Journal of Financial Economics 10 (1982) 433–466. North-Holland Publishing Company

POTENTIAL PERFORMANCE AND TESTS OF PORTFOLIO EFFICIENCY

J.D. JOBSON and Bob KORKIE*

University of Alberta, Edmonton, Canada T6G 2G1

Received September 1981, revised version received September 1982

The potential performance of an asset set may be obtained by choosing the portfolio proportions to maximize the Sharpe (1966) performance measure. If a portfolio has a Sharpe measure equivalent to the potential performance of the underlying set of assets, then it is efficient Multivariate statistical procedures for comparing potential performance and testing portfolio efficiency are developed and then evaluated using simulations. Two likelihood ratio statistics are then used to compare stock and bond indices against sets of 20 and 40 portfolios The procedures are also compared to the Gibbons (1982) methodology for testing financial models

1. Introduction

Mean-standard deviation efficiency and the related concept of performance evaluation has been of significant interest to financial economists since the originating work of Markowitz (1952) and Tobin (1958). This paper extends the available set of evaluation techniques by proposing a performance measurement procedure with reasonable statistical properties which utilizes the efficient set constants of Merton (1972) and Roll (1977). The procedure is consistent with the Sharpe (1966) performance measure, is a generalization of the work of Ross (1980), and is an extension of Jobson and Korkie (1980, 1981). It also employs the likelihood ratio methodology, first applied to financial economics by Gibbons (1982).

The distinctive feature of the proposed evaluation is its comparison of the maximum attainable Sharpe performance (henceforth potential performance) of an asset set with the potential performance of an asset subset. Some apparent uses of the technique are the quantification of the performance contribution made by additional assets, the efficiency evaluation of a portfolio or market index, and for tests of multifactor capital asset pricing models. A disadvantage of the proposed technique is that it requires knowledge of the history of returns on the individual components of the

*The authors are grateful for the helpful comments of Giovanni Barone Adesi, Stephen Beveridge, Steve Hindmarch and the referee, Michael Gibbons. Part of this research was completed while J D. Jobson was a visiting scholar in the Department of Biostatistics, University of North Carolina.

0304–405X/82/0000–0000/$02.75 © North-Holland

portfolio. However, this may not present an encumbrance in many appraisals.

Section 2 derives the potential performance measure from efficient set mathematics and discusses some hypotheses germane to mean-standard deviation efficiency and portfolio performance. Section 3 derives the test statistics for the hypotheses and relates the procedures to other methodologies for testing financial models. Section 4 presents the results of a Monte Carlo investigation into the small sample behavior of the proposed test statistics. Section 5 illustrates the performance measures by determining the effects of bonds and stock indices on potential performance and by testing the efficiency of selected portfolios. Section 5 also compares the new procedures to the methodology of Gibbons (1982) using similar data. Section 6 concludes the paper.

2. Potential performance and the efficiency hypotheses

2.1. Potential performance

The efficient set of portfolios is comprised of the portfolios that minimize portfolio variance for a given mean excess return μ_p, subject to the constraint that investment proportions sum to one. In the presence of a riskless asset, the efficient set becomes the set of linear combinations of the riskless asset and a unique risky asset portfolio m.

Given a population of N assets with mean excess return vector $\boldsymbol{\mu}_{(N\times 1)}$ and covariance matrix $\boldsymbol{\Sigma}_{(N\times N)}$, the vector $\boldsymbol{X}_{m(N\times 1)}$ of risky asset proportions is obtained from minimization of the Lagrangian

$$L := \boldsymbol{X}_m'\boldsymbol{\Sigma}\boldsymbol{X}_m - \lambda_1(\boldsymbol{X}_m'\boldsymbol{\mu} - \mu_p) - \lambda_2(\boldsymbol{X}_m'\boldsymbol{e} - 1),$$

where λ_1 and λ_2 are the multipliers and $\boldsymbol{e}_{(N\times 1)}$ is the unit vector.

The first extremum conditions provide the proportions vector

$$\boldsymbol{X}_m = \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}/\boldsymbol{e}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu},$$

which forms the familiar tangency portfolio in mean-standard deviation space. The mean excess return on the tangency portfolio and its return variance are therefore

$$\mu_m = \boldsymbol{\mu}'\boldsymbol{X}_m = \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}/\boldsymbol{e}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} = a/b,$$

and

$$\sigma_m^2 = \boldsymbol{X}_m'\boldsymbol{\Sigma}\boldsymbol{X}_m = \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}/(\boldsymbol{e}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})^2 = a/b^2,$$

where the efficient set constants are

$$a=\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \quad \text{and} \quad b=\boldsymbol{e}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}.$$

The Sharpe measure of performance for any portfolio p with proportions $\boldsymbol{X}_p$ is

$$Sh_p=\boldsymbol{X}_p'\boldsymbol{\mu}/(\boldsymbol{X}_p'\boldsymbol{\Sigma}\boldsymbol{X}_p)^{\frac{1}{2}}=\mu_p/\sigma_p.$$

The vector $\boldsymbol{X}_m$ is also the value of $\boldsymbol{X}_p$ which maximizes the Sharpe performance over all portfolios is given by

$$Sh_m=\mu_m/\sigma_m=(a/b)(b/\sqrt{a})=\sqrt{a},$$

which is illustrated in fig. 1. Note that the investment proportions of p or of m are not directly required for the performance calculations.

In conclusion for any set of assets, the square of the reward to variability ratio of the efficient portfolio of risky assets is given simply by the efficient set constant 'a'. In the remainder of the paper, the performance measure 'a' or $\sqrt{a}$ is referred to as the potential performance of a one-period buy and hold portfolio which is constructed from the N asset population.

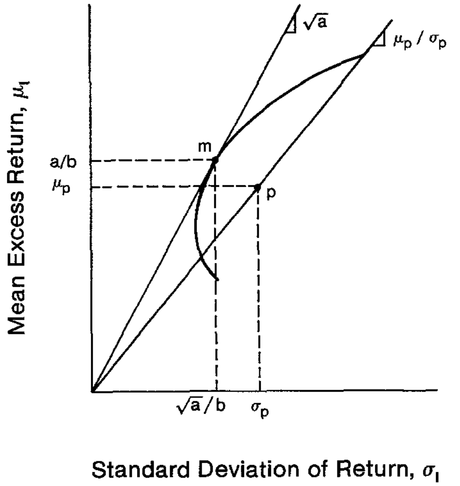

Fig 1 Slopes of rays through the origin measuring the potential performance $\sqrt{a}$ and Sharpe's performance μ_p/σ_p, in a mean-standard deviation space

2.2. The efficiency hypotheses

The distinction has been made between the Sharpe performance of an N-asset portfolio and the potential performance of the N-asset set. This distinction provides a useful framework for the analysis of portfolio efficiency.

The portfolio p will have poor parametric performance, relative to its potential, if it is not the unique buy and hold portfolio m with proportions X_m. Thus, μ_p/σ_p is less than $\sqrt{a}$ if the entire N-asset set is not in the portfolio or the N assets are not held in the correct proportions X_m. The portfolio p cannot have performance exceeding $\sqrt{a}$ unless the portfolio is actively managed in the holding period. However, active portfolio management, via timing and selectivity, is not fruitful if the conditional mean vector and covariance matrix are time stationary.

The cause of a portfolio's inefficiency may be ascertained providing the composition of the portfolio is known. Denote the N_1 non-zero weight assets of the portfolio as the set Γ_1, which is a proper subset of the set Γ of all N assets. Similarly, denote the potential performance of the asset sets Γ_1 and Γ as a_1 and 'a', respectively. The first question of interest is whether the potential performances of the two asset sets are identical. That is, a test of the comparative potential performance hypothesis,

$$H_{01}:a_1=a,$$

is required. The hypothesis H_{01} determines if the cause of the inefficiency is due to the selection of assets. If H_{01} is accepted then we may conclude that the N_1 assets are jointly efficient with respect to the complete set of N assets. In section 3, it is demonstrated that under certain assumptions the hypothesis H_{01} is related to the hypothesis tested by Gibbons (1982) with reference to a general class of financial models. The hypothesis H_{01} can also be related to a test of the arbitrage pricing theory as in Jobson (1982a).[1]

A second hypothesis of interest is whether the portfolio performance μ_p^2/σ_p^2 is equivalent to the potential performance, or

$$H_{02}:\mu_p^2/\sigma_p^2=a.$$

This hypothesis determines if the inefficiency is due to an incorrect selection of weights (including some that are zero) for the N assets. Hypothesis H_{02} may be used to determine the relative efficiency of a pseudo market index I.

[1]Under certain assumptions the test of H_{01} is equivalent to a test of the multi-factor capital asset pricing model. Given a set of k factors or portfolios constructed from the N assets a new set of assets Γ consisting of the k factors and a subset of $(N-k)$ of the original assets can be constructed The comparison of the performance potential of the k factors to the performance potential of the set Γ is a test of a multi-factor capital asset pricing model

That is, I is efficient relative to the set of N assets if it is impossible to construct a portfolio combining I and the N assets such that the resulting potential performance $\sqrt{a}$ exceeds the index's performance μ_I/σ_I. It is demonstrated in section 3 that a test of H_{02} under certain assumptions is equivalent to a test of the equivalence of the Jensen (1968) or Treynor (1965) measures of the N assets.

There are a variety of other questions that may be answered with H_{01} and H_{02}. For example, depending on the test outcomes of H_{01} and H_{02} the cause of a portfolio's inefficiency may be determined [see Korkie (1983)]. Also by defining one set of assets to include bonds as well as stocks the effect of the addition of bonds on potential performance may be tested with hypothesis H_{01}. A test which compares the performance potential of a set of portfolios to the set of portfolios plus stock and bond indices is presented in section 5. The next section, section 3, develops test procedures for the hypotheses H_{01} and H_{02}.

3. Derivation of the hypothesis testing procedures

Testing the hypothesis, that an asset subset has equivalent potential performance to the set of assets from which it was taken, is of considerable interest. If the two tangency portfolios formed from the asset set and subset have equivalent Sharpe performance, then the subset is sufficient to maximize performance. This section develops several test statistics for testing the hypotheses H_{01} and H_{02} discussed in section 2. The development of the tests requires that additional notation be introduced.

The population of N assets is partitioned into two mutually exclusive and exhaustive subsets containing N_1 and N_2 assets. The partitioned forms of the mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ are

$$\boldsymbol{\mu}=\begin{bmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{bmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}=\begin{bmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{bmatrix}.$$

Similar partitions will be used for the sample statistics $\bar{\boldsymbol{r}}$ and $\boldsymbol{S}$ defined below. The null hypothesis $H_{01}:a_1=a$ may now be restated as $H_{01}:\boldsymbol{\mu}_1'\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\mu}_1=\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$.

By employing the well known identity for the inverse of partioned matrices [see Morrison (1967)], the expression for $\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ may be written as

$$\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}=\boldsymbol{\mu}_1'\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\mu}_1+[\boldsymbol{\mu}_2-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\mu}_1]'\boldsymbol{\Sigma}_{22\cdot 1}^{-1}[\boldsymbol{\mu}_2-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\mu}_1], \tag{1}$$

where

$$\boldsymbol{\Sigma}_{22\cdot 1}=[\boldsymbol{\Sigma}_{22}-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}].$$

In order for H_{01} to hold, the positive semi-definite quadratic form $[\mu_2-\Sigma_{21}\Sigma_{11}^{-1}\mu_1]'\Sigma_{22\,1}^{-1}[\mu_2-\Sigma_{21}\Sigma_{11}^{-1}\mu_1]$ must be zero. This implies that $[\mu_2-\Sigma_{21}\Sigma_{11}^{-1}\mu_1]=0$, since Σ and $\Sigma_{22\,1}$ are positive definite. The hypothesis H_{01} can be seen therefore to be equivalent to the hypothesis

$$H_{01}:[\mu_2-\Sigma_{21}\Sigma_{11}^{-1}\mu_1]=0. \tag{2}$$

This result is employed to develop test procedures for H_{01}.

To construct test procedures for H_{01}, some assumptions about the sampling process and the multivariate distribution of excess return premiums are required. A random sample of T excess return observations on N assets is assumed available and given by the $(N\times 1)$ vector r_t, $t=1,2,\ldots,T$, where r_t is multivariate normal with mean excess return vector μ and covariance Σ. The maximum likelihood estimators of μ, Σ, Σ^{-1} and 'a' are given by $\bar{r}$, S, S^{-1} and $\hat{a}$, respectively, where

$$\bar{r}=(1/T)\sum_{t=1}^{T} r_t,$$

$$S=(1/T)\sum_{t=1}^{T}(r_t-\bar{r})(r_t-\bar{r})',$$

$$\hat{a}=\bar{r}'S^{-1}\bar{r}.$$

While $\bar{r}$ is unbiased for μ, unbiased estimators for Σ and Σ^{-1} are provided by $(T/(T-1))S$ and $W^{-1}=((T-N-2)/T)S^{-1}$ respectively [see Anderson (1958)]. The unbiased estimator for 'a' has been shown by Jobson and Korkie (1980) to be $\hat{a}^*=((T-N-2)/T)\,\hat{a}-N/T$. In small samples, the bias in the maximum likelihood estimator of $\hat{a}$ will therefore be substantial relative to 'a'. Inferences about 'a' may be made using the statistic $[(T-N)/N]\,\hat{a}$, which follows a non-central F'_{v_1,v_2,v_3} distribution[2] with $v_1=N$ and $v_2=(T-N)$ degrees of freedom and non-centrality parameter $v_3=(T)(a)$.

Test procedures for H_{01} are, however, more conveniently obtained using the alternative form of the hypothesis given by (2). The remainder of this section outlines four alternative test statistics for H_{01} and H_{02}.

[2]Since exhaustive tables of non-cental F' are not always available, a simple approximation to F'_{v_1,v_2,v_3} is available from the central F_{v_4,v_2} distribution [see Johnson and Kotz (1970)]. That is,

$$F'_{v_1,v_2,v_3}\approx(1+v_3/v_1)F_{v_4,v_2} \quad \text{where} \quad v_4=(v_1+v_3)^2/(v_1+2v_3)$$

Unpublished simulation studies by the authors have shown that, for parameters with magnitudes representative of stock market monthly data, this particular central F approximation is excellent for $T=60$ observations

3.1. The Wald statistic

Let $\gamma=[\mu_2-\Sigma_{21}\Sigma_{11}^{-1}\mu_1]$, then an unbiased estimator of γ is given by the maximum likelihood estimator $\hat{\gamma}=[\bar{r}_2-S_{21}S_{11}^{-1}\bar{r}_1]$, with covariance matrix[3] given by

$$\Omega=[((T-2)-Ta_1)/(T-N_1-2)](\Sigma_{22.1}/T).$$

Because $\hat{\gamma}$ is asymptotically normal, $\hat{\gamma}\Omega^{-1}\hat{\gamma}$ is asymptotically χ^2 with N_2 degrees of freedom if H_{01} is true. Since Ω^{-1} is unknown, a_1 and $\Sigma_{22.1}^{-1}$ are replaced by the unbiased estimators $\hat{a}_1^*=((T-N_1-2)/T)\hat{a}_1-N_1/T$ and $((T-N-2)/T)S_{22.1}^{-1}$, respectively, where $S_{22.1}=[S_{22}-S_{21}S_{11}^{-1}S_{12}]$. The resulting test statistic $\hat{\gamma}\hat{\Omega}^{-1}\hat{\gamma}$ is a Wald (1943) statistic[4] and is asymptotically χ^2 if H_{01} is true. The statistic may be written in the form

$$(T-N-2)(T-N_1-2)/[T(1+\hat{a}_1^*)-2][\bar{r}_2-S_{21}S_{11}^{-1}\bar{r}_1]'S_{22.1}^{-1}[\bar{r}_2-S_{21}S_{11}^{-1}\bar{r}_1],$$

or equivalently [using the sample form of (1)] as

$$((T-N-2)(T-N_1-2))/[T(1+\hat{a}_1^*)-2][\bar{r}'S^{-1}\bar{r}-r_1'S_{11}^{-1}\bar{r}_1],$$

which is approximately[5]

$$\phi_{1,1}=((T-N)[\hat{a}-\hat{a}_1])/[1+\hat{a}_1].$$

The behavior of the statistic $\phi_{1,1}$ in small samples is studied in section 4.1.

3.2. The likelihood ratio test

An alternative approach to developing a test for H_{01} is the likelihood ratio test. The likelihood ratio for testing H_{01} is given by (A.6) in the appendix and can be written as

$$\lambda=(1+\hat{a}_1)^{T/2}/(1+\hat{a})^{T/2}.$$

Knowledge of the critical value of the λ statistic requires that its distribution under H_{01} be known. A general approximation for likelihood

[3] The expression for Ω is obtained by observing that r and S are statistically independent and that $E(S_{21}S_{11}^{-1})=\Sigma_{21}\Sigma_{11}^{-1}$ The covariance between rows i and j of the matrix $S_{21}S_{11}^{-1}$ is given by $(1/(T-N_1-2))\psi_{ij}\Sigma_{11}^{-1}$, where ψ_{ij} is the ijth element of $\Sigma_{22.1}$ See Marx and Hocking (1977)

[4] The Wald statistic was used previously by Gibbons (1980) and Jobson and Korkie (1981).

[5] By employing S.N. Roy's (1953) union-intersection principle, a test statistic identical to $\phi_{1,1}$ may be derived.

ratio tests is that $-2\log_e \lambda$ asymptotically follows a χ^2 distribution with degrees of freedom equal to the number of independent restrictions placed on the parameter by the null hypothesis [see Silvey (1970)]. Thus, the statistic

$$T \log_e[(1+\hat{a})/(1+\hat{a}_1)] \tag{3}$$

has an asymptotic χ^2 distribution with $N_2=(N-N_1)$ degrees of freedom, if H_{01} is true.

A number of approximations have been developed for the distribution of λ which may be of use for performance tests. Bartlett (1938) employs the statistic

$$\phi_{1,2}=(T-N_1-N_2/2-1)\log_e[(1+\hat{a})/(1+\hat{a}_1)],$$

which is asymptotically χ^2 with N_2 degrees of freedom under the null hypothesis. This approximation is more precise than (3) because the term $(T-N_1-N_2/2-1)$ is designed to remove second-order terms in the asymptotic expansion. It is worth noting here that, if T is small relative to N_1 and/or N_2, the value of $\phi_{1,2}$ will be considerably smaller than the value given by eq. (3). The superiority of $\phi_{1,2}$ over (3) is discussed in Anderson (1958). The importance of this difference is noted in the simulation results of section 4 and in section 5.

A second approximation for the distribution of λ is due to Rao (1951), who employs an F distribution. The test statistic for H_{01} is given by

$$\phi_{1,3}=[(T-N-1)/N_2][(1-\Lambda)/\Lambda]=[(T-N-1)/N_2][(\hat{a}-\hat{a}_1)/(1+\hat{a}_1)],$$

where $\Lambda=\lambda^{2/T}$ is a form of Wilks Lambda given by (A.5) of the appendix.[6] Under H_{01}, $\phi_{1,3}$ is asymptotically an F distribution with N_2 and $T-N-1$ degrees of freedom. Both approximations $\phi_{1,2}$ and $\phi_{1,3}$ are discussed in Rao (1973, p. 556). In a multivariate regression model the statistic $\phi_{1,3}$ follows the F distribution exactly while the statistic $\phi_{1,2}$ is asymptotically χ^2. In this application the statistic $\phi_{1,3}$ is only asymptotically an F distribution because the return vector $\bar{r}_1$ is not fixed. It is believed, however, that $\phi_{1,3}$ should be a superior approximation in small samples. The small sample behavior of $\phi_{1,2}$ and $\phi_{1,3}$ is studied in section 4.

3.3. The score test

A test statistic for the null hypothesis H_{01} may also be developed from

[6]Under multivariate normality, the likelihood ratio test for the hypothesis that a subset of the regression coefficients are zero in a multivariate regression also results in a form of Wilks Lambda. Using the form of H_{01} given by (1), the hypothesis may be recognized as being equivalent to the hypothesis of a zero intercept vector in a multivariate regression. A more detailed discussion of this is given in Jobson (1982b) and Jobson and Korkie (1982).

Rao's (1947) score test[7] outlined in Rao (1973, pp. 415–420). This test statistic has the same asymptotic distribution (χ^2 with N_2 d.f.) as the previously developed test statistics $\phi_{1,1}$ and $\phi_{1,2}$.

The score test criterion, given by (A.8) in the appendix, is $[T^2/(T-N_1)]$ $[\hat{a}-\hat{a}_1/(1+\hat{a}_1)]$, which is equivalent to $\phi_{1,1}$ multiplied by the factors $T^2/((T-N_1)(T-N))$. Therefore the score test statistic will always exceed $\phi_{1,1}$, and therefore is not studied in the remainder of the paper.

The adjusted or modified score test statistic as given by (A.9) of the appendix is

$$\phi_{1,4}=T[(\hat{a}-\hat{a}_1)/((1+\hat{a}_1)(1+\hat{a}))].$$

This statistic is equivalent to $\phi_{1,1}$ multiplied by the factor $T/[(1+\hat{a})(T-N)]$. In large samples if $\hat{a}$ is large relative to $\hat{a}_1$, the magnitude of $\phi_{1,4}$ should be less than $\phi_{1,1}$. In section 4 this modified score test statistic is studied and compared to the other three test statistics using a Monte Carlo experiment.

3.4. *Tests for portfolio efficiency*

For the special case of H_{01} when $N_1=1$, the potential performance $\sqrt{a}$ of the asset p and its Sharpe performance μ_p/σ_p are identical. Thus, a test of H_{02} in this case is a test of the mean-standard deviation efficiency of the asset p.

The special case of hypothesis test H_{01} is the test addressed by Ross (1980) for testing the efficiency of a portfolio. One large sample test statistic used by Ross was the expression (3) above with $N_1=1$ and $N_2=N-1$. The preferred Bartlett's small sample approximation, as in the case of $\phi_{1,2}$ is given by

$$\phi_{2,2}=(T-N/2-5/2)\log_e[(1+\hat{a})/(1+\bar{r}_p^2/s_p^2)],$$

where $\bar{r}_p$ and s_p^2 are the sample mean and sample variance of the portfolio of interest. $\phi_{2,2}$ is distributed asymptotically as a χ^2 distribution with $(N-1)$ degrees of freedom under H_{02}. In sections 4 and 5, the advantage of $\phi_{2,2}$ over the Ross test is discussed.

By replacing $\bar{r}_1' S_{11}^{-1}\bar{r}_1$ by $\bar{r}_p^2/s_p^2$, the other H_{01} test statistics may also be modified for a test of H_{02}. That is

$$\phi_{2,1}=(T-N)[(\hat{a}-\bar{r}_p^2/s_p^2)/(1+\bar{r}_p^2/s_p^2)],$$

[7]The authors are indebted to Michael Gibbons for suggesting that the score statistic be included as a potential test statistic. Tests based on this statistic have also been called Lagrange Multiplier tests after Silvey (1970). The Wald statistic, the likelihood ratio χ^2 statistic and the Lagrange Multiplier statistic have been compared in Buse (1982). The Buse article also contains a bibliography for comparisons of the three χ^2 statistics for the multivariate regression model.

which under H_{02} is asymptotically distributed as a χ^2 with $(N-1)$ degrees of freedom,

$$\phi_{2,3}=[(T-N-1)/(N-1)][(\hat{a}-\bar{r}_p^2/s_p^2)/(1+\bar{r}_p^2/s_p^2)],$$

which asymptotically follows an F distribution with $(N-1)$ and $(T-N-1)$ degrees of freedom under H_{02}, and finally

$$\phi_{2,4}=(T[\hat{a}-\bar{r}_p^2/s_p^2])/([1+\hat{a}][1+\bar{r}_p^2/s_p^2]),$$

which under H_{02} is asymptotically χ^2 with $(N-1)$ degrees of freedom.

The behavior of these statistics in small samples is examined in section 4 using Monte Carlo simulation. The relationships of H_{02} to hypotheses of mean-variance efficiency, using Jensen or Treynor measures, is discussed next. The next section also relates these test procedures to the recent work of Gibbons (1982).

3.5. Relationships to other test procedures

3.5.1. Equivalent Treynor measures

The hypotheses H_{01} and H_{02} can be related to other available test procedures for mean-variance efficiency. Roll (1978) has shown that a test of the mean-variance efficiency of a portfolio p, constructed from an N asset set, may be carried out using the security market line. In his proposition S4, he states that a given index is mean-variance efficient if the betas of all assets are related to their mean returns by the same linear function.

The Treynor measure μ_i/β_i, of any security i in p, represents the security's return premium contribution $x_i\mu_i$ divided by its fractional risk contribution $x_i\sigma_p^2/\sigma_{ip}$. This marginal rate of substitution between risk and return must be identical for all members of p, if p is efficient (i.e., maximizes return premium per unit of total risk). Otherwise, some investment reallocation would provide a larger return per risk unit implying that p is inefficient. Thus, all N assets will plot on a market line derived from an efficient portfolio p. A test of equality of the Treynor measures, $H_0:\mu_1/\beta_1=\mu_2/\beta_2=\mu_2/\beta_2=\ldots=\mu_N/\beta_N$ is therefore a test of mean-variance efficiency. A χ^2 test for the equality of Treynor measures is developed in Jobson and Korkie (1981).

As outlined at the beginning of this section, the test of the hypothesis H_{01} is equivalent to testing $H_{01}:\boldsymbol{\mu}_2=\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\mu}_1$. In the case of a single portfolio p, with $N_1=1$, H_{01} becomes $H_{02}:\boldsymbol{\mu}_2=\boldsymbol{\sigma}_2\mu_p/\sigma_p^2$ or $H_{02}:\boldsymbol{\mu}_2=\mu_p\boldsymbol{\beta}$, where $\boldsymbol{\sigma}_2$ is the $(N\times 1)$ vector of covariances between N assets and the portfolio p and $\boldsymbol{\beta}$ is the $(N\times 1)$ vector of security betas computed with respect to p. The hypothesis H_{02} is therefore equivalent to the hypothesis of equal Treynor measures.

3.5.2. Equivalent Jensen measures

The hypothesis H_{02} may also be related to the Jensen performance measure. Ross (1977) has shown that a test of equality of the Jensen performance measures, $H_0:\alpha_1=\alpha_2=\ldots=\alpha_N$ for N assets, is also a test of the mean-variance efficiency of a portfolio p. The $(N\times 1)$ vector $\boldsymbol{\alpha}$, with elements α_j, $j=1,2,\ldots,N$, is the intercept vector for the linear model

$$\boldsymbol{r}_t=\boldsymbol{\alpha}+\boldsymbol{\beta}r_{pt}+\boldsymbol{\varepsilon}_t, \qquad t=1,2,\ldots,T,$$

where $\boldsymbol{r}_t$ is an $(N\times 1)$ vector of return premiums, r_{pt} is the return premium of a portfolio of the N assets, $\boldsymbol{\beta}$ is the $(N\times 1)$ vector of betas between p and the N assets and $\boldsymbol{\varepsilon}_t$ is an error term. This regression model was applied in Black, Jensen and Scholes (1972) to tests of the capital asset pricing model.

As suggested in footnote 6, the vector $(\boldsymbol{\mu}_2-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\mu}_1)$ from H_{01} is recognized as the intercept vector in the multivariate regression of the vector $\boldsymbol{r}_{2t}$ $(N_2\times 1)$, on the $(N_1\times 1)$ vector $\boldsymbol{r}_{1t}$, $t=1,2,\ldots,T$, where $\boldsymbol{r}_t'=(\boldsymbol{r}_{1t}'\boldsymbol{r}_{2t}')$ is multivariate normal. In the special case of H_{02}, with $N_1=1$, the intercept vector becomes $[\boldsymbol{\mu}_2-\boldsymbol{\sigma}_2\mu_p/\sigma^2]$ which is equivalent to $\boldsymbol{\alpha}=[\boldsymbol{\mu}_2-\boldsymbol{\beta}\mu_p]$. Since the portfolio is assumed to be a linear combination of the N assets, there is no loss of generality in letting $\boldsymbol{\mu}_2$ refer only to $N_2=(N-1)$ assets provided that the asset omitted from $\boldsymbol{\mu}$ has a non-zero weight in the portfolio p. The test of H_{02} is therefore equivalent to a test of equality of the Jensen measures provided that the index portfolio is derived from the N assets.

3.5.3. Gibbons multivariate approach

In a recent paper by Gibbons (1982), a multivariate approach to testing the capital asset pricing model is outlined. Gibbons begins with the model $r_{it}^*=\alpha_i+\beta_i r_{mt}^*+\varepsilon_{it}$, $t=1,2,\ldots,T$, $i=1,2,\ldots,N$, where r_{it}^* denotes the return (not the return premium) on asset i at time t and r_{mt}^* denotes the return on the market portfolio at time t. Under the assumption that the N assets' returns are distributed as a multivariate normal, Gibbons develops a likelihood ratio test of the null hypothesis, $\alpha_i=\gamma_0(1-\beta_i)$, $i=1,2,\ldots,N$, where γ_0 is the mean return on a zero beta portfolio orthogonal to m. Because of the nonlinearity of the hypothesis, a Gauss–Newton approximation is employed to solve the likelihood equations. This approximation requires preliminary estimates of γ_0 and the β_i, $i=1,2,\ldots,N$, that are consistent. The likelihood equations for the vector $\boldsymbol{\beta}$ of elements β_i, $i=1,2,\ldots,N$, are given in Gibbons (1980) as

$$\boldsymbol{\beta}^*=[(\boldsymbol{r}_m-\gamma_0\boldsymbol{e})'(\boldsymbol{r}_m-\gamma_0\boldsymbol{e})]^{-1}\begin{bmatrix}(\boldsymbol{r}_1^*-\gamma_0\boldsymbol{e})'(\boldsymbol{r}_m^*-\gamma_0\boldsymbol{e})\\ \vdots \\ (\boldsymbol{r}_N^*-\gamma_0\boldsymbol{e})'(\boldsymbol{r}_m^*-\gamma_0\boldsymbol{e})\end{bmatrix},$$

where $\boldsymbol{r}_m^*$ and $\boldsymbol{r}_i^*$, $i=1,2,\ldots,N$, are $(T\times 1)$ vectors of returns and $\boldsymbol{e}$ is a $(T\times 1)$ vector of unities.

If $\gamma_0\boldsymbol{e}$ is replaced by a vector of risk-free rates then the maximum likelihood estimator of $\boldsymbol{\beta}$, given by $\boldsymbol{\beta}^*$, is equivalent to our maximum likelihood estimator $\hat{\boldsymbol{\beta}}$ given in the appendix. Our likelihood ratio test of the efficiency of m is based on an available time series of the risk-free rate, while Gibbons' likelihood ratio test assumes a constant expected zero beta rate which must be estimated from the data. Gibbons' χ^2 statistic therefore has one less degree of freedom than ours.

In section 5, the $\phi_{1,2}$ χ^2 and $\phi_{1,3}$ F statistics are used to test the efficiency of equal weight portfolios relative to sets of 20 and 40 portfolios. The test procedures are also compared to those of Gibbons (1982).

4. The sampling experiment

A Monte Carlo simulation was designed to investigate the small sample behavior of the asymptotic test statistics outlined in section 3. That is, the small sample properties of the χ^2 statistics $\phi_{1,1}$, $\phi_{1,2}$, $\phi_{1,4}$ and the F statistic $\phi_{1,3}$ are examined for tests of the performance potential hypothesis H_{01}. In addition, the sampling properties of the χ^2 and F statistics $\phi_{2,1}$, $\phi_{2,2}$, $\phi_{2,4}$ and $\phi_{2,3}$ are investigated for tests of the portfolio efficiency hypothesis H_{02}. Of particular interest are the moments and the tail areas of the small sample distributions and their correspondence with the theoretical distributions. The powers of the hypotheses tests are also computed for several arbitrary hypotheses.

The simulation was parameterized with a set Γ of $N=50$ randomly selected NYSE stocks having 360 continuous monthly returns commencing January, 1950. The observed mean vector and covariance matrix for the stocks were treated as the population parameters $\boldsymbol{\mu}$ and Σ. The riskless interest rate was assumed to be zero without perceived loss of generality. To provide two mutually exclusive stock subsets, the 50 stock population was randomly partitioned into $N_1=20$ and $N_2=30$ stock subpopulations, denoted by Γ_1 and Γ_2, respectively. The parameters of these subsets are denoted by $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$, Σ_{11} and Σ_{22}. The partitioned forms of $\boldsymbol{\mu}$ and Σ are therefore denoted by

$$\boldsymbol{\mu}=\begin{bmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{bmatrix} \quad\text{and}\quad \Sigma=\begin{bmatrix}\Sigma_{11}\Sigma_{12}\\ \Sigma_{21}\Sigma_{22}\end{bmatrix}.$$

The efficient set constants 'a', 'b', and 'c', are given in table 1 for the three sets of assets Γ_1, Γ_2, Γ_3. The values of 'a' for the three populations are denoted by a, a_1 and a_2, respectively. These three asset populations are employed in this section to evaluate the sampling properties of the four test statistics for both hypotheses H_{01} and H_{02}.

4.1. Tests of equivalent potential performance H_{01}

In order to examine the behavior of the test statistics when the null hypothesis H_{01} is true, two additional 50 stock populations Γ_A and Γ_B were constructed by modifying the mean vector $\boldsymbol{\mu}$ of Γ, and preserving the covariance matrix. The mean vectors for Γ_A and Γ_B are denoted by

$$\boldsymbol{\mu}_A=\begin{bmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_{A2}\end{bmatrix} \quad\text{and}\quad \boldsymbol{\mu}_B=\begin{bmatrix}\boldsymbol{\mu}_{B1}\\ \boldsymbol{\mu}_2\end{bmatrix},$$

where $\boldsymbol{\mu}_{A2}=\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\mu}_2$ and $\boldsymbol{\mu}_{B1}=\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\mu}_1$ satisfy the form of H_{01} given by (1). The pair of populations Γ_1 and Γ_A now satisfy H_{01} as do the populations Γ_2 and Γ_B. The respective 'a' values from table 1 are $a_A=a_1=0.096$ and $a_B=a_2=0.157$.

In order to examine the power of the test statistics when the null hypothesis H_{01} is not true, two alternatives were studied. The population Γ was compared to each of the subpopulations Γ_1 and Γ_2. From table 1 the respective population 'a' values are $a=0.191$, $a_1=0.096$ and $a_2=0.157$.

Two hundred random samples for each sample size $T=60$, 120 and 240 return observations were generated from the parameters of the Γ, Γ_A and Γ_B 50 stock populations, the 20 stock subset Γ_1 and the 30 stock subset Γ_2.[8] For each sample, the maximum likelihood estimates of the mean vector $\bar{\boldsymbol{r}}$, covariance matrix $\boldsymbol{S}$ and its inverse $\boldsymbol{S}^{-1}$ were computed for Γ, Γ_A, Γ_B, Γ_1 and Γ_2.[9]

The $\phi_{1,i}$ test statistics, for a test of the equivalence of potential performances of the 20 stock subset Γ_1 and the 50 stock set Γ_A, were computed across all samples. Similarly the $\phi_{1,i}$ test statistics, for equivalent potential performance of the 30 stock subset Γ_2 and the 50 stock set Γ_B, were computed over the 200 samples. The $\phi_{1,i}$ statistics' mean, variance and right tail areas were calculated over the 200 samples and compared to their

[8]Return samples were generated using a double precision version of the multivariate normal random number generator from the IMSL Subroutine Package. The 200 replications seemed sufficiently large to provide accurate indications of the small sample distributions of the statistics. For example, the $\phi_{1,3}$ small sample distributions had standard errors from 1 to 3 percent of the sampling mean All calculations in the simulation were performed in double precision Fortran

[9]Checks were made in each sample to ensure that the global minimum mean return $\bar{r}_0=\hat{b}/\hat{c}$ exceeded the riskless rate $r_f=0$. Samples in which r_f exceeded $\bar{r}_0$ were discarded and a new sample was drawn See Merton (1972) for a discussion of the consequences of $r_f>\bar{r}_0$

Table 1

Efficient set constants a, b, and c for three randomly selected $N=50$, $N_1=20$, and $N_2=30$ stock populations with a zero riskless rate, where μ is the mean premium return vector, e is the unit vector, and Σ^{-1} is the covariance matrix inverse. Parameters are formed from 360 monthly returns commencing January 1950.

Number of stocks	Efficient set constants		
	$a=\mu'\Sigma^{-1}\mu$	$b=e'\Sigma^{-1}\mu$	$c=e'\Sigma^{-1}e$
$N=50$	0 191	0.111	0.128
$N_1=20$, $N_1\subset N$	0.096	0.065	0.079
$N_2=30$, $N_2\subset N$	0.157	0.103	0.112

theoretical counterparts under a true H_{01}. The results are shown in tables 2 and 3.

The preceding statistics were generated under true null hypotheses $H_{01}:a_A=a_1$ and $H_{01}:a_B=a_2$. The power of the test statistics, under a false H_{01}, were observed by computing 200 $\phi_{1,i}$ test statistics from the Γ and Γ_1 stock sets as well as from the Γ and Γ_2 stock sets. The powers of the $\phi_{1,i}$ test statistics were computed by counting the number of samples, out of 200, for which a $\phi_{1,i}$ statistic exceeded the $\alpha=0.05$ critical value in a test of the equivalence of the potential performances of the 50 stock set Γ and the 20 stock subset Γ_1. The power calculation was repeated for the 50 stock set Γ versus the 30 stock subset Γ_2. The powers are listed in table 4.

In general, the best statistic appears to be the $\phi_{1,3}$ F statistic. Its sampling mean, variance and tail areas closely correspond to the values of a theoretical F distribution at all three sample sizes and for both simulated populations presented in tables 2 and 3. For example, consider the comparison of the 20 stock and 50 stock population contained in the first panel of table 2, where the $T=60$ sampling distribution of the $\phi_{1,i}$, $i=1,\dots,4$, statistics and the theoretical χ^2 and F statistics are presented. The mean and variance of a theoretical F distribution, with 30 and 9 degrees of freedom, are 1.25 and 0.66 while the sampling mean and variance of the $\phi_{1,3}$ statistic are 1.15 and 0.54. Similar comparison of the right tail areas shows a very close correspondence between the theoretical F and the $\phi_{1,3}$ statistic. The $\phi_{1,1}$, $\phi_{1,2}$, and $\phi_{1,4}$ χ^2 statistics do not perform as well until at least sample size $T=120$. At $T=120$, the $\phi_{1,2}$ statistic is superior to the other χ^2 statistics since there is closer correspondence between a theoretical χ^2 with 30 degrees of freedom and the $\phi_{1,2}$ statistic. As shown in table 3, the power of the $\phi_{1,3}$ statistic under a false H_{01} increases from 0.05 at $T=60$ to 0.57 at $T=240$ for the 20 stock case and from 0.03 at $T=60$ to 0.24 at $T=240$ for the 30 stock case. Thus the $\phi_{1,3}$ test statistic does not demonstrate any powerful

Table 2

Sampling mean, variance and right tail areas of the χ^2 statistics $\phi_{1,1}$, $\phi_{1,2}$ and $\phi_{1,4}$ and the F statistic $\phi_{1,3}$ for tests of the equivalent performance (H_{01} $a_1=a_A$) of $N_1=20$ stocks versus $N=50$ stocks, for sample sizes $T=60$, 120 and 240 Each experiment is based on 200 replications, under a true null hypothesis

Statistic	Mean	Variance	Right tail areas								
			Sample size $T=60$								
χ^2	30.0	60.0	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
$\phi_{1,1}$	38.3	597.0	0.94	0.88	0.82	0.71	0.57	0.43	0.35	0.29	0.21
$\phi_{1,2}$	35.5	97.5	1.00	0.98	0.96	0.91	0.71	0.48	0.30	0.20	0.08
$\phi_{1,4}$	27.2	30.1	1.00	0.96	0.91	0.64	0.34	0.08	0.03	0.01	0.00
F	1.25	0.66	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
$\phi_{1,3}$	1.15	0.54	0.98	0.94	0.90	0.71	0.43	0.21	0.07	0.03	0.01
			Sample size $T=120$								
χ^2	30.0	60.0	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
$\phi_{1,1}$	31.7	103.9	0.99	0.94	0.88	0.71	0.55	0.34	0.19	0.12	0.05
$\phi_{1,2}$	31.0	67.7	0.99	0.95	0.91	0.76	0.55	0.30	0.13	0.08	0.02
$\phi_{1,4}$	28.0	41.8	0.99	0.92	0.88	0.72	0.42	0.15	0.03	0.02	0.00
F	1.03	0.10	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
$\phi_{1,3}$	1.04	0.11	0.99	0.95	0.91	0.72	0.51	0.27	0.12	0.05	0.02
			Sample size $T=240$								
χ^2	30.0	60.0	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
$\phi_{1,1}$	30.0	69.1	0.98	0.94	0.89	0.73	0.49	0.28	0.12	0.07	0.01
$\phi_{1,2}$	29.8	58.5	0.98	0.95	0.90	0.74	0.49	0.23	0.09	0.05	0.02
$\phi_{1,4}$	27.1	43.7	0.98	0.93	0.84	0.65	0.37	0.14	0.03	0.01	0.00
F	1.01	0.08	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
$\phi_{1,3}$	1.00	0.08	0.98	0.94	0.89	0.73	0.47	0.24	0.10	0.05	0.01

ability to distinguish potential performance differences between the 50 stock population and the 20 and 30 stock subsets at $T=60$. It is worth emphasizing that this lack of power stems from the natural variability of the stock market data, rather than from departures from the assumed small sample distribution. The χ^2 statistics $\phi_{1,1}$, and $\phi_{1,2}$ seem to be more powerful that the $\phi_{1,3}$ F at $T=60$. However, this result is due to the excess skewness of the χ^2 statistics above the theoretical χ^2, as shown in tables 1 and 2. In contrast, the $\phi_{1,4}$ statistic tends to be less skewed than a theoretical χ^2 distribution and as a result exhibits less power than the $\phi_{1,3}$ statistic.

It is also important to note that the statistic given by (3) in section 2, which uses the multiplier T rather than $(T-N_1-N_2/2-1)$, would always be

Table 3

Sampling mean, variance and right tail areas of the χ^2 statistics $\phi_{1,1}$, $\phi_{1,2}$ and $\phi_{1,4}$ and the F statistic $\phi_{1,3}$ for tests of the equivalent performance ($H_{01}: a_2 = a_B$) of $N_2 = 30$ stocks versus $N = 50$ stocks, for sample sizes $T = 60$, 120 and 240 Each experiment is based on 200 replications, under a true null hypothesis

Statistic	Mean	Variance	Right tail areas								
			Sample size $T=60$								
χ^2	20.0	40.0	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
$\phi_{1,1}$	26.2	361.6	0.97	0.89	0.83	0.70	0.57	0.44	0.36	0.28	0.18
$\phi_{1,2}$	22.7	60.5	0.99	0.97	0.93	0.83	0.64	0.41	0.23	0.11	0.04
$\phi_{1,4}$	17.7	23.4	1.00	0.92	0.84	0.67	0.36	0.12	0.02	0.00	0.00
F	1.25	0.73	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
$\phi_{1,3}$	1.18	0.73	0.98	0.92	0.88	0.70	0.46	0.22	0.08	0.02	0.01
			Sample size $T=120$								
χ^2	20.0	40.0	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
$\phi_{1,1}$	20.2	52.6	0.99	0.93	0.89	0.71	0.47	0.28	0.16	0.10	0.01
$\phi_{1,2}$	19.8	39.0	0.99	0.94	0.91	0.74	0.46	0.26	0.11	0.05	0.00
$\phi_{1,4}$	16.9	24.3	0.99	0.91	0.81	0.57	0.30	0.10	0.01	0.00	0.00
F	1.03	0.14	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
$\phi_{1,3}$	1.00	0.13	0.99	0.94	0.89	0.71	0.46	0.25	0.10	0.04	0.00
			Sample size $T=240$								
χ^2	20.0	40.0	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
$\phi_{1,1}$	21.0	44.7	0.99	0.97	0.94	0.80	0.53	0.29	0.13	0.07	0.01
$\phi_{1,2}$	20.7	38.7	0.99	0.97	0.94	0.81	0.53	0.27	0.11	0.06	0.01
$\phi_{1,4}$	18.0	27.4	0.99	0.94	0.89	0.67	0.37	0.11	0.03	0.01	0.01
F	1.01	0.11	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
$\phi_{1,3}$	1.04	0.11	0.98	0.95	0.90	0.74	0.51	0.25	0.11	0.06	0.02

larger than $\phi_{1,2}$. This would indicate that $\phi_{1,2}$ is superior to (3) in conforming to a χ^2 distribution. An examination of simulation results in Gibbons (1982) also reflects that the statistic, given by (3), tends to have a larger mean and variance than the theoretical χ^2. The excessive values of the mean χ^2 in Gibbons are approximately equivalent to the excess expected from employing T rather than the Bartlett correction in $\phi_{1,2}$.

4.2. *Tests of portfolio efficiency H_{02}*

To examine the behavior of the test statistics when the null hypothesis H_{02} is true, the true proportions vectors X_m, X_{m1} and X_{m2} were determined using the true parameters of the 50, 20 and 30 stock populations Γ, Γ_1 and Γ_2, respectively. Using the same random samples generated in section 4.1, the

Table 4

Powers of the test statistics $\phi_{1,1}$, $\phi_{1,2}$, $\phi_{1,3}$ and $\phi_{1,4}$ in an $\alpha=0.05$ test of potential performance differences between $N=50$ and $N_1=20$ stock sets (H_{01} $a=a_1$) and between $N=50$ and $N_2=30$ stock sets ($H_{01}\cdot a=a_2$) for sample sizes $T=60$, 120 and 240 Each experiment is based on 200 replications, with parameters $a=0\,191$, $a_1=0\,096$ and $a_2=0\,157$, for stock sets Γ, Γ_1, and Γ_2

Statistic	Test powers					
	$T=60$		$T=120$		$T=240$	
	$N_2=30, \Gamma_2$	$N_1=20, \Gamma_1$	$N_2=30, \Gamma_2$	$N_1=30, \Gamma_1$	$N_2=20, \Gamma_2$	$N_1=20, \Gamma_1$
$\phi_{1,1}$	0.30[a]	0 34	0 14	0 37	0.27	0 64
$\phi_{1,2}$	0 16	0 22	0.09	0 27	0.25	0.58
$\phi_{1,3}$	0 03	0 05	0.08	0 24	0 24	0 57
$\phi_{1,4}$	0 01	0 01	0 02	0 07	0 08	0.36

[a]The probability of rejecting with $\phi_{1,1}$ the (false) null hypothesis, that the potential performance 'a' of the $N=50$ stock set Γ is equal to the potential performance a_2 of the $N_2=30$ stock set Γ_2, is 0 30, with $T=60$ observations The powers may be compared among statistics $\phi_{1,i}$ among sample sizes T, and between the number of stocks N_j

squared Sharpe performance measures $\bar{r}_m^2/s_m^2=(X_m'\bar{r})^2/(X_m'SX_m)$, $\bar{r}_{m1}^2/s_{m1}^2=(X_{m1}'\bar{r}_1)^2/(X_{m1}'S_{11}X_{m1})$ and $\bar{r}_{m2}^2/s_{m2}^2=(X_{m2}'\bar{r}_2)^2/(X_{m2}'S_{22}X_m)$ were computed for each sample. The $\phi_{2,i}$ statistics for testing H_{02} were then computed for each sample. The mean, variance and right tail areas of the test statistics, based on the 200 samples, were determined and are compared to their theoretical counterparts in tables 5, 6 and 7.

The mean, variance and tail areas of $\phi_{2,3}$ are very similar to the mean, variance and tail areas of a theoretical F distribution. The good behavior of the $\phi_{2,3}$ statistic does not seem to be affected by sample size or number of assets. With the exception of the 50 asset case with sample size $T=60$, the statistic $\phi_{2,2}$ was comparable in behavior to a theoretical χ^2 distribution. At $T=60$ for the 50 asset population, $\phi_{2,2}$ was far more skewed than the χ^2 distribution. The test statistic $\phi_{2,1}$ was in general more skewed than the χ^2 distribution, while the statistic $\phi_{2,4}$ was in general less skewed than the χ^2 distribution. The behavior of $\phi_{2,1}$ and $\phi_{2,4}$ were unaffected by sample size or number of assets.

To evaluate the power of the $\phi_{2,i}$ test statistics under a false H_{02}, two inefficient portfolios were formed from each of the two sets of assets Γ and Γ_1. The first pair of inefficient portfolios were the equal weight portfolios with true squared Sharpe measures $(\mu'e)^2/(e'\Sigma e)=0.065$ and $(\mu_1'e)^2/(e_1'\Sigma_{11}e_1)=0.055$, respectively. The second pair of inefficient portfolios were partial equal weight portfolios with equal weights over the first 10 stocks and zero weights over the remaining 40 and 10 stocks, respectively. The theoretical

Table 5

Sampling mean, variance and right tail areas of the χ^2 statistics $\phi_{2,1}$, $\phi_{2,2}$ and $\phi_{2,4}$ and the F statistic $\phi_{2,3}$ for tests of the efficiency ($H_{02}:a=\mu_m^2/\sigma_m^2$) of the $N=50$ stock tangency portfolio m, for sample sizes $T=60$, 120 and 240 Each experiment is based on 200 replications, under a true null hypothesis.

Statistic	Mean	Variance	Right tail areas								
			Sample size $T=60$								
χ^2	49.0	98.0	0.99	0 95	0 90	0 75	0 50	0.25	0.10	0.05	0.01
$\phi_{2,1}$	61.9	1403.6	0.91	0.82	0.77	0 70	0 58	0 43	0.39	0 35	0.26
$\phi_{2,2}$	60.9	191.8	1.00	0.99	0.98	0.94	0.81	0 64	0.42	0.34	0.16
$\phi_{2,4}$	41.4	24.1	0.99	0.93	0.86	0.48	0.06	0.01	0 00	0 00	0.00
F	1.25	0.61	0.99	0 95	0.90	0 75	0.50	0.25	0.10	0 05	0.01
$\phi_{2,3}$	1.14	0.46	0 98	0 94	0.89	0.71	0.42	0.20	0 04	0 02	0.01
			Sample size $T=120$								
χ^2	49.0	98 0	0.99	0.95	0 90	0.75	0.50	0 25	0.10	0.05	0.01
$\phi_{2,1}$	50 1	181.9	0.98	0.91	0.83	0.73	0.49	0.32	0 17	0.11	0.05
$\phi_{2,2}$	49.4	102.9	0.99	0 96	0.90	0.76	0.49	0.27	0.10	0.06	0.02
$\phi_{2,4}$	40 7	49.2	0.97	0 86	0 70	0.41	0 13	0 03	0 00	0 00	0.00
F	1 03	0 08	0 99	0 95	0.90	0 75	0.50	0 25	0.10	0 05	0.01
$\phi_{2,3}$	1 01	0 07	0.99	0 94	0.90	0 73	0 47	0 22	0 09	0 05	0 02
			Sample size $T=240$								
χ^2	49.0	98.0	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0 05	0.01
$\phi_{2,1}$	50.0	117.5	1.00	0.96	0.92	0.76	0.50	0.29	0.15	0.07	0.03
$\phi_{2,2}$	49.4	90 3	1.00	0.98	0.94	0.77	0 49	0.28	0.11	0.06	0.01
$\phi_{2,4}$	41.7	55.9	0.98	0.86	0 73	0 46	0 18	0.05	0.01	0 00	0.00
F	1 01	0.03	0.99	0.95	0.90	0.75	0.50	0.25	0 10	0.05	0.01
$\phi_{2,3}$	1.02	0.05	1.00	0.98	0.93	0.77	0.50	0.28	0 07	0 04	0.01

squared Sharpe measures for both sets was 0.041.[10] Using the same 200 samples as above, the $\phi_{2,i}$ statistics were computed assuming H_{02} to be true. The power in each case was determined by counting the number of samples for which a $\phi_{2,i}$ statistic value exceeded the $\alpha=0.05$ critical value. The power fractions are given in table 8.

Inspection of table 8 reveals that the power of the test statistics $\phi_{2,3}$ and $\phi_{2,2}$ are comparable with the exception of the 50 asset case at $T=60$. In this case, the extreme skewness in $\phi_{2,2}$ results in a larger power than $\phi_{2,3}$. In general, the statistic $\phi_{2,1}$ displays larger powers than $\phi_{2,3}$ reflecting the extreme skewness in the distribution of $\phi_{2,1}$. In contrast, the statistic $\phi_{2,4}$

[10]The squared Sharpe measure for the typical stock is about 0 015 Thus, the equal weight portfolios have squared Sharpe measures of approximately three to four times the typical stock. In contrast, the squared potential performance is about two to three times as large as the equal weight squared Sharpe performance. Thus, these equal weight portfolios represent portfolio inefficiency in the mid-range of the typical stock and maximum potential performance

Table 6

Sampling mean, variance and right tail areas of the χ^2 statistics $\phi_{2,1}$, $\phi_{2,2}$ and $\phi_{2,4}$ and the F statistic $\phi_{2,3}$ for tests of the efficiency (H_{02}: $a_2=\mu_{m2}^2/\sigma_{m2}^2$) of the $N_2=30$ stock tangency portfolio $m2$, for sample sizes $T=60$, 120 and 240. Each experiment is based on 200 replications under a true null hypothesis.

Statistic	Mean	Variance	Right tail areas								
			Sample size $T=60$								
χ^2	29.0	58.0	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
$\phi_{2,1}$	30.3	134.8	0.98	0.90	0.84	0.70	0.51	0.35	0.19	0.11	0.05
$\phi_{2,2}$	29.0	56.7	1.00	0.97	0.89	0.74	0.50	0.26	0.10	0.04	0.01
$\phi_{2,4}$	25.0	25.2	1.00	0.93	0.84	0.61	0.26	0.04	0.10	0.01	0.00
F	1.07	0.18	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
$\phi_{2,3}$	1.01	0.15	0.99	0.94	0.86	0.70	0.45	0.20	0.07	0.02	0.01
			Sample size $T=120$								
χ^2	29.0	58.0	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
$\phi_{2,1}$	29.9	80.9	1.00	0.94	0.89	0.75	0.51	0.29	0.15	0.11	0.04
$\phi_{2,2}$	29.1	56.60	1.00	0.95	0.91	0.76	0.50	0.25	0.12	0.06	0.01
$\phi_{2,4}$	25.0	32.2	0.99	0.92	0.83	0.58	0.28	0.07	0.01	0.01	0.00
F	1.03	0.11	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
$\phi_{2,3}$	1.02	0.09	1.00	0.97	0.91	0.76	0.48	0.23	0.10	0.04	0.01
			Sample size $T=240$								
χ^2	29.0	58.0	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
$\phi_{2,1}$	29.2	64.1	0.98	0.95	0.90	0.74	0.51	0.25	0.11	0.06	0.02
$\phi_{2,2}$	28.8	54.6	0.98	0.96	0.90	0.75	0.51	0.23	0.08	0.04	0.01
$\phi_{2,4}$	25.1	38.1	0.98	0.91	0.80	0.58	0.28	0.07	0.03	0.01	0.00
F	1.01	0.08	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
$\phi_{2,3}$	1.00	0.08	0.98	0.96	0.90	0.74	0.51	0.24	0.09	0.04	0.01

displays much lower powers than $\phi_{2,3}$ reflecting the fact that $\phi_{2,4}$ has less skewness than the theoretical χ^2 distribution. The power of the $\phi_{2,3}$ statistic increases from 0.03 and 0.09 at $T=60$ to 0.67 and 0.40 at $T=240$.

On the basis of the simulation results of this section, two likelihood ratio statistics $\phi_{2,3}$ and $\phi_{2,2}$ are superior to the other χ^2 statistics. For samples of at least $T=120$, both likelihood ratio statistics perform equally well. At $T=60$ the $\phi_{2,3}$ F statistic is well behaved but the $\phi_{2,2}$ χ^2 statistic is not. In addition, at $T=60$ the power of the F test against several alternatives was found to be poor. Once again, it is worth noting that the likelihood ratio χ^2 statistic $\phi_{2,2}$ is superior to the statistic given by (3) in section 3. The statistic $\phi_{2,2}$ therefore is superior to Ross's likelihood ratio χ^2.

In the next section of the paper, the two likelihood ratio test statistics χ^2 and F will be used to compare sets of assets including stocks, government bonds and corporate bonds.

Table 7

Sampling mean, variance and right tail areas of the χ^2 statistics $\phi_{2,1}$, $\phi_{2,2}$ and $\phi_{2,4}$ and the F statistic $\phi_{2,3}$ for tests of the efficiency ($H_{02}\cdot a_1=\mu_{m1}^2/\sigma_{m1}^2$) of the $N_1=20$ stock tangency portfolio $m1$, for sample sizes $T=60$, 120 and 240 Each experiment is based on 200 replications, under a true null hypothesis

Statistic	Mean	Variance	Right tail areas								
			Sample size $T=60$								
χ^2	19.0	38.0	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
$\phi_{2,1}$	20.6	62.1	0.99	0.93	0.91	0.77	0.52	0.34	0.22	0.14	0.03
$\phi_{2,2}$	19.4	36.2	1.00	0.94	0.92	0.79	0.51	0.29	0.11	0.05	0.01
$\phi_{2,4}$	17.7	21.1	1.00	0.94	0.91	0.74	0.44	0.14	0.02	0.00	0.00
F	1.05	0.18	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
$\phi_{2,3}$	1.06	0.16	1.00	0.94	0.91	0.77	0.51	0.28	0.10	0.04	0.01
			Sample size $T=120$								
χ^2	19.0	38.0	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
$\phi_{2,1}$	18.7	42.6	0.98	0.92	0.85	0.77	0.47	0.22	0.09	0.05	0.02
$\phi_{2,2}$	18.2	33.4	0.99	0.93	0.86	0.77	0.45	0.20	0.07	0.04	0.01
$\phi_{2,4}$	16.8	24.4	0.98	0.91	0.84	0.71	0.32	0.12	0.04	0.01	0.00
F	1.02	0.13	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
$\phi_{2,3}$	0.97	0.11	0.99	0.93	0.88	0.78	0.44	0.18	0.06	0.03	0.01
			Sample size $T=240$								
χ^2	19.0	38.0	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
$\phi_{2,1}$	20.0	49.7	1.00	0.96	0.94	0.78	0.57	0.30	0.13	0.09	0.03
$\phi_{2,2}$	19.7	43.2	1.00	0.96	0.94	0.78	0.56	0.28	0.12	0.07	0.02
$\phi_{2,4}$	18.1	32.8	0.99	0.95	0.90	0.71	0.45	0.20	0.06	0.03	0.01
F	1.01	0.12	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
$\phi_{2,3}$	1.05	0.14	1.00	0.96	0.95	0.78	0.52	0.28	0.12	0.07	0.02

5. Some illustrative tests

Two likelihood ratio test statistics have been developed for tests of the potential performance equivalence of different asset sets and for tests of the efficiency of portfolios. Three interesting financial applications of the test statistics are:

(a) the improvement in potential performance from the inclusion of bonds in a stock portfolio,
(b) the efficiency of a portfolio, and the related issue of,
(c) the validity of the Sharpe–Lintner, Black, or multifactor versions of the capital asset pricing model.

The purpose of this section is to examine these issues while illustrating the use of the test statistics.

Table 8

Powers of the $\phi_{2,1}$, $\phi_{2,2}$, $\phi_{2,3}$ and $\phi_{2,4}$ test statistics in an $\alpha=0.05$ test of the efficiency of the equal weight (upper part) and partial equal weight (lower part) portfolios in the $N=50$, Γ stock set (H_{02} $a=\mu_p^2/\sigma_p^2$) and in the $N_1=20$, Γ_1 stock set (H_{02} $a_1=\mu_{p1}^2/\sigma_{p1}^2$). Each experiment is based on 200 replications, with parameters $a=0.191$, $a_1=0.096$, $\mu_p^2/\sigma_p^2=(0.065$ for the equal weight case and 0.041 for the partial equal weight case) and $\mu_{p1}^2/\sigma_{p1}^2=(0.055$ for the equal weight case and 0.041 for the partial equal weight case)

Statistic	Test powers					
	$T=60$		$T=120$		$T=240$	
	$N=50, \Gamma$	$N_1=20, \Gamma_1$	$N=50, \Gamma$	$N_1=20, \Gamma_1$	$N=50, \Gamma$	$N_1=20, \Gamma_1$
$\phi_{2,1}$	0.42	0.21	0.41	0.18	0.74	0.46
$\phi_{2,2}$	0.41	0.10	0.28	0.13	0.63	0.39
$\phi_{2,3}$	0.03	0.09	0.25	0.13	0.67	0.40
$\phi_{2,4}$	0.00	0.00	0.06	0.06	0.41	0.28
$\phi_{2,1}$	0.43[a]	0.23	0.48	0.27	0.82	0.60
$\phi_{2,2}$	0.42	0.14	0.33	0.16	0.77	0.55
$\phi_{2,3}$	0.03	0.13	0.31	0.16	0.78	0.55
$\phi_{2,4}$	0.00	0.05	0.08	0.12	0.61	0.48

[a]The probability of rejecting with $\phi_{2,1}$ the (false) null hypothesis that the potential performance 'a' of the $N=50$ stock set Γ is equal to the performance μ_p^2/σ_p^2 of the equal weight portfolio of 10 stocks is 0.43 with $T=60$ return observations

Four time periods, each containing $T=60$ months, were chosen to coincide with the last four time periods in Gibbons (1982). The subperiods are January 1956 — December 1960, January 1961 — December 1965, January 1966 — December 1970 and January 1971 — December 1975. The monthly returns for all securities listed on the CRSP tape, without missing returns, were obtained for each of the four subperiods. The beta's, estimated in a subperiod for all stocks using the CRSP equally weighted index, were ranked and used to group the stocks into forty equally weighted portfolios each containing the same number of stocks. This beta grouping procedure is not necessary or necessarily best but is used in order that our results may be compared with those of Gibbons.[11] In addition, the same stock set was used to form twenty portfolios according to the same criterion.

[11]The objective, in reducing the dimensionality of the assets, should be to choose an asset set ($N<T$) which has an efficient set arbitrarily close to the global efficient set That is, the chosen set of N assets should have maximum potential performance 'a' given the constraints of N and T. Since 'a' is a function of both $\boldsymbol{\mu}$ and Σ, grouping assets by betas may not achieve the desired results If the asset choice is by 'a', then it is important that the assets be selected outside the test period to avoid biasing the test A possible approach to estimating the dominant set of assets is provided by Jobson and Korkie (1982)

The Ibbotson and Sinquefield (1979) historical monthly return series on U.S. Treasury bills, long-term corporate and Government bonds were obtained for each of the subperiods.[12] The input data employed also included the returns on the CRSP value weighted stock index over the four subperiods.

5.1. The potential performance contributions of bond and stock indices

We are concerned with whether there is any significant improvement in potential performance in two situations: (1) the addition of a bond index to a set consisting of stocks and a stock index and (2) the addition of a stock index to a set consisting of stocks and a bond index. In order to answer these questions, three asset sets were defined as:

(1) forty stock portfolios, the equally weighted index of the corporate bond and the government bond indices, and the CRSP value weighted index;
(2) forty stock portfolios, and the CRSP value weighted index; and
(3) forty stock portfolios and the equally weighted index of corporate bond and the government bond indices.

The monthly excess (above the Treasury bill rate) return samples were used to estimate the potential performances $\hat{a}_t$ of the three asset sets in each of the four subperiods and are shown in table 9. Comparisons of the estimated potential performances in any period indicate that the potential performance increases with the addition of either the bond index or the CRSP index. For example, in the period 1956/1–1960/12, the bond index increases performance from 2.93 to 3.27 while the CRSP index increases performance from 2.70 to 3.27. With the exception of the period 1961/1–1965/12, the CRSP index generally increases performance more than the bond index. The next question is whether these increases in potential performance are significant at conventional type I error levels.

The likelihood χ^2 and F statistics ($\phi_{1,2}$ and $\phi_{1,3}$) for these tests were computed and are reported in table 9, with significant statistics indicated. No significant differences in potential performance appear until the last two periods, when the CRSP index makes a significant (at the $\alpha=0.05$ level) increase in potential performance, and the bond index makes a significant (at the $\alpha=0.10$ level) increase only in the final period.

[12]The Ibbotson–Sinquefield monthly return long term corporate bond series was checked for autocorrelation by computing the sample autocorrelations over 24 lags. Returns were identified as a serially uncorrelated series, which means that historical means and standard deviations might be used to estimate the single month holding parameters In contrast, Treasury bill returns are highly serially correlated suggesting homogeneous nonstationarity and therefore excess bond and stock returns will also be theoretically nonstationary. However, because the Treasury bill return variance is small relative to the return variance of the other assets, the nonstationarity in the excess returns is difficult to observe

Table 9 also presents the results for similar tests based on twenty portfolios of stocks rather than the forty portfolio tests. The asset sets are labelled (4), (5) and (6). The CRSP index makes a significant contribution to performance in every subperiod, whereas the bond index makes no significant contribution. Our result that the bond index adds significantly to the potential performance of 40 portfolios but not significantly to 20 portfolios, is due to the change in the degrees of freedom for the statistical test.

These tests seem to indicate that a small number of stock portfolios (i.e., $N \leqq 40$) are insufficient representatives of the efficient set of assets, in that a broadly based index of stock returns adds significantly to the performance of

Table 9

Statistics for tests of the equivalent potential performance (H_{01} $a_i = a_j$) of asset sets (1) and (2), (1) and (3), (4) and (5), and (4) and (6) in four subperiods using the χ^2 statistic $\phi_{1,2}$ and F statistic $\phi_{1,3}$.

Asset set	Statistic	Subperiod results 1956/1–1960/12	1961/1–1965/12	1966/1–1970/12	1971/1–1975/12
(1) 40 portfolios of stocks, bond index, CRSP index[a]	$\hat{a}_1$	3.27	3.75	2 10	1.61
(2) 40 portfolios of stocks, CRSP index	$\hat{a}_2$	2.93	3.47	1.87	1 19
	$\phi_{1,2}$	1.45	1.06	1 35	3.07[b]
	$\phi_{1,3}$	1.47	1 06	1 36	3.26[b]
(3) 40 portfolios of stocks, bond index	$\hat{a}_3$	2.70	3.71	1.39	0 88
	$\phi_{1,2}$	2.51	0.15	4 55[c]	5.74[c]
	$\phi_{1,3}$	2 62	0.14	5.33[c]	6.60[c]
(4) 20 portfolios of stocks, bond index, CRSP index	$\hat{a}_4$	1.61	0.85	0.50	0.58
(5) 20 portfolios of stocks, CRSP index	$\hat{a}_5$	1.51	0.85	0.46	0 52
	$\phi_{1,2}$	1.47	0 00	1.01	1.45
	$\phi_{1,3}$	1.47	0.00	1.01	1 46
(6) 20 portfolios of stocks, bond index	$\hat{a}_6$	1.39	0.61	0.32	0 31
	$\phi_{1,2}$	3.30[b]	5.21[c]	4.79[c]	7.02[c]
	$\phi_{1,3}$	3.41[b]	5.52[c]	5 05[c]	7.63[c]

[a]The CRSP value weight index and an equal weight index of the Ibbotson–Sinquefield corporate and Government bond indices.

[b]Significant difference, at the $\alpha = 0.10$ level, between the potential performances of the asset set (1) and either asset set (2) or (3) or asset set (4) and either asset set (5) or (6)

[c]Significant difference at the $\alpha = 0.05$ level.

the portfolios. A bond index, while increasing the set's performance, does not consistently add significantly to performance.

The tests have investigated the mean-standard deviation dominance of one set of assets over another set. The next section investigates whether the stock and bond indices are efficient in an asset set based on a relatively small number of portfolios.

5.2. The efficiency of bond and stock portfolios

The choice of a market portfolio proxy and its efficiency is the nexus between financial theory and tests of the efficiency of the true market portfolio or the validity of the capital asset pricing model (CAPM). If the true market portfolio m, comprised of N^* assets, is identical to the unique tangency portfolio in mean excess return–standard deviation space, then the CAPM is true. That is, the mean excess returns are exactly linearly related to their betas, as in $\mu = \beta\mu_m$.

The efficiency of the market portfolio and its CAPM result present two approaches to tests of the underlying financial theory. First, one can choose what is believed to be a proxy to the unobservable true market portfolio and then test if an exact relationship exists between mean excess returns and betas. Second, one can simply test if the proxy market portfolio has risk–return parameters equal to the tangency portfolio.

Since the market portfolio is unobservable and the number of assets N^* is seemingly limitless, rejection of CAPM for a given proxy or equivalently rejection of the efficiency of the proxy may be caused by an inefficient true market portfolio or a bad proxy. On the other hand, acceptance of CAPM may be due to an efficient true market portfolio or an efficient proxy in too small an asset subspace. Thus, a good or bad proxy can cause symmetrical decision errors in hypothesis testing. By testing the efficiency of market proxies in an $N \ll N^*$ asset space, we are testing CAPM conditionally on the risk-free rate and the set of assets.

For the asset set comprised of forty portfolios, the CRSP value weighted index, and an equally weighted bond index (of the corporate and government bond indices), we are interested in the efficiency of the CRSP stock index and the bond index. Results are also obtained with twenty portfolios replacing the forty portfolios. Excess returns (above the Treasury bill rates) were used to estimate the squared Sharpe performance of the two indices in each subperiod. Maximum likelihood estimates ($\hat{a} = \bar{r}_p^2/s_p^2$) are displayed in table 10 (rows labelled $\hat{a}$) together with the potential performance of the asset set. The χ^2 and F statistics, $\phi_{2,2}$ and $\phi_{2,3}$, are reported in the table with significant values ($\alpha = 0.05$ or 0.10) noted by asterisks.

Table 10

Statistics for tests of the efficiency ($H_{02}: a_i = a_j$) of the CRSP value weighted index and Ibbotson–Sinquefield bond index in four subperiods using the χ^2 statistic $\phi_{2,2}$ and F statistic $\phi_{2,3}$. Comparisons are made between asset sets (1) and (2), (1) and (3), (4) and (5), and (4) and (6).

Asset set	Statistic	Subperiod results 1956/1–1960/12	1961/1–1965/12	1966/1–1970/12	1971/1–1975/12
(1) 40 portfolios of stocks, bond index, CRSP index	$\hat{a}_1$	3.27	3.75	2.10	1.61
	$\hat{a}_2$	0.034	0.061	0.000	0.000
(2) CRSP index	$\phi_{2,2}$	53.18[a]	56.21[a]	42.43	35.98
	$\phi_{2,3}$	1.30	1.41	0.87	0.67
	$\hat{a}_3$	0.004	0.001	0.025	0.001
(3) Bond index	$\phi_{2,2}$	52.29	58.39[b]	41.50	35.94
	$\phi_{2,3}$	1.35	1.53	0.86	0.67
(4) 20 portfolios of stocks, bond index, CRSP index	$\hat{a}_4$	1.61	0.85	0.50	0.58
	$\hat{a}_5$	0.034	0.061	0.000	0.000
(5) CRSP index	$\phi_{2,2}$	43.98[b]	26.41	19.06	21.73
	$\phi_{2,3}$	2.69[b]	1.31	0.88	1.02
	$\hat{a}_6$	0.004	0.001	0.025	0.001
(6) Bond index	$\phi_{2,2}$	45.38[b]	29.17	17.90	21.73
	$\phi_{2,3}$	2.82[b]	1.49	0.82	1.02

[a]Significance at the $\alpha = 0.10$ level.
[b]Significance at the $\alpha = 0.05$ level.

In the period 1956/1–1960/12, the CRSP index and bond index have squared Sharpe performances of 0.034 and 0.004 compared to the potential performance of 3.27 for the entire asset set. The CRSP index is significantly inefficient using the χ^2 but not using the F statistic. The bond index is not significant. When twenty portfolios, rather than forty are used, table 10 also shows that the CRSP and bond index are significantly inefficient. This seemingly anomalous result is due to the change in the size of N relative to T and its affect on the sample statistics and degrees of freedom.

No other significant inefficiencies in the CRSP or bond indices arise except in the period 1961/1–1965/12 where the χ^2 statistic shows inefficient CRSP and bond indices. This conclusion holds despite the large differences in performances between an index and the asset set.

The tests indicate that the market proxy is not significantly inefficient in a small asset universe ($N \leq 42$), although the difference in estimated potential performance and an index's measured performance is quite large. Since potential performance increases with the number of assets N, significant differences with larger asset sets might be observed. Unfortunately, this also requires larger sample sizes $T>N$ to preserve the full rank of the covariance matrix. The key to better tests with small N and T seems to be a better procedure for obtaining a small number of assets that generate an efficient set which is comparable to the so-called true efficient set (see footnote 11).

The next section provides some comparative tests with Gibbons' test of CAPM.

5.3. Sharpe–Lintner and Black versions of CAPM

The preceding test of the efficiency of the CRSP value weighted index was also a test of the Sharpe–Lintner version of CAPM, since the risk-free rate on Treasury bills was used to compute excess returns. Gibbons provides a test of the Black CAPM where he uses the Black, Jensen and Scholes estimator $\hat{\gamma}_{\mathrm{BJS}}$ for the subperiod zero beta mean return as well as his one-step Gauss–Newton estimator $\hat{\gamma}^*$. We have calculated the average Treasury bill rate for each subperiod $\bar{r}_f$ and presented it in table 11 along with $\hat{\gamma}_{\mathrm{BJS}}$ and $\hat{\gamma}^*$, as determined by Gibbons. In general, the average T-bill rates are increasing in time whereas the estimated zero beta rates are not increasing.

Table 11

χ^2 statistic $\phi_{2,2}$ from tests of the Black CAPM, using the Treasury bill return r_{ft}, the Black, Jensen and Scholes zero beta return $\hat{\gamma}_{\mathrm{BJS}}$, the Gibbons zero beta return $\hat{\gamma}^*$ and zero, over four subperiods. Tests are conducted using forty portfolios of stocks and the CRSP equal weight index.

	Zero beta mean returns			$\phi_{2,2}$ χ^2 statistics using				Revised[a] Gibbons χ^2
Period	$\bar{r}_f$	$\hat{\gamma}_{\mathrm{BJS}}$	$\hat{\gamma}^*$	r_{ft}	$\hat{\gamma}_{\mathrm{BJS}}$	$\hat{\gamma}^*$	Zero	
1956/1–1960/12	0.21	1.29	0.81	44.58	43.58	41.34	47.28	43.29
1961/1–1965/12	0.25	0.70	0.68	51.73	53.66[b]	53.49[b]	52.45[b]	35.69
1966/1–1970/12	0.44	0.45	0.01	31.54	31.54	34.05	34.30	29.15
1971/1–1975/12	0.47	0.96	0.61	22.76	27.89	23.71	23.44	32.59

[a]χ^2 values from Gibbons (1982, table 1, p. 13) adjusted with Bartlett's correction.
[b]Significant inefficiency for the CRSP index or rejection of the CAPM at $\alpha=0.10$

The subperiod excess returns on the CRSP equal weight index and on the forty portfolios, described earlier, were obtained by subtracting zero, r_{ft} (not $\bar{r}_f$), $\hat{\gamma}_{\mathrm{BJS}}$ or $\hat{\gamma}^*$ from the portfolio and index returns. The $\phi_{2,2}$ χ^2 statistics for tests of CRSP index efficiency were computed for each subperiod and are presented in table 11. The Black version of CAPM is rejected only in the second subperiod 1961/1–1965/12 at the $\alpha=0.10$ level, using either $\hat{\gamma}^*$ or $\hat{\gamma}_{\mathrm{BJS}}$ as the mean zero beta return. Rejection of the models in other subperiods does not occur even when a zero value of the zero beta rate is used.

These results contrast sharply with Gibbons (1982, table 1, p. 13), where the Black version of CAPM is rejected in three of the four subperiods at the $\alpha=0.10$ level. The excess power of the Gibbons test might be due to the nonconformity of his χ^2, with a theoretical χ^2 when N is large relative to T (see our sections 3 and 4). We have found that the χ^2 statistic [eq. (3)], which is not adjusted by Bartlett's correction factor, exhibits larger skewness than a theoretical χ^2 if N is large relative to T. This is the case in Gibbons, where $N=40$ and $T=60$, and the Black CAPM is rejected. The corrected value of Gibbons' χ^2 (using Bartlett's correction) is shown in the last column of table 11, where no significant statistics appear.[13] Thus the Black CAPM is generally not rejected when an appropriate adjustment is made for the excess skewness in the sample χ^2 statistic.

6. Conclusion

The concept of potential performance of a set of assets was defined as the maximum Sharpe measure attainable from a portfolio of the assets. The concept provides a useful linkage to multivariate tests of the potential performance contribution of additional assets, the efficiency of market proxy portfolios, and the Sharpe–Lintner and Black asset pricing models.

Two likelihood ratio statistics, χ^2 and F, were found to be well behaved in small samples ($T \geqq 60$) of monthly returns on stocks. The F statistic was particularly well behaved and suited for multi-variate test of financial models.

The potential performance concept was illustrated by tests which showed that the potential performance of forty portfolios of stocks was significantly less than the potential performance of the forty portfolios and the CRSP value weighted index combined. The portfolio efficiency test was illustrated by tests which showed that inefficiency in the CRSP value weighted index could not be detected. This was also a direct test of the Sharpe–Lintner version of the CAPM. A test of the Black CAPM, using Black, Jensen and

[13]It is also interesting to note that the periods, in which Gibbons rejects the CAPM, seem to have a global minimum variance return which is either less than the $\hat{\gamma}^*$ or very close to it This implies that the tangency point is on the lower boundary of the efficient set or on the extreme upper right boundary

Scholes' and Gibbons' zero beta returns, resulted in model rejection in only one of four subperiods. This contrasts with the results of Gibbons who rejects the model in three of the four subperiods replicated here.

There are at least two problems that require further investigation. First, the comparative power of Gibbons and our methodologies needs to be examined in more detail. Our initial judgement is that the power of the Gibbons' procedure is overstated due to the excessive positive skewness in his χ^2 statistic when the number of assets is large relative to the sample size. Second, methods of increasing the power, by choosing portfolios that maximize potential performance, need to be investigated.

Appendix

A.1. Derivation of the likelihood ratio test for $H_{01}:\boldsymbol{\mu}_2=\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\mu}_1$

For T independent samples from the multivariate normal $\boldsymbol{r}_t \sim N(\boldsymbol{\mu}, \Sigma)$ where $\boldsymbol{\mu}(N\times 1)$, $\boldsymbol{\Sigma}(N\times N)$, $\boldsymbol{r}_t(N\times 1)$, $t=1,2,\ldots,T$, the logarithm of the likelihood function is given by

$$\log L=\frac{-NT}{2}\log 2\pi-\frac{T}{2}\log|\boldsymbol{\Sigma}|-\frac{1}{2}\sum_{t=1}^{T}(\boldsymbol{r}_t-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{r}_t-\boldsymbol{\mu}).$$

If there are no restrictions on the parameters the maximum likelihood estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are given by [Anderson (1958)]

$$\bar{\boldsymbol{r}}=\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{r}_t \quad\text{and}\quad \boldsymbol{S}=\frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{r}_t-\bar{\boldsymbol{r}})(\boldsymbol{r}_t-\bar{\boldsymbol{r}})'.$$

The logarithm of the likelihood function evaluated at $\bar{\boldsymbol{r}}$ and $\boldsymbol{S}$ is given by [Anderson (1958)]

$$\log \tilde{L}=(-NT/2)\log 2\pi-(T/2)\log|\boldsymbol{S}|-(NT/2). \tag{A.1}$$

To test $H_0:\boldsymbol{\mu}_2=\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\mu}_1$ we require the maximum likelihood estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ under this restriction. Using the conventional identity for the inverse of a partitioned matrix and the fact that $|\boldsymbol{\Sigma}|=|\boldsymbol{\Sigma}_{11}||\boldsymbol{\Sigma}_{22\cdot 1}|$ we may write

$$\log L=(-NT/2)\log 2\pi-(T/2)\log|\Sigma_{11}|-(T/2)\log|\Sigma_{22\,1}|$$

$$-\frac{1}{2}\sum_{t=1}^{T}(r_{t1}-\mu_1)'\Sigma_{11}^{-1}(r_{t1}-\mu_1)$$

$$-\frac{1}{2}\sum_{t=1}^{T}[(r_{t2}-\mu_2)-\Sigma_{21}\Sigma_{11}^{-1}(r_{t1}-\mu_1)]'$$

$$\times\Sigma_{22.1}^{-1}[(r_{t2}-\mu_2)-\Sigma_{21}\Sigma_{11}^{-1}(r_{t1}-\mu_1)], \tag{A.2}$$

where

$$\Sigma_{22\,1}=[\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}].$$

Define the matrix $\Theta=\Sigma+\mu\mu'$ and partition Θ to conform to the partitioning of Σ. We may therefore write $\Theta_{11}=\Sigma_{11}+\mu_1\mu_1'$ and hence by employing an identity,[14] Θ_{11}^{-1} may be written

$$\Theta_{11}^{-1}=\Sigma_{11}^{-1}-(\Sigma_{11}^{-1}\mu_1\mu_1'\Sigma_{11}^{-1})/(1+\mu_1'\Sigma_{11}^{-1}\mu_1).$$

After multiplying through by Θ_{21} and Θ_{12}, obtain

$$\Theta_{22.1}=\Sigma_{22.1}+((\mu_2-\Sigma_{21}\Sigma_{11}^{-1}\mu_1)(\mu_2-\Sigma_{21}\Sigma_{11}^{-1}\mu_1)')/(1+\mu_1'\Sigma_{11}^{-1}\mu_1), \tag{A.3}$$

and therefore under H_{01}, $\Theta_{22.1}=\Sigma_{22\,1}$. Defining $\beta=\Theta_{21}\Theta_{11}^{-1}$, then under H_{01}, the log likelihood may be written as

$$\log L=(-NT/2)\log 2\pi-(T/2)\log|\Sigma_{11}|-\frac{1}{2}\sum_{t=1}^{T}(r_{t1}-\mu_1)'\Sigma_{11}^{-1}(r_{t1}-\mu_1)$$

$$-(T/2)\log|\Theta_{22.1}|-\frac{1}{2}\sum_{t=1}^{T}(r_{t2}-\beta r_{t1})'\Theta_{22\,1}^{-1}(r_{t2}-{}^{B}r_{t1}).$$

Note that $\log L=f_1(\mu_1,\Sigma_{11})+f_2(\beta,\Theta_{22\,1})$, where $f_1(\mu_1,\Sigma_{11})$ denotes the marginal likelihood for r_{t1} and $f_2(\beta,\Theta_{22\,1})$ is related to the conditional distribution of r_{t2} given r_{t1}. The maximum likelihood estimators of μ_1, Σ_{11}, β

[14] Given a full rank matrix B $(N\times N)$ and a vector d $(N\times 1)$, then

$$(B+dd')^{-1}=B^{-1}-B^{-1}dd'B^{-1}/(1+d'Bd).$$

See Rao (1973, p 33, no 2 8)

and $\boldsymbol{\Theta}_{22.1}$ are given by

$$\bar{\boldsymbol{r}}_1,\ \boldsymbol{S}_{11}, \hat{\boldsymbol{\beta}}=\left[\sum_{t=1}^{T} \boldsymbol{r}_{1t}\boldsymbol{r}'_{2t}\right]\left[\sum_{t=1}^{T} \boldsymbol{r}_{1t}\boldsymbol{r}'_{1t}\right]^{-1},$$

and

$$\boldsymbol{V}_{22.1}=\sum_{t=1}^{T} \boldsymbol{r}_{2t}\boldsymbol{r}'_{2t}-\left[\sum_{t=1}^{T} \boldsymbol{r}_{1t}\boldsymbol{r}'_{t1}\right]^{-1} \hat{\boldsymbol{\beta}},$$

respectively.

Substituting the maximum likelihood estimators in the likelihood function we obtain

$$\log \hat{L}=(-NT/2)\log 2\pi-(T/2)\log|\boldsymbol{S}_{11}|-(NT/2)-(T/2)\log|\boldsymbol{V}_{22\ 1}|.$$

The logarithm of the ratio of the likelihood functions is therefore given by

$$\log \lambda=\log \hat{L}-\log \tilde{L}=(T/2)\log|\boldsymbol{S}|-(T/2)\log|\boldsymbol{S}_{11}|-(T/2)|\boldsymbol{V}_{22\ 1}|$$

$$=\log(T/2)(|\boldsymbol{S}_{22.1}|/|\boldsymbol{V}_{22.1}|),$$

since

$$|\boldsymbol{S}|=|\boldsymbol{S}_{11}||\boldsymbol{S}_{22.1}|.$$

A special case of Wilk's Lambda is given by

$$\Lambda=|\boldsymbol{S}_{22\ 1}|/|\boldsymbol{V}_{22\ 1}|. \tag{A.4}$$

This ratio expression of Λ may be simplified by exploiting the relationship between $\boldsymbol{S}_{22.1}$ and $\boldsymbol{V}_{22.1}$. Let $\boldsymbol{V}$ denote the matrix given by $(1/T)\sum_{t=1}^{T} \boldsymbol{r}_t\boldsymbol{r}'_t$ and let $\boldsymbol{V}$ be partitioned to conform to that of $\boldsymbol{S}$ and $\boldsymbol{\Sigma}$. The matrices $\boldsymbol{V}$ and $\boldsymbol{S}$ are therefore, related by $\boldsymbol{V}=\boldsymbol{S}+\bar{\boldsymbol{r}}\bar{\boldsymbol{r}}'$ and $\boldsymbol{V}_{11}=\boldsymbol{S}_{11}+\bar{\boldsymbol{r}}_1\bar{\boldsymbol{r}}'_1$. In the same fashion as (A.3), $\boldsymbol{V}_{22\ 1}$ may be written as

$$\boldsymbol{V}_{22\ 1}=\boldsymbol{S}_{22\ 1}+((\bar{\boldsymbol{r}}_2-\boldsymbol{S}_{21}\boldsymbol{S}_{11}^{-1}\bar{\boldsymbol{r}}_1)(\bar{\boldsymbol{r}}_2-\boldsymbol{S}_{21}\boldsymbol{S}_{11}^{-1}\bar{\boldsymbol{r}}_1)')/(1+\bar{\boldsymbol{r}}'_1\boldsymbol{S}_{11}^{-1}\bar{\boldsymbol{r}}_1).$$

The ratio expression for Λ given by (A.4) may now be written as

$$\Lambda=(1+\bar{\boldsymbol{r}}'_1\boldsymbol{S}_{11}^{-1}\bar{\boldsymbol{r}}_1)/(1+\bar{\boldsymbol{r}}'\boldsymbol{S}^{-1}\bar{\boldsymbol{r}}), \tag{A.5}$$

which is obtained by using the sample form of eq. (1) in section 3 and the identity [see Anderson (1958, ch. 8)]

$$\left|V_{22.1}\right| = \left|S_{22.1} + \frac{(\bar{r}_2 - S_{21}S_{11}^{-1}\bar{r}_1)(\bar{r}_2 - S_{21}S_{11}^{-1}\bar{r}_1)'}{(1+\bar{r}_1'S_{11}^{-1}\bar{r}_1)}\right|$$

$$= \begin{vmatrix} 1 & -(\bar{r}_2 - S_{21}S_{11}^{-1}\bar{r}_1)/(1+\bar{r}_1'S_{11}^{-1}\bar{r}_1)^{\frac{1}{2}} \\ (\bar{r}_2 - S_{21}S_{11}^{-1}\bar{r}_1)/(1+\bar{r}_1'S_{11}^{-1}\bar{r}_1)^{\frac{1}{2}} & S_{22.1} \end{vmatrix}$$

$$= \left|S_{22.1}\right| \left[\frac{(\bar{r}_2 - S_{21}S_{11}^{-1}\bar{r}_1)'S_{22.1}^{-1}(\bar{r}_2 - S_{21}S_{11}^{-1}\bar{r}_1)}{(1+\bar{r}_1'S_{11}^{-1}\bar{r}_1)} + 1\right]$$

The likelihood ratio λ is therefore given by

$$\lambda = (1+\bar{r}_1'S_{11}^{-1}\bar{r}_1)^{T/2}/(1+\bar{r}'S^{-1}\bar{r})^{T/2}. \tag{A.6}$$

This expression is employed in section 3.2 of the paper to develop a likelihood ratio test statistic for H_{01}.

A.2. Derivation of the score test for $H_{01}:\mu_2 = \Sigma_{21}\Sigma_{11}^{-1}\mu_1$

For a vector of parameters Θ, the score test statistic for $H_0:\Theta=\Theta^*$ is given by $D'(\Theta^*)I^{-1}(\Theta^*)D(\Theta^*)$, where $D(\Theta^*)$ is the efficient score vector evaluated at $\Theta=\Theta^*$ and $I(\Theta^*)$ is the information matrix evaluated at $\Theta=\Theta^*$. The efficient score vector is the vector of first partial derivatives of the sample log likelihood function with respect to the elements of Θ. The information matrix is the negative of the expectation of the matrix of second partial derivatives for all possible pairs of parameters in Θ. In the exponential family of distributions, the inverse of the information matrix is equivalent to the covariance matrix.

To test a hypothesis $H_0:\Theta_1=\Theta_1^*$ only the portion of $D(\Theta^*)$ pertaining to Θ_1 is used and is written as $D_1(\Theta_1^*,\Theta_2)$. The subvector D_1 may contain some parameters, say Θ_2, not specified by the hypothesis and hence are replaced by their maximum likelihood estimators denoted by $\tilde{\Theta}_2$. The portion of the inverse of the information matrix pertaining to Θ_1 is denoted by $I^{11}(\Theta_1^*,\tilde{\Theta}_2)$, where again $\tilde{\Theta}_2$ denotes the maximum likelihood estimator of parameters not specified under the null hypothesis. The score test statistic therefore becomes

$$D_1'(\Theta_1^*,\tilde{\Theta}_2)\, I^{11}(\Theta_1^*,\tilde{\Theta}_2)\, D_1(\Theta_1^*,\tilde{\Theta}_2).$$

A modification of the score test statistic, when there are nuisance parameters, is obtained if $I^{11}(\Theta_1^*, \tilde{\Theta}_2)$ is replaced by $I_{11}^{-1}(\Theta_1^*, \tilde{\Theta}_2)$, where the latter matrix represents the inverse of only the portion of the information matrix that pertains to Θ_1. In this case the score test statistic is given by

$$D_1'(\Theta_1^*, \tilde{\Theta}_2)\; I_{11}^{-1}(\Theta_1^*, \tilde{\Theta}_2)\; D_1'(\Theta_1^*, \tilde{\Theta}_2).$$

The score test statistic and modified score test statistic for testing H_{01} are developed next.

From (A.2) the logarithm of the likelihood function may be written

$$\begin{aligned}\log L = &(-NT/2)\log 2\pi - (T/2)\log|\Sigma_{11}| - (T/2)\log|\Sigma_{22\,1}| \\ &-\frac{1}{2}\sum_{t=1}^{T}(r_{t1}-\mu_1)'\Sigma_{11}^{-1}(r_{t1}-\mu_1) \\ &-\frac{1}{2}\sum_{t=1}^{T}\{[(r_{t2}-\Sigma_{21}\Sigma_{11}^{-1}r_{t1})-(\mu_2-\Sigma_{21}\Sigma_{11}^{-1}\mu_1)]'\Sigma_{22\,1}^{-1} \\ &\times[(r_{t2}-\Sigma_{21}\Sigma_{11}^{-1}r_{t1})-(\mu_2-\Sigma_{21}\Sigma_{11}^{-1}\mu_1)]\}.\end{aligned}$$

The efficient score vector for the parameter $(\mu_2-\Sigma_{21}\Sigma_{11}^{-1}\mu_1)$ may therefore be written as

$$\begin{aligned}\partial\log L/\partial[\mu_2-\Sigma_{21}\Sigma_{11}^{-1}\mu_1] = &-T\Sigma_{22\,1}^{-1}(\bar{r}_2-\Sigma_{21}\Sigma_{11}^{-1}\bar{r}_1) \\ &+T\Sigma_{22.1}^{-1}(\mu_2-\Sigma_{21}\Sigma_{11}^{-1}\mu_1),\end{aligned} \tag{A.7}$$

which under H_{01} becomes $-T\Theta_{22\,1}^{-1}(\bar{r}_2-\eta\bar{r}_1)$ where $\eta=\Sigma_{21}\Sigma_{11}^{-1}$. The covariance matrix for the maximum likelihood estimator $[\bar{r}_2-S_{21}S_{11}^{-1}\bar{r}_1]$, using Ω defined in section 3.1, is given by $[((T-2)+Ta_1)/(T-N_1-2)]\,\Sigma_{22\,1}$, which in the exponential family corresponds to the portion of the inverse of the information matrix corresponding to the parameter $[\mu_2-\Sigma_{21}\Sigma_{11}^{-1}\mu_1]$. The value of this matrix under H_{01} is $[((T-2)+Ta)/(T-N_1-2)]\Theta_{22\,1}$. The score statistic is therefore given by

$$T[((T-2)+Ta)/(T-N_1-2)](\bar{r}_2-\eta\bar{r}_1)'\Theta_{22.1}^{-1}\Theta_{22\,1}\Theta_{22.1}^{-1}(\bar{r}_2-\eta\bar{r}_1),$$

which is approximately

$$(T^2/(T-N_1))(\bar{r}_2-\eta\bar{r}_1)'\Theta_{22.1}^{-1}(\bar{r}_2-\eta\bar{r}_1)(1+a).$$

Replacing the parameters η and $\Theta_{22\,1}$ by their maximum likelihood estimators under H_{01} yields the statistic

$$
\begin{aligned}
&(T^2/(T-N_1))(\bar{r}_2-V_{21}V_{11}^{-1}\bar{r}_1)'V_{22\cdot 1}^{-1}(\bar{r}_2-V_{21}V_{11}^{-1}\bar{r}_1)(1+\hat{a}) \\
&=(T^2/(T-N_1))([\hat{a}-\hat{a}_1]/[1+\hat{a}_1]),
\end{aligned}
\tag{A.8}
$$

where

$$
V=\frac{1}{T}\sum_{t=1}^{T} r_t r_t' = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}
$$

conforms to the partitioning of $\boldsymbol{\Sigma}$ and $\boldsymbol{\Theta}$.

The adjusted or modified score statistic is obtained by employing the inverse of that portion of the information matrix that pertains to the parameters specified by the null hypothesis. If the parameters not specified by the null hypothesis were known, then the adjusted score statistic would be equivalent to the score statistic. From eq. (A.7) the matrix of second partials of $\log L$ with respect to the elements of $[\boldsymbol{\mu}_2-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\mu}_1]$ is given by $T\boldsymbol{\Sigma}_{22\cdot 1}^{-1}$, which under H_{01} is equivalent to $T\boldsymbol{\Theta}_{22.1}^{-1}$. The efficient score criterion using the modified score statistic, therefore becomes $T(\bar{r}_2-\eta\bar{r}_1)'\boldsymbol{\Theta}_{22\cdot 1}^{-1}(\bar{r}_2-\eta\bar{r}_1)$. Replacing the parameters η and $\boldsymbol{\Theta}_{22\cdot 1}$ by their maximum likelihood estimators yields the statistic

$$
T(\bar{r}_2-V_{21}V_{11}^{-1}\bar{r}_1)'V_{22.1}^{-1}(\bar{r}_2-V_{21}V_{11}^{-1}\bar{r}_1),
$$

which can be written as

$$
T([\hat{a}-\hat{a}_1]/[1+\hat{a}][1+\hat{a}_1]). \tag{A.9}
$$

The score statistics (A.8) and (A.9) are discussed further in section 3.3.

References

Anderson, T.W., 1958, An introduction to multivariate analysis (Wiley, New York).

Bartlett, M.S., 1938, Further aspects of the theory of multiple regression, Proceedings of the Cambridge Philosophical Society 34, 33–47.

Black, F., 1972, Capital market equilibrium with restricted borrowing, Journal of Business 45, 444–454.

Black, F., M.C. Jensen and M Scholes, 1972, The capital asset pricing model· Some empirical tests, in: Studies in the theory of capital markets (Praeger, New York)

Buse, A., 1982, The likelihood ratio, Wald and Lagrange multiplier test: An expository note, American Statistician 36, 153–157

Gibbons, M.R., 1980, Econometric methods for testing a class of financial models — An application of the nonlinear multivariate regression model, Ph.D dissertation (Department of Economics, University of Chicago, Chicago, IL)

Gibbons, M.R , 1982, Multivariate tests of financial models, Journal of Financial Economics 10, 3–27.

Ibbotson, R.G. and R.A. Sinquefield, 1979, Stocks, bonds, bills and inflation: Historical returns (1926–1978), (Financial Analysts Research Foundation, University of Virginia, Charlottesville, VA).
Jensen, M.C., 1968, The performance of mutual funds in the period 1945–1964, Journal of Finance 23, 389–416.
Jobson, J.D., 1982a, A multivariate linear regression test for the arbitrage pricing theory, Journal of Finance 37, 1037–1042.
Jobson, J.D., 1982b, Testing for zero intercept in multivariate normal regression using the univariate multiple regression F test, American Statistician, forthcoming.
Jobson, J.D. and R.M. Korkie, 1980, Estimation for Markowitz efficient portfolios, Journal of the American Statistical Association 75, 544–554.
Jobson, J.D. and R.M. Korkie, 1981, Performance hypothesis testing with the Sharpe and Treynor measures, Journal of Finance 36, 889–908.
Jobson, J.D. and R.M. Korkie, 1982, Statistical inference in two parameter portfolio theory with multiple regression software, Working paper (University of Alberta, Edmonton).
Johnson, N.L. and S. Kotz, 1970, Continuous univariate distributions — 2 (Houghton-Mifflin, Boston, MA).
Korkie, R.M., 1983, Evaluating the internal and external performance of portfolios, Journal of Portfolio Management, forthcoming.
Lintner, J., 1965, Security prices, risk and maximal gains from diversification, Journal of Finance 20, 587–616.
Markowitz, H., 1952, Portfolio selection, Journal of Finance 7, 77–91.
Marx, D.L. and R.R. Hocking, 1977, Moments of certain functions of elements in the inverse Wishart matrix, Paper presented at the meeting of the American Statistical Association, Chicago, IL.
Merton, R., 1972, An analytical derivation of the efficient portfolio frontier, Journal of Financial and Quantitative Analysis 7, 1851–1872.
Morrison, D.F., 1967, Multivariate statistical methods (McGraw-Hill, New York).
Rao, C.R., 1947, Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation, Proceedings of Cambridge Philosophical Society 44, 50–57.
Rao, C.R., 1951, An asymptotic expansion of the distribution of Wilks criterion, Bulletin of the Institute of International Statistics 33, 33–38.
Rao, C.R., 1973, Linear statistical inferences and its applications (Wiley, New York).
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.
Roll, R., 1978, Ambiguity when performance is measured by the securities market line, Journal of Finance 33, 1051–1069.
Roll, R. and S.A. Ross, 1980, An empirical investigation of the arbitrage pricing theory, Journal of Finance 35, 1073–1103.
Ross, S.A., 1977, A note on the capital asset pricing model (CAPM): Short-selling restrictions and related issues, Journal of Finance 32, 177–183.
Ross, S.A., 1980, A test of the efficiency of a given portfolio, Paper presented at a world econometrics meeting, Aix-en-Provence.
Roy, S.N., 1953, On a heuristic method of test construction and its use in multivariate analysis, Annals of Mathematics and Statistics 24, 220–238.
Sharpe, W.F., 1966, Mutual fund performance, Journal of Business, Suppl., no. 1, part 2, 119–138.
Silvey, S.D., 1970, Statistical inference (Penguin Books, Middlesex).
Tobin, J., 1958, Liquidity preference as behavior towards risk, Review of Economic Studies 25, 65–85.
Treynor, J.L., 1965, How to rate management funds, Harvard Business Review XLIII, 63–75.
Wald, A., 1943, Tests of statistical hypotheses concerning several parameters when the number of observations is large, Transactions of the American Mathematical Society 54, 426–482.