weights used in institutional portfolios, and an optimized prediction rule, unbiased with respect to these weights, will, in practice, be very similar to one optimized with respect to the weights in the individual portfolio.

The first condition turns out to be equivalent to the requirement that the predictor be unbiased relative to the characteristics that are employed in the selection rule. As a consequence, introducing these characteristics as ingredients in a fundamental prediction rule for individual securities has the additional advantage of allowing us to use the individual forecasts unhesitatingly as ingredients for the portfolio forecast.

The estimated fundamental prediction rule suggested that shares which have recently exhibited high (low) share turnover rates will exhibit higher (lower) betas in the near future than would otherwise be expected. This implies that a prediction rule that did not employ this characteristic would be downward biased for currently popular stocks. To confirm this expectation, an experiment was carried out using data subsequent to that employed in the previously mentioned study. Among the 700 largest capitalized industrial firms, the twenty-five that exhibited the highest share turnover ratios (STO) in the year 1972 were grouped into a portfolio, as were the twenty-five that had exhibited the lowest STO. It was expected that the historical beta for the high STO portfolio would underpredict the actual future beta for that portfolio, while that for the low STO portfolio would overpredict. Historical betas were taken from the Merrill Lynch beta file, as of December 1972. Returns for the portfolios were observed for the first six months of 1973, a period in which the average stock in the sample of 700 stocks fell by 23 percent. The high STO portfolio had a high estimated historical beta of 1.27. Therefore, reliance on the technical prediction rule, using information-available on December 31, 1972, would have forecast a six-month decline of 28.8 percent for the portfolio, conditional on the market return. The portfolio actually declined by 39.1 percent, exhibiting a much higher beta than the historical beta. This was expected on the basis of the high STO rate in 1972. The low STO portfolio had a low estimated historical beta of 0.725. Relying on this technical forecast, the portfolio would have been expected to decline by 15.1 percent. It actually exhibited a still lower beta, declining by only 6 percent! This experiment shows the magnitude of the biases inherent in a technical rule that ignores recent fundamental changes in the nature of the firm.
III. A MORE DETAILED TREATMENT OF THE "BAYESIAN ADJUSTMENT"

In this section, the properties of fitted minimum mean square error prediction rules will be analyzed in greater detail. The reader who is only interested in the qualitative properties of predictions of risk may omit this section with little loss; it is intended for those who want to understand the statistical theory somewhat more deeply.

III.a. Combining Technical and Fundamental Approaches: The Special Case of the Bayesian Adjustment

To simplify the exposition and to introduce the concept of Bayesian adjustment, consider the simplest possible kind of fundamental information about beta—namely, the knowledge that the average beta, weighted by market capitalizations, $\bar{\beta}$, is one. This is fundamental knowledge in the sense that it stems from the concept of beta as a market response and takes into account our knowledge of a relationship across all betas. What does this information tell us? It tells us that all betas must be distributed in some manner with a weighted average at one:

Here, $\bar{\beta} = \sum_n c_n \beta_n$, where $c_n$ is the proportion of total market capitalization in stock $n$. This information leads to a very simple fundamental prediction rule—namely, the "null" prediction rule that beta equals its average value. Thus, the prediction for stock $n$ is $\hat{\beta}_n = \bar{\beta}$. Accordingly, the error is $\hat{\beta}_n - \beta_n = \bar{\beta} - \beta_n$. The capitalization-weighted bias is $\sum_n c_n E(\hat{\beta}_n - \beta_n) = \sum_n c_n \bar{\beta} - \sum_n c_n \beta_n = \bar{\beta} - \bar{\beta} = 0$. Thus, the "null" predictor is unbiased relative to capitalization weights across the sample of all stocks. Also, the weighted average squared error for any set of weights $a_1, \ldots, a_N$, $\Sigma a_n = 1$, is

$$\sum_n a_n E(\hat{\beta}_n - \beta_n)^2 = \sum_n a_n (\bar{\beta} - \beta_n)^2 = V_\beta,$$

where $V_\beta$ is the weighted variance of $\beta$. In other words, the null predictor that beta is equal to its mean has an average error of zero and a mean square error that is equal to the variance of the
population of betas. The greater the variance of this population, the poorer is the null predictor.

Next, consider an alternative prediction rule, based upon a technical estimator of historical beta. Let that estimator for security \( n \) be given by \( \beta_{n}^h \). If this was the output of a regression designed to estimate the true average beta in the historical period, \( \beta_{n}^h \), then it is safe to assume that it was unbiased. That is,

\[
E(\beta_{n}^h) = \beta_{n}^h.
\]

How good the estimator was as an estimator of the historical average depends on its variance. Let this variance be denoted by \( V_n \):

\[
VAR(\beta_{n}^h) = E(\beta_{n}^h - \beta_{n}^h)^2 = V_n.
\]

The regression estimator itself provides an associated estimate of \( V_n \), say, \( \hat{V}_n \), which is unbiased if the regression is properly specified. Thus, we may be able to estimate the accuracy of the technical estimator, at least insofar as it was an estimator of the historical average.

But will the historical estimator be an equally good predictor of the future value? Only if the future value is identical to the historical average. On the contrary, firms' risk characteristics are continually changing, and it is unlikely that the future value will be equal to the past. For one thing, firms that had extreme risk positions in the past are likely to approach the normal value, either as a result of deliberate management policy or as the consequence of the disappearance of extraordinary circumstances that were associated with an extreme risk position. Therefore, it is reasonable to build in some tendency for the risk of the firm to converge toward the norm. This is accomplished in the model in which the future beta, \( \beta_n \), is given by

\[
\beta_n = \beta + k(\beta_{n}^h - \beta) = (1-k)\beta + k\beta_{n}^h.
\]

Here, \( k \) measures the rate of "memory" of the process. Thus, if \( k = 1 \), the future value of beta, \( \beta_n \), will be identical to the historical value, \( \beta_{n}^h \), and the process has "perfect memory." If \( k \) equals zero, the future value, \( \beta_n \), will be \( \beta \); thus, the future value of \( \beta \) equals the population average, and the process has "zero memory" of the historical value, \( \beta_{n}^h \). Alternatively, \( (1-k) \) is the rate of convergence toward the population mean.

When this convergence process is repeated over several periods, all the \( \beta_n \) will eventually converge to the mean (unless \( k = 1 \) so that no convergence occurs). We know that this is contrary to reality, for \( V_b \), the variance of the cross-sectional distribution of betas, appears to have remained
nearly constant over the past fifty years. Thus, there must have been other events that offset the
tendency to converge to the norm. We may represent these by random error terms, \( e_n \), that shock the
individual security betas. Then the model becomes:

\[
\beta_n = \bar{\beta} + k(\beta_n^h - \bar{\beta}) + e_n,
\]

where \( E(e_n) = 0 \), \( \text{VAR}(e_n) = \sigma^2_e \), and where we assume that the random term, \( e_n \), is uncorrelated with \( \beta_n^h \).

In this model, what will be the properties of \( H\beta \) as a predictor for \( \beta \)? First, the predi-
cction error is:

\[
(H\beta_n - \beta_n) = (H\beta_n - (1-k)\bar{\beta} - k\beta_n^h - e_n) = (H\beta_n - \beta_n^h) + (1-k)(\beta_n^h - \bar{\beta}) - e_n.
\]

Thus, the prediction error has three parts: error in estimating the historical beta, failure to cap-
ture the tendency to converge toward the norm, and failure to predict the random shift in beta. The
expected error, for any stock \( n \), is:

\[
E(H\beta_n - \beta_n) = E(H\beta_n - \beta_n^h) + E[(1-k)(\beta_n^h - \bar{\beta})] + E(-e_n)
\]

\[
= (1-k)(\beta_n^h - \bar{\beta}) = \text{BIAS}_n.
\]

The expected value of the first term disappears, because \( H\beta \) is an unbiased estimator of the histori-
ical average. The third term disappears because the expected random shift is zero. But the second
term remains, since there is an expected tendency to converge toward the norm. Thus, the historical
beta is an upward-biased predictor for those stocks that were historically above the norm and, con-
versely, a downward-biased predictor for those stocks that were historically below the norm. However,
these biases average out to zero, since

\[
\sum_n c_n \text{BIAS}_n = \sum_n c_n (1-k)(\beta_n^h - \bar{\beta}) = (1-k)\left(\sum_n c_n \beta_n^h - \sum_n c_n \bar{\beta}\right)
\]

\[
= (1-k)(\beta_n^h - \bar{\beta}) = 0.
\]

Thus, the predictor remains unbiased on average across all stocks.

The mean square error of the predictor, for any stock \( n \), is:

\[
E\left[(H\beta_n - \beta_n)^2\right] = \text{BIAS}_n^2 + \text{VAR}(H\beta_n - \beta_n)
\]

\[
= (1-k)^2(\beta_n^h - \bar{\beta})^2 + \sigma^2_e + \sigma^2_e.
\]
since the variance of the error is just the sum of the variance of the historical estimation error and the variance of the independent random shift in beta.

Thus far, the two alternative prediction rules, the "null" fundamental predictor and the technical predictor, have been treated separately. Now consider the possibility of combining them to obtain a weighted average of the two predictors that may possibly be superior to both. The weighted average predictor will be of the form:

\[ \tilde{\beta}_n = \kappa_n \beta_n + \lambda_n H \beta_n. \]

Here, \( \lambda_n \) is the weight given to \( H \beta_n \), the technical estimator, and \( \kappa_n \) is the weight given to the null predictor. If this combined predictor is to be superior, it should have a lower mean square error. The mean square error of the predictor, in terms of the as-yet-underdetermined weights, is:

\[
\text{MSE}[\tilde{\beta}_n] = E[(\tilde{\beta}_n - \beta_n)^2] = E[(\kappa_n \beta_n + \lambda_n H \beta_n - \beta_n)^2]
\]

\[
= E[(\kappa_n \beta_n + \lambda_n H \beta_n - (1-k)\beta_n - e_n)^2]
\]

\[
= E \left[ \left( (\kappa_n - 1 + \lambda_n) \beta_n + \lambda_n (H \beta_n - \beta_n) + (\lambda_n - k) \beta_n - e_n \right)^2 \right]
\]

\[
= E \left[ \left( (\lambda_n - k) \beta_n - e_n \right)^2 \right]
\]

Thus, the mean square error is the sum of four terms. The third term is the only one that involves \( \kappa_n \), so we may choose the value of \( \kappa_n \) so as to minimize it. The term is a square, and so can never be smaller than zero. It will equal zero if \( \kappa_n = (1-k) \beta_n \), that is, if \( \kappa_n = (1-k) \beta_n \). Thus, we find that the optimal value for \( \kappa_n \) is \( (1-k) \beta_n \), or, equivalently, that the optimal predictor is a weighted average of the two constituent prediction rules.

\[ \beta_n = (1-k) \beta_n + \lambda_n H \beta_n. \]

This is the first important step in finding the best weighted combination of the two predictors.
The predictor can also be written as:

\[ \tilde{\beta}_n = \bar{\beta} + \lambda_n (\tilde{\beta} - \bar{\beta}). \]

The optimal adjustment factor \( \lambda \) will be found to lie between zero and one. Thus, the predictor takes the form of a conservative adjustment of the historical beta, in which the technical estimator is multiplied by a factor less than one, that draws the prediction in toward the average. This is called "the Bayesian Adjustment," after the statistician, Bayes.

The Bayesian approach to statistical inference relies on a "prior" probability distribution reflecting all knowledge before measurement of the phenomenon to be estimated. This prior distribution is then modified to a "posterior" distribution incorporating the information obtained in the measurement, as well as the prior information. The procedure for incorporating the measurement is known as Bayes' law. In the present case, the distribution of true betas is the prior distribution, \( \beta_n \) is the measurement, and \( \tilde{\beta} \) will, in fact, be the mean value of the posterior distribution. From this perspective, the philosophical limitations of the simplistic Bayesian Adjustment are clear. It presumes that prior information is limited to the knowledge that this firm is randomly drawn from the population of all firms. All other information about the firm's characteristics that might be treated legitimately as prior information is ignored. Thus, the simple rule that has come to be termed "the Bayesian Adjustment" is only one possible form of Bayesian adjustment—that form which employs the minimum of prior information.

Returning to the analysis of this estimator, its expected value, in terms of the as-yet-undetermined adjustment factor, will be:

\[ E[\tilde{\beta}] = (1-\lambda_n)\bar{\beta} + \lambda_n E[\beta_n] = (1-\lambda_n)\bar{\beta} + \lambda_n \beta_n^h, \]

since \( \beta_n^h \) is an unbiased estimator of \( \beta_n^h \). Therefore, the bias of this predictor for an individual stock is:

\[ \text{BIAS}[\tilde{\beta}_n] = E[\tilde{\beta}_n - \beta_n] = (1-\lambda_n)\bar{\beta} + \lambda_n \beta_n^h - (1-k)\bar{\beta} - k\beta_n^h = (\lambda_n - k)(\beta_n^h - \bar{\beta}). \]

The mean square error of this weighted average predictor reduces to:

\[ \text{MSE}[\tilde{\beta}_n] = (\lambda_n - k)^2 (\beta_n^h - \bar{\beta})^2 + \lambda_n^2 \nu + v_e. \]

The last term is independent of the chosen weight, for it is the unavoidable variance arising from the unpredictable shift in beta. The optimal value of \( \lambda_n \) will therefore be the value that minimizes the remaining terms.
Although it is not difficult to derive an optimal \( \lambda_n \) for each security \( n \), the procedure involves complexities beyond the scope of this paper, so let us instead find the optimal value \( \lambda^* \), such that the same value \( \lambda^* \) is to be used for all securities. This second approach suffices to introduce the concepts that are of primary interest.

The problem thus becomes one of choosing that value of \( \lambda^* \) which minimizes some weighted average of the mean square prediction errors for the various securities. For convenience, we assume that the criterion is to minimize the sum of the mean square errors, weighted by some weights \( a_1, \ldots, a_N \). Thus, \( \lambda^* \) is the value that minimizes:

\[
\sum_n a_n \text{MSE}[\tilde{\beta}_n] = \sum_n a_n (\lambda-k)^2 (\tilde{\beta}_n - \beta_n)^2 + \lambda^2 \bar{v}_n + \text{a constant}
\]

\[
= (\lambda-k)^2 \sum_n a_n (\tilde{\beta}_n - \beta_n)^2 + \lambda^2 \sum_n a_n \bar{v}_n + \text{a constant}.
\]

But \( \sum_n a_n (\tilde{\beta}_n - \beta_n)^2 \) is just the weighted cross-sectional variance of the true historical beta. It is reasonable to assume that this variance is the same as the cross-sectional variance of the future beta, or \( \bar{v}_\beta \). Similarly, \( \sum_n a_n \bar{v}_n \) is the appropriately weighted average of estimation errors in the historical betas, which may be denoted by \( \bar{V} \). Therefore, \( \lambda^* \) minimizes:

\[
S(\lambda) = (\lambda-k)^2 v_\beta + \lambda^2 \bar{V}_{\beta}.
\]

Since this is a quadratic function of \( \lambda \) with positive coefficients, the extreme value is a minimum. It is found by setting the first derivative equal to zero:

\[
\frac{dS(\lambda)}{d\lambda} = 2(\lambda-k)v_\beta + 2\lambda \bar{V} = 0
\]

or

\[
\lambda^* = k \frac{v_\beta}{v_\beta + \bar{V}}.
\]

Since \( \bar{V} \) is positive, the optimal weight \( \lambda^* \) will be greater than zero and less than \( k \). The greater is the memory \( k \), the greater will be the weight \( \lambda_n \) on the historical estimator. The greater is the variance \( \bar{V} \) of the historical estimators, the smaller will be the weight on those estimators. The greater is the variance of true beta, the greater is the weight placed on the historical estimator.

Is this Bayesian adjusted predictor unbiased? The two original predictors, from which it is combined, were each unbiased relative to a capitalization-weighted average, so this remains unbiased.
relative to that average. But what about relative to any other weighting or any other characteristic? To know this, we must ask whether the original estimators were themselves unbiased with respect to these characteristics.

The bias in the null predictor is just the true beta less the average, $\beta_n - \bar{\beta}$. Hence, if any characteristic is correlated with this difference, then the predictor will be biased relative to that characteristic. In fact, many variables are so correlated, and they include, among others, industrial classification, specific risk, and a large number of balance sheet items. Consequently, the null predictor will be biased relative to all these characteristics, and such characteristics should appear in a fundamental prediction rule. In the case of the technical estimator, the errors in estimating the past average are the result of random specific events in the past that we can expect to be largely uncorrelated with characteristics now being used to select stocks. As previously discussed, however, there may be bias resulting from changes from past values to future values that may be correlated with some characteristic being used in a selection rule. The weighted-average estimator formed from the technical and null estimators will exhibit biases that are exactly equal to the weighted average of the biases of the separate rules.

Before proceeding to discuss the estimation of the parameters $k$, $V_0$, and $V$ that determine the optimal Bayesian adjustment, this section will be concluded with a numerical illustration of the concepts developed thus far. For simplicity, let $\bar{\beta} = 1$ and assume that all betas are constant over time so that $k = 1$ and $V_e = 0$. Then,

$$\bar{\beta}_n = (1 - \lambda^*) + \lambda^* \beta_n$$

$$\text{BIAS}(\bar{\beta}_n) = (\lambda^* - 1)(\beta_n - \bar{\beta})$$

$$\text{MSE}(\bar{\beta}_n) = (\lambda^* - 1)^2 (\beta_n - \bar{\beta})^2 + \lambda^2 V_n$$

To obtain a numerical example, assume that $V_n = 0.13$ for every stock $n$, and that the cross-sectional variance of beta is $V_\beta = 0.16$. Then, $\bar{V} = 0.13$, as well, and $\lambda^* = 0.16/(0.13 + 0.16) = 0.55$.

Table 4 provides the bias, variance, and mean square errors of the alternative predictor for stocks with true betas of 2.0, 1.0, and 0.50. The Bayesian adjusted beta turns out to be a good predictor for the stocks with true betas lying near one, but an inferior predictor for stocks with true betas lying far from one, because the conservative adjustment introduces a serious bias for such stocks. The point is, of course, that we don't know which stocks have betas near one and which do not; all that we do know is the estimate $\bar{\beta}_n$ and the fact that the Bayesian adjustment is the optimal
### TABLE 4

<table>
<thead>
<tr>
<th>$\beta_n$</th>
<th>$E(\tilde{\beta}_n)$</th>
<th>BIAS($\tilde{\beta}_n$)</th>
<th>MSE($\tilde{\beta}_n$)</th>
<th>VAR($\tilde{\beta}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 1</td>
<td>2.0</td>
<td>1.550</td>
<td>-0.450</td>
<td>0.2418</td>
</tr>
<tr>
<td>Case 2</td>
<td>1.0</td>
<td>1.000</td>
<td>0.000</td>
<td>0.0393</td>
</tr>
<tr>
<td>Case 3</td>
<td>0.5</td>
<td>0.725</td>
<td>0.225</td>
<td>0.0900</td>
</tr>
</tbody>
</table>

$\lambda = 0.55$; $\bar{\beta} = 1$; $HV_n = \bar{V} = 0.13$. 

![Diagram](image-url)
III.b. Estimating the Optimal Bayesian Adjustment by Regression

There are two different approaches to estimating the parameters needed for the Bayesian Adjustment. The first and simplest is by means of linear regression. This approach provides the easiest and most robust means of constructing the Bayesian adjustment. An understanding of it also clarifies the relationship between the simple Bayesian adjustment and the more sophisticated fundamental prediction rules considered earlier.

Because the Bayesian Adjusted Predictor minimizes the mean square error in prediction, it is closely related to a regression estimator. Let $\hat{\beta}_{nt}$ denote the value of $\beta_n$ in some period $t$. Let $\hat{HB}_{nt}$ denote the historical beta estimator available at the start of period $t$, that can be used to form a prediction rule. In the regression

$$\hat{\beta}_{nt} = b_0 + b_1 \hat{HB}_{nt},$$

the coefficients estimated by least squares would minimize the sum of squared errors in predicting $\hat{\beta}_{nt}$. If the data were weighted by the $a_n$, the least squares regression would minimize the same variance that defines the desired Bayesian adjustment, so it must be the case that $\hat{b}_0 = (1-\lambda^*)$ and $\hat{b}_1 = \lambda^*$. The reason put simply is that the regression fits the estimated coefficients so as to minimize the weighted sum of squared errors, which is just the criterion we wish to use in choosing a prediction rule.

The problem with the above regression is that it is not possible to carry it out, because the variable $\hat{\beta}_{nt}$ is not observed. However, the ratio of the excess security return to the excess market return, $\frac{(r_{nt} - r_{Ft})}{(r_{Mt} - r_{Ft})}$, which we do observe, has an expected value equal to $\hat{\beta}_{nt}$. Substituting this into the regression, we have:

$$E\left[\frac{r_{nt} - r_{Ft}}{r_{Mt} - r_{Ft}}\right] = b_0 + b_1 \hat{HB}_{nt}.$$
variables, shown in brackets, are each equal to the product of the market return with either the "null" predictor of one or the technical beta fitted to previous data. With the appropriate weights, the resulting estimates of $b_0$ and $b_1$ will yield the Bayesian adjustment.\footnote{14}

This approach can be generalized to fit any prediction rule for beta. One can attempt to explain the excess return on stocks successfully, by using as a prediction the observed return on the market multiplied by a factor that is the fitted prediction rule for beta. The great advantage of this approach is that it can be used to incorporate as many sources of information as desired. Thus the following regression relation,

\[
E[r_{nt} - r_{Ft}] = b_0 \{r_{Mt} - r_{Ft}\} + b_1 \{E_0 \{r_{Mt} - r_{Ft}\}\} + b_2 \{X_{2nt} (r_{Mt} - r_{Ft})\} + \ldots + b_k \{X_{knt} (r_{Mt} - r_{Ft})\}
\]

where $X_2, X_3, \ldots, X_k$ are various descriptors useful in predicting beta, will provide a prediction rule that is the optimal Bayesian prediction rule employing these sources of information, in the sense of minimizing the appropriately-weighted mean square prediction error.

In discussing the regression approach, the question of the appropriate weights, $a_n$, $n=1, N$ has been sidestepped. In fact, these are determined by two considerations which may or may not conflict, depending on the correctness of the specification of the descriptors used in the prediction rule. One consideration is that, for purposes of statistical efficiency, the weights should be chosen in inverse proportion to the residual variance of return. When this is done, the weighted residuals all possess equal variances, that is, are "homoscedastic," and least squares regression is optimally efficient. Suppose, for example, that betas can be predicted solely on size. In the hypothetical illustration given in figure 3a, it can be seen that beta declines with size. But in addition the variance of the errors in the prediction of beta are smaller for large companies than for small companies. Consequently, in order to achieve homoscedasticity the larger companies should be weighted more heavily, although it is very unlikely that the weighting will be in exact proportion to their absolute size.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{3a_3b.png}
\caption{Figure 3a}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{3a_3b.png}
\caption{Figure 3b}
\end{figure}
Now, the next consideration that influences the choice of weights, applies only if accuracy for one class of firms requires the sacrifice of accuracy for another class. When such a trade-off exists, the weights should be set in proportion to the relative importance of the two classes. In other words, if it is more important to achieve accuracy for large firms than for small, then the weights given to large firms should be greater, so that the fitted mean square error prediction rule will indeed place heavier emphasis on accurate prediction for large than for small firms. These weights will not necessarily be identical to the previous ones. Therefore, it becomes imperative to understand when there is an implicit trade-off of this kind.

The answer lies in the correctness of the specification of the prediction rule. If the best prediction rule is actually a linear function of the descriptors, then the weights that achieve statistical efficiency will also maximize predictive accuracy for all classes of securities. In other words, by setting the weights to achieve homoscedasticity we simultaneously obtain the best available predictors for all classes of securities. This can be seen in the above figure 3a, where the problem in attaining an accurate prediction rule is to estimate the linear relationship. The better that is estimated, the more accurate will be the prediction rule for all classes.

On the other hand, if the best prediction rule is not a linear function of the descriptors, but instead some other functional form, there will be a trade-off in accuracy between classes that lie at different points on the nonlinear relationship. This phenomenon is illustrated in figure 3b, where a nonlinear relationship between beta and size is drawn. There it is seen that fitting any form of linear relationship will necessarily introduce errors for some class of securities. If all data points are weighted equally, the fitted line drawn in crosses will be obtained: this will entail, among other things, substantial underprediction for very large firms. On the other hand, if the large firms are given the dominant weight, the dashed line will be obtained, and there will be accurate prediction for large firms but severe underprediction for small firms. Thus a trade-off in accuracy exists.

The moral of this discussion is quite simple: the correct solution is to identify, as part of the statistical analysis, the actual functional form of the relationship—linear or nonlinear. Once this is accomplished, the true functional form can be estimated (either by transformed descriptors that are nonlinear functions of the original descriptors, by dummy variables that fit a step function, or by nonlinear regression), the conflict between accuracy in different sectors disappears while accuracy in all sectors increases, and the appropriate weights are once more those that achieve homoscedasticity. This solution minimizes mean square prediction error. Moreover, when the functional
form is correctly specified, fundamental betas are unbiased relative to all captured characteristics: that is, prediction error is uncorrelated with these captured characteristics, regardless of the weights used in defining the bias. On the other hand, these fundamental betas may yet be biased relative to any uncaptured characteristics, to the extent that these uncaptured characteristics are uncorrelated with the captured ones.

III.c. Estimation of the Underlying Variances that Determine the Bayesian Adjustment

The regression approach directly estimated the ratio

$$\lambda^* = k \frac{V_\beta}{V_\beta + \overline{V}}.$$ 

The alternative approach is to estimate the three parameters $k$, $V_\beta$, and $\overline{V}$ and compute their ratio. In practice, those who have used this approach have ignored $k$, implicitly assuming that it was equal to one, that is, that there was no tendency to converge toward the norm. This rather dangerous simplification leaves only two parameters to be estimated: $V_\beta$ the cross-sectional variance of the true betas, and $\overline{V}$ the average estimation error variance in the historical beta estimates.

The procedure to estimate these is quite straightforward. First, the cross-sectional distribution of the historical estimates $\beta^h_n$ is observed, and the variance is computed. Each estimate is the sum of the true underlying historical beta, $\beta^h_n$, and the estimation error, $\beta^*_n - \beta^h_n$. Since the underlying value is unobservable, there is no direct way of identifying these two components. However, the regression estimator is such that $\beta^*_n - \beta^h_n$ is independent of $\beta^h_n$. Therefore

$$E(\beta^*_n - \overline{\beta})^2 = E(\beta^*_n - \beta^h_n + \beta^h_n - \overline{\beta})^2 = E(\beta^*_n - \beta^h_n)^2 + (\beta^h_n - \overline{\beta})^2 = V_n + (\beta^h_n - \overline{\beta})^2.$$ 

Taking the weighted sum across the securities,

$$\sum_n a_n E(\beta^*_n - \overline{\beta})^2 = \sum_n a_n V_n + \sum_n a_n (\beta^h_n - \overline{\beta})^2$$

or

$$\text{VAR}(\beta^*_n) = \overline{V} + V_\beta.$$ 

Thus the variance of the estimates $\beta^*_n$ is equal to the sum of the variance of the underlying betas, and the average variance of measurement error. In other words, the distribution of these estimates
will be more widely spread than the underlying distribution as a result of the addition of these errors to the spread of the underlying distribution.

Now, each regression estimator carries along with it an estimator of the variance of estimation error in $\beta$. This estimated variance may be denoted by $HV_n$. The weighted average of these estimators is an estimator of $\overline{HV}$:

$$\overline{HV} = \sum_n a_n HV_n.$$

This is an estimator of one of the parameters that we need. Moreover, by subtracting this estimator from the variance of the distribution of estimates, $VAR(\beta)$, we obtain an estimator of the variance on the distribution of the underlying betas, $\nu_\beta$. In other words, the estimator $\overline{HV}$ tells us how much of the spread in the distribution of the $\beta$'s which we see, is due to measurement error. The remainder must be due to the spread of the underlying betas, that is, to $\nu_\beta$.

Notice that everything in this approach hinges on the estimator $HV$. If the individual estimators $HV_n$ are unbiased, then their average value should be an excellent estimator of $\nu$, so that a good estimate of the accuracy of the technical predictor is obtained. Moreover, by subtracting this estimate from the variance of observed betas, a good estimate of $\nu_\beta$, the accuracy of the null predictor is obtained. Hence, the choice of $\lambda^*$ will be a good one. On the other hand, if the accuracy of $\overline{HV}$ is poor, the choice of $\lambda^*$ will be poor. In the next section, a case study will be used to illustrate this potential problem.

III.d. A Case Study in Bayesian Adjustment

The following description outlines a procedure that is currently used in investment management.

1. For each stock $n$ in a universe of $N$ available stocks thirteen-week cumulative returns are computed. These are returns computed over thirteen-week intervals. The first such thirteen-week return is computed beginning at some date $t_0$, another is computed beginning a week later, and so on. In this way returns are computed over thirteen-week moving periods that overlap one another, so that
the events in any week will appear in thirteen different moving returns. With five years of data, 248
data points would be obtained. Moving thirteen-week returns are then computed for the market index
over matching periods. The technical estimate of beta, the raw beta, \( \hat{\beta}_n \), is then defined as the re-
gression of these security returns on thirteen-week moving market returns. Using the conventional
formula for least squares regression, an estimator \( \hat{\nu}_n \) of the variance of the estimation error in
this beta estimator is also calculated.

**Discussion:** This method of computing the technical beta is viewed as a means of dealing with mismatch-
ing between the timing of security returns and market returns. In other words, the calculation of re-
turns over longer time periods has the advantage of grouping within each period the responses of the
security to the market during that period. Whatever the advantages derived from this, it is unfortunately the case that the use of moving overlapping periods disastrously reduces the efficiency of the
regression estimator. The problem is that ordinary regression is designed for the case where residuals
are independent from one observation to another. In this procedure, on the other hand, there is extre-
me high correlation between the residuals in successive observations, because the same week’s
data and hence the specific events for each week appear in thirteen successive observations. As a
result, assuming the random walk hypothesis, this procedure can be shown to have an error variance
that is approximately \( 8.5 \text{ times}^{15} \) as great as for a conventional regression using 260 weekly obser-
vations of nonoverlapping observations over the same five-year period. Thus an eightfold increase in
variance is the price paid for the dubious value of overcoming the mistiming problem.

However, in estimating the error variance by the standard formula that presumes the independ-
dence of residuals which is absent in this case, a systematic error is made. In fact the estimator
\( \hat{\nu}_n \) of the error variance will be, on average, equal to less than one-eighth\(^{16} \) of the correct value.
Consequently, the increase in estimation error variance due to the overlapping periods is concealed
by the inappropriate use of a formula that assumes away this problem. Presumably, this explains the
previous failure to detect the inefficiency of this estimator.

(2) The **equal-weighted** mean and variance of the distribution of estimated beta values are com-
puted, \( \mu_\beta^* \) and \( \text{VAR}(\hat{\beta}) \) respectively. Then, the variance of the underlying population of betas is
estimated by subtracting from the variance of the estimated betas that part which is thought to be due
to estimation error, namely

\[
\hat{\nu}_\beta = \text{VAR}(\hat{\beta}) - \frac{\sum_{n=1}^{N} \hat{\nu}_n}{N}.
\]

**Discussion:** The primary error here is that the correction being subtracted for the estimation error
is downwards biased by a factor of eight, so that the implied variance, \( \hat{\nu}_\beta \), for the population of true
betas will be much larger than the true $\beta^*$. A little discussion on this point will be helpful. As a result of the use of inefficient estimators of the betas, the dispersion of the distribution of estimated betas will be wider than for true betas. If an unbiased estimator of the estimation error variance for these estimated betas were used, this increased dispersion would be correctly attributed to the poor estimators, and an unbiased estimator of the variance of the true betas would be obtained. However, since the estimation error is understated, the variance of the underlying betas is overestimated. This erroneous conclusion is important, because it will determine the weight that should be given to the null predictor.

(3) For each security, the optimal Bayesian adjustment is computed as

$$\hat{\beta}_n = \mu_\beta + \left( \frac{\hat{\beta}_n}{\hat{\beta}_n + \hat{\nu}_n} \right) (\hat{\beta}_n - \mu_\beta).$$

**Discussion:** There are two major problems in this predictor. First, as a result of the previous errors in computing the technical estimate of $\beta$, the weight given the technical beta is entirely inappropriate. Since the error variance, $\hat{\nu}_n$, for the technical beta is underestimated, the technical beta will be weighted excessively; moreover, since the variance of the underlying distribution of actual betas is consequently overestimated leading to an erroneous down-weighting of the null predictor, the technical estimator is given a further excessive weight. Finally, the method presumes that there is no beta convergence toward the norm, i.e., that $k=1$, and therefore omits the appropriate down-weighting of the historical estimator to reflect this phenomenon. The eventual Bayesian adjustment is less than one-tenth of its optimal size.

The second major problem is that the mean value is taken as the equal weighted mean, with all of the stocks in a large universe being included in the sample. As a result, $\mu_\beta$ is greater than 1.5, because the numerous stocks will small capitalizations and typically high betas are weighted equally with the small number of large companies that typically have low betas. This is entirely the wrong set of weights for the typical institutional client, and results in some nonsensical implications. Suppose that the beta of a stock with a large capitalization is estimated to be 1.4. Normally, the actual beta would be considered to be lower and this estimate would be adjusted downwards. In this case, however, the estimate is adjusted upwards. In short, the prediction rule is strongly biased relative to capitalization, and will tend to predict excessively high betas for large capitalization companies and excessively low betas for small capitalization companies. The effects of this problem are, however, minimized by the previous errors, which tend to wipe out the Bayesian adjustment, so that the use of an appropriate norm becomes a moot point.
It is instructive to note that these errors would have shown up instantly in a regression of the form

\[ r_{nt} - r_{ft} = b_0 (r_{Mt} - r_{ft}) + b_1 u_{nt} (r_{Mt} - r_{ft}). \]

The regression approach computes the optimal Bayesian adjustment directly, whereas the series of steps carried out in the application under study only lead to the optimal adjustment if the estimators \( \hat{\beta}_n \) are unbiased, and they will fail to be unbiased if the regression equation is misspecified. In statistical terms, the regression approach is "robust" against misspecification of the historical beta regressions, but the approach using the estimators \( \hat{\beta}_n \) completely fails in robustness against this potential problem.

III.e. Bayesian Adjustment for the Prediction of Portfolio Risk

This concluding section reexamines the appropriately adjusted beta predictors from the standpoint of their use in portfolio beta prediction. Ideally, the optimal predictor for portfolio beta should equal the weighted average of the optimal predictors for constituent securities, with the weights being the investment proportions. When the Bayesian adjusted predictors can be aggregated in this way to obtain minimum MSE prediction for portfolio beta, the "optimal aggregation property" may be said to hold. This aggregation property does hold approximately, but only if the Bayesian adjusted prediction rule is unbiased with respect to the information used to select the portfolio. For this reason, the simple Bayesian adjustment developed thus far provides a terrible predictor of portfolio risk, because it is severely biased with respect to any fundamental and industry characteristics that are correlated with beta.

For simplicity, suppose that there is a portfolio of \( J \) securities with an equal amount invested in each security. Suppose further that the predicted beta for the \( j \)th security is \( \hat{\beta}_j \) and that the aggregated prediction of risk for the portfolio based on these individual security predictors is \( \hat{\beta}_p = 1/J \sum_j \hat{\beta}_j \). In terms of the prediction error on individual securities, the error in predicting the portfolio is \( \sum_j (\hat{\beta}_j - \beta_j) \). Therefore we find that the mean square error in prediction is

\[
\text{MSE}(\hat{\beta}_p) = \text{BIAS}(\hat{\beta}_p)^2 + \text{VAR}(\hat{\beta}_p - \beta_p)
\]

\[
= \left( \frac{1}{J} \sum_j \text{BIAS}(\hat{\beta}_j) \right)^2 + \sum_j \frac{1}{J^2} \text{VAR}(\hat{\beta}_j - \beta_j)
\]

\[+ 2 \sum_{j < k} \frac{1}{J^2} \text{COV}(\hat{\beta}_j - \beta_j, \hat{\beta}_k - \beta_k).\]
In discussing the magnitude of the mean square error, we are handicapped by the absence of information on the covariances of errors in predicting individual beta. Although these covariances may well be nonzero because of the contribution of common factors to the returns on many securities, we will assume that they are zero because this will simplify the exposition and will not change the qualitative conclusions of this section.

\[ \text{MSE}[\hat{\beta}_j] = \left( \frac{1}{j} \text{BIAS}[\hat{\beta}_j] \right)^2 + \frac{1}{j^2} \text{VAR}[\hat{\beta}_j] \]

\[ = \left( \text{AVERAGE BIAS} \right)^2 + \frac{\text{AVERAGE VARIANCE}}{j} \]

Just by examining this expression, it is apparent that the variance in estimating the individual security risk disappears, as \( J \) increases, because these random errors average out. Will the bias in predicting these individual betas average out as well? This depends on whether the rule used to put the securities in the portfolio is correlated with the bias. If it is not, then the average bias in the securities in the portfolio will be the same as the average bias in all securities in the universe and hence zero. If there is a correlation, however, the average bias will be nonzero by virtue of this correlation, and so no matter how large \( J \) is, the error from this bias will persist.

As a cogent illustration of this point, consider the optimal Bayesian adjustment for the portfolio beta prediction. As an additional simplification let us assume that the estimation error in variance in the technical estimator, \( \hat{\nu}_n \), is identical to \( \nu \) for all \( n \). In this case the variance in predicting \( \beta_n \) for any Bayesian prediction rule will also be the same for all firms. Let \( \hat{\beta}_P \) be the technical historical beta estimator, given by \( \hat{\beta}_P = 1/J \sum_j \hat{\beta}_j \). The variance of this estimator is \( \nu_P = \nu/J \). Let \( \nu_{BP} \) be the variance of the cross-sectional distribution of portfolio betas across the institutional investment managers where the prediction rule is to be applied. Then, by following exactly the same argument as was presented for individual securities in III.a, but this time applied to the population of portfolios, the minimum mean square error predictor for the typical portfolio is

\[ \beta_P^* = (1-\lambda_P^*)\hat{\beta}_P + \lambda_P^* \beta_{BP} \]

\[ \lambda_P^* = \frac{\nu_{BP}}{\nu_{BP} + \nu_P} \]

We next proceed to compare this predictor with the predictors for individual securities. Suppose first that institutional portfolios are drawn at random, by throwing ill-sighted darts at the security list. Then the variance of the beta of a portfolio containing \( J \) stocks will be just \( 1/J \)
times the variance of the beta of individual stocks, because the portfolio beta is just the average of  
J randomly drawn betas:

\[ V_{BP} = \text{VAR} \left( \frac{\sum_{j=1}^{J} \beta_{j}}{J} \right) = \frac{1}{J} \sum_{j=1}^{J} \text{VAR} (\beta_{j}) = \frac{1}{J} \sum_{j=1}^{J} J V_{\beta} = \frac{V_{\beta}}{J}. \]

When this fact is substituted into the formula for \( \lambda^* \), it is found to be equal to the previously derived \( \lambda^* \):

\[ \lambda^*_p = k \frac{V_{BP}}{V_{BP} + V_p} = k \frac{V_{\beta}/J}{V_{\beta}/J + V_p/J} = k \frac{V_{\beta}}{V_{\beta} + V_p} = \lambda^*. \]

In other words, if institutional portfolios are randomly drawn from the population of stocks, then the optimally Bayesian adjusted predictors for individual securities can be aggregated to obtain an optimal Bayesian adjusted predictor of the portfolio beta:

\[ \hat{\beta}_p = (1-\lambda^*_p) \bar{\beta} + \lambda^*_p \hat{\beta}_P = (1-\lambda^*_p) \bar{\beta} + \lambda^*_p \hat{\beta}_P = \frac{1}{J} \sum_{j=1}^{J} \left[ (1-\lambda^*_p) \bar{\beta}_j + \lambda^*_p \hat{\beta}_j \right] = \frac{1}{J} \sum_{j=1}^{J} \hat{\beta}_j. \]

There is, of course, a catch in this derivation. The derived Bayesian adjustment is correct only because the portfolios were drawn at random, and as a consequence their true beta variance is tiny.

For example, in the numerical example already developed in III.a, portfolios of fifty stocks would exhibit a variance of \( V_{BP} = V_{\beta}/50 = 0.16/50 = 0.003 \), or a standard deviation of true portfolio betas of only 0.055, implying that only one portfolio in forty would have a beta as great as 1.11 or as low as 0.89. For institutions with portfolios of 100 stocks the corresponding range is from 0.94 to 1.06.

For any portfolio constructed so as to have a beta different from one, or selected by other criteria that are correlated with beta, it is highly probable that the actual beta will be well beyond this range. Table 5 shows the bias and mean square error of the above predictor for portfolio betas of .85, 1.0, and 1.3. Note that for a portfolio with beta of .85, the Bayesian adjusted predictor is seriously inferior to use of the historical beta, with a bias of .06 and a mean square error that is much larger. For a beta of 1.3, such as is quite commonly encountered in institutional portfolios, the Bayesian adjusted predictor exhibits five times the mean square error of the historical beta, all due to the bias.

The explanation for the terrible performance of the Bayesian adjusted predictor is that it presumes that the expected bias in the prediction of beta for stocks in the portfolio is zero.
TABLE 5

<table>
<thead>
<tr>
<th>Case 1</th>
<th>$\beta_p$</th>
<th>$E(\tilde{\beta}_p)$</th>
<th>$\text{BLAS}(\tilde{\beta}_p)$</th>
<th>$\text{MSE}(\tilde{\beta}_p)$</th>
<th>$\text{VAR}(R^2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.30</td>
<td>1.1650</td>
<td>0.1350</td>
<td>0.019435</td>
<td>0.004</td>
</tr>
<tr>
<td>Case 2</td>
<td>1.00</td>
<td>1.0000</td>
<td>0.0000</td>
<td>0.001210</td>
<td>0.004</td>
</tr>
<tr>
<td>Case 3</td>
<td>0.85</td>
<td>0.9175</td>
<td>0.0675</td>
<td>0.005766</td>
<td>0.004</td>
</tr>
</tbody>
</table>

$\lambda = 0.55$; $\bar{v} = 1.0$; $\overline{\nu} = 0.004$. 

![Graph showing probability distributions for different cases]
Therefore if the criterion for selecting stocks for the portfolio is correlated with beta, and there-
fore negatively correlated with this bias, the bias in the predictor will be much greater than ex-
pected and will swamp the tiny reduction in variance obtained by reducing the contribution of the
estimation error variance in Hβ.

However, interestingly enough, the optimal aggregation property does hold, provided that the
characteristic that is employed to select the portfolio is included in the fitted prediction rule.
For example, suppose that an institution constructs a high beta portfolio by picking stocks that are
high in Hβ. The stocks that pass this criterion will tend to have higher true β_n^h and also higher
estimation errors Hβ - β_n^h. It turns out that, as a result, there will be a bias in the average esti-
mation error Hβ - β_n^h, that offsets the bias from the conservative Bayesian adjustment (1-λ)(β_n^h-β),
so that the Bayesian adjusted predictor remains optimal for the portfolio.

Thus, if a fundamental prediction rule incorporates the characteristics used to select the
portfolio, so that it is unbiased with respect to these characteristics, then the fitted optimal pre-
diction rule for individual securities will also be the optimal prediction rule for portfolios. This
fact underscores once more the crucial requirement that whatever information is used for selection be
captured in the prediction rule.

IV. SUMMARY

The systematic risk (or beta) of a security arises from the dependence of the security return,
on the one hand, and the market return, on the other, on economic events. Beta is a measure of this
joint dependence that is determined by the relative responses of security return and market return to
economic events and by the variance of those events. As such, beta will change if the fundamental
characteristics of the firm change, thereby altering responsiveness relative to economic events, or if
the variance in the economy is redistributed.

The estimates of beta that are now commonly available rely on historical security price behav-
ior. Such estimates are appropriate as estimators of risk assumed in the past for purposes of measur-
ing past performance, but they are not necessarily appropriate for portfolio management, where a pre-
diction of risk in the planning period is required.

For a prediction rule to result in a portfolio-management process that achieves effective con-
trol of beta, two criteria should be met. First, the prediction rule should be unbiased relative to
any characteristics of securities that may be used to select portfolios; thus, the average expected
error for any group of securities likely to show up in a portfolio should be zero, with the average being weighted by capitalization. Second, the mean square prediction error for each stock, and on the average across all stocks, should be minimized. These criteria are not drawn arbitrarily from a statistics textbook but are deduced from exigencies of portfolio management.

The historical estimate of beta is suboptimal on both these counts. Fundamental data provide more timely and accurate information concerning the responsiveness of the security return to economic events. When the response coefficients of the firm, due to fundamental variables, are forecast, but the variance of economic events is assumed to remain constant, the result is an unconditional forecast of risk. When the variance of economic events is forecast as well, a conditional forecast of risk is obtained.

The relationship between future risk and currently available data may be explored by regression. When the regression is properly specified, the outcome is an unconditional prediction rule that optimally exploits current information, in the sense of satisfying the above criteria to the highest possible degree. A previous study, employing this approach, demonstrated that fundamental information provided superior predictive performance when compared with the traditional historical estimate or the recently introduced "Bayesian adjustments" of that estimate.

It is natural to combine the fundamental and technical approaches. The strength of the technical approach is that it directly estimates the past value of the item of interest, namely, beta. The strength of fundamental prediction is that it responds to the current condition of the company and exhibits less measurement error. Fortunately, the two techniques are easily coalesced in a single prediction rule by including the historical beta estimate as information in the regression analysis. In the resulting prediction rule, fundamental and technical data are optimally weighted according to their predictive potential.

As an illustration of this concept, the simplest "Bayesian adjusted" prediction rule, now in extensive use among professional investors, is considered in detail in the last section of the paper. It is seen to be a special case of the combination of fundamental and technical information. In fact, it is the special case that uses the minimum of fundamental information. It is shown to provide a suboptimal prediction of individual security risk, and a very poor prediction of portfolio risk, precisely because it fails to reflect other fundamental information that may be used in portfolio construction. This omission leads to bias with respect to that information and to greater mean square error. Two alternative methods of estimating the Bayesian adjustment are developed: the regression approach mentioned previously, and the alternative of estimating the historical sources of error variance. The dangers of the latter approach are highlighted by a case study of an ill-conceived--
although commercially successful—adjustment procedure.

In conclusion, this paper has attempted to place the prediction of beta in a sound economic perspective and to show that a methodology for improved prediction of risk, employing fundamental information, exists.

---

FOOTNOTES

1 In statistical terms, $\beta_{nt} = \frac{\text{COV}(r_{nt}, r_{Mt})}{\text{VAR}(r_{Mt})}$ is the regression coefficient of security return on market return. It gives the slope of the line relating the expected value for $r_{nt}$ to $r_{Mt}$. If the intercept of the line is zero, the expected value of $r_{nt}/r_{Mt}$ will equal $\beta_{nt}$, and the statement in the text will be strictly true. If we rephrase the example given in the text, so that the security return and market return are given as returns in excess of the return on the riskless asset, then the basic Sharpe-Lintner-Mossin capital asset pricing model implies that the intercept term will indeed be zero. Otherwise, the statement must be rephrased to read $\beta_{nt} = \frac{E}{r_n - r_{nt}}$, where $E_{nt}$ is the intercept of the relationship.

2 The variance of the market due to energy-related events is $1/3(6^2+0^2+6^2) = 24$, and due to inflation-related events it is $1/3(3^2+0^2+3^2) = 6$. Consequently, the total variance of the market is 30, and 24/30 or 4/5 of this is explained by energy-related events.

3 A formal proof of this equation is given as follows: Let the $j$th factor influencing the market be denoted by $f_{jt}$, with $r_{Mt} = \sum_{j} f_{jt}$ and $\text{VAR}(f_{jt}) = \bar{V}_{jt}$, $\text{COV}(f_{jt}, f_{ik}) = 0$ for $i \neq j$. (Without loss of generality, the factors are standardized so that the market response coefficient is 1.)

Then $r_{nt} = \sum_{j} \gamma_{nt} f_{jt} + u_{nt}$, where $\gamma_{nt}$ is the response to factor $j$, and $u_{nt}$ is the specific component of return for security $n$, independent of the factors. Therefore,

$$\beta_{nt} = \frac{\text{COV}(r_{nt}, r_{Mt})}{\text{VAR}(r_{Mt})}$$

$$= \frac{\text{COV}(\sum_{j} \gamma_{nt} f_{jt} + u_{nt}, \sum_{j} f_{jt})}{\text{VAR}(\sum_{j} f_{jt})}$$

$$= \frac{\sum_{j} \gamma_{nt} \bar{V}_{jt} + \sum_{i} \sum_{j \neq i} 0}{\bar{V}_{jt} + \sum_{j} \sum_{i \neq j} 0}.$$

This equation can be viewed in another light. $\gamma_{nt}$ 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.
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 naïve belief that we know with certainty whether the market will rise or fall. We can never be so certain. To reduce the exposure of 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.

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.

The derivation of this formula is simple:

\[
\text{MSE}\left(\hat{\beta}_n\right) = E[\hat{\beta}_n - \beta_n]^2 = E[\hat{\beta}_n - E(\hat{\beta}_n)] + E(\hat{\beta}_n) - \beta_n]^2.
\]

\[
= E[\hat{\beta}_n - \beta_n + \beta_n - \beta_n]^2
\]

\[
= E[\hat{\beta}_n - \beta_n]^2 + E[\beta_n - \beta_n]^2 + 2E[\hat{\beta}_n - \beta_n][\hat{\beta}_n - \beta_n].
\]

Now, \(E[\hat{\beta}_n - \beta_n]^2 = \text{VAR}(\hat{\beta}_n)\), by definition

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

and

\[E[\hat{\beta}_n - \beta_n][\hat{\beta}_n - \beta_n] = [\hat{\beta}_n - \beta_n]E[\hat{\beta}_n - \beta_n]
\]

\[= [\hat{\beta}_n - \beta_n]E[\hat{\beta}_n - \beta_n] = 0.\]
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:

\[
\text{VAR}(r_n) = \beta_n^2 \text{VAR}(r_M) + \text{VAR}(u_n),
\]

where \( \text{VAR}(u_n) \) 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:

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

and

\[
\text{VAR}(r_p) = \sum_{n=1}^{N} W_n^2 \beta_n^2 \text{VAR}(r_M) + \sum_{n=1}^{N} W_n^2 \text{VAR}(u_n).
\]

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

\[
\text{VAR}(r_p) = \text{VAR}(r_M) \sum_{n=1}^{N} W_n^2 \beta_n^2 \approx \beta_p^2 \text{VAR}(r_M).
\]

Also,

\[
\beta_p = \frac{\text{COV}(r_p, r_M)}{\text{VAR}(r_M)} = \frac{\text{COV} \left( \sum_{n=1}^{N} W_n r_n, r_M \right)}{\text{VAR}(r_M)} = \frac{\sum_{n=1}^{N} W_n \text{COV}(r_n, r_M)}{\text{VAR}(r_M)} = \sum_{n=1}^{N} W_n \beta_n.
\]

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.

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 \( \hat{\theta}_a \) is proportional to the capitalization of that asset.

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 \( \hat{\mu} \) and variance \( \hat{\sigma} \) of the return to the underlying common stock, has the properties of the integral

\[
\hat{V} = \int_{x_0}^{\infty} x \exp\left(-\frac{1}{2} \frac{(x - \hat{\mu})^2}{\hat{\sigma}}\right) dx.
\]

The integral is a nonlinear function of \( \frac{\hat{\mu}}{\hat{\sigma}} \) and \( \hat{\sigma} \), 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 \( \hat{\sigma} \). Indeed, the criterion can only be written in the form of an integral equation.

The form of the regression equation that is commonly used is as follows:

\[
R_{nt} = \hat{\alpha} + \hat{\beta}_n R_{Mt} + u_t \quad \text{where} \quad E(u_t) = 0 \]

\[
E(u_t u_s) = \sigma_u^2 \quad t = s
\]

\[
= 0 \quad t \neq s.
\]

If one believes in the market model of Sharpe, Lintner, and Mossin, then one would also believe that the above equation was misspecified. In this case, the following regression would be run:

\[
[R_{nt} - R_{mt}] = \hat{\alpha}_n + \hat{\beta}_n [R_{Mt} - R_{Mt}] + u_t \quad \text{where} \quad E(u_t) = 0 \]

\[
E(u_t u_s) = \sigma_u^2 \quad t = s
\]

\[
= 0 \quad t \neq s.
\]

The minimum variance unbiased estimates of the historical average beta would be slightly but not significantly different. In the first method, beta would equal \( \text{COV}(R_{Mt}, R_{nt}) / \text{VAR}(R_{Mt}) \), while in the second method, it would equal \( \text{COV}(R_{Mt} - R_{Mt}, R_{nt} - R_{Mt}) / \text{VAR}(R_{Mt} - R_{Mt}) \).

An optimal prediction rule of this kind is formulated as follows. Let \( P_0(\beta, \sigma^2) \) be a prior distribution for the two risk parameters of the firm, equal to the cross-sectional distribution of these parameters in the population of firms. Let \( \hat{\beta}_n, \hat{\sigma}_n^2 \) denote estimators for historical averages of these two parameters for security \( n \), obtained from the historical regression. Then we obtain a posterior distribution incorporating the prior distribution and the information from the regression: \( P_1(\beta_n, \sigma_n^2) \). The mean value of \( \hat{\beta}_n \) from this posterior distribution is the optimal predictor. The
problem takes a relatively simple form when $\beta$ and $\sigma^2$ are independent in the prior distribution $P_0$ (c.f., Arnold Zellner, An Introduction to Bayesian Inference in Econometrics, John Wiley & Sons, New York, 1971, Chapter 4, Section 4.2, and Appendix 2). The solution relying on this simplification is taken, for example, in Vasicek (1973) and in Part III.d, below. However, in actuality, $\beta$ and $\sigma^2$ are highly correlated under $P_0$, so that the simplistic assumption of prior independence is damaging; it leads to undesirably large reductions in predicted beta for high beta firms. Discussion of the fully general approach would take us too far afield and will not be presented here.

13 According to the Sharpe,Lintner, and Mossin model:

$$E[r_{nt} | r_{Mt}] = r_{Fe} (1 - \beta) + \beta r_{Mt}.$$  

Consequently, by rearrangement we can derive:

$$\beta = \frac{E[r_{nt} - r_{Fe}]}{E[r_{Mt} - r_{Fe}]} r_{Mt}.$$  

14 In this footnote, the fact that the regression derives the Bayesian adjustment will be explained. To simplify the exposition, it is assumed that there is only a single period, so that the time subscript "t" can be omitted, and that $r_{Fe}$ is zero. Then the model becomes:

$$r_n = \beta r_M + \epsilon_n.$$  

The regression to be fitted is $r_n = b_0 r_M + b_1 H \beta_n r_M$. Since $r_M$ is a constant, we can divide through both sides to obtain the equivalent regression,

$$\frac{r_n}{r_M} = b_0 + b_1 H \beta_n.$$  

This form allows an easy derivation, since it is simple regression with a constant term and a single explanatory variable. (A similar derivation, relying on matrix algebra, is presented for the fully general case in section III of Barr Rosenberg and Vinay Marathe, "Tests of Capital Asset Pricing Hypotheses," IBER Working Paper in Finance, No. 32, 1975.)

From the familiar formula for linear regression, the estimated slope coefficient is:

$$b_1 = \frac{\text{COV}(r_n, r_M)}{\text{VAR}(H \beta_n)}.$$  

where, since this is a capitalization-weighted regression, the covariance and variance are themselves
capitalization-weighted. Thus, to evaluate the estimator, we need to evaluate \( \text{COV}(\tau_n / \tau_M, \hat{H}_n) \) and \( \text{VAR}(\hat{H}_n) \).

Now, since the residual return on stock \( n \) is uncorrelated with beta,

\[
\text{COV}(\tau_n / \tau_M, \hat{H}_n) = \text{COV}(\beta_n \tau_n / \tau_M, \hat{H}_n) = \text{COV}(\beta_n, \hat{H}_n).
\]

Thus, as we desire, the regression of \( \tau_n / \tau_M \) on \( \hat{H}_n \) accomplishes the same purpose as a regression
of \( \beta_n \) on \( \hat{H}_n \). Next, substituting the model for \( \beta_n \) and writing the \( \hat{H}_n \) as the sum of the true
historical beta and the estimation error, we obtain:

\[
\text{COV}(\beta_n, \hat{H}_n) = \text{COV}(1 - k\overline{\beta} + k\beta_n, \epsilon_n, \beta_n + (\hat{H}_n - \beta_n^h)).
\]

Since \( \overline{\beta} \) is a constant, it makes no contribution to the covariance. Since \( \epsilon_n \) is the unpredictable
shift in beta, it is uncorrelated with the historical terms and makes no contribution to the covar-
iance. Finally, the estimation error \( (\hat{H}_n - \beta_n^h) \) is uncorrelated with the true historical value \( \beta_n^h \),
so it makes no contribution to the covariance. Thus, we have:

\[
\text{COV}(\beta_n, \hat{H}_n) = \text{COV}(k\beta_n^h, \beta_n^h) = kV_{\beta}.
\]

Next, to evaluate the denominator of \( b_1 \), notice that \( \hat{H}_n \) is the sum of the true \( \beta_n^h \) and the esti-
mation error \( (\hat{H}_n - \beta_n^h) \). Since these two terms are uncorrelated with one another,

\[
\text{VAR}(\hat{H}_n) = \text{VAR}(\beta_n^h) + \text{VAR}(\hat{H}_n - \beta_n^h) = V_{\beta} + \overline{\nu}.
\]

Therefore, we have indeed that:

\[
b_1 = \frac{\text{COV}(\beta_n, \hat{H}_n)}{\text{VAR}(\hat{H}_n)} = k \frac{V_{\beta}}{V_{\beta} + \overline{\nu}} = \lambda^*.
\]

With this derivation completed, it is an easy matter to verify that:

\[
b_0 = \text{MEAN}(\tau_n / \tau_M) - b_1 \text{MEAN}(\hat{H}_n) = 1 - b_1 = 1 - \lambda^*.
\]

Thus, the regression estimates the Bayesian adjustment directly. In practice, the regression is
run using information from many periods in the modified form obtained by multiplying through by \( \tau_M \)
on both sides,
The loss of efficiency is most easily demonstrated under the following assumptions: Let
\[ z_{nt} = \ln \left( \frac{r_{nt}}{r_{pt}} + 1 \right) \]
denote the excess logarithmic return on stock \( n \) in week \( t \), and let \( z_{Mt} \) denote the excess logarithmic market return, \( \ln \left( \frac{r_{Mt}}{r_{pt}} + 1 \right) \). The model for return is
\[ z_{nt} = \alpha + \beta z_{Mt} + u_n. \]
Ignoring possible mistiming in the security response, the random-walk hypothesis implies that
\[ \text{COV}(z_{Ms}, z_{Mt}) = 0 = \text{COV}(u_s, u_t) \]
for \( s \neq t \); and, also, since \( u \) is a residual return, \( \text{COV}(u_s, z_{Mt}) = 0 \)
for all \( s, t \). Let \( \sigma^2 = \text{VAR}(u_s) \) for all \( s \), and let \( \sigma_M^2 = \text{VAR}(z_{Mt}) \) for all \( t \).

The ordinary least squares regression then implies
\[ \hat{\beta} = \left( \frac{T}{\sum_{t=1}^{T} \hat{z}_{Mt}^2} \frac{T}{\sum_{t=1}^{T} \hat{z}_{Mt}^2} \right)^{-1} \left( \frac{T}{\sum_{t=1}^{T} \hat{z}_{Mt}^2} \right), \]
where the \( \hat{\cdot} \) indicates the difference from the sample mean. For a sample of as many as 260 weeks, there is a negligible error from assuming that the sample mean equals the population mean, and this will be done below to simplify the derivation. The variance of estimation error in this equation is
\[ \text{VAR}(\hat{\beta} - \beta) = \frac{\sigma^2}{\left( \frac{T}{\sum_{t=1}^{T} \hat{z}_{Mt}^2} \right)} \]
For large \( T \), the typical value is equal to the probability limit
\[ \frac{\sigma^2}{\left( \frac{T}{\sum_{t=1}^{T} \hat{z}_{Mt}^2} \right)} = \frac{\sigma^2}{(T \sigma_M^2)}. \]

By contrast, the regression on overlapping cumulative thirteen-week logarithmic returns is given by:
\[ \hat{\beta} = \frac{T-12}{\sum_{t=1}^{T-12} \left( \sum_{s=t}^{t+12} \hat{z}_{Ms} \right) \left( \sum_{s=t}^{t+12} \hat{z}_{Ms} \right)^2}, \]

After the model for returns is substituted into the numerator, we obtain:
\[ \hat{\beta} = \beta + \frac{T-12}{\sum_{t=1}^{T-12} \left( \sum_{s=t}^{t+12} \hat{z}_{Ms} \right) \left( \sum_{s=t}^{t+12} \hat{z}_{Ms} \right)^2} \]

The second term constitutes the estimation error. Its mean value is zero, so the estimator is unbiased. The estimation error variance is just the expectation of the square of this term. To evaluate this, the first step is to rearrange terms to obtain:
\[ \bar{b} - b = \frac{T}{t=1} u_t s_t (\hat{z}_M) \left( \frac{\sum_{t=1}^{T} n_t z_t^2}{T} \sum_{s=1}^{n_t} n_t s_t z_t M_t z_M s_t M_s z_M s_t M_s} \right), \]

where the newly-defined terms are as follows:

1. \( n_{tt} \) is the number of appearances of \( z_M^2 \) in the product in the denominator:

   \[
   n_{tt} = \begin{cases} 
   t & \text{for } t < 13 \\
   13 & \text{for } 13 \leq t \leq T - 12 \\
   T - t + 1 & \text{for } t > T - 12 
   \end{cases}
   \]

2. \( 2n_{cs} \) is the number of appearances of \( z_M z_s \) in the product in the denominator:

   \[
   n_{cs} = \begin{cases} 
   \min\{t,s\} & \text{if } t < 13, s < 13 \\
   \min\{T-t+1, T-s+1\} & \text{if } s > T - 12, t > T - 12 \\
   13 - |t-s| & \text{if } |t-s| < 13, \text{ and } \{13 < t < T - 12 \text{ or } 13 s < T - 12\} \\
   0 & \text{if } |t-s| \geq 13
   \end{cases}
   \]

3. \( s_{(z)M} = \frac{s = \max\{T, t+12\}}{s = \min\{1, t-12\}} n_{cs} z_M. \)

The expected value of the square of the denominator is:

\[
E \left[ \left( \sum_{t=1}^{T} u_t s_t (\hat{z}_M) \right)^2 \right] = \sigma^2 \sum_{t=1}^{T} E[\langle \hat{z}_M \rangle^2] \]

\[ = \sigma^2 \sum_{t=1}^{T} \sum_{s=1}^{n_t} n_{ts} a_M^2 \]

\[ = \sigma^2 \sum_{t=1}^{T} n_{tt} \sigma_M^2 + 2(6448). \]

The expected value of the square of the numerator is:

\[
E \left[ \left( \sum_{t=1}^{T} (n_t z_t^2 + a_{M,H}) \sum_{s=1}^{n_t} n_t s_t z_t M_t z_M M_s) \right)^2 \right] = \left( \sum_{t=1}^{T} n_t \sigma_M^2 \right)^2 + \sigma_M^4 (k_{H} - 1) \sum_{t=1}^{T} n_{tt}^2 + 4a_M^4 \sum_{t=1}^{T} \sum_{s=t+1}^{n_{ts}^2} \sigma_M^2. \]
\[ \sigma^4_M \left( 13(T-12) \right)^2 + (169(T-24) + 1300)(k_{M-1}) + 2600(T-24) + 18,928, \]

where \( k_M \) is the kurtosis of the distribution of market returns. A typical value for the squared ratio is equal to the ratio of the expected values of the squared numerator and the squared denominator. Substituting \( T = 260 \), this expression becomes:

\[ \frac{\sigma^2 \sigma^2_M(359,580)}{\sigma^2_M(10,552,308 + 10,184 (k_{M-1}))}. \]

Note that the kurtosis makes a minor contribution. Substituting a typical value of 12 for the kurtosis of logarithmic market returns, the ratio becomes:

\[ \frac{\sigma^2 \sigma^2_M(359,580)}{\sigma^2_M(11,087,729)} = \frac{\sigma^2}{30.84 \sigma^2_M}. \]

Comparing this with the typical variance of the ordinary least squares estimator, it is found to be roughly 8.5 times as great:

\[ \frac{(\sigma^2/30.84 \sigma^2_M)/(\sigma^2/260 \sigma^2_M)} = 8.43. \]

This derivation relied on several assumptions, but essentially the same conclusions will be reached if raw returns are substituted for logarithmic returns, or if a plausible amount of miscoding is incorporated, or if heteroscedasticity in \( u \) is introduced, or if nonstationarity in the parameters \( \alpha \) and \( \beta \) is allowed.

16 Under the erroneous assumption that the residuals in the regression are serially uncorrelated, the estimator for \( \text{VAR}(\hat{\beta} - \beta) \) is given by:

\[ \frac{1}{T-13} \sum_{t=1}^{T-12} \left( \sum_{\alpha=t}^{t+12} \frac{z_{m\alpha} - \hat{\beta} \sum_{\alpha=t}^{t+12} z_{M\alpha}}{T-12} \right)^2. \]

To a close approximation, the expected value of the numerator is \( 13\sigma^2 \), and the expected value of the denominator is \( (T-12)(13\sigma^2_M) \). Thus, a typical value for the ratio is \( \sigma^2/\sigma^2_M(T-12) \) or \( \sigma^2/248\sigma^2_M \) if \( T = 260 \).

17 This expression for the error in prediction for the portfolio beta presumes that the investment proportions will remain constant at the current values \( w_j \). In fact, as the securities
exhibit future returns, the investment proportions, and hence the portfolio beta, will change. However, the changes are random and are as likely to reduce as to increase the weight on each security. It may therefore be shown that, to an excellent approximation, the minimum mean square error prediction for the future portfolio beta, allowing for possible changes in investment proportions due to future security return, retains the current weights.

REFERENCES


Zellner, Arnold. An Introduction to Bayesian Inference in Econometrics.