A Test of the Efficiency of a Given Portfolio

Author(s): Michael R. Gibbons, Stephen A. Ross and Jay Shanken

A TEST OF THE EFFICIENCY OF A GIVEN PORTFOLIO

By MICHAEL R. GIBBONS, STEPHEN A. ROSS, AND JAY SHANKEN[1]

A test for the ex ante efficiency of a given portfolio of assets is analyzed. The relevant statistic has a tractable small sample distribution. Its power function is derived and used to study the sensitivity of the test to the portfolio choice and to the number of assets used to determine the ex post mean-variance efficient frontier.

Several intuitive interpretations of the test are provided, including a simple mean-standard deviation geometric explanation. A univariate test, equivalent to our multivariate-based method, is derived, and it suggests some useful diagnostic tools which may explain why the null hypothesis is rejected.

Empirical examples suggest that the multivariate approach can lead to more appropriate conclusions than those based on traditional inference which relies on a set of dependent univariate statistics.

KEYWORDS: Asset pricing, CAPM, multivariate test, portfolio efficiency.

1. INTRODUCTION

The modern theory of finance has always been rooted in empirical analysis. The mean-variance capital asset pricing model (CAPM) developed by Sharpe (1964) and Lintner (1965) has been studied and tested in more papers than can possibly be attributed here. This is only natural; the quality and quantity of financial data, especially stock market price series, are the envy of other fields in economics.

The theory is generally expressed in terms of its first-order conditions on the risk premium. Expected returns on assets are linearly related to the regression coefficients, or betas, of the asset returns on some index of market returns. In other words, risk premiums in equilibrium depend on betas. The standard tests of the CAPM are based on regression techniques with various adaptations. For some notable examples, see Black, Jensen, and Scholes (1972) and Fama and MacBeth (1973). Usually, cross-sectional regressions are run of asset returns on estimated beta coefficients, and estimates of the slope are reported. Often the data are grouped to reduce measurement errors, and sometimes the estimation is done at a sequence of time points to create a time series of estimates from which the precision of the overall average can be determined.

Roll (1977, 1978), among others, has raised serious doubts whether these procedures are, in fact, tests of the CAPM. Insofar as proxies are used for the market portfolio, the Sharpe-Lintner theory is not being tested. Furthermore, as Roll emphasizes, the regression tests are probably of quite low power, and

grouping may lower the power further. These objections leave the empirical testing of the CAPM in an odd state of limbo. If the proxy is not a valid surrogate, then as tests of the CAPM the existing empirical investigations are somewhat beside the point.[2] On the other hand, if the proxy is valid, then the small sample distribution and power of the tests are unknown.

This is unfortunate and indicative of a missed opportunity. The CAPM is one of many financial theories which suggest quite specific hypotheses couched in terms of observables. The rich data available for testing these hypotheses are an incentive to develop tests which are explicitly directed at them. In this paper we develop a canonical example of such a test using multivariate statistical methods. The problem we consider is the central one addressed in tests of the CAPM. Since the theory is equivalent to the assertion that the market portfolio is mean-variance efficient, we wish to test whether any particular portfolio is ex ante mean-variance efficient.

While the paper is organized into seven sections, it also can be viewed as consisting of three parts. The first part (Sections 2 through 4) considers a multivariate statistic for testing mean-variance efficiency and examines the properties of such a test. The second part (Sections 5 and 6) studies the relation between this multivariate test and alternative approaches based on a set of univariate statistics. The third part (Sections 7 and 8) concludes the paper by extending the framework to related hypotheses and providing suggestions for future research. A more detailed summary of each section follows.

In Section 2 we recall a necessary condition for the efficiency of some portfolio. We use this implication as a null hypothesis that can be tested using a statistic which has a tractable finite sample distribution under both the null and alternate hypotheses. In addition, we relate this statistic to three alternative approaches which are based on asymptotic approximations. In the third section the multivariate test is given a geometric interpretation in the mean-standard deviation space of portfolio theory. The method and geometry are then applied to a data set from one of the classic empirical papers in modern finance; we reaffirm and complement the findings of Black, Jensen, and Scholes (1972). The fourth section turns to issues relating to the power of the test. Here we consider the sensitivity of the test to the choice of the portfolio which is examined for efficiency and the effect of the number of assets used to determine the ex post efficient frontier. A new data base is analyzed in this section, and we demonstrate that one's conclusions regarding the efficiency of a given index can be altered by the type of assets used to construct the ex post frontier.

The fifth section attempts to contrast actual empirical results when the multivariate method is used versus informal inference based on a set of dependent univariate statistics. Here we provide examples where the multivariate test rejects even though none of the univariate statistics seem to be significant. We also have

---

[2] Recent work by Kandel and Stambaugh (1987) and Shanken (1987b) do consider tests of the CAPM conditional on an assumption about the correlation between the proxy and the true market portfolio.

the reverse situation where there are a seemingly large number of "significant" univariate statistics; yet, the multivariate test fails to reject at the traditional levels of significance. In this section we also introduce another data set which allows us to re-examine the size-effect anomaly. Section 6 develops an alternative interpretation of the multivariate test. The statistic is equivalent to the usual calculation for a t statistic on an intercept term in a *univariate* simple regression model, with the ex post efficient portfolio used as the dependent variable and the portfolio whose ex ante efficiency is under examination as the explanatory variable. This section also develops some useful diagnostics for explaining why the null hypothesis may not be consistent with the data. Most of the empirical work in this section focuses on the size effect only in the month of January.

Section 7 extends the analysis to a case where one wishes to investigate the potential efficiency of some linear combination of a set of portfolios, where the weights in the combination are not specified. This turns out to be a minor adaptation of the work in Section 2.

2. TEST STATISTIC FOR JUDGING THE EFFICIENCY OF A GIVEN PORTFOLIO

We assume throughout that there is a given riskless rate of interest, R_{ft}, for each time period. Excess returns are computed by subtracting R_{ft} from the total rates of return. Consider the following multivariate linear regression:

$$(1) \qquad \tilde{r}_{it} = \alpha_{ip} + \beta_{ip}\tilde{r}_{pt} + \tilde{\varepsilon}_{it} \qquad \forall i = 1,\ldots, N,$$

where $\tilde{r}_{it} \equiv$ the excess return on asset i in period t; $\tilde{r}_{pt} \equiv$ the excess return on the portfolio whose efficiency is being tested; and $\tilde{\varepsilon}_{it} \equiv$ the disturbance term for asset i in period t. The disturbances are assumed to be jointly normally distributed each period with mean zero and nonsingular covariance matrix Σ, conditional on the excess returns for portfolio p. We also assume independence of the disturbances over time. In order that Σ be nonsingular, $\tilde{r}_{pt}$ and the N left-hand side assets must be linearly independent.

If a particular portfolio is mean-variance efficient (i.e., it minimizes variance for a given level of expected return), then the following first-order condition must be satisfied for the given N assets:

$$(2) \qquad \mathscr{E}(\tilde{r}_{it}) = \beta_{ip}\mathscr{E}(\tilde{r}_{pt}).$$

Thus, combining the first-order condition in (2) with the distributional assumption given by (1) yields the following parameter restriction, which is stated in the form of a null hypothesis:

$$(3) \qquad H_0: \alpha_{ip} = 0, \qquad \forall i = 1,\ldots, N.$$

Testing the above null hypothesis is essentially the same proposal as in the work by Black, Jensen, and Scholes (1972), except that they replace $\tilde{r}_{pt}$ by a portfolio which they call the market portfolio and refer to their test as a test of the CAPM. In addition, they do not report the joint significance of the estimated

values for α_{ip} across all N equations; instead, they report N univariate t statistics based on each equation.

Given the normality assumption, the null hypothesis in (3) can be tested using "Hotelling's T^2 test," a multivariate generalization of the univariate t-test (e.g., see Malinvaud (1980, page 230)). A brief derivation of the equivalent F test is included for completeness and as a means of introducing some notation that will be needed later. If we estimate the multivariate system of (1) using ordinary least squares for each individual equation, the estimated intercepts have a multivariate normal distribution, conditional on r_{pt} ($\forall t = 1, \ldots, T$), with

$$(4) \qquad \sqrt{T/(1 + \hat{\theta}_p^2)} \, \hat{\alpha}_p \sim N\left\{ \sqrt{T/(1 + \hat{\theta}_p^2)} \, \alpha_p; \Sigma \right\},$$

where $T \equiv$ number of time series observations on returns; $\hat{\alpha}_p' \equiv (\hat{\alpha}_{1p} \hat{\alpha}_{2p} \ldots \hat{\alpha}_{Np})$; $\hat{\theta}_p \equiv \bar{r}_p / s_p$; $\bar{r}_p \equiv$ sample mean of $\tilde{r}_{pt}$; and $s_p^2 \equiv$ sample variance of $\tilde{r}_{pt}$ without an adjustment for degrees of freedom. Furthermore, $\hat{\alpha}_p$ and $\hat{\Sigma}$ are independent with $(T - 2)\hat{\Sigma}$ having a Wishart distribution with parameters $(T - 2)$ and Σ. These facts imply (see Morrison (1976, page 131)) that $(T(T - N - 1)/N(T - 2))W_u$ has a noncentral F distribution with degrees of freedom N and $(T - N - 1)$, where

$$(5) \qquad W_u \equiv \hat{\alpha}_p' \hat{\Sigma}^{-1} \hat{\alpha}_p / (1 + \hat{\theta}_p^2)$$

and $\hat{\Sigma} \equiv$ unbiased residual covariance matrix.[3] (The corresponding statistic based on the maximum likelihood estimate of Σ will be denoted as W.) The noncentrality parameter, λ, is given by

$$(6) \qquad \lambda \equiv \left[T/(1 + \hat{\theta}_p^2) \right] \alpha_p' \Sigma^{-1} \alpha_p.$$

Under the null hypothesis that α_p equals zero, $\lambda = 0$, and we have a central F distribution. More generally, the distribution under the alternative provides a way to study the power of the test; more will be said about this in a later section. It is also interesting to note that under the null hypothesis the W_u statistic has a central F distribution unconditionally, for the parameters of this central F do not depend on $\tilde{r}_{pt}$ in any way. However, we do not know the unconditional distribution of $\hat{\alpha}_p$ or W_u under the alternate, for the conditional distribution depends on the sample values of $\tilde{r}_{pt}$ through $\hat{\theta}_p^2$.

Generally, the normality assumption has been viewed as providing a "good working approximation" to the distribution of monthly stock returns (see Fama (1976, Chapter 1) for a summary of the relevant empirical work). There is some evidence, however, that the true distributions are slightly leptokurtic relative to the normal distribution. While departures from normality of the disturbances in (1) will affect the small-sample distribution of the test statistic, simulation evidence by MacKinlay (1985) suggests that the F test is fairly robust to such misspecifications.[4] This is important, since the application of standard asymptotic tests to the efficiency problem can result in faulty inferences, given the sample sizes often used in financial empirical work.

---

[3] We assume that N is less than or equal to $T - 2$ so that $\hat{\Sigma}$ is nonsingular.
[4] Tests for normality of the residuals of the size and industry portfolios, which are used below, do reveal excess kurtosis and some skewness as well. These results are available on request to the authors.

TABLE 1

A COMPARISON OF FOUR ASYMPTOTICALLY EQUIVALENT TESTS OF EX ANTE EFFICIENCY OF A
GIVEN PORTFOLIO. THE W STATISTIC IS DISTRIBUTED AS A TRANSFORM OF A CENTRAL F
DISTRIBUTION IN FINITE SAMPLES. THE WALD TEST, THE LIKELIHOOD RATIO TEST (LRT),
AND THE LAGRANGE MULTIPLIER TEST (LMT) ARE MONOTONE TRANSFORMS OF W, AND
EACH IS DISTRIBUTED AS CHI-SQUARE WITH N DEGREES OF FREEDOM AS T APPROACHES
INFINITY.

N	T	P-Value Using Exact Distribution of W	P-Values Using Asymptotic Approximations		
			Wald	LRT	LMT
10	60	.05	.008	.027	.071
20	60	.05	.000	.007	.094
40	60	.05	.000	.000	.173
58	60	.05	.000	.000	.403
10	120	.05	.023	.038	.060
20	120	.05	.005	.023	.070
40	120	.05	.000	.003	.094
58	120	.05	.000	.000	.122
118	120	.05	.000	.000	.431
10	240	.05	.035	.044	.055
20	240	.05	.109	.035	.059
40	240	.05	.003	.017	.069
58	240	.05	.000	.006	.079
118	240	.05	.000	.000	.123
238	240	.05	.000	.000	.451
10	60	.10	.025	.061	.122
20	60	.10	.000	.019	.146
40	60	.10	.000	.000	.216
58	60	.10	.000	.000	.404
10	120	.10	.056	.081	.111
20	120	.10	.017	.053	.122
40	120	.10	.000	.010	.147
58	120	.10	.000	.000	.175
118	120	.10	.000	.000	.432
10	240	.10	.076	.090	.106
20	240	.10	.048	.075	.111
40	240	.10	.009	.041	.122
58	240	.10	.001	.018	.133
118	240	.10	.000	.000	.178
238	240	.10	.000	.000	.452

Note: N is the number of assets used together with portfolio p to construct the ex post frontier, and T is the number of time series observations.

Table I illustrates this problem for the Wald, likelihood ratio, and Lagrange multiplier tests, each of which is asymptotically distributed as chi-square with N degrees of freedom as $T \rightarrow \infty$.[5] Since the small-sample distribution of W is known (assuming normality), the implied realization of W can be inferred from the information in the first three columns of Table I (i.e., N, T, and the hypothetical p-value). The implied asymptotic p-values given in the last three

[5] Jobson and Korkie (1982) also discuss these three tests using a simulation. They approximate the distribution of the likelihood ratio test with an F distribution based on Rao's (1951) work. In their 1985 paper they recognize that a small sample distribution is available under the null hypothesis.

columns are then obtained using the fact that each test statistic is a monotonic function of W.[6]

Consistent with the results of Berndt and Savin (1977), the p-values are always lowest for the Wald test and highest for the Lagrange multiplier test with the likelihood ratio test in between. Clearly, the asymptotic approximation becomes worse as the number of assets, N, approaches the number of time series observations, T. Shanken (1985) reaches similar conclusions based on an approximation when the riskless asset is not observable.

3. A GEOMETRIC INTERPRETATION OF THE TEST STATISTIC, W

So far, the primary motivation for the W statistic has been its well-known distributional properties. For rigorous statistical inference such results are an absolute necessity. Just as important, though, is the development of a measure which allows one to examine the economic significance of departures from the null hypothesis. Fortunately, our test has a nice geometric interpretation.

It is shown in the Appendix that:

$$(7) \qquad W = \left[\frac{\sqrt{1 + \hat{\theta}^{*2}}}{\sqrt{1 + \hat{\theta}_p^2}} \right]^2 - 1 \equiv \psi^2 - 1,$$

where $\hat{\theta}^*$ is the ex post price of risk (i.e., the maximum excess sample mean return per unit of sample standard deviation) and $\hat{\theta}_p$ is the ratio of ex post average excess return on portfolio p to its standard deviation (i.e., $\hat{\theta}_p \equiv \bar{r}_p / s_p$). Note that ψ cannot be less than one since $\hat{\theta}^*$ is the slope of the ex post frontier based on all assets used in the test (including portfolio p).

The curve in Figure 1a represents the (ex post) minimum-variance frontier of the risky assets. When a riskless investment is available, the frontier is a straight line emanating from the origin and tangent to the curve at m. $\hat{\theta}^*$ is the slope of the tangent line whereas $\hat{\theta}_p$ is the slope of the line through p.

An examination of (7) suggests that ψ^2 should be close to one under the null hypothesis. When $\hat{\theta}^*$ is sufficiently greater than $\hat{\theta}_p$, the return per unit of risk for portfolio p is much lower than the ex post frontier tradeoff, and we will reject the hypothesis that portfolio p is ex ante mean-variance efficient. In Figure 1a ψ is just the distance along the ex post frontier up to any given risk level, σ, divided by the similar distance along the line from the origin through p.

The reader may wonder why the test is based on the square of the slopes as opposed to the actual slopes. The reason is straightforward. Our null hypothesis only represents a necessary condition for ex ante efficiency. This condition is satisfied even if portfolio p is on the negative sloping portion of the minimum-variance frontier for all assets (including the risk-free security). Thus, only the

---

[6] The relations are $LRT = T \ln(1 + W)$ and $LMT = TW/(1 + W)$. Shanken (1985) has discussed this result for the case where the riskless asset does not exist. A proof of the result in the case where the riskless asset does exist is available upon request to the authors. Berndt and Savin (1977) discuss similar relationships among alternative asymptotic tests in a more general setting.
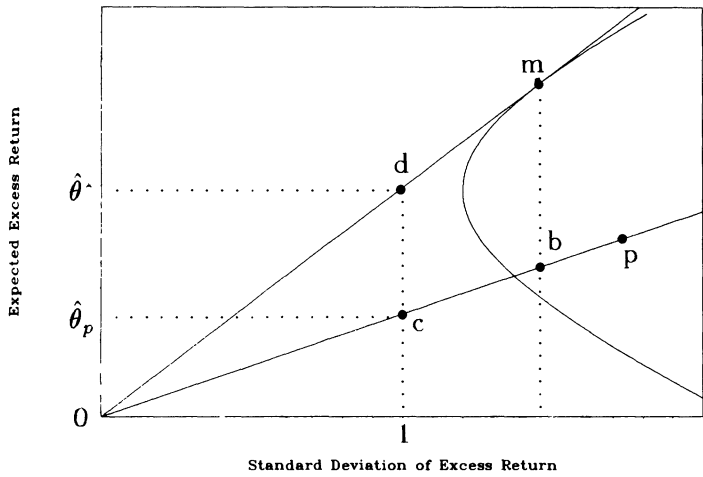
1a.) Geometric intuition for W. Note the distance Oc is $\sqrt{1 + \hat{\theta}_p^2}$, and the distance Od is $\sqrt{1 + \hat{\theta}*^2}$.
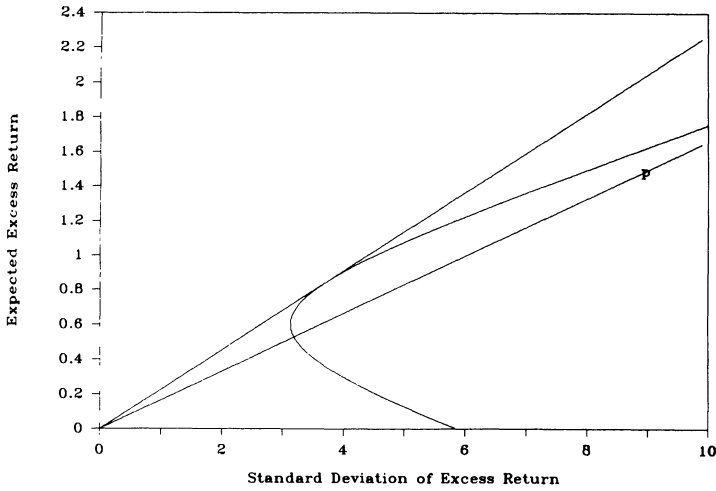
1b.) *Ex post* efficient frontier based on 10 beta-sorted portfolios and the CRSP Equal-Weighted Index using monthly data, 1931-1965. Point p represents the CRSP Equal-Weighted Index.

FIGURE 1.—Various plots of ex post mean variance efficient frontiers.
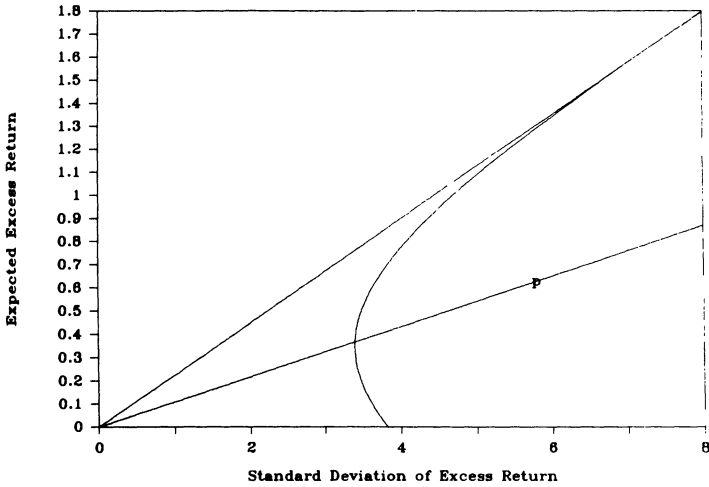
1c.) *Ex post* efficient frontier based on 12 industry portfolios and the CRSP Value-Weighted Index using monthly data, 1926-1982. Point p represents the CRSP Value-Weighted Index.

1d.) *Ex post* efficient frontier based on 10 size-sorted portfolios and the CRSP Value-Weighted Index using monthly data, 1926-1982. Point p represents the CRSP Value-Weighted Index.

FIGURE 1.—Continued.

TABLE II

SUMMARY STATISTICS ON BETA-SORTED PORTFOLIOS BASED ON MONTHLY DATA, 1931–65
($T = 420$). ALL SIMPLE EXCESS RETURNS ARE NOMINAL AND IN PERCENTAGE FORM, AND THE
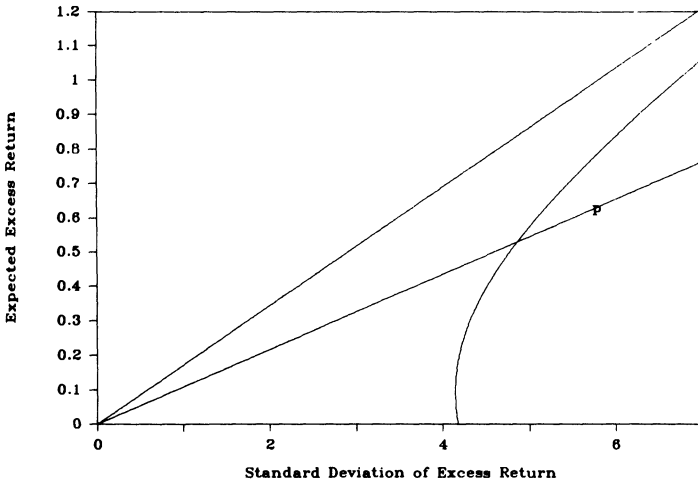CRSP EQUAL-WEIGHTED INDEX IS PORTFOLIO p. THE FOLLOWING PARAMETER ESTIMATES
ARE FOR THE REGRESSION MODEL: $\tilde{r}_{it} = \alpha_{ip} + \beta_{ip}\tilde{r}_{pt} + \tilde{\varepsilon}_{it}$ $\forall i = 1,\ldots,10$ AND $\forall t = 1,\ldots,420$,
WHERE R_i^2 IS THE COEFFICIENT OF DETERMINATION FOR EQUATION i.

Portfolio Number	$\hat{\alpha}_{ip}$	$s(\hat{\alpha}_{ip})$	$\hat{\beta}_{ip}$	$s(\hat{\beta}_{ip})$	R_i^2
1	-0.19	0.17	1.54	0.02	0.94
2	-0.19	0.10	1.37	0.01	0.97
3	-0.06	0.09	1.24	0.01	0.98
4	-0.09	0.07	1.17	0.01	0.98
5	-0.06	0.07	1.06	0.01	0.98
6	0.05	0.07	0.92	0.01	0.97
7	0.03	0.07	0.86	0.01	0.97
8	0.12	0.07	0.74	0.01	0.96
9	0.14	0.08	0.63	0.01	0.92
10	0.22	0.09	0.51	0.01	0.85

NOTE: For this sample period $\hat{\theta}_p$ and $\hat{\theta}^*$ are 0.166 and 0.227, respectively. These imply a value for W_u equal to 0.023, which has a p-value of 0.476. Under the hypothesis that the CRSP Equal-Weighted Index is efficient, $\mathscr{E}(\hat{W}_u)$ is 0.024 and $SD(\hat{W}_u)$ is 0.011.

absolute value of the slope is relevant for our null hypothesis, and our test is then based on the squared values.

Figure 1b is based on a data set that is very similar to the one used by Black, Jensen, and Scholes (1972) (hereafter, BJS).[7] Using monthly returns on 10 beta-sorted portfolios from January, 1931 through December, 1965, $\hat{\theta}^* = 0.227$ while the CRSP Equal-Weighted Index, which is portfolio p, has $\hat{\theta}_p = 0.166$. To judge whether these two slopes are statistically different, we can calculate $(\psi^2 - 1)$, which is 0.02333. Based on the results in Section 2, we can use a central F distribution with degrees of freedom 10 and 409 to judge the statistical significance of this difference in slopes. The resulting F statistic is 0.96, which has a p-value of 0.48. Our multivariate test confirms the conclusion reached by BJS for their overall time period in that the ex ante efficiency of the CRSP Equal-Weighted Index cannot be rejected; equivalently, if this Index is taken as the true market portfolio, then the Sharpe-Lintner version of the CAPM cannot be rejected. Table II provides some summary statistics on the beta-sorted portfolios that were used for Figure 1b. Table II, when compared with Table II in BJS, verifies that our data base is very similar to the one used by BJS.

BJS provide various scatter plots of average returns versus estimated betas to judge the fit of the data to the expected linear relation if the CRSP Equal-Weighted

---

[7] While BJS relied on the data from the Center for Research in Security Prices (hereafter, CRSP) at the University of Chicago, it is not possible to replicate their data. The CRSP tapes are continually revised to reflect data errors, and one would need the same version of the CRSP file to perfectly duplicate a data base. For example, we were able to find more firms per year than reported in Table 1 of BJS because of corrections to the data base. Also we relied on Ibbotson and Sinquefield (1979) for the return of US Treasury Bills as the riskless rate. This latter data base was not used by BJS. However, we followed the grouping procedure outlined in BJS in forming the 10 portfolios that were used in constructing Figure 1 and Table II.

Index is efficient. We view figures like our Figure 1b as complementary to these scatter plots, for they summarize the multivariate test in a manner familiar to financial economists. The advantage of the scatter plots in BJS is that they may provide some information as to which asset or which set of assets is least consistent with the hypothesis that the index is efficient; figures like Figure 1b really do not provide such information. On the other hand, the scatter plots in BJS can be difficult to interpret due to heteroscedasticity across the different portfolios as well as contemporaneous cross-sectional dependence. Section 6 will suggest some other types of diagnostic information based on the multivariate framework.

To understand further the behavior of our measure of efficiency, ψ^2, its small sample distribution given in Section 2 is helpful. Since a linear transform of ψ^2 has a central F distribution with degrees of freedom N and $(T - N - 1)$, we can use the first two moments of the central F to calculate:

$$(8) \qquad \mathscr{E}\left(\psi^2 - 1\right) = \left[\frac{N}{T - N - 3}\right]$$

and

$$(9) \qquad SD\left(\psi^2 - 1\right) = \left[\frac{1}{T - N - 3}\right]\sqrt{\frac{2N(T - 3)}{T - N - 5}} \ .$$

The first moment for ψ^2 only exists if $T > N + 3$ while the second moment for ψ^2 only exists if $T > N + 5$. These last two equations for the moments can be applied to the BJS data set for 1931–1965 where $N = 10$ and $T = 420$, so $\mathscr{E}(\psi^2 - 1)$ and the standard deviation of ψ^2 are 0.024 and 0.011, respectively. As the realized value of $\psi^2 - 1$ is less than its expectation, it is not surprising that the ex ante efficiency of the Equal-Weighted Index cannot be rejected for this time period.

This measure, ψ, is a new variant of the geometry developed to examine portfolio performance. In past procedures the efficient frontier has been taken as given, and a distance such as mb in Figure 1a has been used as a measure of p's performance. Note that mb is simply the return differential of the ex post optimal portfolio over p, computed at the sample standard deviation of the ex post optimal portfolio. Another suggestion has been to use the difference in their slopes $\hat{\theta}^* - \hat{\theta}_p$ as a measure of p's relative performance. How the true ex ante frontier is to be known is unclear, and if the ex post frontier is used, then we face the statistical problem of this paper.

4. THE POWER OF THE MULTIVARIATE TEST FOR EFFICIENCY

The empirical illustration in the previous section fails to reject the ex ante efficiency of the Equal-Weighted Index when using 10 beta-sorted portfolios as in BJS.[8] Such a result may occur because the null hypothesis is in fact true, or it

---

[8] We have also examined our data base using the same subperiods as in BJS. When we aggregate the results of the multivariate test across these four subperiods, we can reject ex ante efficiency at usual levels of significance. This confirms the conclusions reached by BJS.

may be due to the use of a test which is not powerful enough to detect economically important deviations from efficiency of the Index. Questions of power for various types of test statistics have been a long standing concern among financial economists (e.g., see Roll (1977), among others). This section will focus on the power of the multivariate test.

From Section 2 we know that under both the null and alternate hypotheses a simple transform of W, or ψ^2, has an F distribution with degrees of freedom N and $T - N - 1$. The F distribution is noncentral with the noncentrality parameter given by equation (6); under the null hypothesis the noncentrality parameter is zero. It deserves emphasis that the F distribution under the alternative is conditional on the returns of portfolio p since the noncentrality parameter depends on $\hat{\theta}_p^2$. Thus, we will be studying the power function conditional on a value for $\hat{\theta}_p^2$, not the unconditional power function.

The probability of rejecting a false null hypothesis increases as the noncentrality parameter increases, holding constant the numerator and denominator degrees of freedom (Johnson and Kotz (1970, page 193)). Studying the factors that affect the noncentrality parameter, λ, will give some guidance about the power of the multivariate test. From equation (6) we can see that λ is a weighted sum of squared deviations about the point $\alpha_p = 0$. The weighting matrix is the inverse of the covariance matrix of the ordinary least squares estimators for α_p. Thus, estimated departures from the null are weighted according to the variability of the estimator and the cross-sectional dependence among the estimators.

The noncentrality parameter can also be given an intuitive economic interpretation. The derivation of equation (23) in the Appendix would hold for the population counterparts of the sample estimates, so it is also true that $\alpha_p' \Sigma^{-1} \alpha_p = \theta^{*2} - \theta_p^2$. It follows directly that

$$\lambda = \left[T/\left(1 + \hat{\theta}_p^2 \right) \right] \left(\theta^{*2} - \theta_p^2 \right).$$

Not surprisingly, the power of the test will increase as the ex ante *in*efficiency of portfolio p increases as measured in terms of the slope of the relevant opportunity sets. If $\hat{\theta}_p^2$ increases, the precision of the estimator for α_p declines, so the power of the test decreases.

Figure 2 summarizes how the power of the test is affected by θ^* and θ_p. When the proportion of potential efficiency (i.e., θ_p/θ^*) is equal to one, the null hypothesis is true. As this proportion approaches zero, the given portfolio is becoming less efficient. Figure 2 is based on values for the significance level, N, and T that are common for existing empirical work on asset pricing models; we have used $N = 10$ or 20 and $T = 60$ or 120 and a five percent significance level. Empirical work on the CRSP Indexes reports estimates of θ_p between 0 and 0.4. We have used this range to guide our selection of a grid for θ_p and θ^*. In addition, Figure 2 is based on the assumption that $\hat{\theta}_p = \theta_p$ to eliminate one of the parameters that affect λ; this assumption suggests that our calculations of power are for situations where the sample is representative of the underlying population.

Even within the range of parameters that we consider, the probability of rejecting the null hypothesis ranges from five percent to nearly 100 percent depending on the difference between the two relevant measures of slope. For
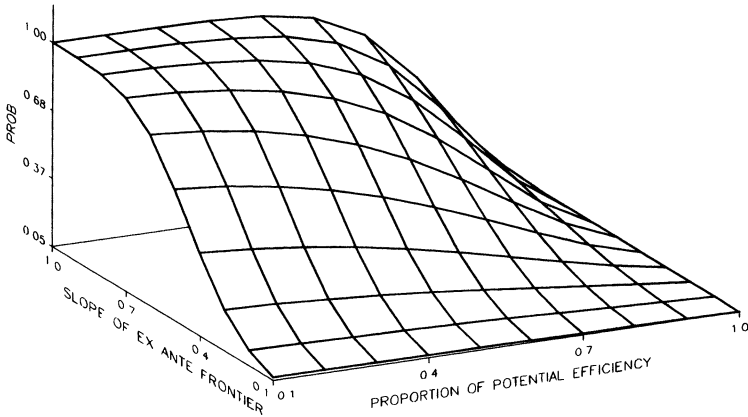
2a.) N = 10 and T = 60.
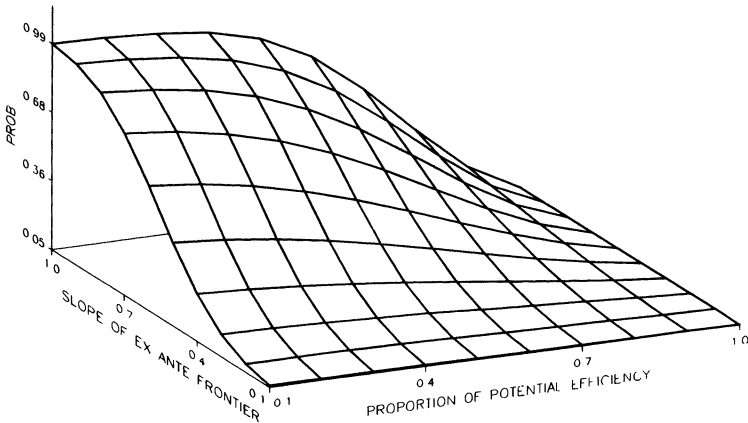
2b.) N = 20 and T = 60.

FIGURE 2.—Sensitivity of the power of the test to the choice of the index. Each figure is based on a different combination of the number of assets (N) and the number of time series observations (T). In all cases a critical level of five percent is used.

example, if θ_p equals .2 (which is high relative to the average from 1926–1982) and if $N = 20$ and $T = 60$, then the probability of rejecting a false null hypothesis ranges from ten percent (for $\theta^* = .3$) to 98 percent (for $\theta^* = 1.0$).

Given the data bases that are available, an empiricist is always faced with the question of the appropriate sizes for N and T. For example, with the CRSP monthly file we have a data base which extends back to 1926 for every firm on the New York Stock Exchange. This would permit the empiricist to use around 700 time series observations (i.e., T) and well over 2000 firms (i.e., N). However, the actual N used may be restricted by T to keep estimates of covariance matrices nonsingular, and the actual T used is constrained by concerns over
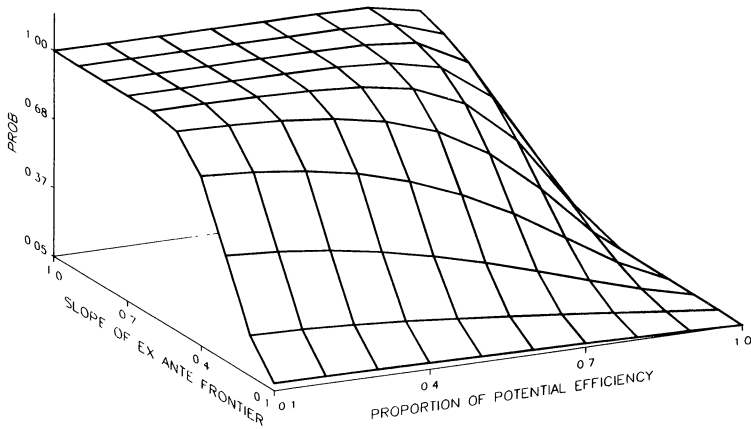
2c.) N = 10 and T = 120.
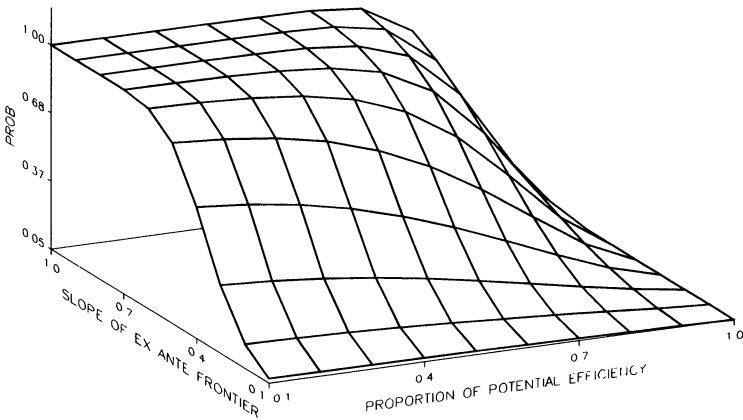
2d.) N = 20 and T = 120.

FIGURE 2.—Continued.

stationarity. It is not uncommon to see published work where T is around 60 monthly observations and N is between 10 and 20. While these numbers for N and T are common, we are not aware of any formal attempts to study the appropriate values to select. We will now examine this issue in the context of the specific hypothesis of ex ante efficiency. While the analysis is focused on an admittedly special case, our hope is that it may shed some light on other cases as well.

To get more intuition about the impact of N on equation (6), consider a case in which Σ has a constant value down the diagonal and a constant (but different) value for all off-diagonal elements. Since Σ represents the contemporaneous covariances across assets after the "market effect" has been removed, such a

structure includes the Sharpe (1963) diagonal model as a special case when the off-diagonal terms are zero. The more general case where the off-diagonal terms are constant but are nonzero is motivated by the work of Elton and Gruber (1973) and Elton, Gruber, and Urich (1978).[9] Under this structure we can parameterize Σ as:

$$(10) \qquad \Sigma = (1 - \rho)\omega^2 I_N + \rho\omega^2 \iota_N \iota_N',$$

where $\rho \equiv$ the correlation between $\tilde{\varepsilon}_{it}$ and $\tilde{\varepsilon}_{jt}$; $\omega^2 \equiv$ the variance of $\tilde{\varepsilon}_{it}$; $I_N \equiv$ an identity matrix of dimension N; and $\iota_N' \equiv$ a $1 \times N$ vector of 1's. The inverse[10] of this patterned matrix is (Graybill (1983)):

$$\Sigma^{-1} = \frac{1}{(1 - \rho)\omega^2}\left[I_N - \frac{\rho}{1 + (N - 1)\rho}\iota_N \iota_N'\right].$$

Substituting the above equation for Σ^{-1} in equation (6) gives:

$$(11) \qquad \lambda = \frac{T/\left(1 + \hat{\theta}_p^2\right)}{(1 - \rho)\omega^2}\left[N\mu_2 - \frac{N^2\rho}{(1 - \rho) + N\rho}\mu_1^2\right],$$

where $\mu_1 \equiv (\iota_N'\alpha_p)/N$ and $\mu_2 \equiv (\alpha_p'\alpha_p)/N$. One could view μ_1 as a measure of the "average" misspecification across assets while μ_2 indicates the noncentral dispersion of the departures from the null hypothesis across assets.

When N is relatively large and ρ is not equal to zero, equation (11) implies:

$$(12) \qquad \lambda/N \cong \frac{T/\left(1 + \hat{\theta}_p^2\right)}{(1 - \rho)\omega^2}\left(\mu_2 - \mu_1^2\right) = \frac{T/\left(1 + \hat{\theta}_p^2\right)}{(1 - \rho)\omega^2}\text{VAR}(\alpha_p),$$

where $\text{VAR}(\alpha_p)$ denotes the cross-sectional variance of the elements of α_p. Thus, λ is approximately proportional to N and T.[11] Alternatively, if either $\rho = 0$ or $\mu_1 = 0$, then λ is exactly proportional to N and T.[12] Unfortunately, this is still not adequate to determine the impact of changing N and T, for these two parameters affect not only the noncentrality parameter but also the degrees of freedom.

We have evaluated the power of the multivariate test for various combinations of λ, N, and T.[13] These numerical results provide some guidance on the proper

---

[9] Strictly speaking, the Sharpe diagonal model allows for heteroscedasticity in the disturbances of the market model equations; our formulation assumes homoscedasticity. Also, the constant correlation model of Elton and Gruber is usually applied to the correlation matrix for total returns; we are assuming constant correlation after the market effect has been removed.

[10] Necessary and sufficient conditions for this inverse to exist are that $\rho \neq 1$ and $\rho \neq (1 - N)^{-1}$; see Graybill (1983, page 190–191). In addition, the matrix should be positive definite; this would require that $\rho > -1/(N - 1)$.

[11] In general, since $\rho < 1$, λ/N is less than or equal to the right side of (12) when $\rho \geq 0$.

[12] If the Equal-Weighted Index is portfolio p, then we would expect μ_1 to approach zero as N becomes large.

[13] These numerical calculations require evaluation of a noncentral F distribution and an inverse of a central F distribution. The latter calculation relied on the MDFI subroutine provided by IMSL. The former calculations are based on a subroutine written by J. M. Bremner (1978), and a driver program written by R. Bohrer and T. Yancey of the University of Illinois at Champaign-Urbana. Each subroutine was checked by verifying its output with the published tables reported in Tang (1938) and Titu (1967).
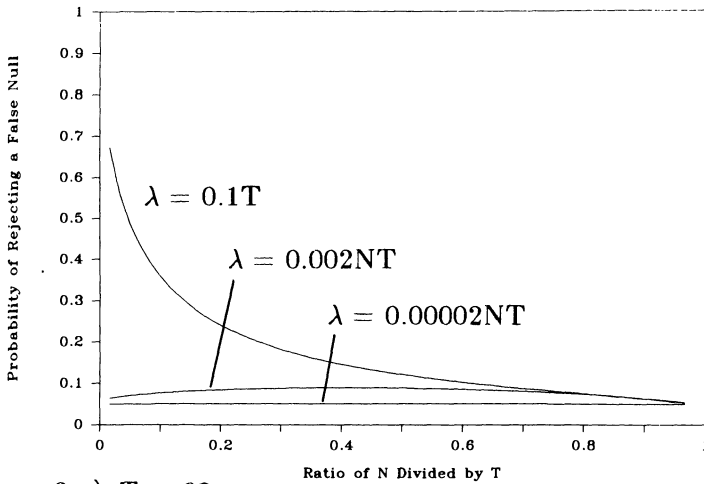
3a.) $T = 60$.
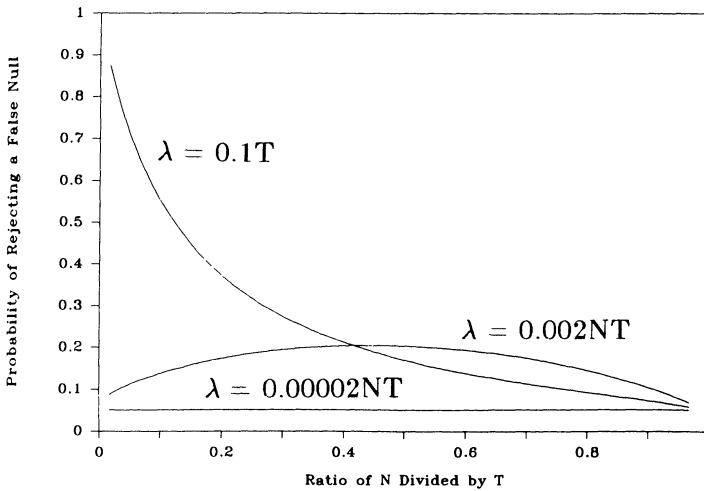
3b.) $T = 120$.

FIGURE 3.—Sensitivity of the power of the test to the choice of the number of assets (N) gives a fixed number of time series observations (T).

choice of N and T. We assumed that λ is proportional to NT, and Figure 3 provides various values for the constant of proportionality.[14] We selected this constant of proportionality based on equation (11) when $\rho = 0$. In this case, the constant is $\mu_2/[\omega^2(1 + \hat{\theta}_p^2)]$. We then replaced μ_2 and ω^2 with the cross-sectional averages of $\hat{\alpha}_{ip}^2$ and $\hat{\sigma}_i^2$, respectively, from an actual data set. We also know that $\hat{\theta}_p$ is 0.166 for the CRSP Equal-Weighted Index (1931–1965) and 0.109 for the

[14] MacKinlay (1987) studies the power of the test using alternative parameterizations of the noncentrality parameter.
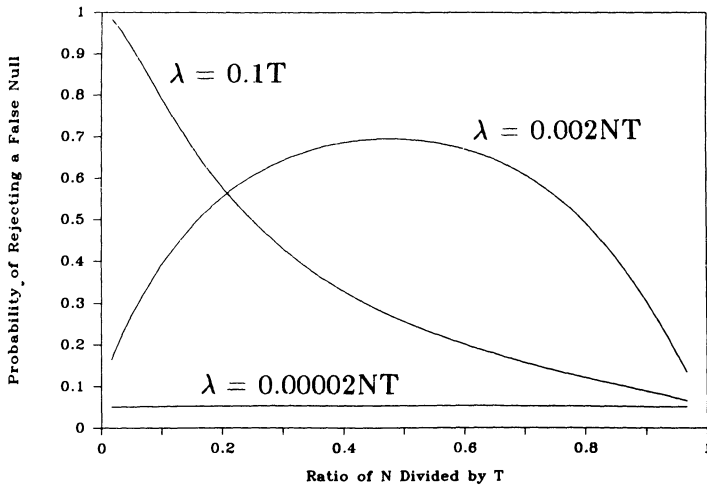
3c.) $T = 240$.

FIGURE 3.—Continued.

CRSP Value-Weighted Index (1926–1982). This provides a rough guide to typical values for the constant of proportionality. The constant is 0.004 using the beta-sorted portfolios, and it is 0.002 using a set of industry portfolios. For size-sorted portfolios the constant is 0.004 using all months and 0.763 for monthly data only using January. (The details on how the industry portfolios and size-sorted portfolios were created will be provided later in this paper.)

In Figure 3, we look at cases where the constant is 0.00002 and 0.002, which are small relative to the above calculations. For purposes of comparison, Figure 3 also includes a case where λ is not affected by N; instead we set $\lambda = .1T$. This represents a situation where an investigator has one asset that violates the null hypothesis, and all the remaining assets that are added are consistent with the efficiency of some given portfolio p. While Figure 3 is based on specific values for the constant of proportionality, the general pattern that is observed is consistent with a wide range of choices that we tried but did not report here.[15]

For a fixed number of time series observations, Figure 3 demonstrates that there may be an important decision to be made by the empiricist. Even though the noncentrality parameter increases as N increases, it is not necessarily appropriate to choose the maximum N possible. Given our particular parameterization of the problem, it appears that N should be roughly a third to one half of T, or when five years of monthly data are used, 20 to 30 assets may be appropriate. When the constant is very low, the power is so small for all possible values of N

---

[15] We were not able to evaluate the noncentral F for very high values of λ, so we have little knowledge about the shape of the power function when the constant of proportionality is high. If the constant is large enough, it is conceivable that a corner solution of setting $N = T - 2$ may be appropriate.

that it is not an important decision. Alternatively, if the noncentrality parameter is proportional to T and not affected by N, clearly setting $N = 1$ is the preferred strategy. In this case adding securities does not provide more information about departures from the null hypothesis; however, additional securities increase the number of unknown parameters to be estimated. It deserves emphasis that these conclusions about the proper choice of N may not be appropriate for all possible situations and models.

The choices of N and T are not the only decisions facing the empiricist in designing the econometric analysis. Since N must always be less than T (unless highly structured covariance matrices are entertained), the empiricist must also decide how to select the assets to maximize the power of the test. Given N and T we wish to maximize the quadratic form $\alpha_p' \Sigma^{-1} \alpha_p$, or equivalently θ^*; however, these parameters are unobservable. A common approach is to use beta-sorted portfolios. While dispersion in betas is useful in decreasing the asymptotic standard error in estimates of the expected return on the zero-beta asset (Gibbons (1980) and Shanken (1982)), such sorting need not maximize departures from the null hypothesis as measured by λ.[16]

Empirical examples presented below illustrate the effect that different asset sets can have on the outcome of the test. First, we consider a set of 12 industry portfolios.[17] An industry grouping seems reasonable on economic grounds and also captures some of the important correlations among stocks. To measure the return from a "buy-and-hold" investment strategy, the relative market values of the securities are used to weight the returns. Almost every monthly return on the CRSP tape from 1926 through 1982 is included, which should minimize problems with survivorship bias.[18] Table III provides some summary statistics on the industry portfolios.

The multivariate F statistic rejects the hypothesis of ex ante efficiency at about the one percent significance level. The relevant F statistic is 2.13 with degrees of freedom 12 and 671; its p-value is 0.013.[19] To complement these numerical results, Figure 1c, which is similar to Figures 1a and 1b, provides a geometrical summary.

[16] In fact, for a given set of N securities, the multivariate test is invariant to how we group these assets into N portfolios; we could form N portfolios so that they have very little dispersion in their beta values with no impact on the power. This follows from the well-known result in the multivariate statistics literature that our test is invariant to linear transformations of the data (Anderson (1984, pages 321–323)). Of course, the selection of the original subset of assets to be analyzed is important even though the way they are aggregated into portfolios is not (given that the number of portfolios is the same as the number of original assets).

[17] For the details of the data base, see Breeden, Gibbons, and Litzenberger (1987), who developed these data for tests of the consumption-based asset pricing model. The industry grouping closely follows a classification used by Sharpe (1982).

[18] However, all firms with a SIC number of 39 (i.e., miscellaneous manufacturing industries) are excluded to avoid any possible problems with a singular covariance matrix when the CRSP Value-Weighted Index is used as portfolio p.

[19] While not reported here, we also analyzed this data set across various subperiods. Based on five year subperiods, the p-value for the F statistic is less than five percent in 7 out of 11 cases, is less than 10 percent in 9 out of 11 cases, and rejects when aggregated across the subperiods. Thus, the rejection of the overall period is confirmed by the subperiods as well.

To understand this low p-value, consider the fact that for this time period $\hat{\theta}_p = 0.109$ while the slope of the opportunity set using the ex post optimal portfolio, $\hat{\theta}^*$, is more than double with a value of 0.224. With these numbers we can calculate ψ^2 as 1.038. For $N = 12$ and $T = 684$, $\mathscr{E}(\tilde{\psi}^2)$ is 1.018 with $SD(\tilde{\psi}^2)$ of 0.007. Thus, the realized value of ψ^2 is nearly three standard deviations from its expected value if the CRSP Value-Weighted Index is truly ex ante efficient.

Perhaps of greater interest is the fact that the multivariate test rejects the null hypothesis at the one percent level even though all 12 univariate t statistics fail to reject at even the five percent level. The next section builds on such contrasting results by analyzing why univariate test may be difficult to summarize across different assets.

5. THE PROBLEM WITH UNIVARIATE TESTS

Table II suggests that high beta portfolios earn too little and low beta portfolios too much if the Equal-Weighted Index is presumed to be efficient; similar evidence was used by BJS to garner support for the zero-beta version of the CAPM. Yet, this pattern is difficult to interpret. The upper triangular portion of Table IV provides the sample correlation matrix of the market model residuals based on the regressions that are summarized in Table II. A very distinctive pattern emerges in that the residuals of portfolios with similar betas are positively correlated while those of portfolios with very different betas are negatively correlated. Based on the variance-covariance matrix for $\hat{\alpha}_p$ in equation (4), it is clear that the estimators for α_{ip} will have the same pattern of correlation. Thus, it is difficult to infer whether the observed pattern in estimated values of α_{ip}'s is

TABLE III

SUMMARY STATISTICS ON INDUSTRY-SORTED PORTFOLIOS BASED ON MONTHLY DATA, 1926–82 ($T = 684$). ALL SIMPLE EXCESS RETURNS ARE NOMINAL AND IN PERCENTAGE FORM, AND THE CRSP VALUE-WEIGHTED INDEX IS PORTFOLIO p. THE FOLLOWING PARAMETER ESTIMATES ARE FOR THE REGRESSION MODEL: $\tilde{r}_{it} = \alpha_{ip} + \beta_{ip}\tilde{r}_{pt} + \tilde{\varepsilon}_{it}$ $\forall i = 1,\dots,12$ AND $\forall t = 1,\dots,684$, WHERE R_i^2 IS THE COEFFICIENT OF DETERMINATION FOR EQUATION i.

Industry Portfolio	$\hat{\alpha}_{ip}$	$s(\hat{\alpha}_{ip})$	$\hat{\beta}_{ip}$	$s(\hat{\beta}_{ip})$	R_i^2
Petroleum	0.17	0.14	0.93	0.02	0.69
Financial	−0.05	0.09	1.19	0.02	0.89
Consumer Durables	0.03	0.09	1.29	0.02	0.90
Basic Industries	0.00	0.00	1.09	0.01	0.94
Food and Tobacco	0.12	0.07	0.76	0.01	0.83
Construction	−0.17	0.12	1.20	0.02	0.85
Capital Goods	0.10	0.08	1.08	0.01	0.91
Transportation	−0.17	0.14	1.20	0.02	0.78
Utilities	0.05	0.09	0.74	0.02	0.76
Trade and Textiles	0.00	0.00	0.94	0.02	0.77
Services	0.43	0.37	0.80	0.06	0.19
Recreation	−0.03	0.13	1.22	0.02	0.78

NOTE: For this sample period $\hat{\theta}_p$ and $\hat{\theta}^*$ are 0.109 and 0.224, respectively. These imply a value for W_u equal to 0.038, which has a p-value of 0.013. Under the hypothesis that the CRSP Value-Weighted Index is efficient, $\mathscr{E}(\tilde{W}_u)$ is 0.018 and $SD(\tilde{W}_u)$ is 0.007.

TABLE IV

SAMPLE CORRELATION MATRIX OF RESIDUALS FROM MARKET MODEL REGRESSIONS USING EXCESS RETURNS.

THE UPPER TRIANGULAR PORTION OF THE TABLE IS BASED ON 10 BETA-SORTED PORTFOLIOS FOR THE DEPENDENT VARIABLES AND THE CRSP EQUAL-WEIGHTED INDEX FOR PORTFOLIO p. ALL MONTHLY DATA FROM 1931–65 ($T = 420$) ARE USED. TABLE II SUMMARIZES THE OTHER PARAMETER ESTIMATES FOR THIS REGRESSION MODEL. THE LOWER TRIANGULAR PORTION OF THE TABLE IS BASED ON 10 SIZE-SORTED PORTFOLIOS FOR THE DEPENDENT VARIABLES AND THE CRSP VALUE-WEIGHTED INDEX FOR PORTFOLIO p. ALL MONTHLY DATA FROM 1926–82 ($T = 684$) ARE USED. TABLE V SUMMARIZES THE OTHER PARAMETER ESTIMATES FOR THIS REGRESSION MODEL.

					Portfolio Number:				
1	2	3	4	5	6	7	8	9	10
	.52	.39	.03	−.32	−.51	−.64	−.60	−.64	−.50
.62		.38	.01	−.16	−.35	−.50	−.52	−.53	−.55
.72	.68		.08	−.06	−.25	−.37	−.43	−.46	−.49
.70	.66	.75		−.06	−.07	−.11	−.13	−.32	−.22
.63	.66	.70	.72		.21	.25	.12	.03	−.16
.63	.61	.68	.65	.72		.34	.36	.26	.06
.52	.41	.57	.55	.62	.67		.43	.46	.23
.39	.39	.51	.46	.50	.59	.52		.49	.27
.28	.35	.21	.27	.26	.36	.27	.30		.56
−.54	−.59	−.68	−.66	−.68	−.68	−.61	−.66	−.38	

NOTE: For the upper triangular portion of the table, portfolio 1 consists of firms with the highest values for historical estimates of beta while portfolio 10 contains the firms with the lowest values. For the lower triangular portion of the table, portfolio 1 is a value-weighted portfolio of firms whose market capitalization is in the lowest decile of the NYSE while portfolio 10 contains firms in the highest decile.

due to correlation in the estimation error or to the actual pattern in the true parameters.

Other examples from empirical work in financial economics could also be cited where univariate tests are difficult to interpret. Since the work of Banz (1981) and Reinganum (1981), the "size effect" has received a great deal of attention. (For more information about this research see Schwert (1983), who summarizes the existing evidence and also provides a useful bibliography.) While most of the research in this area now focuses on returns in January, we begin by looking at the original evidence which did not distinguish between January and non-January returns.

We have created a data base of monthly stock returns using the CRSP file. Firms were sorted into 10 portfolios based on the relative market value of their total equity outstanding. In other words, we ranked firms by their market values in December, 1925 (say), and we then formed 10 portfolios where the first portfolio contains all those firms in the lowest decile of firm size and the tenth portfolio consists of companies in the highest decile of firm size on the New York Stock Exchange. Each of the ten portfolios is value-weighted, and the firms are not resorted by their market values for five years. Thus, the returns on these 10 portfolios from January, 1926 through December, 1930 represent the returns from a buy-and-hold strategy without any rebalancing for five years; this portfolio formation was adopted to represent a low transaction cost investment strat-

TABLE V

SUMMARY STATISTICS ON SIZE-SORTED PORTFOLIOS BASED ON MONTHLY DATA, 1926–82 ($T = 684$). ALL SIMPLE EXCESS RETURNS ARE NOMINAL AND IN PERCENTAGE FORM, AND THE CRSP VALUE-WEIGHTED INDEX IS PORTFOLIO p. THE FOLLOWING PARAMETER ESTIMATES ARE FOR THE REGRESSION MODEL: $\tilde{r}_{it} = \alpha_{ip} + \beta_{ip}\tilde{r}_{pt} + \tilde{\varepsilon}_{it}$ $\forall i = 1, \ldots, 10$ AND $\forall t = 1, \ldots, 684$, WHERE R_i^2 IS THE COEFFICIENT OF DETERMINATION FOR EQUATION i.

Portfolio Number	$\hat{\alpha}_{ip}$	$s(\hat{\alpha}_{ip})$	$\hat{\beta}_{ip}$	$s(\hat{\beta}_{ip})$	R_i^2
1	0.28	0.24	1.59	0.04	0.68
2	0.34	0.18	1.45	0.03	0.78
3	0.25	0.14	1.40	0.02	0.84
4	0.18	0.13	1.36	0.02	0.85
5	0.19	0.10	1.27	0.02	0.89
6	0.18	0.09	1.25	0.02	0.91
7	0.08	0.07	1.18	0.01	0.93
8	0.08	0.06	1.17	0.01	0.95
9	0.00	0.00	1.16	0.01	0.96
10	− 0.01	0.05	0.94	0.00	0.98

NOTE: Portfolio 1 is a value-weighted portfolio of firms whose market capitalization is in the lowest decile of the NYSE while portfolio 10 contains firms in the highest decile. For this sample period $\hat{\theta}_p$ and $\hat{\theta}^*$ are 0.109 and 0.172, respectively. These imply a value for W_u equal to 0.017, which has a p-value of 0.301. Under the hypothesis that the CRSP Value-Weighted Index is efficient, $\mathscr{E}(\tilde{W}_u)$ is 0.015 and $SD(\tilde{W}_u)$ is 0.007.

egy. The resorting and rebalancing occurred in December of 1925, 1930, ..., 1980. Table V summarizes the behavior of the returns on these portfolios for the entire time period.

Given the existing evidence on the size effect, some readers may find it somewhat surprising that, in the overall period from 1926 through 1982, the multivariate test fails to reject efficiency of the CRSP Value-Weighted Index at the usual levels of significance. The first row of Table VI reports the statistic and its corresponding p-value; Figure 1d provides a geometrical interpretation for this overall period.[20]

The correlation matrix of the market model residuals of the size portfolios exhibits a distinctive pattern. The lower triangular portion of Table IV provides this information based on the overall period. However, the pattern is identical across every ten year subperiod reported in Table VI, and a similar pattern is also described by Brown, Kleidon, and Marsh (1983, page 47) and Huberman and Kandel (1985b). The correlation is positive and high among the low decile firms. The correlation declines as one compares portfolios from very different deciles. Even more striking is the fact that the highest decile portfolio has negative sample correlation with all other decile portfolios. (In some of the subperiods, this negative correlation occurred for the ninth decile as well.) Thus, if we observe that the lowest decile performs well (i.e., estimated alphas that are positive), we would then expect that the highest decile would do poorly (and vice versa). This is the case, for example, in the period 1946–1955, where five out of

---

[20] The subperiod results in Table VI are consistent with the conclusions of Brown, Kleidon, and Marsh (1983) who find the size effect is not constant across all subperiods.

TABLE VI

TESTING THE EX ANTE EFFICIENCY OF THE CRSP VALUE-WEIGHTED INDEX (I.E., PORTFOLIO p) RELATIVE TO 10 SIZE-SORTED PORTFOLIOS. ALL SIMPLE EXCESS RETURNS ARE NOMINAL AND IN PERCENTAGE FORM. OVERALL PERIOD IS BASED ON ALL MONTHLY DATA FROM 1926–82. THE FOLLOWING MODEL IS ESTIMATED AND TESTED: $\tilde{r}_{it} = \alpha_{ip} + \beta_{ip}\tilde{r}_{pt} + \tilde{\varepsilon}_{it}$ $\forall i = 1,\ldots,10$ AND $\forall t = 1,\ldots,T$. H_0: $\alpha_{ip} = 0 \ \forall i = 1,\ldots,10$.

Time Period (T)	$\hat{\theta}_p$	$\hat{\theta}^*$	W_u (P-Value)	Number of $\lvert t(\hat{\alpha}_{ip}) \rvert \geq 1.96$
1926–1982	0.109	0.172	0.018	1
(684)			(0.301)	
1926–1935	0.065	0.354	0.119	0
(120)			(0.227)	
1936–1945	0.146	0.286	0.059	0
(120)			(0.765)	
1946–1955	0.308	0.469	0.113	5
(120)			(0.264)	
1956–1965	0.216	0.604	0.302	1
(120)			(0.001)	
1966–1975	−0.019	0.408	0.164	0
(120)			(0.065)	
1976–1982	0.093	0.538	0.275	9
(84)			(0.039)	

NOTE: $\hat{\theta}_p$ is the ratio of the sample average excess return on the CRSP Value-Weighted Index divided by its sample standard deviation, and $\hat{\theta}^*$ is the maximum value possible of the ratio of the sample average excess return divided by the sample standard deviation. $W \equiv (\hat{\theta}^{*2} - \hat{\theta}_p^2)/(1 + \hat{\theta}_p^2)$, and it is distributed as a transform of a central F distribution with degrees of freedom 10 and $T - 11$ under the null hypothesis. W should converge to zero as T approaches infinity if the CRSP Value-Weighted Index is ex ante efficient. By converting the p-values for the W_u statistics to an implied realization for a standardized normal random variable, the results across the 6 subperiods can be summarized by summing up the 6 independent and standardized normals and dividing by the square root of 6 as suggested in Shanken (1985). This quantity is 2.87 which implies a rejection across the subperiods at the usual levels of significance.

ten portfolios have significant alphas (at the five percent level), but the multivariate test cannot reject the efficiency of the Value-Weighted Index.

Even though summarizing the results of univariate tests can be difficult, applied empirical work continues to report such statistics. This is only natural, for univariate tests are more intuitive (perhaps because they are used more) and seem to give more diagnostic information about the nature of the departure from the null hypothesis when it is rejected. Part of the goal of this paper is to provide some intuition behind multivariate tests. Section 3 has already done this to some extent by demonstrating that the multivariate test can be viewed as a particular measurement in mean-standard deviation space of portfolio theory. The next section shows that the multivariate test is equivalent to a "t test" on the intercept in a particular regression which should be intuitive. A way to generate diagnostic information about the nature of the departures from the null hypothesis is also provided.

6. ANOTHER INTERPRETATION OF THE TEST STATISTIC, W

The hypothesis that $\alpha_{ip} = 0 \ \forall i$ is violated if and only if some linear combination of the α's is zero; i.e., if and only if some portfolio of the N assets has a

nonzero intercept when its excess returns are regressed on those of portfolio p. With this in mind, it is interesting to consider the portfolio which, in a given sample, maximizes the square of the usual t statistic for the intercept. It is well known in the literature on multivariate statistics that this maximum value is Hotelling's T^2 statistic, our TW_u. In this section we focus on the composition of the maximizing portfolio, a, and its economic interpretation.

Thus, let $\tilde{r}_{at} \equiv a'\tilde{r}_{2t}$, where $\tilde{r}_{2t}$ is an $N \times 1$ vector with typical element $\tilde{r}_{it}$ $\forall i = 1, \ldots, N$. Let $\hat{\alpha}_p$ be the $N \times 1$ vector of regression intercept estimates. Then $\hat{\alpha}_a = a'\hat{\alpha}_p$ and $\widehat{VAR}(\hat{\alpha}_a) = (1 + \hat{\theta}_p^2) a'\hat{\Sigma} a / T$ by (4) above.[21] Therefore,

$$(13) \qquad t_a^2 = \left[\frac{\hat{\alpha}_a}{\widehat{SD}(\hat{\alpha}_a)} \right]^2 = \frac{(a'\hat{\alpha}_p)^2}{\widehat{VAR}(a'\hat{\alpha}_p)} = \frac{T(a'\hat{\alpha}_p)^2}{(1 + \hat{\theta}_p^2) a'\hat{\Sigma} a}.$$

Since we can multiply a by any scalar without changing the value of t_a^2, we shall adopt the normalization that $a'\hat{\alpha}_p = c$, where c is any constant different from zero. With this normalization, $T(a'\hat{\alpha}_p)^2$ and $(1 + \hat{\theta}_p^2)$ in (13) are fixed given the sample. Hence, maximizing t_a^2 is equivalent to the following minimization problem:

$$\min_a : \quad a'\hat{\Sigma} a$$

$$\text{subject to:} \quad a'\hat{\alpha}_p = c.$$

Since the above problem is similar to the standard portfolio problem, the form of the solution is:

$$a = \frac{c}{\hat{\alpha}_p'\hat{\Sigma}^{-1}\hat{\alpha}_p} \hat{\Sigma}^{-1}\hat{\alpha}_p.$$

Substituting this solution for a into equation (13), t_a^2 becomes:

$$t_a^2 = \frac{T\hat{\alpha}_p'\hat{\Sigma}^{-1}\hat{\alpha}_p}{1 + \hat{\theta}_p^2}.$$

Combining this equation with (5) establishes that $t_a^2 = TW_u$. Not surprisingly, the distribution of t_a is not Student t, for portfolio a was formed after examining the data.

The derivation of t_a^2 suggests some additional information to summarize empirical work on ex ante efficiency. Given the actual value of a based on the sample, one will know the particular linear combination which led to the rejection of the null hypothesis. If the null hypothesis is rejected, then a may give us some constructive information about how to create a better model.

Portfolio a has an economic basis as well. When this portfolio is combined properly with portfolio p, the combination turns out to be ex post efficient. In

[21] Since we are working with returns in excess of the riskless rate, $\iota_N' a$ need not equal 1, for the riskless asset will be held (long or short) so that all wealth is invested.

other words, for some value of k,

$$(14) \qquad \tilde{r}_t^* = (1-k)\tilde{r}_{pt} + k\tilde{r}_{at},$$

where $\tilde{r}_t^*$ is the return on this ex post efficient portfolio. For convenience, we set c so that the sample means of $\tilde{r}_{pt}$, $\tilde{r}_{at}$, and $\tilde{r}_t^*$ are all equal. The equivalence of these three means requires that:

$$c \equiv \bar{r}_p \frac{\hat{\alpha}_p'\hat{\Sigma}^{-1}\hat{\alpha}_p}{\hat{\alpha}_p'\hat{\Sigma}^{-1}\bar{r}_2}.$$

For the remainder of the paper, we will refer to portfolio a as the "active" portfolio. In many applications of our methodology, portfolio p will be a "passive" portfolio, i.e., a buy-and-hold investment strategy. While our methods are applicable to situations where portfolio p is not passive, certainly in its application to tests of the CAPM, portfolio p will be passive. In such a setting portfolio a is naturally interpreted as an active portfolio, for it represents a way to improve the efficiency of portfolio p. The terminology of "active" and "passive" has been used by Treynor and Black (1973), among others.

To establish this relation between the ex post efficient portfolio and portfolios a and p, we first recall the equation for the weights of an efficient portfolio, w^*. Using equation (22) in the Appendix and setting m in that equation to $\bar{r}_p$:

$$(15) \qquad w^* = \frac{\bar{r}_p}{\bar{r}'\hat{V}^{-1}\bar{r}}\hat{V}^{-1}\bar{r}.$$

$\hat{V}$ can be parameterized as:

$$\hat{V} = \begin{bmatrix} s_p^2 & s_p^2\hat{\beta}_p' \\ s_p^2\hat{\beta}_p & \hat{\beta}_p\hat{\beta}_p's_p^2 + \hat{\Sigma} \end{bmatrix}.$$

Using the formula for the inverse of a partitioned matrix (see equation (24) in the Appendix) on the last expression and substituting this into equation (15) for $\hat{V}^{-1}$, equation (14) can be derived after some tedious, but straightforward, algebra.

The previous paragraphs have established that the square of the usual t statistic for the estimated intercept, $\hat{\alpha}_a$, equals the T^2 test statistic, TW_u. A similar result can be established as well for the ex post efficient portfolio with the same sample mean as portfolio p, i.e., the estimated intercept, $\hat{\alpha}^*$, from regressing $\tilde{r}_t^*$ on $\tilde{r}_{pt}$ has a squared t statistic, t^{*2}, which is identical to t_a^2. Since we will not use this result in what follows, we only note the fact here without proof.[22]

To illustrate the usefulness of the active portfolio interpretation, we return to the example of Section 5 where the size effect (across all months) is examined. The second column of Table V is roughly consistent with the findings of Brown,

---

[22] The two key facts used in the proof are that $\alpha^* = k\alpha_a$ and $SD(\tilde{\varepsilon}_t^*) = kSD(\tilde{\varepsilon}_{at})$, where $\tilde{\varepsilon}_t^*$ is the disturbance in the regression of $\tilde{r}_t^*$ on $\tilde{r}_{pt}$; these equalities hold for the estimates as well as the parameters. Since t^* is essentially a ratio of $\hat{\alpha}^*$ and $\widehat{SD}(\hat{\varepsilon}_t^*)$, k cancels. The two key facts can be established by working with the moments of $\tilde{r}_t^*$ based on equation (14).

TABLE VII

DESCRIPTIVE INFORMATION ABOUT THE ACTIVE PORTFOLIO, a. THESE STATISTICS ARE BASED ON SIZE-SORTED PORTFOLIOS USING MONTHLY RETURNS, 1926–82 ($T = 684$). ALL SIMPLE EXCESS RETURNS ARE NOMINAL AND IN PERCENTAGE FORM, AND THE CRSP VALUE-WEIGHTED INDEX IS PORTFOLIO p.

	Monthly Returns in all Months ($T = 684$)	Monthly Returns in only January ($T = 57$)	Monthly Returns Excluding January ($T = 627$)
$\hat{\alpha}_a$	0.05	1.36	0.05
t_a^2	11.97	71.57	11.09
k	7.56	0.87	7.95
a_1	-0.02	0.09	-0.04
a_2	0.04	0.07	-0.01
a_3	0.04	0.08	0.03
a_4	0.00	0.07	-0.02
a_5	0.06	0.03	0.12
a_6	0.08	-0.04	0.06
a_7	0.01	0.10	0.03
a_8	0.10	0.04	0.09
a_9	0.04	-0.52	0.01
a_{10}	0.60	-0.16	0.63
a_{RF}	0.12	1.25	0.09
Σa_t	1.00	1.00	1.00

NOTE: Portfolio 1 is a value-weighted portfolio of firms whose market capitalization is in the lowest decile of the NYSE while portfolio 10 contains firms in the highest decile. The portion of wealth invested in the riskless asset is denoted by a_{RF}.

Kleidon, and Marsh in that the estimated alphas are approximately monotonic in the decile size rankings. However, such a result does not imply that an optimal portfolio should give large weight to small firms. As Dybvig and Ross (1985) point out, alphas only indicate the direction of investment for marginal improvements in a portfolio. The portfolio that is globally optimal may have a very different weighting scheme than is suggested by the alphas. A comparison of Tables V and VII verifies this.

For example, the portion of the active portfolio invested in the portfolio of the smallest firms (i.e., a_1) has a sign which is opposite that of its estimated alpha. Furthermore, the active portfolio suggests spreading one's investment fairly evenly across the portfolios in the bottom 9 deciles and then investing a rather large proportion in the portfolio of large firms, not small firms. Table VII also reports $\hat{\alpha}_a$, t_a^2, and k for the overall period. Note that as k is much greater than one ($k = 7.56$), the ex post efficient portfolio has a huge short position in the value-weighted index. Since this index is dominated by the largest firms, the net large firm position in the efficient portfolio is therefore actually negative. It is interesting that ex post efficiency is achieved by avoiding (i.e., shorting) large firms rather than aggressively investing in small firms.

The reader should keep in mind that Tables IV through VI and the second column of Table VII have examined the size effect across all months. Based on just these results, the size effect seems to be less important than perhaps originally thought. However, if the data are sorted by January returns versus

TABLE VIII

SUMMARY STATISTICS ON SIZE-SORTED PORTFOLIOS BASED ON JANUARY RETURNS, 1926–82 ($T = 57$). ALL SIMPLE EXCESS RETURNS ARE NOMINAL AND IN PERCENTAGE FORM, AND THE CRSP VALUE-WEIGHTED INDEX IS PORTFOLIO p. THE FOLLOWING PARAMETER ESTIMATES ARE FOR THE REGRESSION MODEL: $\tilde{r}_{it} = \alpha_{ip} + \beta_{ip}\tilde{r}_{pt} + \tilde{\varepsilon}_{it} \ \forall i = 1, \ldots, 10$, WHERE R_i^2 IS THE COEFFICIENT OF DETERMINATION FOR EQUATION i.

Portfolio Number	$\hat{\alpha}_{ip}$	$s(\hat{\alpha}_{ip})$	$\hat{\beta}_{ip}$	$s(\hat{\beta}_{ip})$	R_i^2
1	6.12	0.87	1.67	0.18	0.62
2	4.60	0.67	1.52	0.14	0.69
3	3.43	0.47	1.44	0.10	0.80
4	2.88	0.49	1.44	0.10	0.80
5	1.79	0.34	1.19	0.07	0.85
6	1.79	0.30	1.21	0.06	0.88
7	0.85	0.20	1.16	0.04	0.94
8	0.80	0.23	1.17	0.05	0.92
9	0.31	0.17	1.05	0.03	0.95
10	−0.52	0.10	0.94	0.02	0.98

NOTE: Portfolio 1 is a value-weighted portfolio of firms whose market capitalization is in the lowest decile of the NYSE while portfolio 10 contains firms in the highest decile. For this sample period $\hat{\theta}_p$ and $\hat{\theta}^*$ are 0.259 and 1.197, respectively. These imply a value for W_u equal to 1.256, which has a p-value of 0.000. Under the hypothesis that the CRSP Value-Weighted Index is efficient, $\mathscr{E}(\tilde{W}_u)$ is 0.219 and $SD(\tilde{W}_u)$ is 0.111.

non-January returns, the multivariate approach confirms the importance of the size effect—at least for the month of January. Table VIII summarizes the sample characteristics of our 10 size-sorted portfolios when using only returns in January from 1926 through 1982. A comparison of Tables V and VIII reveals that the size effect is much more pronounced in January than in other months; this is consistent with the work by Keim (1983). This impression from the univariate statistics is confirmed by the multivariate test of ex ante efficiency, for the F test is 5.99 with a p-value of zero to three decimal places. In contrast, the F test based on all months excluding January is 1.09 with a p-value of 0.36. The weights of the active portfolio, a, are presented in the last two columns of Table VII for January versus non-January months. As in the first column of Table VII, the active portfolio is not dominated by small firms. For the month of January, one's investment should be evenly spread (roughly speaking) across the eight portfolios in the bottom deciles (or smaller firms); however, firms in the top two deciles (or larger firms) should be shorted.[23] Results for non-January months are similar to those based on all monthly data.

The evidence in Table VII suggests that the optimal active portfolio is not dominated by small firms even in the month of January—at least based on the ex post sample moments. Nevertheless, in the marketplace we see the development of mutual funds which specialize in holding the equities of just small firms.[24]

---

[23] The active portfolio for the month of January involves a rather large position in the riskless asset (a_{RF} equals 1.25). This investment in the riskless security is necessary to maintain a sample mean return on the active portfolio equal to that of the CRSP Value-Weighted Index.

[24] Examples of such funds include The Small Company Portfolio of Dimensional Fund Advisors and the Extended Market Fund of Wells Fargo Investment Advisors.

Such funds suggest that efficient portfolios may be achieved by combining indexes like the S&P 500 (or the CRSP Value-Weighted Index) with a portfolio of small firms. We now turn to an examination of the ex ante efficiency of such a linear combination in the next section. A multivariate statistical test of such an investment strategy turns out to be a simple extension of the test developed in Section 2.

7. TESTING THE EFFICIENCY OF A PORTFOLIO OF L ASSETS

If a portfolio of L other portfolios is efficient, then there exist parameter restrictions on the joint distribution of excess returns similar to those considered earlier. Specifically, if $\tilde{r}_{xt} = \sum_{j=1}^{L} x_j \tilde{r}_{jt}$ (where $\sum_{j=1}^{L} x_j = 1$) and if $\tilde{r}_{xt}$ is efficient, then

$$(16) \qquad \mathscr{E}(\tilde{r}_{it}) = \sum_{j=1}^{L} \delta_{ij} \mathscr{E}(\tilde{r}_{jr}),$$

where the δ_{ij}'s are the coefficients in the following regression:

$$(17) \qquad \tilde{r}_{it} = \delta_{i0} + \sum_{j=1}^{L} \delta_{ij} \tilde{r}_{jt} + \tilde{\eta}_{it} \qquad \forall i = 1, \ldots, N.$$

(We will assume that the stochastic characteristics of $\tilde{\eta}_{it}$ are the same as those of $\tilde{\varepsilon}_{it}$ in equation (1).) Conversely, (16) implies that some portfolio of the given L portfolios is on the minimum variance frontier (Jobson and Korkie (1982)). Thus, a necessary condition for the efficiency of a linear combination $(\tilde{r}_{1t}, \tilde{r}_{2t}, \ldots, \tilde{r}_{Lt})$ with respect to the total set of $N + L$ risky assets is:

$$(18) \qquad H_0: \qquad \delta_{i0} = 0 \qquad \forall i = 1, \ldots, N.$$

The above null hypothesis follows when the parameter restriction given by (16) is imposed on (17).

In this case, $[T/N][(T-N-L)/(T-L-1)](1 + \bar{r}_p' \hat{\Omega}^{-1} \bar{r}_p)^{-1} \hat{\delta}_0' \hat{\Sigma}^{-1} \hat{\delta}_0$ has a noncentral F distribution with degrees of freedom N and $(T-N-L)$, where $\bar{r}_p$ is a vector of sample means for $\tilde{r}_{pt} \equiv (\tilde{r}_{1t}, \tilde{r}_{2t}, \ldots, \tilde{r}_{Lt})$, $\hat{\Omega}$ is the sample variance-covariance matrix for $\tilde{r}_{pt}$, δ_0 has a typical element δ_{i0}, and $\hat{\delta}_0$ is the least squares estimator for δ_0 based on the N regression equations in (17) above. Further, the noncentrality parameter is given by $[T/(1 + \bar{r}_p' \hat{\Omega}^{-1} \bar{r}_p)] \delta_0' \Sigma^{-1} \delta_0$. Under the null hypothesis (18), the noncentrality parameter is 0.

For an application of the methodology developed in this section, we return to the results based on the size-sorted portfolios using returns only during the month of January. In the previous section, we found that we could reject the ex ante efficiency of the CRSP Value-Weighted Index. It could be that there exists a linear combination of the lowest decile portfolio and the Value-Weighted Index which is efficient. To consider such a case, we set $L = 2$ and $N = 9$. (Since portfolio 1 has become a regressor in a system like (17), we can no longer use it as a dependent variable.) The F statistic to test hypothesis (18) is 1.09 with a

p-value of 0.39, so we cannot reject efficiency of this combination at the usual levels of significance. Of course, this inference ignores the obvious pre-test bias.

Throughout this paper we have assumed that there is an observable riskless rate of return, in which case the efficient frontier is simply a line in mean-standard deviation space. Suppose, now, that we wish to determine whether a set of $L + 1$ portfolios ($L \geq 1$) spans the minimum-variance frontier determined by these portfolios and the N other assets. The $N + L + 1$ asset returns are assumed to be linearly independent. If we observe the return on the "zero-beta" portfolio (which in practice we do not), this spanning hypothesis (with $L = 1$) naturally arises in the context of the zero-beta version of the CAPM due to Black (1972).[25]

To formulate the test for spanning for any $L \geq 1$, consider the system of regression equations,

$$(19) \qquad R_{it} = \delta_{i0} + \sum_{j=1}^{L+1} \delta_{ij} \tilde{R}_{jt} + \bar{\eta}_{it} \qquad \forall i = 1, \ldots, N,$$

where $\tilde{R}_{it}$ denotes total returns, not excess returns. Huberman and Kandel (1985a) observe that the spanning hypothesis is equivalent to the following restrictions:

$$(20) \qquad \delta_{i0} = 0 \qquad \forall i = 1, \ldots, N$$

and

$$(21) \qquad \sum_{j=1}^{L+1} \delta_{ij} = 1 \qquad \forall i = 1, \ldots, N.$$

Imposing (21) on the parameters in (19) and letting $\tilde{r}_{it}$ denote returns in excess of the returns on portfolio $L + 1$, we derive (17). Thus, the problem of testing (20) in the context of (17) is identical to that of testing (18) in the riskless case above. All we have learned about testing the riskless asset case is equally relevant to the spanning problem, provided that "excess returns" are interpreted appropriately. Perhaps most importantly, the exact distribution of our test statistic is known under both the null and alternative hypotheses, permitting evaluation of the power of the test. Note that this test of spanning imposes (21) and then assesses whether the intercepts in the resulting regression model are equal to zero.[26] In contrast, Huberman and Kandel (1985a) propose a joint F test of (20)

[25] More generally, suppose the $L + 1$st portfolio is uncorrelated with each of the first L portfolios and has minimum variance among all such orthogonal portfolios. A simple generalization of the argument in Fama (1976, page 373) establishes that $\delta_{iL+1} = 1 - \sum_{j=1}^{L} \delta_{ij}$ for all i. It then follows (details are available on request) from the results of Huberman and Kandel (1985a) that the $L + 1$ portfolios span the minimum-variance frontier if and only if some combination of the first L portfolios is on the frontier. Thus, a test of the latter hypothesis can be conducted as in this section provided that the minimum-variance orthogonal portfolio is observable.

[26] An intermediate approach would be to first test (21) directly and then, provided the null is not rejected, proceed to test (20). Once again, the test of (21) is an F test, and the exact distribution under the alternative may be determined along the lines of our earlier analysis. Of course, this test statistic does require that we observe the return on the $L + 1$ spanning portfolios.

and (21) against an unrestricted alternative; however, the distribution of this statistic has not been studied under the alternative.

8. SUMMARY AND FUTURE RESEARCH

While this paper focuses on a particular hypothesis from modern finance, this apparently narrow view is adopted to gain better insight about a broad class of financial models which have a very similar structure to the one that we examine. The null hypothesis of this paper is a central hypothesis common to all risk-based asset pricing theories.[27] The nature of financial data and theories suggests the use of multivariate statistical methods which are not necessarily intuitive. We have attempted to provide some insight into how such tests function and to explain why they may provide different answers relative to univariate tests that are applied in an informal manner. In addition, we have studied the power of our suggested statistic and have isolated factors which will change the power of the test. There are at least two natural extensions of this work, and we now discuss each in turn.

First, the multivariate test considered here requires that the number of assets under study always be less than the number of time series observations. This restriction is imposed so that the sample variance-covariance matrix remains nonsingular. A test statistic which could handle situations with a large number of assets would be interesting.[28]

Second, we have not been very careful to specify the information set on which the various moments are conditioned. Gibbons and Ferson (1985), Grossman and Shiller (1982), and Hansen and Singleton (1982, 1983) have emphasized the importance of this issue for empirical work on positive models of asset pricing. Our methods provide a test of the ex ante *unconditional* efficiency of some portfolio—that is, when the opportunity set is constructed from the unconditional moments, not the conditional moments. When the riskless rate is changing (as it is in all of our data sets), then our methods provide a test of the *conditional* efficiency of some portfolio given the riskless rate. Of course, such an interpretation presumes that our implicit model for conditional moments given the riskless rate is correct. Ferson, Kandel, and Stambaugh (1987) and Shanken (1987a) provide more detailed analysis of testing conditional mean-variance efficiency.[29]

[27] If there is no riskless asset, then the null hypothesis becomes nonlinear in the parameters, for the intercept term is proportional to $(1 - \beta_{ip})$. Gibbons (1982) has explored this hypothesis using statistics which only have asymptotic justification. These statistics have been given an elegant geometric interpretation by Kandel (1984). While we still do not have a complete characterization of the small sample theory, Shanken (1985, 1986) has provided some useful bounds for the finite sample behavior of these tests.

[28] See Affleck-Graves and McDonald (1988) for some preliminary work on this problem.

[29] As Hansen and Richard (1987) emphasize, efficiency relative to a given information set need not imply efficiency relative to a subset. This implication does hold given some additional (and admittedly restrictive) assumptions, however. Let the information set, I, include the riskless rate, and let p be efficient, given I. Assume betas conditional on I are constant and $\mathscr{E}(\tilde{r}_{it}|r_{pt}, I)$ is linear in r_{pt}. It follows that $\mathscr{E}(\tilde{r}_{it}|r_{pt}, I) = \beta_{ip}r_{pt}$, and by iterated expectations $\mathscr{E}(\tilde{r}_{it}|r_{pt}, R_{ft}) = \beta_{ip}r_{pt}$, where R_{ft} is the riskless rate. Thus, p is on the minimum-variance frontier, given R_{ft}, and the methods of this paper are applicable.

Graduate School of Business, Stanford University, Stanford, CA, U.S.A.,

School of Organization and Management, Yale University, New Haven, CT, U.S.A.,

and

Simon School of Business Administration, University of Rochester, Rochester, NY, U.S.A.

APPENDIX

DERIVATION OF EQUATION (7)

To understand the derivation of (7), first consider the basic portfolio problem:

$$\text{min:} \quad w'\hat{V}w$$

subject to $w'\bar{r} = m$, a mean constraint, where $w \equiv$ the vector of $N + 1$ portfolio weights; $\hat{V} \equiv$ the variance-covariance matrix of $N + 1$ assets; and $\bar{r} \equiv$ the vector of $N + 1$ sample mean excess returns. Without loss of generality, we assume that p itself is the first component of our excess return vector. Thus, $\bar{r}' = (\bar{r}_p, \bar{r}_2')$ where $\bar{r}_2$ is a column vector of mean excess returns on the original N assets. The first-order conditions for this problem are:

$$(22) \qquad w = \varphi \hat{V}^{-1}\bar{r}$$

and

$$\varphi = \frac{m}{\bar{r}'\hat{V}^{-1}\bar{r}},$$

where φ is the Lagrange multiplier. Hence,

$$\left[\frac{\text{mean}}{\text{standard deviation}}\right]^2 = \frac{m^2}{w'\hat{V}w}$$

$$= \frac{m^2}{\left[\dfrac{m}{\bar{r}'\hat{V}^{-1}\bar{r}}\right]^2 \bar{r}'\hat{V}^{-1}\bar{r}}$$

$$= \bar{r}'\hat{V}^{-1}\bar{r}$$

$$\equiv \hat{\theta}*^2.$$

Finally, to arrive at (7) we need to establish that:

$$(23) \qquad \hat{\alpha}_p' \hat{\Sigma}^{-1} \hat{\alpha}_p = \hat{\theta}*^2 - \hat{\theta}_p^2,$$

where in contrast to the rest of the paper $\hat{\Sigma}$ is now the maximum likelihood estimator. The last equality follows from rewriting the elements of $\hat{V}$ in terms of s_p^2, $\hat{\beta}_p$, and $\hat{\Sigma}$ and then finding $\hat{V}^{-1}$ using the formula for a partitioned inverse. These steps lead to:

$$(24) \qquad \hat{V}^{-1} = \begin{bmatrix} s_p^{-2} + \hat{\beta}_p' \hat{\Sigma}^{-1} \hat{\beta}_p & -\hat{\beta}_p' \hat{\Sigma}^{-1} \\ -\hat{\Sigma}^{-1} \hat{\beta}_p & \hat{\Sigma}^{-1} \end{bmatrix}.$$

Then straightforward algebra yields:

$$\bar{r}'\hat{V}^{-1}\bar{r} = \left(\bar{r}_p^2/s_p^2\right) + \left[\left(\bar{r}_2 - \hat{\beta}_p\bar{r}_p\right)'\hat{\Sigma}^{-1}\left(\bar{r}_2 - \hat{\beta}_p\bar{r}_p\right)\right].$$

Since $\hat{\alpha}_p = \bar{r}_2 - \hat{\beta}_p \bar{r}_p$ and since the first term on the left-hand side of the above equation is $\hat{\theta}^{*2}$ and the first term on the right-hand side is $\hat{\theta}_p^2$, we can rewrite the last equation as:

$$\hat{\theta}^{*2} = \hat{\theta}_p^2 + \hat{\alpha}_p' \hat{\Sigma}^{-1} \hat{\alpha}_p$$

or

$$\hat{\alpha}_p' \hat{\Sigma}^{-1} \hat{\alpha}_p = \hat{\theta}^{*2} - \hat{\theta}_p^2.$$

Thus,

$$W = \frac{\hat{\theta}^{*2} - \hat{\theta}_p^2}{1 + \hat{\theta}_p^2} = \psi^2 - 1,$$

and the equality given in (7) has been justified.

REFERENCES

AFFLECK-GRAVES, J., AND B. MCDONALD (1988): "Multivariate Tests of Asset Pricing: The Comparative Power of Alternative Statistics," Working Paper, University of Notre Dame.

ANDERSON, T. W. (1984): *An Introduction to Multivariate Statistical Analysis*, Second Edition. New York: John Wiley.

BANZ, R. W. (1981): "The Relationship between Return and Market Value of Common Stock," *Journal of Financial Economics*, 9, 3–18.

BERNDT, E. R., AND N. E. SAVIN (1977): "Conflict among Criteria for Testing Hypotheses in the Multivariate Linear Regression Model," *Econometrica*, 45, 1263–1277.

BLACK, F. (1972): "Capital Market Equilibrium with Restricted Borrowing," *The Journal of Business*, 45, 444–454.

BLACK, F., M. C. JENSEN, AND M. SCHOLES (1972): "The Capital Asset Pricing Model: Some Empirical Findings," in *Studies in the Theory of Capital Markets*, ed. by M. C. Jensen. New York: Praeger.

BREEDEN, D. T., M. R. GIBBONS, AND R. H. LITZENBERGER (1987): "Empirical Tests of the Consumption-Oriented CAPM," Research Paper #879, Graduate School of Business, Stanford University, Stanford, CA.

BREMNER, J. M. (1978): "Mixtures of Beta Distributions, Algorithm AS 123," *Applied Statistics*, 27, 104–109.

BROWN, P., A. W. KLEIDON, AND T. A. MARSH (1983): "New Evidence on the Nature of Size-Related Anomalies in Stock Prices," *Journal of Financial Economics*, 12, 33–56.

DYBVIG, P. H., AND S. A. ROSS (1985): "The Analytics of Performance Measurement Using a Security Market Line," *Journal of Finance*, 40, 401–416.

ELTON, E. J., AND M. J. GRUBER (1973): "Estimating the Dependence Structure of Share Prices— Implications for Portfolio Selection," *Journal of Finance*, 28, 1203–1232.

ELTON, E. J., M. J. GRUBER, AND T. J. URICH (1978): "Are Betas Best?" *Journal of Finance*, 33, 1375–1384.

FAMA, E. F. (1976): *Foundations of Finance*. New York: Basic Books.

FAMA, E. F., AND J. D. MACBETH (1973): "Risk, Return, and Equilibrium: Empirical Tests," *Journal of Political Economy*, 81, 607–636.

FERSON, W., S. KANDEL, AND R. STAMBAUGH (1987): "Tests of Asset Pricing with Time-Varying Expected Risk Premiums and Market Betas," *Journal of Finance*, 42, 201–220.

GIBBONS, M. R. (1980): "Estimating the Parameters of the Capital Asset Pricing Model— A Minimum Expected Loss Approach," Research Paper #565, Graduate School of Business, Stanford University, Stanford, CA.

——— (1982): "Multivariate Tests of Financial Models: A New Approach," *Journal of Financial Economics*, 10, 3–28.

GIBBONS, M. R., AND W. FERSON (1985): "Testing Asset Pricing Models with Changing Expectations and an Unobservable Market Portfolio," *Journal of Financial Economics*, 14, 217–236.

GRAYBILL, F. A. (1983): *Matrices with Applications in Statistics*, Second Edition. Belmont, CA: Wadsworth.

GROSSMAN, S., AND R. SHILLER (1982): "Consumption Correlatedness and Risk Measurement in Economies with Non-Traded Assets and Heterogeneous Information," *Journal of Financial Economics*, 10, 195–210.

HANSEN, L., AND S. RICHARD (1987): "The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models," *Econometrica*, 55, 587–614.

HANSEN, L., AND K. SINGLETON (1982): "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, 50, 1269–1286.

——— (1983): "Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Stock Market Returns," *Journal of Political Economy*, 91, 249–265.

HUBERMAN, G., AND S. KANDEL (1985A): "Likelihood Ratio Tests of Asset Pricing and Mutual Fund Separation," Graduate School of Business, University of Chicago, Chicago, IL.

——— (1985b): "A Size Based Stock Returns Model," Graduate School of Business, University of Chicago, Chicago, IL.

IBBOTSON, R., AND R. SINQUEFIELD (1979): *Stocks, Bonds, Bills and Inflation: Historical Returns (1926–1978)*. Charlottesville, VA: Financial Analysts Research Foundation.

JOBSON, J. D., AND B. KORKIE (1982): "Potential Performance and Tests of Portfolio Efficiency," *Journal of Financial Economics*, 10, 433–466.

——— (1985): "Some Tests of Linear Asset Pricing with Multivariate Normality," *Canadian Journal of Administrative Sciences*, 2, 114–138.

JOHNSON, N. L., AND S. KOTZ (1970): *Continuous Univariate Distributions*, Volume 2. New York: John Wiley.

KANDEL, S. (1984): "The Likelihood Ratio Test Statistic of Mean-Variance Efficiency without a Riskless Asset," *Journal of Financial Economics*, 13, 575–592.

KANDEL, S., AND R. STAMBAUGH (1987): "On Correlations and Inferences about Mean-Variance Efficiency," *Journal of Financial Economics*, 18, 61–90.

KEIM, D. B. (1983): "Size-Related Anomalies and Stock Return Seasonality: Further Empirical Evidence," *Journal of Financial Economics*, 12, 13–32.

LINTNER, J. (1965): "The Valuation of Risk Assets and the Selection of Risky Investment in Stock Portfolios and Capital Budgets," *Review of Economics and Statistics*, 47, 13–37.

MACBETH, J. (1975): "Tests of Two Parameter Models of Capital Market Equilibrium," Ph.D. Dissertation, Graduate School of Business, University of Chicago, Chicago, IL.

MACKINLAY, A. C. (1985): "An Analysis of Multivariate Financial Tests," Ph.D. Dissertation, Graduate School of Business, University of Chicago, Chicago, IL.

——— (1987): "On Multivariate Tests of the CAPM," *Journal of Financial Economics*, 18, 341–372.

MALINVAUD, E. (1980): *Statistical Methods of Econometrics*, Third Edition. Amsterdam: North Holland Publishing.

MERTON, R. (1973): "An Intertemporal Capital Asset Pricing Model," *Econometrica*, 41, 867–887.

MORRISON, D. F. (1976): *Multivariate Statistical Methods*, Second Edition. New York: McGraw-Hill.

RAO, C. R. (1951): "An Asymptotic Expansion of the Distribution of Wilks Criterion," *Bulletin of the Institute of International Statistics*, 33, 33–38.

REINGANUM, M. R. (1981): "Misspecification of Capital Asset Pricing: Empirical Anomalies Based on Earnings Yields and Market Values," *Journal of Financial Economics*, 9, 19–46.

ROLL, R. (1977): "A Critique of the Asset Pricing Theory's Tests—Part 1: On Past and Potential Testability of the Theory," *Journal of Financial Economics*, 4, 129–176.

——— (1978): "Ambiguity When Performance is Measured by the Securities Market Line," *Journal of Finance*, 33, 1051–1069.

SCHWERT, G. W. (1983): "Size and Stock Returns, and Other Empirical Regularities," *Journal of Financial Economics*, 12, 3–12.

SHANKEN, J. (1982): "An Analysis of the Traditional Risk-Return Model," Ph.D. Dissertation, Graduate School of Industrial Administration, Carnegie-Mellon University, Pittsburgh, PA.

——— (1985): "Multivariate Tests of the Zero-Beta CAPM," *Journal of Financial Economics*, 14, 327–348.

——— (1986): "Testing Portfolio Efficiency when the Zero-Beta Rate Is Unknown: A Note," *Journal of Finance*, 41, 269–276.

——— (1987a): "The Intertemporal Capital Asset Pricing Model: An Empirical Investigation," Working Paper, University of Rochester.

——— (1987b): "Multivariate Proxies and Asset Pricing Relations: Living with the Roll Critique," *Journal of Financial Economics*, 18, 91–110.

SHARPE, W. F. (1963): "A Simplified Model of Portfolio Analysis," *Management Science*, 9, 277–293.

——— (1964): "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk," *Journal of Finance*, 19, 425–442.

——— (1982): "Factors in New York Stock Exchange Security Returns, 1931–79," *Journal of Portfolio Management*, 8, 5–19.

SILVEY, S. D. (1975): *Statistical Inference*. London: Chapman and Hall.

STAMBAUGH, R. F. (1982): "On the Exclusion of Assets from Tests of the Two-Parameter Model: A Sensitivity Analysis," *Journal of Financial Economics*, 10, 237–268.

TANG, P. C. (1938): "The Power Function of the Analysis of Variance Tests with Tables and Illustrations of Their Use," *Statistical Research Memoirs*, 2, 126–150.

TIKU, M. L. (1967): "Tables of the Power of the F-Test," *Journal of the American Statistical Association*, 62, 525–539.

TREYNOR, J. L., AND F. BLACK (1973): "How to Use Security Analysis to Improve Portfolio Selection," *Journal of Business*, 46, 66–86.