Roy D. Henriksson and Robert C. Merton
Massachusetts Institute of Technology

On Market Timing and Investment Performance. II. Statistical Procedures for Evaluating Forecasting Skills*

The evaluation of the performance of investment managers is a much studied problem in finance. Based upon the model developed in Part I of this paper, the statistical framework is derived for both parametric and nonparametric tests of market-timing ability. If the manager's forecasts are observable, then the nonparametric test can be used without further assumptions about the distribution of security returns. If the manager's forecasts are not observable, then the parametric test can be used under the assumption of either a capital asset pricing model or a multifactor return structure. The tests differ from earlier work because they permit identification and separation of the gains of market-timing skills from the gains of micro stock-selection skills.

I. Introduction

In Merton (1981; hereafter referred to as Part I), one of us developed a basic model of market-timing forecasts where the forecaster predicts when stocks will outperform bonds and when bonds will outperform stocks but does not predict the magnitude of the superior performance. In that analysis, it was shown that the pattern of returns from successful market timing has an isomorphic correspondence to the pattern of returns from following certain option investment strategies where the implicit prices paid for the options are less than their "fair" or market values. This isomorphic correspondence was used to drive an equilibrium theory of value for market-timing forecasting skills. By analyzing how investors would use the market timer's forecast to modify their probability beliefs about stock returns, it was shown that the conditional probabilities of a correct forecast (conditional on the return on the market) provide both necessary and sufficient conditions for such forecasts to have a positive value.

In the analysis presented here, we use the

* Earlier versions of the paper were presented in seminars at Berkeley, Carnegie-Mellon, University of Chicago, Dartmouth, Harvard, University of Southern California, and Vanderbilt; we thank the participants for their comments. Aid from the National Science Foundation is gratefully acknowledged.

(*Journal of Business*, 1981, vol. 54, no. 4)

0021-9398/81/5404-0005$01.50

basic model of market timing derived in Part I to develop both parametric and nonparametric statistical procedures to test for superior forecasting skills.

The evaluation of the performance of investment managers is a topic of considerable interest to both practitioners and academics. To the former, such evaluations provide a useful aid for the efficient allocation of investment funds among managers. To the latter, significant evidence of superior forecasting skills would violate the Efficient Markets Hypothesis.[1] Such violations, if found, would have far-reaching implications for the theory of finance with respect to optimal portfolio holdings of investors, the equilibrium valuation of securities, and many decisions in corporate finance. With so much at stake, it is not surprising that much has been written on this subject. Indeed, a major application of modern capital market theory has been to provide a structural specification to measure investment performance. Within this structure, it is the practice to partition forecasting skills into two components (see Fama 1972): (1) "microforecasting," which forecasts price movements of individual stocks relative to stocks generally, and (2) "macroforecasting," which forecasts price movements of the general stock market relative to fixed income securities. The former is frequently called "security analysis" and the latter is referred to as "market timing." Moreover, this partitioning of forecasting skills takes on added significance through the work of Treynor and Black (1973), who have shown that investment managers can effectively separate actions related to security analysis from those related to market timing.

Most of the recent empirical studies of investment performance focus on microforecasting and are based on a mean-variance capital asset pricing model (CAPM) framework[2] where the 1-period excess return on security i can be written as

$$Z_i(t) - R(t) = \alpha_i + \beta_i[Z_M(t) - R(t)] + \epsilon_i(t), \tag{1}$$

where $Z_i(t)$ is the 1-period return per dollar on security i, α_i is the expected excess return from microforecasting, β_i is the ratio of the covariance of the return on security i with the market divided by the variance of the return on the market, and $\epsilon_i(t)$ has the property that its expectation, conditional on knowing the outcome for the market return

1. As a tautology, superior forecasting skills must be based on information that is not reflected in security prices. Therefore, if such information is obtainable, then security prices will not reflect all available information and the market will not be efficient. Fama (1970) provides an excellent discussion of both the Efficient Markets Theory and various attempts to test it.

2. The capital asset pricing model (CAPM) refers to the equilibrium relationships among security prices which result when investors have homgeneous beliefs and choose their portfolios based on a mean-variance criterion function. For the original derivations, see Sharpe (1964), Lintner (1965), and Mossin (1966). For a comprehensive review of the model, see Jensen (1972*a*).

$Z_M(t)$, is equal to its unconditional expectation which is zero. That is, $E[\epsilon_i(t) \mid Z_M(t)] = E[\epsilon_i(t)] = 0$.

Using this specification, both Fama (1972) and Jensen (1972*b*) develop theoretical structures for the evaluation of micro- and macroforecasting performance of investment managers where the basis for the evaluation is a comparison of the ex post performance of the manager's fund with the returns on the market. In the Jensen analysis, the market timer is assumed to forecast the actual return on the market portfolio, and the forecasted return and the actual return on the market are assumed to have a joint normal distribution. Jensen shows that under these assumptions, a market timer's forecasting ability can be measured by the correlation between the market timer's forecast and the realized return on the market.[3] However, Jensen also shows that the separate contributions of micro- and macroforecasting cannot be identified using the structure of (1) unless for each period, the market-timing forecast, the portfolio adjustment corresponding to that forecast, and the expected return on the market are known.

Grant (1977) explains how market-timing actions will affect the results of empirical tests that focus only on microforecasting skills. He shows that market-timing ability will cause the regression estimate of α_i in (1) to be a downward-biased measure of the excess returns resulting from microforecasting ability.

Treynor and Mazuy (1966) add a quadratic term to (1) to test for market-timing ability. In the standard CAPM regression equation, a portfolio's return is a linear function of the return on the market portfolio. However, they argue that if the investment manager can forecast market returns, he will hold a greater proportion of the market portfolio when the return on the market is high and a smaller proportion when the market return is low. Thus, the portfolio return will be a nonlinear function of the market return. Using annual returns for 57 open-end mutual funds, they find that for only one of the funds can the hypothesis of no market-timing ability be rejected with 95% confidence.

Kon and Jen (1979) use the Quandt (1972) switching regression technique in a CAPM framework to examine the possibility of changing levels of market-related risk over time for mutual fund portfolios. Using a maximum likelihood test, they separate their data sample into different risk regimes and then run the standard regression equation for each such regime. They find evidence that many mutual funds do have discrete changes in the level of market-related risk they choose which is consistent with the view that managers of such funds do attempt to incorporate market timing into their investment strategies.

3. See Jensen (1972*b*), pp. 317–18. This result is also derived in Treynor and Black (1973).

The model of market-timing forecasts presented here differs from those of these earlier studies in that we assume that our forecasters follow a more qualitative approach to market timing. Namely, we assume that they either forecast that $Z_M(t) > R(t)$ or forecast that $Z_M(t) \leq R(t)$. The forecasters in our model are less sophisticated than those hypothesized in, for example, the Jensen (1972*b*) formulation where they do forecast how much better the forecast superior investment will perform. However, as is shown in Part I, when this simple forecast information is combined with a prior distribution for returns on the market, a posterior distribution is derived which would permit probability statements about the magnitudes of the superior investment's performance.

A brief formal description of our forecast model is as follows: Let $\gamma(t)$ be the market timer's forecast variable where $\gamma(t) = 1$ if the forecast, made at time $t - 1$, for time period t is that $Z_M(t) > R(t)$ and $\gamma(t) = 0$ if the forecast is that $Z_M(t) \leq R(t)$. We define the probabilities for $\gamma(t)$ conditional upon the realized return on the market by

$$\begin{aligned} p_1(t) &\equiv \text{prob}\,[\gamma(t) = 0 \,|\, Z_M(t) \leq R(t)] \\ 1 - p_1(t) &= \text{prob}\,[\gamma(t) = 1 \,|\, Z_M(t) \leq R(t)] \end{aligned} \tag{2a}$$

and

$$\begin{aligned} p_2(t) &\equiv \text{prob}\,[\gamma(t) = 1 \,|\, Z_M(t) > R(t)] \\ 1 - p_2(t) &= \text{prob}\,[\gamma(t) = 0 \,|\, Z_M(t) > R(t)]. \end{aligned} \tag{2b}$$

Therefore, $p_1(t)$ is the conditional probability of a correct forecast given that $Z_M(t) \leq R(t)$, and $p_2(t)$ is the conditional probability of a correct forecast given that $Z_M(t) > R(t)$. It is assumed that $p_1(t)$ and $p_2(t)$ do not depend upon the magnitude of $|Z_M(t) - R(t)|$. Hence, the conditional probability of a correct forecast depends only on whether or not $Z_M(t) > R(t)$. Under this assumption, it was shown in Part I that the sum of the conditional probabilities of a correct forecast, $p_1(t) + p_2(t)$, is a sufficient statistic for the evaluation of forecasting ability.

Unlike the earlier studies of market timing, this formulation of the problem permits us to study market timing without assuming a CAPM framework. Indeed, provided that the market timer's forecasts are observable, we derive in Section II of this paper a nonparametric test of forecasting ability which does not require any assumptions about either the distribution of returns on the market or the way in which individual security prices are formed. Although the substantive context of the test presented there is market timing, the same test could be used to evaluate forecasting ability between any two securities.

If the market timer's forecasts are not directly observable, then to test market timing requires further assumptions about the structure of equilibrium security prices. In Section III, we derive such a test using the assumption that the CAPM holds. However, in contrast to the

Jensen formulation, our parametric test permits us to identify the separate contributions from micro- and macroforecasting to a portfolio's return using as our only data set the realized excess returns on the portfolio and on the market. Although the test specification in Section III assumes a CAPM framework, it can easily be adopted to a multifactor pricing model as described in Merton (1973) and Ross (1976).

II. A Nonparametric Test of Market Timing

In Part I, it was shown that a necessary and sufficient condition for a forecaster's predictions to have no value is that $p_1(t) + p_2(t) = 1$. Under this condition, an investor would not modify his prior estimate of the distribution of returns on the market portfolio as a result of receiving the prediction and therefore would pay nothing for the prediction. It follows that a necessary condition for market-timing forecasts to have a positive value is that $p_1(t) + p_2(t) \neq 1$. As shown in Part I, a sufficient condition for a positive value is that $p_1(t) + p_2(t) > 1$. For example, a perfect forecaster who is always correct will have $p_1(t) = 1$ and $p_2(t) = 1$; therefore, $p_1(t) + p_2(t) = 2 > 1$. Formally, forecasts with $p_1(t) + p_2(t) < 1$ can be shown to have a negative value because such forecasts are systematically incorrect. However, such forecasts are perverse in the sense that the contrary forecasts with $p_1'(t) \equiv 1 - p_1(t)$ and $p_2'(t) \equiv 1 - p_2(t)$ would satisfy $p_1'(t) + p_2'(t) > 1$ and therefore have positive value. For example, a market timer who is always wrong will have $p_1(t) + p_2(t) = 0$. However, such forecasts have all the informational content of a forecaster who is always right because by following a strategy of always doing the opposite of the forecasts that are always wrong, one will always be right. Thus, one can reasonably argue that forecasts with $p_1(t) + p_2(t) < 1$ have positive value as well, provided one is aware that the forecasts are perverse.

Therefore, a test of a forecaster's market-timing ability is to determine whether or not $p_1(t) + p_2(t) = 1$. Of course, if $p_1(t)$ and $p_2(t)$ were known, then such a test is trivial. However, $p_1(t)$, $p_2(t)$, or their sum, are rarely, if ever, observable. Generally, it will be necessary to estimate $p_1(t) + p_2(t)$, and then use these estimates to determine whether one can reject the natural null hypothesis of no forecasting skills. That is, H_0: $p_1(t) + p_2(t) = 1$ where the conditional probabilities of a correct forecast are not known. Essentially, this is a test of independence between the market timer's forecast and whether or not the return on the market portfolio is greater than the return from riskless securities.

The nonparametric test constructed around this null hypothesis takes advantage of the fact that conditional probabilities of a correct forecast are sufficient statistics to measure forecasting ability and yet they do not depend on the distribution of returns on the market or on any particular model for security price valuation. The essence of the

test is to determine the probability that a given outcome from our sample came from a population that satisfies the null hypothesis. To determine this probability, we proceed as follows. First, we define the following variables: $N_1 \equiv$ number of observations where $Z_M \leq R$; $N_2 \equiv$ number of observations where $Z_M > R$; $N \equiv N_1 + N_2 =$ total number of observations; $n_1 \equiv$ number of successful predictions, given $Z_M \leq R$; $n_2 \equiv$ number of unsuccessful predictions, given $Z_M > R$; $n \equiv n_1 + n_2 =$ number of times forecast that $Z_M \leq R$.

By definition, $E(n_1/N_1) = p_1$ and $E(n_2/N_2) = 1 - p_2$ where E is the expected value operator. From the null hypothesis, we have that

$$E(n_1/N_1) = p_1 \underset{H_0}{=} 1 - p_2 = E(n_2/N_2), \tag{3}$$

and from (3), it follows that

$$E[(n_1 + n_2)/(N_1 + N_2)] = E(n/N) \underset{H_0}{=} p_1 \equiv p. \tag{4}$$

Both n_1/N_1 and n_2/N_2 have the same expected value under our null hypothesis, namely, p, and both are drawn from independent subsamples. Hence, only one or the other need be estimated.

Both n_1 and n_2 are sums of independently and identically distributed random variables with binomial distributions. Therefore, the probability that $n_i = x$ from a subsample of N_i drawings can be written as

$$p(n_i = x \mid N_i,p) = \binom{N_i}{x} p^x(1 - p)^{N_i - x}; \; i = 1,2. \tag{5}$$

Given the null hypothesis, we can use Bayes's Theorem to determine the probability that $n_1 = x$ given N_1, N_2, and n, that is, $P(n_1 = x \mid N_1,N_2,n)$. Denote the event that our market timer forecasts m times that $Z_M \leq R$ (i.e., $n = m$) as A and the event that, of the times he forecasts that $Z_M \leq R$, he is correct x times and incorrect $m - x$ times (i.e., $n_1 = x$ and $n_2 = m - x$) as B. Then $P(n_1 = x \mid N_1,N_2,m) = P(B \mid A)$, and by Bayes's Theorem, we have that

$$\begin{aligned} P(B \mid A) &= \frac{P(B + A)}{P(A)} = \frac{P(B)}{P(A)} \\ &= \frac{\binom{N_1}{x}\binom{N_2}{m - x} p^x(1 - p)^{N_1 - x} p^{m-x}(1 - p)^{N_2 - m + x}}{\binom{N}{m} p^m(1 - p)^{N-m}} \\ &= \frac{\binom{N_1}{x}\binom{N_2}{m - x}}{\binom{N}{m}}. \end{aligned} \tag{6}$$

Hence, under the null hypothesis, the probability distribution for n_1—the number of correct forecasts, given that $Z_M \leq R$—has the form of a hypergeometric distribution and is independent of both p_1 and p_2. Therefore to test the null hypothesis, it is unnecessary to estimate either of the conditional probabilities. So, provided that the forecasts are known, all the variables necessary for the test are directly observable. Given N_1, N_2, and n, the distribution of n_1 under the null hypothesis is determined by (6) where the feasible range for n_1 is given by:

$$\underline{n}_1 \equiv \max(0, n - N_2) \leq n_1 \leq \min(N_1, n) \equiv \bar{n}_1. \tag{7}$$

Equations (6) and (7) can be used in a straightforward fashion to establish confidence intervals for testing the null hypothesis of no forecasting ability. For a standard two-tail test with a probability confidence level of c, one would reject the null hypothesis if $n_1 \geq \bar{x}(c)$ or if $n_1 \leq \underline{x}(c)$, where $\bar{x}$ and $\underline{x}$ are defined to be the solutions to the equations[4]

$$\sum_{x=\bar{x}}^{\bar{n}_1} \binom{N_1}{x}\binom{N_2}{n-x}\Big/\binom{N}{n} = (1-c)/2 \tag{8a}$$

and

$$\sum_{x=\underline{x}_1}^{\underline{x}} \binom{N_1}{x}\binom{N_2}{n-x}\Big/\binom{N}{n} = (1-c)/2. \tag{8b}$$

However, we would argue that a one-tail test (or at least one which weights the right-hand tail much more heavily than the left) is more appropriate in this case. If forecasters are rational, then it will never be true that $p_1(t) + p_2(t) < 1$, and a very small n_1 would simply be the "luck of the draw" no matter how unlikely. It seems most unlikely to us that a "real world" forecaster who had the talents to generate significant forecasting information would not have the talent to recognize that his forecasts were systematically perverse, while at the same time, we as outside observers of those forecasts can clearly see the errors of his ways. For such a one-tail test with a probability confidence level of c, one would reject the null hypothesis if $n_1 \geq x^*(c)$ where $x^*(c)$ is defined as the solution to

$$\sum_{x=x^*}^{\bar{n}_1} \binom{N_1}{x}\binom{N_2}{n-x}\Big/\binom{N}{n} = 1 - c. \tag{9}$$

By inspection of (8a) and (9), $x^*(c) < \bar{x}(c)$, and therefore, given an observation in the right tail, a one-tail test is, of course, more likely to

4. Because the hypergeometric distribution is discrete, the strict equalities of eqq. (8a) and (8b) will not, in general, be attainable. Therefore, in (8a), $\bar{x}$ should be interpreted as the lowest value of x for which the summation does not exceed $(1 - c)/2$. In (8b), $\underline{x}$ should be interpreted as the highest value of x for which the summation does not exceed $(1 - c)/2$.

reject the null hypothesis than a two-tail one for any fixed confidence level c. However, this fact in no way implies a greater likelihood of rejecting the null hypothesis when it is true by using a one-tail test.

Computation of the confidence intervals for either the two-tail or one-tail test using (8) or (9) is straightforward when the sample size is small. However, for large samples, the factorial or gamma function computations can become quite cumbersome. Fortunately, for those large samples where such computations become a problem, the hypergeometric distribution can be accurately approximated by the normal distribution.[5] The parameters used for this normal approximation are the mean and variance for the hypergeometric distribution given in (6), which can be written as

$$E(n_1) = \frac{nN_1}{N} \tag{10a}$$

and

$$\sigma^2(n_1) = [n_1N_1(N - N_1)(N - n)]/[N^2(N - 1)]. \tag{10b}$$

Tables 1, 2, and 3 give values of n_1 for a one-tail test that reject the null hypothesis at the 99% confidence level for different values of N_1, N_2, and n. As would be expected, the required estimated value of $p_1(t) + p_2(t)$ decreases as the size of the total sample increases. Tables 1–3 also demonstrate that the normal distribution can be an excellent approximation for determining the confidence intervals for the hypergeometric distribution, even for observation samples as small as 50.[6]

By focusing on the conditional frequencies of correct forecasts, the test procedure described in (6)–(10) takes into account the possibility that the market timer may not have the same skill in forecasting up markets as down markets. That is, $p_1(t)$ need not be equal to $p_2(t)$. However, if one knows that the forecaster whose predictions are being tested has equal ability with respect to both types of markets, then the conditional probabilities of a correct forecast, $p_1(t)$ and $p_2(t)$, are equal to each other, and therefore, each is equal to the unconditional probability of a correct forecast, $p(t)$. That is, $p_1(t) = p_2(t) = p(t)$. In that case, one need only measure the unconditional frequency of a correct forecast to test for market-timing ability where the null hypothesis of

5. The large-sample cases where direct computation of the confidence intervals using (8) or (9) are most cumbersome are when $N_1 \approx N_2$ or $n \approx N/2$. In these cases, the normal approximation will be quite good for even moderately large samples. See Lehmann (1975, theorem 19) for a general proof. The normal approximation will not be a good one even for quite large samples in those cases where there are substantial differences between N_1 and N_2 or between n and $N/2$. However, it is precisely in these latter cases where direct computation using (8) or (9) is not cumbersome even for very large samples.

6. As discussed in 5, this excellent approximation should only be expected to obtain when $N_1 \approx N_2$ and $n \approx N/2$.

TABLE 1 **Required Outcomes to Reject the Null Hypothesis with 99% Confidence**

			True Distribution			Normal Approximation		
N	N_1	n	Required Value of n_1	Total Correct Forecasts	$\frac{n_1}{N_1} + \frac{N_2 - n_2}{N_2}$	Required Value of n_1	Total Correct Forecasts	$\frac{n_1}{N_1} + \frac{N_2 - n_2}{N_2}$
50	25	25	17	34	1.36	17	34	1.36
50	35	25	22	34	1.43	22	34	1.43
50	15	25	12	34	1.43	12	34	1.43
50	25	15	12	34	1.36	12	34	1.36
50	25	35	22	34	1.36	22	34	1.36
50	15	15	9	38	1.43	9	38	1.43
50	15	35	15	30	1.43	15	30	1.43
50	35	15	15	30	1.43	15	30	1.43
50	35	35	29	38	1.43	29	38	1.43

TABLE 2 **Required Outcomes to Reject the Null Hypothesis with 99% Confidence**

			True Distribution			Normal Approximation		
N	N_1	n	Required Value of n_1	Total Correct Forecasts	$\frac{n_1}{N_1} + \frac{N_2 - n_2}{N_2}$	Required Value of n_1	Total Correct Forecasts	$\frac{n_1}{N_1} + \frac{N_2 - n_2}{N_2}$
100	50	50	32	64	1.28	32	64	1.28
100	75	50	43	61	1.29	44	63	1.35
100	25	50	18	61	1.29	19	63	1.35
100	50	75	43	61	1.22	44	63	1.26
100	50	25	18	61	1.22	19	63	1.26
100	25	25	12	74	1.31	12	74	1.31
100	25	75	24	48	1.28	24	48	1.28
100	75	25	24	48	1.28	24	48	1.28
100	75	75	62	74	1.31	62	74	1.31

TABLE 3 **Required Outcomes to Reject the Null Hypothesis with 99% Confidence**

			True Distribution			Normal Approximation		
N	N_1	n	Required Value of n_1	Total Correct Forecasts	$\frac{n_1}{N_1} + \frac{N_2 - n_2}{N_2}$	Required Value of n_1	Total Correct Forecasts	$\frac{n_1}{N_1} + \frac{N_2 - n_2}{N_2}$
200	100	100	59	118	1.18	60	120	1.20
200	50	100	33	116	1.21	34	118	1.24
200	150	100	83	116	1.21	84	118	1.24
200	100	150	83	116	1.16	84	118	1.18
200	100	50	33	116	1.16	34	118	1.18
200	50	50	20	140	1.20	20	140	1.20
200	50	150	44	88	1.17	45	90	1.20
200	150	50	44	88	1.17	45	90	1.20
200	150	150	120	140	1.20	120	140	1.20

no forecasting skills is that $p(t) = .5$. The distribution of outcomes drawn from a population that satisfies this null hypothesis is the binomial distribution which can be written as

$$P(k \mid N,p) = \binom{N}{k} p^k (1-p)^{N-k} = \binom{N}{k} (.5)^N, \tag{11}$$

where k is the number of correct predictions and N is the total number of observations. One can use (11) in an analogous fashion to (6) to construct either one-tail or two-tail confidence intervals for rejecting the null hypothesis. While the simplicity of this test may be attractive, the reader should be warned that a test which uses (11) instead of (6) is only appropriate if there is strong reason to believe that $p_1(t) = p_2(t)$.

As is discussed at length in Part I, the unconditional probability of a correct forecast cannot, in general, be used as a measure of market-timing ability. Specifically, it is shown that an unconditional probability of a correct forecast greater than one-half, $p(t) > .5$, is neither a necessary nor a sufficient condition for a forecaster's market-timing ability to have positive value. To see why it is not sufficient, consider the case of a forecaster who *always* predicts that the return on the market will exceed the return on riskless securities. Such completely predictable forecasts, like a stopped clock, clearly have no value. However, if the historical frequency with which the returns on the market exceeded the returns on riskless securities were significantly greater than one-half, then this forecaster's unconditional probability of a correct forecast would exceed one-half, and the null hypothesis would be rejected. Indeed, if a two-tail test were used, then all that would be required to reject the null hypothesis is that the historical frequency of up markets versus down markets be significantly different than one-half. However, if this "stopped clock" forecaster were evaluated by the test procedure described in (6)–(10), then, independent of the relative frequencies of up and down markets, the null hypothesis of no forecasting ability would not be rejected because for any sample of observations, $p_1(t) = 0$ and $p_2(t) = 1$, and hence, $p_1(t) + p_2(t) = 1$. Therefore, by using the unconditional probability procedure in (11), one is actually testing the joint null hypothesis of no market-timing ability and $p_1(t) = p_2(t)$.

In summary, we have derived a nonparametric procedure for testing market-timing ability which takes into account the possibility that forecasting skills are different for up markets than for down markets. Because the critical statistic for the test is $p_1(t) + p_2(t)$, it is not essential that the individual conditional probabilities be stationary through time. Rather, the critical stationarity property for the validity of equation (6) is that their sum, $p_1(t) + p_2(t)$, be stationary, which is,

of course, true under the null hypothesis of no market-forecasting ability. Moreover, it is straightforward to show that this same procedure can be used to test forecasters who have more confidence in their predictions during some periods than they do in other periods. One such application can be found in Lessard, Henriksson, and Majd (1981), where the predictions of some foreign-exchange forecasters are tested using our procedure. However, it is essential to our test procedure that the forecasts of market-timing be observable. We will therefore turn to the development of a procedure to test market-timing ability when such forecasts cannot be observed.

III. Parametric Tests of Market Timing

To use the nonparametric procedures to test investment performance, the predictions of the forecaster must be observable. However, it is frequently the case when measuring managed portfolio performance that the examiner only has access to the time series of realized returns on the portfolio and does not have the investment manager's market-timing forecasts themselves. While under certain conditions it is possible to infer from the portfolio return series alone what the manager's forecasts were, such inferences will, in general, provide noisy estimates of the forecasts. These estimates will be especially noisy if the manager's portfolio positions are influenced by his microforecasts for individual securities. In this section, we derive procedures which permit the testing of timing ability using return data alone. Of course, there is a "cost" of not having the time series of forecasts, which is that these test procedures require the assumption of a specific generating process for returns on securities. Thus, these procedures are parametric tests of the joint hypothesis of no market-timing ability and the assumed process for the returns on securities.

As noted earlier, most of the recent empirical studies of investment performance assume a pattern of equilibrium security returns which is consistent with the Security Market Line of the CAPM in addition to some assumptions about the market-timing behavior. The standard-regression-equation specification for portfolio returns used in these studies can be written as

$$Z_p(t) - R(t) = \alpha + \beta x(t) + \epsilon(t), \tag{12}$$

where $Z_p(t)$ is the realized return on the portfolio, $x(t) \equiv Z_M(t) - R(t)$ is the realized excess return on the market, and $\epsilon(t)$ is a residual random term which is assumed to satisfy the conditions

$$\begin{aligned} &E[\epsilon(t)] = 0 \\ &E[\epsilon(t) \,|\, x(t)] = 0 \\ &E[\epsilon(t) \,|\, \epsilon(t-i)] = 0, \quad i = 1,2,3, \ldots . \end{aligned} \tag{13}$$

Provided that the investment manager does not attempt (or, at least, is unsuccessful at) forecasting market returns, the standard least-squares estimation of (12) can be used to test for microforecasting skills. However, Jensen (1972*b*) shows that it is impossible to use this structural specification to separate the incremental performance due to stock selection from the increment due to market timing when the return data alone are used. The tests derived here do permit such a separation.

As in the earlier studies, we also assume that securities are priced according to the CAPM, although the tests can easily be adapted to accommodate a multifactor model provided that the factors are known. We further assume that as a function of his forecast, discretely different systematic risk levels for the portfolio are chosen by the forecaster. For example, in the case we analyze in detail here, it is assumed that there are two target risk levels which depend on whether or not the return on the market portfolio is forecast to exceed the return on riskless securities. That is, the investment manager is assumed to have one target beta when he predicts $Z_M(t) > R(t)$ and another target beta when he predicts that $Z_M(t) \leq R(t)$. In Section IV, we indicate how the test procedures can be adapted to the more general case of multiple target risk levels.

Let η_1 denote the target beta chosen for the portfolio by the manager when his forecast is that $Z_M(t) \leq R(t)$ and let η_2 denote the target beta chosen when his forecast is that $Z_M(t) > R(t)$. If $\beta(t)$ denotes the beta of the portfolio at time t, then $\beta(t) = \eta_1$ for a down-market forecast and $\beta(t) = \eta_2$ for an up-market forecast. If the forecaster is rational, then $\eta_2 > \eta_1$. Of course, if $\beta(t)$ were observable at each point in time, then, as discussed in Fama (1972), the market-timing forecast is observable, and one could simply apply the nonparametric tests of the previous section. However, if beta is not observable, then $\beta(t)$ is a random variable. Under the assumption that beta is not observable, let b denote the unconditional (on the forecast) expected value of $\beta(t)$. Then

$$b = q[p_1\eta_1 + (1 - p_1)\eta_2] + (1 - q)[p_2\eta_2 + (1 - p_2)\eta_1], \quad (14)$$

where q is equal to the unconditional (on the forecast) probability that $Z_M(t) \leq R(t)$. In Part I, the distribution from which q is computed was called the prior distribution. If we define the random variable $\theta(t)$ as equal to $[\beta(t) - b]$, then $\theta(t)$ is the unanticipated component of beta, and its distribution, conditional on the realized excess return on the market, $x(t)$, can be written as:

conditional on $x(t) \leq 0$:

$$\underline{\theta} = \underline{\theta}_1 \text{ where}$$

$$\underline{\theta}_1 = (\eta_1 - \eta_2)[1 - qp_1 - (1 - q)(1 - p_2)] \text{ with prob} = p_1 \quad (15a)$$

$$= (\eta_2 - \eta_1)[qp_1 + (1 - q)(1 - p_2)] \text{ with prob} = 1 - p_1$$

and

conditional on $x(t) > 0$:

$\underline{\theta} = \underline{\theta}_2$ where

$$\underline{\theta}_2 = (\eta_2 - \eta_1)[qp_1 + (1 - q)(1 - p_2)] \text{ with prob} = p_2 \tag{15b}$$
$$= (\eta_1 - \eta_2)[1 - qp_1 - (1 - q)(1 - p_2)] \text{ with prob} = 1 - p_2.$$

From (15), it follows that the conditional (on $x[t]$) expected value of θ can be written as

$$E(\theta \mid x) = \bar{\theta}_1 \tag{16a}$$
$$= (1 - q)(p_1 + p_2 - 1)(\eta_1 - \eta_2), \text{ for } x(t) \leqslant 0$$

and

$$E(\theta \mid x) = \bar{\theta}_2 \tag{16b}$$
$$= q(p_1 + p_2 - 1)(\eta_2 - \eta_1), \text{ for } x(t) > 0.$$

The per period return on the forecaster's portfolio can be written as

$$Z_p(t) = R(t) + [b + \theta(t)]x(t) + \lambda + \epsilon_p(t), \tag{17}$$

where λ is the expected increment to the return on the portfolio from microforecasting or security analysis and $\epsilon_p(t)$ is assumed to satisfy the standard CAPM conditions given in (13).

Under the posited return process for the portfolio given in (17), a least-squares regression analysis can be used to identify the separate increments to performance from microforecasting and macroforecasting. The regression specification can be written as

$$Z_p(t) - R(t) = \alpha + \beta_1 x(t) + \beta_2 y(t) + \epsilon(t) \tag{18}$$

where $y(t) \equiv \max[0, R(t) - Z_M(t)] = \max[0, -x(t)]$.

The motivation behind the specification given in (18) comes from the analysis of the value of market timing presented in Section IV of Part I. There it was shown that up to an additive noise term, the returns per dollar invested in a portfolio using the market-timing strategy described here will be the same as those that would be generated by pursuing a partial "protective put" option investment strategy where for each dollar invested in this strategy, $[p_2\eta_2 + (1 - p_2)\eta_1]$ dollars are invested in the market; $(p_1 + p_2 - 1)(\eta_2 - \eta_1)$ put options on the market portfolio are purchased with an exercise price (per dollar of the market) equal to $R(t)$; and the balance is invested in riskless securities. The value of the market timing (per dollar of assets managed) is that the $(p_1 + p_2 - 1)(\eta_2 - \eta_1)$ puts are obtained, in effect, for no cost. Note that $y(t)$ as defined in (18) is exactly the return on one such put option.

From (17), the expected return on the portfolio, conditional on $x > 0$, can be written as

$$E(Z_p \mid x > 0) = R + (b + \bar{\theta}_2)E(x \mid x > 0) + \lambda, \tag{19a}$$

and the expected return, conditional on $x \leq 0$, can be written as

$$E(Z_p \mid x \leq 0) = R + (b + \bar{\theta}_1)E(x \mid x \leq 0) + \lambda, \tag{19b}$$

where bars over random variables denote expected values.

For the analysis of the regression coefficients and error term in (18), it will be convenient to express the relevant variances and covariances of the regression variables in terms of the expected values and variances of the random variables $x_1(t)$ and $x_2(t)$ which are defined by

$$\begin{aligned} x_1(t) &\equiv \min[0,x(t)] \\ x_2(t) &\equiv \max[0,x(t)]. \end{aligned} \tag{20}$$

If $\text{var}[x_i(t)] \equiv \sigma_i^2$, $i = 1,2$, then we can write the variances and covariances of the regression variables as

$$\begin{aligned} \text{var}[y(t)] &\equiv \sigma_y^2 = \sigma_1^2 \\ \text{var}[x(t)] &\equiv \sigma_x^2 = \sigma_1^2 + \sigma_2^2 - 2\bar{x}_1\bar{x}_2 \\ \text{cov}[x(t),y(t)] &\equiv \sigma_{xy} = \bar{x}_1\bar{x}_2 - \sigma_1^2 \\ \text{cov}[Z_p(t),x(t)] &\equiv \sigma_{px} = (b + \bar{\theta}_1)(\sigma_1^2 - \bar{x}_1\bar{x}_2) + (b + \bar{\theta}_2)(\sigma_2^2 - \bar{x}_1\bar{x}_2) \\ \text{cov}[Z_p(t),y] &\equiv \sigma_{py} = (b + \bar{\theta}_2)\bar{x}_1\bar{x}_2 - (b + \bar{\theta}_1)\sigma_1^2. \end{aligned} \tag{21}$$

From (18) and (21), it follows that the large sample least-squares estimates of β_1 and β_2 can be written as

$$\begin{aligned} \text{plim } \hat{\beta}_1 &= \frac{\sigma_{px}\sigma_y^2 - \sigma_{py}\sigma_{xy}}{\sigma_x^2\sigma_y^2 - \sigma_{xy}^2} \\ &= b + \bar{\theta}_2 \\ &= p_2\eta_2 + (1 - p_2)\eta_1 \end{aligned} \tag{22}$$

and

$$\begin{aligned} \text{plim } \hat{\beta}_2 &= \frac{\sigma_{py}\sigma_x^2 - \sigma_{px}\sigma_{xy}}{\sigma_x^2\sigma_y^2 - \sigma_{xy}^2} \\ &= \bar{\theta}_2 - \bar{\theta}_1 \\ &= (p_1 + p_2 - 1)(\eta_2 - \eta_1). \end{aligned} \tag{23}$$

From (22), plim $\hat{\beta}_1 = E[\beta(t) \mid x(t) > 0]$, and it is also equal to the fraction invested in the market portfolio in the option strategy used in Section IV of Part I to replicate the market-timing strategy. The investment manager's market-timing ability is expressed as $\hat{\beta}_2$. The true β_2 will equal zero if either the forecaster has no timing ability—that is, if $p_1(t) + p_2(t) = 1$—or he does not act on his forecasts—that is, if $\eta_2 = \eta_1$. With reference to the replicating option investment strategy, from

(23), plim $\hat{\beta}_2$ is equal to the number of "free" put options on the market provided by the manager's market-timing skills. Indeed, as shown in Part I, the value of market-timing skills per dollar of assets managed is equal to $(p_1 + p_2 - 1)(\eta_2 - \eta_1)g(t)$ where $g(t)$ is the market price of such a put. Thus, $\hat{\beta}_2 g(t)$ is an estimate of the value of the market-timing ability of the manager.

Formally interpreted, a negative value for the regression estimate $\hat{\beta}_2$ would imply a negative value for market timing. However, a true negative value for β_2 would violate the rationality assumptions of $p_1(t) + p_2(t) \geq 1$ and $\eta_2 \geq \eta_1$. Hence, as was discussed for the nonparametric tests in Section II, the reader should consider the relative merits of a one-tail versus the standard two-tail test of significance with respect to rejecting the null hypothesis that $\beta_2 = 0$.

The increment to portfolio performance from microforecasting can also be measured using regression equation (18). The large sample least-squares estimate of α can be written as[7]

$$\text{plim } \hat{\alpha} = E(Z_p) - R - \text{plim } \hat{\beta}_1 \bar{x} - \text{plim } \hat{\beta}_2 \bar{y} = \lambda. \tag{24}$$

Hence, from (23) and (24), regression equation (18) can be used to identify and estimate the separate contributions of microforecasting and macroforecasting to overall portfolio performance.

To complete the analysis of (18), we now investigate the properties of the error term $\epsilon(t)$ being careful to take into account the differences between the actual betas and their estimates. That is, if the forecaster strictly follows the posited behavior of two discrete risk levels η_1 and η_2, then the actual $\beta(t)$ of the fund will never be equal to the estimated β unless, of course, he is a perfect forecaster.

Let us define the following variables: $\Delta_1 = 1$ if $x \leq 0$ and the market timer's forecast is incorrect, $\Delta_1 = 0$ otherwise; $\nabla_1 = 1$ if $x \leq 0$ and the market timer's forecast is correct, $\nabla_1 = 0$ otherwise; $\Delta_2 = 1$ if $x > 0$ and the market timer's forecast is incorrect, $\Delta_2 = 0$ otherwise; $\nabla_2 = 1$ if $x > 0$ and the market timer's forecast is correct, $\nabla_2 = 0$ otherwise. It follows immediately that:

$$\begin{aligned} E(\Delta_1) &= p_1 & E(\nabla_1) &= 1 - p_1 \\ E(\Delta_2) &= 1 - p_2 & E(\nabla_2) &= p_2. \end{aligned} \tag{25}$$

From (18), each period's estimation error ϵ can be written as:

$$\epsilon = \Delta_1(1 - p_1)(\eta_1 - \eta_2)x_1 - \nabla_1 p_1(\eta_1 - \eta_2)x_1 + \Delta_2 p_2(\eta_1 - \eta_2)x_2 - \nabla_2(1 - p_2)(\eta_1 - \eta_2)x_2 + \epsilon_p, \tag{26}$$

7. When testing for forecasting ability, the relevant portfolio returns are those earned before any deduction for management fees. If one uses the returns earned after the deduction of management fees and if the fees charged can be expressed as a fixed percentage of assets, m, then plim $\hat{\alpha} = \lambda - m$.

where ϵ_p includes any error term resulting from microforecasting. Since, by definition, microforecasting is independent of x, ϵ_p is independent of x. It follows by the Law of Large Numbers that as the number of observations, N, gets large,

$$\Sigma \frac{\Delta_1}{N}, \Sigma \frac{\Delta_2}{N}, \Sigma \frac{\nabla_1}{N}, \text{ and } \Sigma \frac{\nabla_2}{N}$$

will all approach their expectation. Therefore, from (26),

$$\begin{aligned} \lim_{N\to\infty} \left[\frac{\Sigma \epsilon}{N}\right] &= [p_1(1-p_1)(\eta_1-\eta_2) - (1-p_1)p_1(\eta_1-\eta_2)]\bar{x}_1 \\ &\quad + [(1-p_2)p_2(\eta_1-\eta_2) - p_2(1-p_2)(\eta_1-\eta_2)]\bar{x}_2 \quad (27) \\ &\quad + \lim_{N\to\infty} \Sigma\epsilon_p/N \\ &= \lim_{N\to\infty} \left[\frac{\Sigma\epsilon_p}{N}\right] = 0. \end{aligned}$$

Thus, for large samples, the coefficient from least-squares estimation of (18), plus the realized excess return on the market, will give us an unbiased estimate of the portfolio return.[8]

As discussed, the motivation behind the regression specification (18) was the analysis in Part I which showed the correspondence between market-timing investment strategies and certain option investment strategies. However, there is alternative, but equivalent, specification which some may find to be more intuitive. Namely, by a linear transformation of (18), we can write this alternative regression equation as

$$Z_p(t) - R(t) = \alpha' + \beta_1' x_1(t) + \beta_2' x_2(t) + \epsilon, \quad (28)$$

where $x_1(t)$ and $x_2(t)$ are as defined in (20). Because $x_1(t) = 0$ and $x_2(t) = x(t)$ if $x(t) > 0$ β_2' has a rather intuitive interpretation as the "up-market" beta of the portfolio. Similarly, because $x_1(t) = x(t)$ and $x_2(t) = 0$ if $x(t) \leq 0$, β_1' can be interpreted as the "down-market" beta of the portfolio. Indeed, the large sample properties of the regression estimators $\hat{\beta}_1'$ and $\hat{\beta}_2'$ fit these intuitive interpretations (at least in the sense of expected values). That is,

$$\begin{aligned} \text{plim } \hat{\beta}_1' &= E[\beta(t) \,|\, x(t) \leq 0] \\ &= p_1\eta_1 + (1-p_1)\eta_2 \end{aligned} \quad (29a)$$

8. Although unbiased, ordinary least squares estimation is not efficient because $\beta(t)$ is not stationary, and therefore the standard deviation of the error term is an increasing function of $|x(t)|$. To improve the efficiency of the estimates, one could use generalized least squares estimation to correct for this heteroscedasticity.

and

$$\begin{aligned} \text{plim } \hat{\beta}_2' &= E\left[\beta(t) \,\middle|\, x(t) > 0\right] \\ &= p_2\eta_2 + (1 - p_2)\eta_1, \end{aligned} \tag{29b}$$

where $E[\beta(t) \,|\, x(t) > 0] = b + \bar{\theta}_2$ and $E[\beta(t) \,|\, x(t) \leq 0] = b + \bar{\theta}_1$. The test for market-timing ability using this specification would be to show that $\hat{\beta}_2'$ is significantly greater than $\hat{\beta}_1$. That is, show that the expected "up-market" beta of the portfolio is greater than the expected "down-market" beta of the portfolio. The large sample properties of $\hat{\alpha}'$ are the same as for $\hat{\alpha}$ in (18): namely, plim $\hat{\alpha}' = \lambda$.

IV. Summary and Extensions

Provided that the forecaster only attempts to predict the sign of $Z_M(t) - R(t)$ but not its magnitude, and provided that his forecasts are observable, a procedure for testing market timing has been derived which does not depend on any distributional assumptions about the returns on securities. The test includes the possibility that the forecaster's confidence in his forecasts as measured by (p_1, p_2) can vary over time, and indeed, if such variations are observable, then the test can be refined to measure his forecasting ability for each such variation.

In the case where the forecaster's predictions are not observable, a parametric test procedure was derived which permits separate measurements of the contributions to portfolio performance from market timing and security analysis. As is apparent from the analysis of the error term in Section III, this test will accommodate the case where the two-target risk levels chosen by the manager vary over time provided that these variations are random around a stationary mean.

The test is also applicable to the case where the forecaster selects from more than two discrete systematic risk target levels, as long as the different levels are based on differing levels of confidence in the forecasts and not differing expectations of the level of the return. In this case, the large sample least-squares estimates of β_1 and β_2 represent a weighted average of the different risk levels.

The test procedures presented here can be extended to evaluate the performance of a market timer who segments his prediction of $x(t)$ into more than two discrete regions. For example, a forecaster might have four possible predictions: that $x(t) < -10\%$, that $-10\% \leq x(t) < 0$, that $0 \leq x(t) < 10\%$, and that $10\% \leq x(t)$. We briefly illustrate how the analysis would be applied to such multiple regions for the parametric test case. As in the two-region case, it is assumed that the probability of a particular forecast will only depend on the region in which $x(t)$ falls. However, there are now more than two possible forecasts.

Specifically, we assume that there are n different regions that the forecaster might predict and define p_{ij} as the probability that the forecaster's prediction was that $x(t)$ would be in the jth region, given that $x(t)$ actually ended up in the ith region. The only constraint on the conditional probabilities is that $\Sigma_{j=1}^{n} p_{ij} = 1, i = 1,2, \ldots, n$. The return on the forecaster's portfolio can be defined as in Section III except that now

$$b = \sum_{i=1}^{n} \delta_i \sum_{j=1}^{n} p_{ij}\eta_j,$$

where δ_i is defined as the probability that $x(t)$ will end up in region i and η_j is the chosen level of systematic risk when the forecast is that $x(t)$ will end up in region j.

The regression equation corresponding to (28) in the two region case can be written as $Z_p - R = \alpha + \Sigma_{i=1}^{n} \beta_i x_i + \epsilon$, where $x_i = x$ if x is in region i, and $x_i = 0$ otherwise.

The large sample least-squares estimates of β_i and α are: plim $\hat{\beta}_i = \Sigma_{j=1}^{n} p_{ij}\eta_j$ and plim $\hat{\alpha} = \lambda$. From this analysis, it follows that for sufficiently finely partitioned regions (that is, for large enough values of n), it is at least in principle, possible to separate the incremental returns from micro- and macroforecasting without any restrictions on the distribution of forecasts. All that is required are the actual returns from the market, the portfolio, and riskless securities.

References

Fama, E. 1970. Efficient capital markets: a review of theory and empirical work. *Journal of Finance* 25, no. 2:383–417.

Fama, E. 1972. Components of investment performance. *Journal of Finance* 27, no. 2:551–67.

Grant, D. 1977. Portfolio performance and the "cost" of timing decisions. *Journal of Finance* 32, no. 2:837–46.

Jensen, M. C. 1972*a*. Capital markets: theory and evidence. *Bell Journal of Economics and Management Science* 3, no. 2:357–98.

Jensen, M. C. 1972*b*. Optimal utilization of market forecasts and the evaluation of investment performance. In G. P. Szego and K. Shell (eds.), *Mathematical Methods of Investment and Finance*. Amsterdam: North-Holland.

Kon, S. J., and Jen, F. C. 1978. The investment performance of mutual funds: an empirical investigation of timing, selectivity, and market efficiency. *Journal of Business* 52, no. 2:263–89.

Lehmann, E. 1975. *Nonparametrics: Statistical Methods Based on Ranks*. San Francisco: Holden-Day.

Lessard, D.; Henriksson, R.; and Majd, S. 1981. Risk and return on foreign currency assets: an evaluation of forecasting ability. Working draft. Cambridge, Mass.: Massachusetts Institute of Technology.

Lintner, J. 1965. Security prices, risk, and maximal gains from diversification. *Journal of Finance* 20 (December): 587–615.

Merton, R. C. 1973. An intertemporal capital asset pricing model. *Econometrica* 41, no. 5:867–87.

Merton, R. C. 1981. On market timing and investment performance. I. An equilibrium theory of value for market forecasts. *Journal of Business* 54 (July): 363–406.

Mossin, J. 1966. Equilibrium in a capital asset market. *Econometrica* 34, no. 4:768–83.
Quandt, R. E. 1972. A new approach to estimating switching regressions. *Journal of the American Statistical Association* 67, no. 338:306–10.
Ross, S. 1976. The arbitrage theory of capital asset pricing. *Journal of Economic Theory* 3 (December): 343–62.
Sharpe, W. F. 1964. Capital asset prices: a theory of market equilibrium under conditions of risk. *Journal of Finance* 19, no. 4:425–42.
Treynor, J., and Black, F. 1973. How to use security analysis to improve portfolio selection. *Journal of Business* 46, no. 1:66–86.
Treynor, J., and Mazuy, F. 1966. Can mutual funds outguess the market? *Harvard Business Review* 44 (July–August): 131–36.