Journal of Financial Economics 26 (1990) 221–254. North-Holland

Bayesian inference in asset pricing tests*

Campbell R. Harvey
Duke University, Durham, NC 27706, USA

Guofu Zhou
Washington University, St. Louis, MO 63130, USA

Received February 1989, final version received July 1990

We test the mean–variance efficiency of a given portfolio using a Bayesian framework. Our test is more direct than Shanken's (1987b), because we impose a prior on all the parameters of the multivariate regression model. The approach is also easily adapted to other problems. We use Monte Carlo numerical integration to accurately evaluate 90-dimensional integrals. Posterior-odds ratios are calculated for 12 industry portfolios from 1926–1987. The sensitivity of the inferences to the prior is investigated by using three different distributions. The probability that the given portfolio is mean–variance efficient is small for a range of plausible priors.

1. Introduction

There are two competing approaches to statistical inference: classical and Bayesian. The fundamental difference between them is the notion of probability. In the classical framework, the probability of an event is defined by the limit of its relative frequency. Estimators and test procedures are evaluated in repeated samples. In the Bayesian framework, probability is defined by a degree of belief.

The Bayesian approach makes it possible to incorporate a belief about the hypothesis being tested and its alternative in the form of a prior-odds ratio. When we look at the data, we get a posterior-odds ratio, which summarizes all the evidence (prior and sample) in favor of the hypothesis or its alterna-

*We have benefited from the comments of Douglas Foster, John Geweke, Michael Hemler, Michael Lavine, Robert McCulloch, Peter Muller, Robert Nau, Tom Smith, Robert Winkler, William Schwert (the editor), and seminar participants at Duke University and Washington University. We thank Peter Rossi for all the suggestions he made in detailed correspondence and conversations. We are especially indebted to Jay Shanken (the referee), who provided numerous insights that greatly improved this paper.

tive. The posterior-odds ratio can be interpreted as the ratio of the probability that the hypothesis is valid to the probability that its alternative is valid.

Many applications in finance involve prior beliefs about the behavior of the data. However, almost all empirical analysis has been carried out in the classical framework. The slow adoption of Bayesian econometrics is a result of two practical difficulties in implementing the approach. One is how to choose a prior. The critical difficulty arises, however, in evaluating the posterior distribution. This may involve high-dimensional integration, which is analytically intractable. For example, some of our empirical work involves 90-dimensional integration. Fortunately, with the recent development of Monte Carlo numerical integration, high-order integration problems are routinely solved with a high degree of accuracy.

This paper examines multivariate tests of the mean–variance efficiency of a given portfolio. Usually, these tests are done in a classical framework.[1] Bayesian inference about mean–variance efficiency has received relatively little attention. An exception is an important paper by Shanken (1987b) who uses a result in Gibbons, Ross, and Shanken (1989) to develop a computationally convenient way to calculate the posterior-odds ratio.

Shanken (1987b) uses the posterior-odds ratio to test the restriction imposed by the Sharpe (1964) – Lintner (1965) capital asset pricing model (CAPM) that the intercepts in the multivariate regression of excess returns on the market excess return are equal to zero. Shanken's test is indirect, however. He replaces the intercepts with a function of the intercepts and tests whether this function is zero. With this method, he can impose a prior only on the function, not on the intercepts. More importantly, Shanken's test cannot easily be applied to other problems because it critically relies on the sampling distribution of the classical F statistic proposed in Gibbons, Ross, and Shanken (1989). For example, we cannot apply Shanken's results to test the restrictions implied in the Black (1972) CAPM. Indeed, Shanken realizes that:

> A more ambitious and much more complicated approach to this problem would start with a joint prior distribution for all parameters in the multivariate linear regression of returns.

Our paper addresses this challenge and proposes a full Bayesian specification of the asset pricing model tests.[2] We use the algorithm suggested by Geweke (1988, 1989) to evaluate the posterior distributions. Posterior-odds ratios are calculated using both a diffuse prior and an informative prior to test the restrictions implied by the Sharpe–Lintner CAPM. Further, we

[1] See for example, Gibbons (1982), Stambaugh (1982), Shanken (1985, 1986), MacKinlay (1987), and Gibbons, Ross, and Shanken (1989).

[2] McCulloch and Rossi (1988, 1989a, b) independently address this challenge. Their focus, however, is on the arbitrage pricing theory. They present an interesting methodology that evaluates hypotheses by means of direct utility loss.

check the sensitivity of the inference to the choice of prior distributions. Finally, we calculate Bayesian confidence intervals for the parameters of interest. Tests are carried out on monthly returns from 1926 to 1987 on 12 industry portfolios. The evidence suggests that the value-weighted New York Stock Exchange (NYSE) market portfolio is not mean-variance efficient.

The paper is organized as follows. In the second section, we present the Bayesian framework for testing the asset pricing restrictions. The empirical results are included in the third section. Some concluding remarks are offered in section 4. A brief introduction to Monte Carlo integration appears in appendix A.

2. Methodology

2.1. Asset pricing restrictions

A test of the Sharpe-Lintner CAPM can be viewed as a test of the mean-variance efficiency of the market portfolio. Consider the multivariate regression model:

$$r_{it} = \alpha_{ip} + \beta_{ip} r_{pt} + \varepsilon_{it}, \qquad i = 1,\ldots,N,$$

where r_{it} is the return on asset i in excess of the return on a Treasury bill, r_{pt} is the excess return on the market portfolio, and ε_{it} is the disturbance, which is assumed to be correlated contemporaneously but not across time:

$$\mathrm{E}\varepsilon_{it}\varepsilon_{js} = \begin{cases} \sigma_{ij}, & s = t, \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

Eq. (1) can be written more compactly as

$$\boldsymbol{R} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{E}, \tag{3}$$

where $\boldsymbol{R}$ is a T (observations) row by N (asset) column matrix of excess returns, $\boldsymbol{X}$ is a T by 2 matrix, with the first column being a vector of ones and the second column the market excess return, $\boldsymbol{B}$ is a 2 by N coefficient matrix with the α coefficients in the first row and the β coefficients in the second row, and $\boldsymbol{E}$ is the T row by N column disturbance matrix.

For both the classical and Bayesian analysis, the disturbances, ε_{it}, are assumed to be uncorrelated with the portfolio return, r_{pt}. They are also assumed to have a normal distribution with a zero mean and covariance matrix $\boldsymbol{I}_T \otimes \boldsymbol{\Sigma}$, where $\boldsymbol{I}_T$ is the identity matrix of order T and $\boldsymbol{\Sigma}$ is the N by N matrix of the σ_{ij} elements in (2). The distinctive feature of the Bayesian

framework is that the parameters, $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\Sigma}$, are viewed as random variables. In the classical set-up, the parameters are constants. Classical statistics attempts to answer the question: 'If the true values of the parameters are $\boldsymbol{B}$ and $\boldsymbol{\Sigma}$, what is the probability we would have observed the data?' Bayesian statistics attempts to answer a different question: 'Given that we have observed the data, what is the probability distribution of the parameters $\boldsymbol{B}$ and $\boldsymbol{\Sigma}$?' Of course, the latter question cannot be answered without considering prior beliefs.

A mean–variance efficient portfolio, r_{pt}, must satisfy the following first-order condition:

$$\mathrm{E}[r_{it}] = \beta_{ip}\mathrm{E}[r_{pt}], \tag{4}$$

for $i = 1, \ldots, N$. This implies the intercept parameters in (3) should be zero for all assets:

$$\alpha_{ip} = 0, \qquad i = 1, \ldots, N. \tag{5}$$

This multivariate restriction of (3) will be tested.

2.2. *Bayesian inference*

Analysis of a model in the classical framework ignores prior information about the distribution of the parameter values. Numerically, the classical and Bayesian approaches may yield similar results if the Bayesian procedure uses a diffuse or 'ignorance' prior on the parameters. The Bayesian approach, however, gives the econometrician the option of incorporating prior knowledge into the estimation. We will analyze the estimation problem using both diffuse and informative priors.

The standard diffuse prior for the multivariate model (3) is the following prior density on the parameters $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Sigma}$:

$$p(\boldsymbol{B}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-(N+1)/2}, \tag{6}$$

where $|\boldsymbol{\Sigma}|$ is the determinant of the covariance matrix $\boldsymbol{\Sigma}$.

Using Bayes' rule and following Zellner (1971), the posterior density for the parameters $\boldsymbol{B}$ and $\boldsymbol{\Sigma}$ is

$$P(\boldsymbol{B},\boldsymbol{\Sigma}) = P(\boldsymbol{B}|\boldsymbol{\Sigma})P(\boldsymbol{\Sigma}),$$

where

$$P(\boldsymbol{B}|\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-1}\exp\left\{-\tfrac{1}{2}\mathrm{tr}\left[(\boldsymbol{B}-\hat{\boldsymbol{B}})'\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{B}-\hat{\boldsymbol{B}})\boldsymbol{\Sigma}^{-1}\right]\right\}$$

and

$$P(\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\mu/2} \exp\left\{-\tfrac{1}{2} \operatorname{tr} \boldsymbol{\Sigma}^{-1} \boldsymbol{S}\right\}, \tag{7}$$

where 'tr' denote the trace of a matrix, $\mu = T - 1 + N$, $\hat{\boldsymbol{B}}$ is the least-squares estimate of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, and $\boldsymbol{S} = (\boldsymbol{R} - \boldsymbol{X}\hat{\boldsymbol{B}})'(\boldsymbol{R} - \boldsymbol{X}\hat{\boldsymbol{B}})$. In the classical framework, the unbiased estimate of $\boldsymbol{\Sigma}$ is $\boldsymbol{S}/(T-2)$. For notational convenience, we have suppressed the dependence of the posterior density on the data. All priors are denoted by lower-case 'p' and posteriors are upper case.

The intercept vector, $\boldsymbol{\alpha}$, is the parameter of interest. Its marginal posterior distribution is multivariate t and given in Zellner (1971) by

$$P(\boldsymbol{\alpha}) \propto \left[\upsilon + (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})' \boldsymbol{H} (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})\right]^{-(\upsilon+N)/2}, \tag{8}$$

where $\upsilon = T - 1 - N$, $\boldsymbol{H} = \upsilon \boldsymbol{S}^{-1}/a$, and a is the (1, 1) element of $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ Appendix B contains the proof. The first and second moments are

$$\mathrm{E}[\boldsymbol{\alpha}] = \hat{\boldsymbol{\alpha}}$$

and

$$\operatorname{var}[\boldsymbol{\alpha}] = \frac{\upsilon}{\upsilon - 2} \boldsymbol{H}^{-1} = \frac{a}{\upsilon - 2} \boldsymbol{S}. \tag{9}$$

By using a multivariate t distribution or by directly integrating the posterior distribution, we can construct a Bayesian confidence region (interval in the one-dimensional case) or highest posterior density region:

$$-h\sqrt{\operatorname{var}[\alpha_{ip}]} + \hat{\alpha}_{ip} < \alpha_{ip} < \hat{\alpha}_{ip} + h\sqrt{\operatorname{var}[\alpha_{ip}]}, \qquad i = 1, \ldots, N, \tag{10}$$

where h is a number chosen such that the integration of the α's marginal posterior density over the region in (10) is $\gamma \in (0,1)$, the given 'significance level'. The Bayesian confidence region (10) can be interpreted as the region into which the parameters, $\boldsymbol{\alpha}$, have a probability γ of falling. It may be used to test the hypothesis that the intercepts, $\boldsymbol{\alpha}$, are zero. As in the classical approach, we reject the hypothesis if zero is outside the confidence region. As shown later, in section 3, this procedure will yield roughly the same results as the classical approach. The interpretation, however, is different. The Bayesian approach asks: 'What is the probability that the intercepts fall into this N-dimensional region?'

These confidence regions are not available in the classical framework. Although there are known methods for evaluating the confidence interval for a linear combination of the intercepts, we are interested in the confidence

intervals for each individual α_i, so that for a given significance level γ, all the intervals will cover the true parameters simultaneously $\gamma\%$ of the times if we draw repeated samples. Consider the simple case in which Σ is known. In this situation, $\hat{\alpha}$ would have a normal distribution: $\hat{\alpha} \sim N(\alpha, a\Sigma)$. Since high-dimensional normal tables are not available, we must use the Monte Carlo approach to evaluate the confidence region. When Σ is not known, however, the problem is much more complicated. Although both the Bonferroni method [Shanken (1990)] and Scheffé's (1977) S-method can be used to construct confidence regions, these regions are conservative in the sense that the overall probability that they will cover the true parameters is at least $(1-\gamma)$ for a given level of significance γ. That is, these confidence intervals will be bigger than the confidence intervals that contain the true parameters with exact $1-\gamma$ probability.

Bayesian posterior analysis reveals how our prior beliefs should change in light of the data. When we study a parameter θ, its posterior density gives us a basis for expressing (posterior) beliefs about possible values of θ. The Bayesian confidence interval is just a tool to quantify such information. Suppose the posterior density of θ vanishes except in the interval $[0,5]$, and the posterior probability for $\theta \geq 3$ is 99%. This implies the probability for $\theta < 2$ is less than 1%. As a result, a null hypothesis that $\theta = 0$ or $\theta = 1$ may be considered highly doubtful and hence be rejected. The Bayesian confidence intervals work on the same principle. This type of testing procedure, however, implicitly assumes the use of a simple loss function that measures the importance of a point by the posterior probability in a certain region containing the point. The classical inference by confidence intervals essentially uses the same loss function. Of course, this only weighs the statistical importance of the hypothesis. The economic significance is not taken into account.

Even if one has a different loss function (knowing the economic context of the problem), the Bayesian posterior analysis may still be useful. With a different loss function, however, the Bayesian confidence intervals themselves will not yield a rejection of the null, but they do offer information so that a decision may be based on the loss function. Consider an example. Suppose θ is a reliability measurement for a machine part. Suppose it is an unacceptable risk to have 1/1,000,000 chance of $\theta = 0.5$. As a result, even if we find the posterior probability of $\theta < 1$ is less than 0.001, we are unable to reject the null $\theta = 0.5$ or $\theta = 0$.

Since the Bayesian posterior analysis is not formulated in terms of the null hypothesis, even when we reject the null under the Bayesian confidence interval we cannot obtain the probability that the null is true because the hypotheses themselves are not defined in the probability space. If one wants to assign probabilities for the null and alternative hypotheses, the analysis focuses on the posterior-odds ratios, which deliver the posterior probability that the null hypothesis is true. This ratio is calculated in the next section.

Bayesian posterior inference for any given function of the parameters can be performed by evaluating certain integrals that represent the desired statistical measure, for example, the mean. These integrals are often analytically intractable but can be evaluated with reliable accuracy by using the Monte Carlo integration approach of Geweke (1988, 1989).

In the context of the mean–variance efficiency problem, we study the following function of the parameters:

$$\lambda = \boldsymbol{\alpha}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\alpha}.$$

This function is of interest for four reasons. First, as demonstrated in Shanken (1987a), it is linked to the correlation between the (efficient) tangency portfolio and the given portfolio. An equivalent geometric interpretation is given in Gibbons, Ross, and Shanken (1989). Second, it is the function (differing by a constant) tested by Shanken (1987b) to evaluate whether the intercepts are zero. Third, it is the unknown parameter of the important W statistic proposed in Gibbons, Ross, and Shanken. Finally, it is good example for comparing the results of the high-dimensional numerical integration with the known analytical solution.

The statistical properties of λ can be obtained by evaluating integrals under the posterior distribution of all of the parameters. In the present case, however, this can be simplified because λ is not a function of $\boldsymbol{\beta}$. As a result, it is sufficient to consider the posterior distribution of $\boldsymbol{\alpha}$ and $\boldsymbol{\Sigma}$ to obtain inference about λ. Appendix B proves that the posterior distribution of $\boldsymbol{\alpha}$ given $\boldsymbol{\Sigma}$ is

$$P(\boldsymbol{\alpha}|\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-1/2} \exp\left\{ -\frac{1}{2a} (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}) \right\}.$$

This is a multivariate normal distribution with covariance matrix $a\boldsymbol{\Sigma}$. Using (12), we can evaluate the moments of the function of interest, λ. Actually, the mean can be evaluated analytically as

$$\bar{\lambda} = \mathrm{E}\lambda = Na + \hat{\boldsymbol{\alpha}}' \hat{\boldsymbol{\Sigma}}^{-1} \hat{\boldsymbol{\alpha}}, \tag{13}$$

where $\hat{\boldsymbol{\Sigma}} = (T-2)^{-1} S$. This is demonstrated in appendix B. However, the variance of λ is analytically intractable.

To further characterize λ, an examination of the posterior distribution is informative. Consider how the Monte Carlo evaluation of the mean of λ works. A single replication of λ_i is obtained by drawing $\boldsymbol{\alpha}_i$ and $\boldsymbol{\Sigma}_i$ from their posterior distribution. The mean and standard error of λ are simply approximated by the average and standard deviation of the $\{\lambda_i\}$ over a given number of replications. The posterior density can be plotted by sorting these λ_i values.

The diffuse prior given in (6) was first introduced into Bayesian multivariate analysis by Geisser and Cornfield (1963) and the idea can be traced back to Jeffreys (1961). It is a prior of 'minimum prior information' [Press (1982)] and its use produces results similar to those obtained with the classical approach. This may be why it is the most widely used prior for multivariate models. If one has prior information on $\boldsymbol{\alpha}$, however, an informative prior should be used. One possible form of the informative prior is

$$p(\boldsymbol{B}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-(N+1)/2} p_0(\boldsymbol{\alpha}),$$

where $p_0(\boldsymbol{\alpha})$ is a multivariate normal distribution for the parameters $\boldsymbol{\alpha}$. If p_0 is set equal to one, we obtain the prior in (6). Then the posterior density for $\boldsymbol{\alpha}$ is

$$P(\boldsymbol{\alpha}) \propto p_0(\boldsymbol{\alpha})[\upsilon + (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})' \boldsymbol{H} (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})]^{-(\upsilon+N)/2}. \tag{15}$$

This is a product of a multivariate t density with a normal density. As Zellner (1971) notes, however, it is much more complicated than it seems. An analytic Bayesian analysis of the α coefficients is very difficult, if not impossible. However, the means, variances, and Bayesian confidence regions can be computed numerically by performing Monte Carlo integration in a N-dimensional parameter space based on the posterior density in (15).

Of course, we need not assume that the prior is a multivariate normal distribution for the parameters. Many other choices of the prior $p_0(\boldsymbol{\alpha})$ can be easily analyzed by the Monte Carlo integration approach. Indeed, the approach suggested here allows the researcher to form and choose a rich set of priors for the analysis, without the constraint that the problem is analytically solvable.

2.3. *An economic interpretation of the λ parameter*

The λ is related to the correlation, ρ, between the tangency portfolio and the given portfolio. Shanken (1987a) shows that

$$\lambda = \theta_p^2(\rho^{-2} - 1), \tag{16}$$

where θ_p^2 is the squared Sharpe measure (ratio of expected excess return to standard deviation of return) for portfolio p. The ρ is the ratio of the Sharpe measure for the given portfolio to the Sharpe measure for the tangency portfolio. If $\lambda = 0$, this implies that $\rho = 1$, which in turn implies that the given portfolio is efficient. Note that θ_p^2 is exogenous to the multivariate regression model. However, conditioning on a given value of θ_p^2, we can

assess the efficiency by examining a plot of the posterior density of ρ. Given θ_p, ρ is a function of λ. As a result,

$$\text{density}(\rho) = \text{density}(\lambda)\left|\frac{d\lambda}{d\rho}\right|,$$

where

$$\left|\frac{d\lambda}{d\rho}\right| = 2\theta_p^2\left(\frac{\lambda}{\theta_p^2}+1\right)^{3/2}.$$

An equivalent representation of (16) is proposed by Gibbons, Ross, and Shanken (1989):

$$\lambda = \theta_t^2 - \theta_p^2, \tag{17}$$

where θ_t is the Sharpe measure for the tangency portfolio. Whereas ρ is a relative measure of the deviation from efficiency, λ is an absolute measure. This can be seen in fig. 1. The correlation measure is the ratio of the Sharpe measures: slope(OB)/slope(OA). The λ is the difference in the squared Sharpe measures. Thus, a λ of zero implies efficiency. Geometrically, λ can be interpreted as the difference in the squared lengths of OA and OB, since $\lambda = (1+\theta_t^2) - (1+\theta_p^2)$.

2.4. *Posterior-odds ratios*

In the Bayesian framework, a test of the hypothesis that the intercepts, $\boldsymbol{\alpha}$, are exactly zero is called a sharp null hypothesis. Testing this particular hypothesis reveals that the econometrician has some prior belief that the null hypothesis may be true. This belief can be incorporated into the prior-odds ratio. Given the data and the theory, a posterior-odds ratio is calculated that allows the econometrician to modify his beliefs in light of the data.

The null hypothesis for testing the efficiency of a given portfolio is

$$H_0: \quad \boldsymbol{\alpha} = \mathbf{0},$$

and the alternative is

$$H_1: \quad \boldsymbol{\alpha} \neq \mathbf{0}.$$

The hypothesis, H_0, relates parameters to a single value. Testing of this type of sharp hypothesis was pioneered by Jeffreys (1961). We employ the stan-

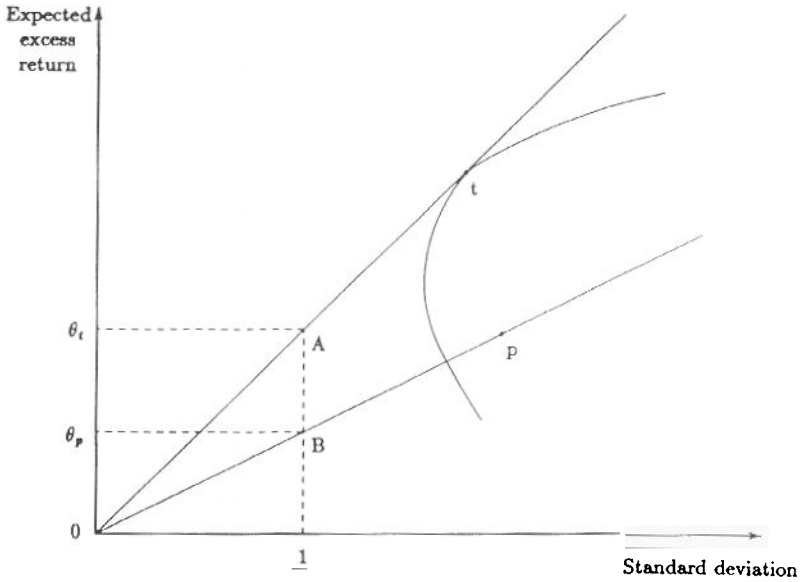


Fig. 1. The geometry of the measures of departures from the null hypothesis that the given portfolio is mean–variance efficient.

The correlation, ρ, between the given portfolio, p, and the tangency portfolio, t, is the ratio of Sharpe measures, θ_p/θ_t. A Sharpe measure of a portfolio is the ratio of its expected excess return to its standard deviation. Geometrically, the correlation is slope(OB)/slope(OA). A correlation of unity implies that the given portfolio is efficient. The other measure of the departure from the null hypothesis, λ, is the difference between θ_t^2 and θ_p^2. Geometrically, the λ is the difference in the squared lengths of OA and OB. If there is no difference in the squared returns, λ is zero and the given portfolio must be efficient.

dard diffuse prior under the null hypothesis, H_0:

$$p(\boldsymbol{\beta}, \boldsymbol{\Sigma}|H_0) \propto |\boldsymbol{\Sigma}|^{-(N+1)/2}. \tag{18}$$

For the alternative hypothesis, H_1, we use the prior:

$$p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Sigma}|H_1) \propto |\boldsymbol{\Sigma}|^{-(N+1)/2} f(\boldsymbol{\alpha}|\boldsymbol{\Sigma}), \tag{19}$$

where $f(\boldsymbol{\alpha}|\boldsymbol{\Sigma})$ is a N-dimensional Cauchy density:

$$f(\boldsymbol{\alpha}|\boldsymbol{\Sigma}) = \frac{c|k\boldsymbol{\Sigma}|^{-1/2}}{\left(1+\boldsymbol{\alpha}'(k\boldsymbol{\Sigma})^{-1}\boldsymbol{\alpha}\right)^{(N+1)/2}}, \tag{20}$$

with $c = \Gamma((N+1)/2)/\pi^{(N+1)/2}$ and $\Gamma(\cdot)$ is the Gamma function. The zero vector and the matrix $k\boldsymbol{\Sigma}$ are the location vector and scale matrix for the Cauchy density. These can be roughly interpreted as the mean and covari-

ance matrices for the Cauchy density in (20). The Cauchy density behaves much like an N-dimensional normal density with zero mean and covariance matrix $k\boldsymbol{\Sigma}$ near the origin point, but the Cauchy density has fatter tails. As a result, a random vector following a Cauchy distribution does not possess finite moments of order greater than or equal to one. Zellner and Siow (1980) extend Jeffrey's method to a univariate regression model. We generalize their approach to the multivariate regression model.

The posterior-odds ratio, K_c for H_0 and H_1 with prior odds 1 : 1, is

$$K_c = \frac{\int\int L(\boldsymbol{\beta}, \boldsymbol{\Sigma}|H_0)\, p(\boldsymbol{\beta}, \boldsymbol{\Sigma}|H_0)\, \mathrm{d}\boldsymbol{\beta}\, \mathrm{d}\boldsymbol{\Sigma}}{\int\int\int L(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Sigma}|H_1)\, p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Sigma}|H_1)\, \mathrm{d}\boldsymbol{\alpha}\, \mathrm{d}\boldsymbol{\beta}\, \mathrm{d}\boldsymbol{\Sigma}}$$

where $L(\boldsymbol{\beta}, \boldsymbol{\Sigma}|H_0)$ is the likelihood function, $L(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Sigma})$, valued at $\boldsymbol{\alpha} = 0$. The posterior-odds ratio is equal to the prior odds times the ratio of averaged likelihoods weighted by the prior densities. This contrasts with the classical likelihood-ratio testing procedure, which uses ratios of maximized likelihoods.

By analytically integrating $\boldsymbol{\beta}$ out, (21) is simplified:

$$K_c = \frac{1}{Q}\left(\frac{|S|}{|S_R|}\right)^{(T-1)/2}$$

where S_R, like S, is the cross-product matrix of the OLS residuals, but of the restricted model where the intercepts equal zero. The other scalar, Q, is defined as

$$Q = \int\int \exp\left[-\frac{1}{2a}(\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})\right] f(\boldsymbol{\alpha}|\boldsymbol{\Sigma})\, P(\boldsymbol{\Sigma})\, \mathrm{d}\boldsymbol{\alpha}\, \mathrm{d}\boldsymbol{\Sigma}, \tag{23}$$

with $f(\boldsymbol{\alpha}|\boldsymbol{\Sigma})$ given by (20) and $P(\boldsymbol{\Sigma})$ in (7). The determinants, $|S|$ and $|S_R|$, are measures of spread, and are sometimes called generalized variance [see Anderson (1984)]. The ratio $|S|/|S_R|$ measures the relative goodness-of-fit of the unrestricted model and the restricted one. A smaller ratio implies that the restricted model is less likely to be valid. This smaller ratio lowers K_c. The scalar Q can be interpreted as a weighting of the likelihoods. It is extremely complicated to evaluate. For example, with $N = 12$ assets, the order of the integration is the number of nonredundant elements in $\boldsymbol{\Sigma}$ plus the number of parameters $\boldsymbol{\alpha}$, which is $[N(N+1)/2 + N] = 90$. Its evaluation is feasible only numerically.

An odds ratio, K_n, is also evaluated for the multivariate normal prior. This prior is similar to the Cauchy prior except that $f(\boldsymbol{\alpha}|\boldsymbol{\Sigma})$ is replaced with a multivariate normal density, $g(\boldsymbol{\alpha}|\boldsymbol{\Sigma})$, for $\boldsymbol{\alpha}$ with covariance matrix $k\boldsymbol{\Sigma}$:

$$g(\boldsymbol{\alpha}|\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^N}|k\boldsymbol{\Sigma}|^{-1/2}\exp\left(-\tfrac{1}{2}\left(\boldsymbol{\alpha}'(k\boldsymbol{\Sigma})^{-1}\boldsymbol{\alpha}\right)\right). \tag{24}$$

This second density checks the sensitivity of the inferences to the choice of prior. The normal distribution has thinner tails than the Cauchy distribution. As a result, the prior mass will be less spread out. Intuitively, large deviations from zero in the intercepts should provide more evidence against the null hypothesis with the normal.

Finally, in their tests of the arbitrage pricing theory, McCulloch and Rossi (1988) use the Savage density approach to obtain posterior-odds ratios. The prior under the alternative is chosen as a density of the parameters $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\Sigma}$, where the marginal distribution of $\boldsymbol{\Sigma}$ is an inverted Wishart while $[\boldsymbol{\beta}, \boldsymbol{\alpha}]$ follows a multivariate normal distribution conditional on $\boldsymbol{\Sigma}$. Letting $\boldsymbol{\alpha} = \mathbf{0}$, a density in terms of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ is obtained. This is the choice of the prior density under the null hypothesis. So, to specify the Savage density completely, it suffices to give only the prior under the alternative:

$$p(\boldsymbol{B}, \boldsymbol{\Sigma}) \propto \left[|\boldsymbol{\Sigma}|^{-1}\exp\left\{-\tfrac{1}{2}\operatorname{tr}\left[(\boldsymbol{B}-\boldsymbol{B}_0)'\boldsymbol{\Psi}_0(\boldsymbol{B}-\boldsymbol{B}_0)\boldsymbol{\Sigma}^{-1}\right]\right\}\right]$$

$$\times\left[|\boldsymbol{\Sigma}|^{-\mu_0/2}\exp\left\{-\tfrac{1}{2}\operatorname{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{S}_0\right\}\right], \tag{25}$$

where $\boldsymbol{S}_0$ is the prior-variance structure for $\boldsymbol{\Sigma}$, and $\boldsymbol{B}_0$ and $\boldsymbol{\Psi}_0$ determine the prior means and variances of $\boldsymbol{B}$ parameters conditioning on $\boldsymbol{\Sigma}$.

Given our choice of prior (denoted with a 0 subscript), one can obtain the posterior density (which is identified with a 1 subscript). It can be verified that both prior and posterior marginal densities of $\boldsymbol{\alpha}$ are multivariate-t, so their densities have the form of (8). That is, the densities are proportional to

$$q_i^{-(v_i+N)/2} = 1 + (\boldsymbol{\alpha}-\bar{\boldsymbol{\alpha}}_i)'\boldsymbol{H}_i(\boldsymbol{\alpha}-\bar{\boldsymbol{\alpha}}_i)/v_i, \qquad i = 0, 1,$$

where v_0, v_1 are the degrees of freedom for marginal prior and posterior distribution of $\boldsymbol{\alpha}$, and $\bar{\boldsymbol{\alpha}}_0, \bar{\boldsymbol{\alpha}}_1$ are prior and posterior means of $\boldsymbol{\alpha}$.

The Savage odds-ratio calculation was first studied by Dickey (1971) in the univariate case and extended by Rossi (1980) to the multivariate case. The odds ratio K_s can be written as

$$K_s = (c_1/c_0)(v_0/v_1)^{N/2}(|\boldsymbol{H}_1|/|\boldsymbol{H}_0|)^{1/2}(q_1/q_0)^{-(v_0+N)/2}q_1^{-(v_1-v_0)/2} \tag{26}$$

where

$$c_i = \frac{\Gamma((v_i + N)/2)}{\Gamma(v_i/2)}, \qquad i = 0, 1,$$

and q_i is evaluated at $\boldsymbol{\alpha} = \mathbf{0}$.

Of course, the variables with the zero subscripts must be determined before the posterior analysis is undertaken. Conceptually, an investigator can assign any values to $\boldsymbol{B}_0$, $\boldsymbol{\Psi}_0$, and $\boldsymbol{S}_0$ to reflect his particular prior belief. One approach is to choose a ten-year subset of the data to form $\boldsymbol{B}_0$, $\boldsymbol{\Psi}_0$, and $\boldsymbol{S}_0$ by Bayesian posterior analysis. The posterior-odds ratios can then be calculated with the remaining data. These estimates may not reflect our prior beliefs about the relative efficiency of the given portfolio, however. We can calculate the mean of λ, $\bar{\lambda}_0$, by using the prior sample in the same way as (13). This mean can be used in (16) to solve for the relative efficiency, ρ, implied in the ten-year subperiod with a value of $\theta_p = 0.5$. This seems to be a reasonable choice suggested by Shanken (1987b). The value implies that a portfolio with an average excess return of 10% will have a standard deviation of 20% per annum. To make the procedure operational, we adjust the initial estimates of $\boldsymbol{B}_0$ and the estimates of the covariance by a scale parameter k to attain three levels of prior efficiency: 50%, 60%, and 70%. Intuitively, when ρ is less than 50% we want to reduce the size of the intercepts and the uncertainty proportionally to raise the prior efficiency to the desired level.

In the Cauchy and normal densities, only a single scale parameter needs to be determined. By using different ten-year samples of the data, one obtains roughly the same scale estimates. Unlike this case, the Savage density requires the prior specification of many variables. As a result, different ten year data sets can deliver quite different estimates even after scaling. To study the sensitivity of our inferences to the variability of the ten-year subperiods, we conduct the Savage analysis by using consecutive ten-year subperiods to obtain the corresponding odds ratios.

The three prior densities requires one's prior belief about the fundamental parameters of the multivariate regression model. These beliefs imply a prior on the relative efficiency of the given portfolio. Unfortunately, it is not clear how to obtain the prior degree of efficiency from the Cauchy or normal prior. For the Savage density, however, it is straightforward to evaluate the relative efficiency implied by the prior under the alternative hypothesis.

Although the Bayesian approach offers a consistent framework for incorporating prior information into the analysis, it opens the door for disagreements over the choice of prior. Traditionally, a prior distribution might have been chosen because it was analytically tractable. In contrast, we have chosen

Table 1

Means, standard deviations, and autocorrelations for the portfolio excess returns[a] of NYSE common stocks sorted by industry and the excess return on the NYSE value-weighted index based on monthly data from February 1926 to December 1987 (743 observations).

			Autocorrelation						
Portfolio	Mean	Std. dev.	ρ_1	ρ_2	ρ_3	ρ_4	ρ_{12}	ρ_{24}	ρ_{36}
Petroleum	0.00757	0.06364	0.021	−0.011	−0.050	0.088	0.015	−0.010	−0.031
Finance & real estate	0.00671	0.06111	0.115	−0.050	−0.138	0.009	0.052	0.025	0.041
Consumer durables	0.00832	0.07662	0.150	−0.003	−0.111	0.008	−0.021	−0.011	0.039
Basic industries	0.00713	0.06502	0.120	0.009	−0.121	0.032	−0.020	0.031	0.015
Food & tobacco	0.00658	0.04939	0.103	−0.037	−0.090	−0.001	0.016	−0.028	0.033
Construction	0.00606	0.07204	0.163	0.039	−0.094	−0.021	−0.021	−0.003	0.012
Capital goods	0.00749	0.06618	0.127	−0.001	−0.095	0.015	0.004	0.015	0.011
Transportation	0.00571	0.07824	0.149	−0.006	−0.155	−0.014	−0.004	0.025	0.035
Utilities	0.00552	0.04830	0.150	−0.035	−0.133	−0.005	−0.014	0.036	0.034
Textiles & trade	0.00628	0.06300	0.150	−0.011	−0.074	0.003	−0.011	−0.022	0.042
Services	0.00689	0.07502	0.017	0.029	−0.003	0.052	0.038	−0.025	0.088
Recreation	0.00710	0.07611	0.204	0.026	−0.076	−0.051	0.021	−0.028	−0.032
NYSE value-weighted	0.00641	0.05727	0.113	−0.010	−0.118	0.022	−0.001	0.017	0.028

[a]All rates of return are in excess of the one-month Treasury-bill rate.

without this constraint. The examination of more than one prior distribution reveals the sensitivity of the inferences to the choice of prior.[3]

3. Empirical results

3.1. *The data and summary statistics*

Twelve industry portfolios are used in the empirical work. The industry groupings follow Sharpe (1982), Breeden, Gibbons, and Litzenberger (1989), Gibbons, Ross, and Shanken (1989), and Ferson and Harvey (1991). The portfolios are value-weighted. The market return is the value-weighted NYSE return available from the Center for Research in Security Prices (CRSP) at the University of Chicago. All returns are in excess of the 30-day Treasury-bill rate available from Ibbotson Associates. These monthly data span the 1926–1987 period.

Means, standard deviations, and autocorrelations of the data are presented in table 1. The means range from 6.8% per year for the utilities industry to 10.5% per year for the consumer-durables industry. The lowest standard

[3]A technical appendix that provides a detailed derivation of each formula in the paper and the FORTRAN programs are available from the authors on request.

deviation is found in utilities and the highest in transportation. There is some evidence of first-order autocorrelation in the returns. There is no evidence of any seasonals in these portfolio returns, however.

3.2. Ordinary-least-squares regressions

Table 2 presents ordinary-least-squares (OLS) regressions of the portfolio excess returns on the market excess return. The market model betas range from 1.3 in consumer durables (industry 3) to a low of 0.7 in utilities (industry 9). The value-weighted market factor is able to explain over 80% (on average) of the variation in the portfolio returns.

The Sharpe–Lintner model restricts the intercept to be zero. Inspection of table 2 reveals that the intercept in the food and tobacco industry (industry 5) is more than two standard errors from zero. Three other industries have intercepts more than 1.2 standard errors from zero. At the bottom of the table, the exact F statistic is calculated for the multivariate test that all the intercepts are zero. This is one of the statistics studied by Gibbons, Ross, and Shanken (1989). The classical probability value of the statistic is 0.025. This would be interpreted as providing evidence against the model's restrictions at the 95% level but not at the 99% level. This evidence is consistent with the probability value of 0.013 reported by Gibbons, Ross, and Shanken for the 1926–1982 period.

3.3. Evaluation of the accuracy of the Monte Carlo integration

The main practical difficulty that has slowed the adoption of Bayesian econometrics is the integration of the posterior distribution. Traditional grid methods of numerical integration can handle only low-dimensional problems. In contrast, with the recent advances in Monte Carlo integration, high-order problems can be solved with considerable accuracy. In fact, the accuracy of the Monte Carlo method does not depend on the order of integration – it depends only on the number of replications.

To evaluate the accuracy of the Monte Carlo integrations, we compare the known analytical mean of α_i given in (9) and the analytical mean of λ given in (13) with the numerical estimates. For the intercepts, the order of integration is only 12. For the λ, the order of integration is 90. Both of these problems would be infeasible using the grid method. For example, for a 10-point grid, numerical evaluation of the λ would require 10^{90} calculations of the integrand.

Table 3 compares the analytical results with the numerical results. The second column reports the analytical calculation of the intercepts. These are the same as the OLS estimates in table 2. The next three columns provide the Monte Carlo evaluation of the 12-dimensional integral for 1,000, 10,000,

Table 2

Ordinary-least-squares estimates of the model:

$$r_{it} = \alpha_i + \beta_i r_{pt} + u_{it}, \qquad i = 1, \ldots, 12,$$

where r_i is the excess return[a] on an industry sorted portfolio, r_p is the excess return on the NYSE value-weighted index, and u_i is an industry-specific disturbance. The regression is estimated over February 1926 to December 1987 (743 observations).

Portfolio	α_i (% per month)	β_i	$\bar{R}_i^{2}$ [c]
Petroleum	0.1589 (0.1273)	0.9342 (0.0221)	0.701
Finance & real estate	0.0201 (0.0695)	1.0153 (0.0121)	0.905
Consumer durables	0.0181 (0.0891)	1.2700 (0.0155)	0.901
Basic industries	0.0100 (0.0614)	1.0976 (0.0107)	0.935
Food & tobacco	0.1539 (0.0750)	0.7863 (0.0130)	0.831
Construction	−0.1290 (0.1084)	1.1490 (0.0188)	0.834
Capital goods	0.0414 (0.0724)	1.1039 (0.0126)	0.912
Transportation	−0.1986 (0.1377)	1.2014 (0.0239)	0.773
Utilities	0.0834 (0.0887)	0.7319 (0.0154)	0.753
Textiles & trade	0.0084 (0.1108)	0.9675 (0.0192)	0.773
Services	0.0418 (0.1766)	1.9517 (0.0307)	0.593
Recreation	−0.0470 (0.1285)	1.1820 (0.0223)	0.791
F statistic[b] for $\alpha_i = 0, i = 1, \ldots, 12$		1.958	
P-value		0.025	

[a]All rates of return are in excess of the one-month Treasury-bill rate. Standard errors are in parentheses.

[b]F statistic is the exact statistic proposed in Gibbons, Ross, and Shanken (1989) for the multivariate test of whether the 12 intercepts are jointly equal to zero.

[c]Coefficient of determination adjusted for degrees of freedom.

Table 3

An evaluation of the accuracy of the Monte Carlo integration. We compare the numerical means and standard errors obtained from Monte Carlo integration with the analytical values. The analytical means and standard errors for the intercepts, α_i, are ordinary-least-squares estimators.[a] The numerical means and standard errors are evaluated from 1,000, 10,000, and 100,000 Monte Carlo draws from the marginal posterior density of the intercepts – which is a multivariate t-distribution. The λ parameter measures the deviation from the null hypothesis that the intercepts are all zero.[b] Although the analytical mean of λ can be calculated,[c] its standard error is intractable. The numerical mean and standard deviation are obtained by 1,000, 10,000, and 100,000 Monte Carlo draws from the posterior density of the vector of intercepts, which is a multivariate t-distribution, and from the posterior density of the covariance parameter matrix, which is an inverted Wishart distribution. The estimation is based on monthly data from February 1926 to December 1987 (743 observations).

Portfolio	α_i analytical (% per month)	α_i numerical 1,000 replications (% per month)	α_i numerical 10,000 replications (% per month)	α_i numerical 100,000 replications (% per month)	Order of integration
Petroleum	0.1589	0.1612	0.1581	0.1583	12
	(0.1284)	(0.1271)	(0.1274)	(0.1278)	
		[0.0040][d]	[0.0013]	[0.0004]	
Finance & real estate	0.0201	0.0210	0.0200	0.0201	12
	(0.0701)	(0.0712)	(0.0707)	(0.0701)	
		[0.0023]	[0.0007]	[0.0002]	
Consumer durables	0.0181	0.0157	0.0187	0.0188	12
	(0.0899)	(0.0862)	(0.0905)	(0.0902)	
		[0.0027]	[0.0009]	[0.0003]	
Basic industries	0.0100	0.0097	0.0093	0.0102	12
	(0.0619)	(0.0625)	(0.0615)	(0.0623)	
		[0.0020]	[0.0006]	[0.0002]	
Food & tobacco	0.1539	0.1540	0.1537	0.1534	12
	(0.0757)	(0.0761)	(0.0750)	(0.0745)	
		[0.0024]	[0.0008]	[0.0002]	
Construction	−0.1299	−0.1280	−0.1291	−0.1290	12
	(0.1094)	(0.1079)	(0.1088)	(0.1087)	
		[0.0034]	[0.0011]	[0.0003]	
Capital goods	0.0414	0.0422	0.0418	0.0414	12
	(0.0730)	(0.0707)	(0.0732)	(0.0731)	
		[0.0022]	[0.0007]	[0.0002]	
Transportation	−0.1986	−0.2008	−0.1977	−0.1990	12
	(0.1389)	(0.1435)	(0.1389)	(0.1390)	
		[0.0045]	[0.0014]	[0.0004]	
Utilities	0.0834	0.0822	0.0828	0.0837	12
	(0.0895)	(0.0895)	(0.0898)	(0.0898)	
		[0.0028]	[0.0009]	[0.0003]	
Textiles & trade	0.0084	0.0087	0.0093	0.0081	12
	(0.1118)	(0.1135)	(0.1116)	(0.1116)	
		[0.0036]	[0.0011]	[0.0004]	
Services	0.0418	0.0445	0.0410	0.0414	12
	(0.1782)	(0.1830)	(0.1758)	(0.1781)	
		[0.0058]	[0.0018]	[0.0006]	
Recreation	−0.0470	−0.0454	−0.0468	−0.0473	12
	(0.1297)	(0.1277)	(0.1298)	(0.1297)	
		[0.0040]	[0.0013]	[0.0004]	

Table 3 (continued)

λ analytical	λ numerical 1,000 replications	λ numerical 10,000 replications	λ numerical 100,000 replications	Order of integration
0.0489	0.0491 (0.0151) [0.00047]	0.0488 (0.0151) [0.00015]	0.0488 (0.0150) [0.00005]	90

[a]The model estimated is $r_{it} = \alpha_i + \beta_i r_{pt} + u_{it}$, $i = 1, \ldots, 12$, where r_i is the excess return on an industry-sorted portfolio, r_p is the excess return on the NYSE value-weighted index, and u_i is an industry-specific disturbance. All rates of return are in excess of the one-month Treasury-bill rate. The standard errors, in parentheses, are slightly larger than the OLS standard errors because the covariance matrix is divided by $T - 2 - N$ in the Bayesian framework.

[b]λ is defined as $\lambda \quad \boldsymbol{\alpha}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\alpha}$, where $\boldsymbol{\alpha}$ is a vector of the intercepts and $\boldsymbol{\Sigma}$ is the covariance matrix.

[c]The analytical mean of λ is given by $\bar{\lambda} = Na + \hat{\boldsymbol{\alpha}}'\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\alpha}}$, where $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\Sigma}}$ are the OLS estimates of the intercepts and covariance matrix, N is the number of assets, a is the $(1,1)$ element of $(\boldsymbol{X}'\boldsymbol{X})^{-1}$, and $\boldsymbol{X}$ is a 2 by T matrix that includes a column of ones and the excess return on the CRSP value-weighted index.

[d]Standard errors of the numerical means in brackets.

and 100,000 replications. The analytical standard errors from (9) and numerical standard errors of the intercepts are in parentheses. The analytical standard errors in table 3 are slightly larger than the OLS standard errors in table 2. In OLS, the covariance matrix is divided by $T - 2$ and standard errors are then calculated. In the Bayesian framework, the covariance matrix is divided by $T - 2 - N$. This accounts for the small difference.

The Monte Carlo calculations are remarkably close to the analytical results with as few as 1,000 replications. With 100,000 replications, the Monte Carlo integration delivers five digits of precision. The formal way of assessing the numerical accuracy is the standard error of the numerical Monte Carlo integration. These measures, which correspond to $\sigma_n/\sqrt{n}$ in appendix A, are presented in square brackets. These standard errors decrease as the replications increase, indicating increased accuracy.

The results for the λ parameter are presented in the bottom panel of table 3. Unlike the evaluation of the intercepts, the evaluation of the λ involves a 90-dimensional integration. The accuracy of the Monte Carlo integration is not affected, however, by the increased dimensionality. As the number of replications increases, the numerical λ approaches the analytic value.[4] The standard errors cannot be compared because the analytic standard error is intractable.

[4]We also ran, but do not report, 1,000,000 replications. With this number of replications, the difference between the numerical value and the analytical value of λ is 0.000013.

In summary, the Monte Carlo method delivers accurate solutions to high-dimensional integration problems that were previously considered infeasible using the grid method. Even with a smaller number of replications, the numerical estimates are close to the analytical ones. The evaluation of the Bayesian confidence intervals and the posterior-odds ratios presented in the next two sections uses 100,000 replications.

3.4. *Bayesian confidence intervals and posterior analysis*

Consider the Bayesian confidence region for the intercepts. For a given h, the region is in a 12-dimensional space as described by (10). Since it is impossible to present the region graphically, we plot the 12 intervals together in fig. 2. The figure has three panels that correspond to three values of h: 1.2, 2.0, and 2.8. The probability value presented by each panel is the probability that each intercept falls into the interval simultaneously.

Intuitively, the smaller the h, the smaller the interval and the smaller the probability of the intercepts falling into the interval simultaneously. With $h = 1.2$, there is only a 6.1% chance that the intercepts lie in the intervals simultaneously. With $h = 2.8$, there is a 94.2% chance that the intercepts lie in the intervals. In the third panel, where $h = 2.8$, the value of zero is covered by all of the portfolios. The portfolios that have most of the area away from zero are: petroleum, food and tobacco, construction, transportation, and utilities. It is these portfolios that are likely to drive a rejection of the null hypothesis that the intercepts are equal to zero.

The λ parameter, from (17), summarizes the absolute level of the departures from the null hypothesis. The posterior density is presented in fig. 3. It is important to realize that the posterior density of the intercepts is a multivariate t that contains zero as an interior point, whereas the posterior density of λ contains zero as a boundary point. This reflects the fact that λ is nonnegative. In terms of the Sharpe measures, $\theta_p \leq \theta_t$.

If the given portfolio is efficient, $\theta_p = \theta_t$ and $\lambda = 0$. In the Bayesian framework, however, all of the parameters including λ are treated as random variables. So if the null hypothesis is indeed true, we can only expect most of the mass to be close to zero. An important question arises: How close should λ be to zero? Fig. 3 suggests that the mass is spread out and has no obvious concentration near zero. The posterior density of λ itself, however, does not seem to offer a test for the null hypothesis,[5] because if the null is indeed true, we still get a posterior density of λ that has all of its mass away from zero. Nevertheless, since λ is the difference between the squared Sharpe measures of the tangency portfolio and the given portfolio, we are still interested in this difference along with its distributional properties, for it may

[5]We are grateful to the referee for bringing this important issue to our attention.

shed additional light on our understanding of the behavior of the given portfolio.

Fig. 4 presents the posterior density of the correlation measure ρ from (17) for three assumptions for the value of θ_p. The first value, $\theta_{p1} = 0.111926$, is obtained from table 1. If annualized, it implies an annual excess return of 3.9% for a standard deviation of 10%. The third choice is an extremely conservative one, $\theta_{p3} = 0.288675$, and is considered by Shanken (1987a) to be 'greater than any conceivable true value'. This choice corresponds to an annualized 10% standard deviation on a 10% excess return. The final selection, θ_{p2}, is the average of θ_{p1} and θ_{p3}.

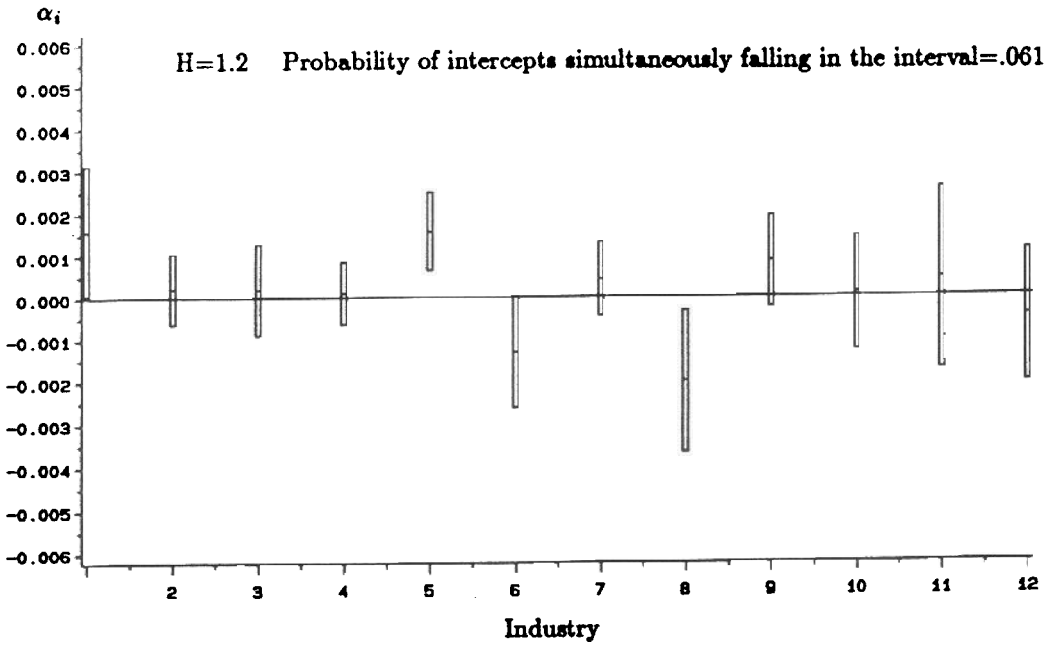
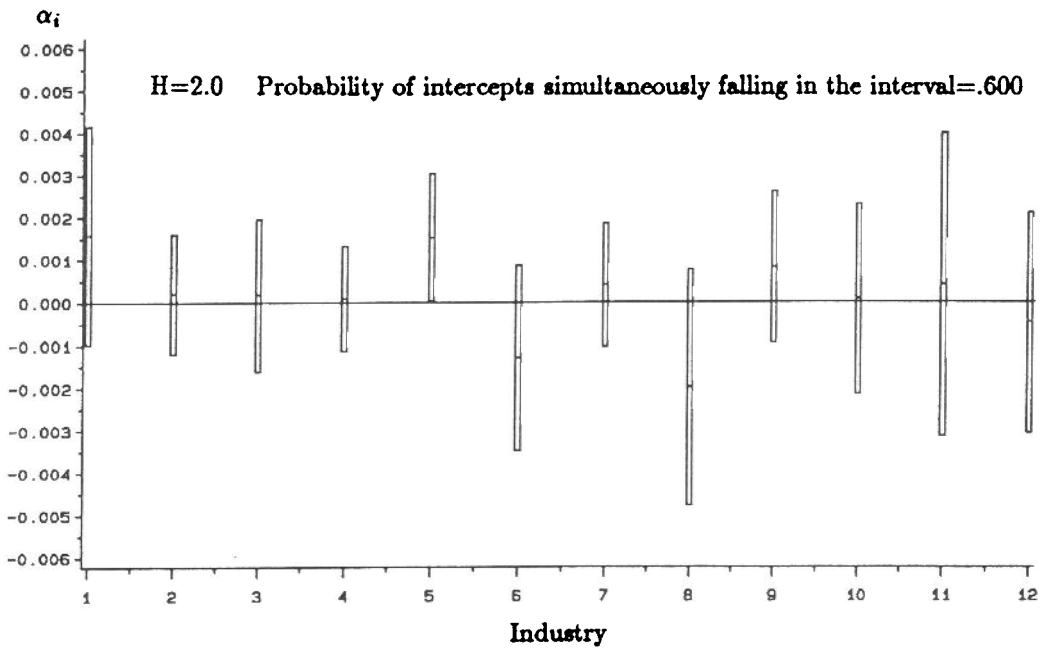


Fig. 2. Bayesian confidence regions for the intercepts, α_i, in the multivariate regression of excess returns on the CRSP value-weighted excess return.

The intervals are constructed using a multivariate t-distribution:

$$-h\sqrt{\text{var}[\alpha_{ip}]} + \hat{\alpha}_{ip} < \alpha_{ip} < \hat{\alpha}_{ip} + h\sqrt{\text{var}[\alpha_{ip}]}, \qquad i = 1,\ldots,N, \tag{10}$$

where $N = 12$ is the number of industry portfolios, $\hat{\alpha}_{ip}$ is the OLS estimator of intercept for industry i, and $h = 1.2, 2.0, 2.8$ is a number chosen such that the integration of the α's marginal posterior density over the region in (10) is $\gamma \in (0,1)$, the given 'significance level'. The confidence region (10) can be interpreted as the region that the parameters, α, have a probability γ of falling in. The estimates are based on monthly data from February 1926 to December 1987 (743 observations). The industry groups are: 1 = petroleum, 2 = finance/real estate, 3 = consumer durables, 4 = basic industries, 5 = food/tobacco, 6 = construction, 7 = capital goods, 8 = transportation, 9 = utilities, 10 = textiles/trade, 11 = services, and 12 = leisure.



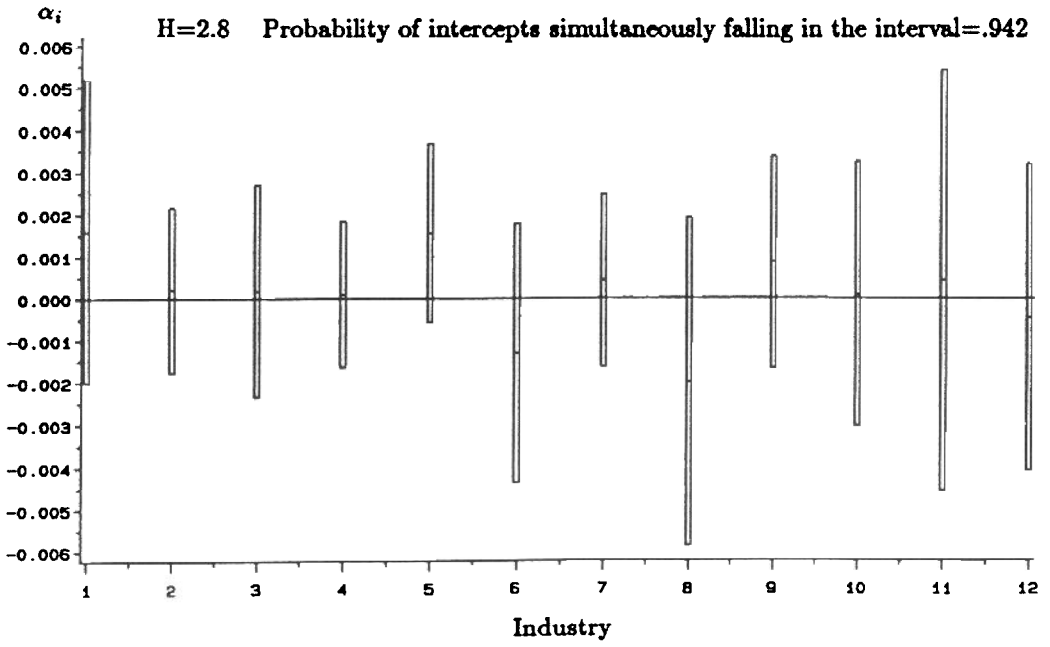


Fig. 2 (continued)

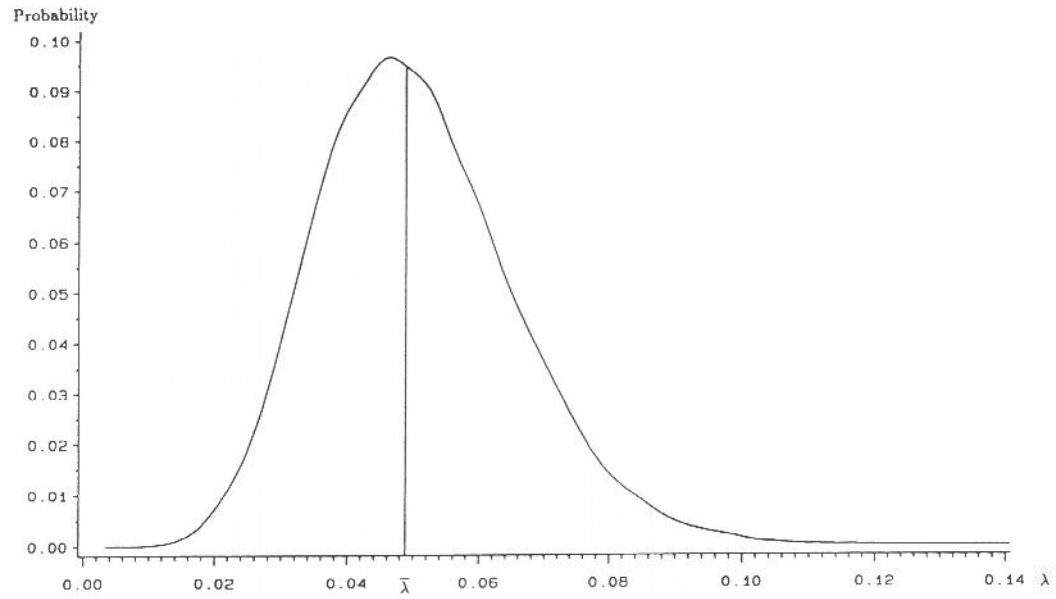


Fig. 3. The posterior density of λ.

The measure λ is the difference between the squared Sharpe measures for the tangency portfolio and the given portfolio. A Sharpe measure of a portfolio is the ratio of its expected excess return to its standard deviation. Let $\boldsymbol{\alpha}$ and $\boldsymbol{\Sigma}$ be the intercepts and covariance matrix in the multivariate regression of 12-industry-portfolio excess returns on the CRSP value-weighted excess return. Based on 100,000 Monte Carlo draws from the posterior density of the vector of intercepts, which is a multivariate t-distribution, and from the posterior density of the covariance parameter matrix, which is an inverted Wishart distribution, the density of λ is obtained by sorting the generated samples of $\lambda = \boldsymbol{\alpha}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\alpha}$. The evaluations are based on monthly data from February 1926 to December 1987 (743 observations).

The correlation measure is the ratio of the Sharpe measures for the given portfolio and the tangency portfolio. Conditioning on λ in (16), as θ_p increases, ρ increases. Intuitively, assigning a higher return to the given portfolio per unit of risk should make the portfolio more efficient. This is seen in fig. 4, where the densities are pushed toward efficiency as the θ_p is increased. Even with the conservative assumption for θ_p, however, there is only a small probability that the given portfolio is more than 90% efficient. The probabilities are evaluated in detail in table 4.

The ρ is a measure of the relative efficiency of the given portfolio. A value of unity implies efficiency. If it is probability 1 that $\rho > 0.9$, the given portfolio is at least 90% efficient. Table 4 presents the probabilities that the given portfolio attains certain minimum levels of efficiency. With the conservative assumption for θ_p, there is 49% probability that the given portfolio is more than 80% efficient. There is a less than one percent chance, however, that the portfolio is more than 90% efficient. This evidence indicates that it is unlikely that the CRSP value-weighted portfolio is efficient.

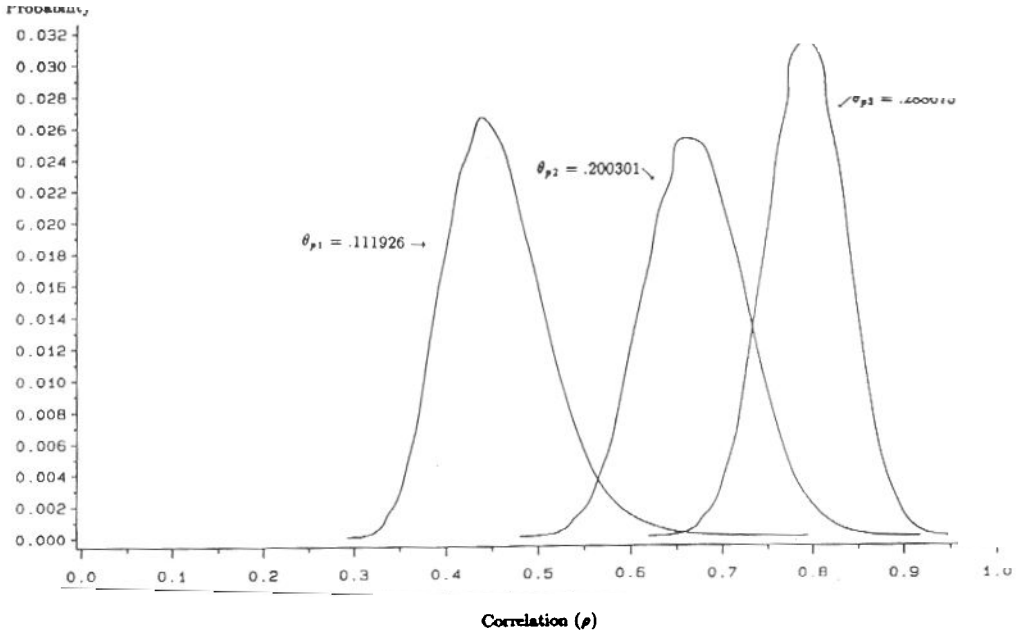


Fig. 4. The posterior density of ρ.

The measure ρ is the ratio of the Sharpe measures of the CRSP value-weighted portfolio to the tangency portfolio, θ_p/θ_t. A Sharpe measure of a portfolio is the ratio of its expected excess return to its standard deviation. Let α and Σ be the intercepts and covariance matrix in the multivariate regression of 12-industry-portfolio excess returns on the CRSP value-weighted excess return. To evaluate the posterior density, we choose values for θ_p. The first corresponds to the mean excess return (0.641% per month) and the standard deviation (5.73% per month) on the NYSE value-weighted index for the sample February 1926 to December 1987. On an annualized basis, this Sharpe measure implies a 3.9% annualized excess return for a 10% standard deviation. The third value is the most conservative, implying an annualized 10% standard deviation for a 10% excess return on the given portfolio. The second value is the average of the first and third values. Based on 100,000 Monte Carlo draws from the posterior density of the vector of intercepts, which is a multivariate t-distribution, and from the posterior density of the covariance parameter matrix, which is an inverted Wishart distribution, the densities are evaluated by sorting the generated samples of $\rho = (\alpha'\Sigma^{-1}\alpha/\theta_p^2 + 1)^{-1/2}$. The evaluation is carried out on monthly data from February 1926 to December 1987 (743 observations).

3.5. *Posterior-odds ratios*

Posterior-odds ratios for three prior distributions are presented in table 5. The posterior-odds ratio can be interpreted as the probability that the null is true divided by the probability that the alternative is true. The null hypothesis states that the intercepts are zero. A low value of the posterior-odds ratio is evidence against the model's restriction that the market portfolio is mean–variance efficient.

Using the fact that the probability that the alternative is true is one minus the probability that the null is true, it is possible to extract the posterior

Table 4

The probability that the given portfolio attains certain minimum levels of efficiency. The efficiency is measured by ρ, which is the ratio of the Sharpe measures[a] of the given portfolio to the tangency portfolio, θ_p/θ_t. Three values are chosen[b] for θ_p to evaluate the posterior probability for ρ to be in a prespecified relative efficiency interval. Based on 100,000 Monte Carlo draws from the posterior density of the vector of intercepts[c] (α), which is a multivariate t-distribution and from the posterior density of the covariance parameter matrix (Σ), which is an inverted Wishart distribution, the probabilities are evaluated by sorting the generated samples of $\rho = (\alpha'\Sigma^{-1}\alpha/\theta_p^2 + 1)^{-1/2}$. The analysis is based on monthly data from February 1926 to December 1987 (743 observations).

Sharpe measure θ_p	Probability $(0.5 < \rho < 1)$	Probability $(0.6 < \rho < 1)$	Probability $(0.7 < \rho < 1)$	Probability $(0.8 < \rho < 1)$	Probability $(0.9 < \rho < 1)$
0.1119	0.2365	0.0178	0.0007	0.0000	0.0000
0.2003	0.9999	0.9206	0.3422	0.0194	0.0001
0.2887	1.0000	1.0000	0.9862	0.4806	0.0080

[a]A Sharpe measure of a portfolio is the ratio of its expected excess return to its standard deviation.

[b]The first value corresponds to the mean excess return (0.641% per month) and the standard deviation (5.73% per month) on the NYSE value-weighted index for the sample February 1926 to December 1987. On an annualized basis, this Sharpe measure implies a 3.9% annualized excess return for a 10% standard deviation. The third value is the most conservative, implying an annualized 10% standard deviation for a 10% excess return on the given portfolio. The second value is the average of the first and third values.

[c]The model is $r_{it} = \alpha_i + \beta_i r_{pt} + u_{it}$, $i = 1, \ldots, 12$, where r_i is the excess return on the ith industry-sorted portfolio, r_p is the excess return on the NYSE value-weighted index, and u_i is an industry-specific disturbance. All rates of return are in excess of the one-month Treasury-bill rate.

probability that the hypothesis is true:

$$P_0 = \frac{K}{1+K}. \tag{27}$$

In the context of the asset pricing tests, this is the posterior probability that the market index is mean–variance efficient. This probability is reported in brackets in table 5.

To evaluate the odds ratios, we need to specify the value of the scaling parameter k for both the conditioning Cauchy prior in (20) and the conditioning normal prior in (24). Since the shape of the Cauchy is similar to the normal around the origin point (with the exception of the fatter tails), the same scaling parameter is used for both prior densities.

Consider the intuition behind the choice of scale for the normal prior in (24). Given Σ, the scale k represents the prior uncertainty about α. Suppose we had a data set X_0 before we conducted the posterior-odds-ratio analysis. In the classical framework, the least-squares estimator of α has a covariance

Table 5

Posterior odds ratios for the efficiency of the CRSP value-weighted index and posterior probabilities that the index is efficient based on two prior distributions: Cauchy[a] and normal[b]. For a range[c] of prior beliefs, we evaluate the odds ratios and the posterior probabilities by Monte Carlo integration based on 100,000 draws. The evaluations are for monthly data from February 1926 to December 1987 (743 observations).

Scale k	Cauchy	Normal
0.0006	0.3549 [26.19%][d]	0.0981 [8.93%]
0.0008	0.3579 [26.36%]	0.1178 [10.54%]
0.0010	0.3610 [26.52%]	0.1461 [12.45%]
0.0012	0.3661 [26.80%]	0.1745 [14.86%]
0.0014	0.3744 [27.24%]	0.1839 [15.53%]

[a]The model is $r_{it} = \alpha_i + \beta_i r_{pt} + u_{it}$, $i = 1, \ldots, 12$, where r_i is the excess return on the ith industry-sorted portfolio, r_p is the excess return on the NYSE value-weighted index, and u_i is an industry-specific disturbance. All rates of return are in excess of the one-month Treasury-bill rate. The Cauchy prior under the null hypothesis is the standard diffuse prior, $|\Sigma|^{-(N+1)/2}$, where $|\Sigma|$ is the determinant of the covariance matrix and N is the number of assets. The prior under the alternative is the product of this diffuse prior and a Cauchy density with mean zero and scale matrix $k\Sigma$, where k is a scaling parameter.

[b]The prior under the null hypothesis is the diffuse prior. The prior under the alternative is the product of the diffuse prior and a normal density with mean zero and covariance matrix $k\Sigma$.

[c]The range is represented by k, which is the scaling parameter in the covariance matrix of the normal density and in the scale matrix of the Cauchy density. A higher k spreads out the prior density concentration about zero, representing additional prior uncertainty.

[d]The posterior probability that the null hypothesis is true is in brackets.

matrix $a_0\Sigma$, where a_0 is a scaling factor that is the (1,1) element of $(X_0'X_0)^{-1}$. The Bayesian posterior analysis in (12) also suggests that the marginal distribution of α conditioned on Σ is multivariate normal with covariance matrix $a_0\Sigma$. So it seems reasonable to choose the covariance matrix for the normal prior to be $k\Sigma$ where k is roughly the same size as a_0. Unfortunately, a data set such as X_0 is not available to us. As a result, we randomly chose a ten-year subset of the data to get a sense of the magnitude of a_0. This examination of some of the data in our sample[6] shows a_0 to be about 0.001, so we choose $k = 0.001$. To examine the sensitivity of the inferences to this scaling, k we allow to vary from 0.0006 to 0.0014, as reported in table 5.

[6]Over the January 1946 to December 1955 period, $a_0 = 0.00092$. Zellner and Siow (1980) use the entire sample to construct the scale factor. In our case, over the full sample, $a_0 = 0.0013$.

The second and third columns report the posterior-odds ratios for the Cauchy and normal-prior distributions based on 100,000 replications. The first column is the value of the scale parameter, k, in (20). The scale parameter changes the variance in the normal and the dispersion in the Cauchy. A larger scale spreads out the prior distribution and may reduce the evidence against the null. This is observed in the empirical results in table 5. The posterior-odds ratios increase monotonically as the scale factor increases.

Both the Cauchy and normal priors provide some evidence against the model's restrictions. The odds ratios are relatively insensitive to the choice of scaling factor. The results for the Cauchy prior indicate that the posterior probability that the market portfolio is mean–variance efficient ranges from 26.2% to 27.2%. The normal prior suggests that the odds against the null hypothesis are greater. The normal prior is expected to deliver a lower odds ratio because its mass is more concentrated at the mean than the Cauchy. The posterior probability that the market portfolio is efficient ranges from 8.9% to 15.5% with this prior.

The final posterior-odds ratio is based on the Savage density. This odds ratio is calculated analytically. We use a ten-year subsample of the data to get the prior intercept and the covariance structure. In addition, we scale them to attain three prior levels of efficiency: 50%, 60%, and 70%. Further, we investigate the sensitivity of the results to the choice of the prior sample by looking at six ten-year subperiods.

Table 6 reports the posterior probabilities that the null hypothesis is true. With a 50% assumed level of prior efficiency, five of the six subsamples show less than 1% chance that the null is true. The exception occurs when the 1926–1935 period is used to construct the prior intercepts and covariance structure. This is, perhaps, not surprising, because this subsample spans the Great Depression. When the prior level of efficiency is increased to 60% or 70%, there is overwhelming evidence against the mean–variance efficiency of the given portfolio. With all prior subsamples, the posterior probabilities that the null is true are less than 1%. Compared with the Cauchy and normal, the Savage prior offers more evidence against the model's restrictions. The odds ratios from all three priors are consistent in favor of the alternative.

The Bayesian approach gives the investigator the option of including prior beliefs in the empirical analysis. Of course, there may be disagreement over the particular prior chosen. We have constructed a number of prior densities to mimic a range of plausible prior beliefs. With the Savage approach, we use a number of ten-year subperiods to form the prior intercepts and covariances. In addition, we scale these parameters to impose certain prior levels of efficiency. Furthermore, we examine two additional prior densities: the Cauchy and the normal. Of course, these are only a small set of the possible prior densities. The techniques introduced in this paper could allow the investigator to choose virtually any prior density.

Table 6

Posterior probabilities for the efficiency of the CRSP value-weighted index based on the Savage density approach. With this approach, we choose ten-year subperiods (called prior samples) to construct six data-based priors.[a] Each prior density is scaled to reflect three prior beliefs about efficiency[b] levels: 50%, 60%, and 70%. The posterior probabilities are calculated over the remaining sample. The evaluations are based on monthly data from February 1926 to December 1987 (743 observations).

	Assumed level of prior efficiency		
Prior sample	50%	60%	70%
1926–1935	0.2889	0.0040	0.0001
1936–1945	0.0047	0.0001	< 0.0001
1946–1955	< 0.0001	< 0.0001	< 0.0001
1956–1965	< 0.0001	< 0.0001	< 0.0001
1966–1975	< 0.0001	< 0.0001	< 0.0001
1976–1985	0.0010	< 0.0001	< 0.0001

[a]The model is $r_{it} = \alpha_i + \beta_i r_{pt} + u_{it}$, $i = 1, \ldots, 12$, where r_i is the excess return on the ith industry-sorted portfolio, r_p is the excess return on the NYSE value-weighted index, and u_i is an industry-specific disturbance. All rates of return are in excess of the one-month Treasury-bill rate. In the Savage approach, the prior under the alternative is the product of an inverted Wishart density on the covariance matrix, Σ, and a conditioning normal density on the intercepts, α, and the slopes, β. The prior means and covariances are obtained from ten-year subperiods (prior samples) and are scaled to attain three levels of prior efficiency. The prior under the null hypothesis is obtained by letting $\alpha = \mathbf{0}$.

[b]The prior efficiency is measured by the ratio of the Sharpe measures of the NYSE value-weighted index to the tangency portfolio. A Sharpe measure is the ratio of expected excess return on a portfolio to its standard deviation.

4. Conclusions

This paper uses the technique of Monte Carlo integration to construct a general Bayesian framework for evaluating the mean–variance efficiency of a given portfolio. The ability of the integration technique to evaluate high-dimensional problems accurately allows us to choose prior densities for the Bayesian analysis freely.

Although we concentrate on the Sharpe–Lintner asset pricing restrictions, the approach developed here can easily be applied to test the restrictions of other asset pricing models such as the Black (1972), Merton (1973), and Long (1974) CAPMs, the finite version [Dybvig (1983) and Grinblatt and Titman (1983)] of Ross's (1976) arbitage pricing theory, and Rubinstein (1976) and Breeden's (1979) consumption CAPM. Indeed, the technique can even be used to evaluate complicated nonlinear constraints. Geweke (1986) shows how to use the Bayesian framework to evaluate inequality constraints; they are impossible to evaluate with the classical approach.

We obtain three main empirical results. First, we show in an asset pricing application that the Monte Carlo integration technique is able to deliver accurate evaluation of high-order (90-dimensional) problems. Second, we show how to calculate Bayesian confidence intervals for the intercepts. In

addition, we evaluate the posterior density of the λ parameter, which measures the absolute level of deviation from efficiency. We also characterize the posterior distribution of the correlation between the efficient portfolio and the given portfolio. The probability that the given portfolio attains certain minimal levels of efficiency is also presented. Third, we construct posterior-odds ratios for the test that the intercepts are zero. The sensitivity of the odds ratios to the scaling of the variance is explored. We also evaluate odds ratios based on different prior distributions.

The Bayesian procedure in this paper complements Shanken's (1987b) approach. One strength of Shanken is that he directly links the prior belief about the relative efficiency to the odds in favor of efficiency. As a result, his method can be used to test approximate efficiency. In addition, his procedure is computationally convenient. In contrast, our approach starts with the joint prior on all the parameters in the multivariate regression model. As such, our method is more computationally demanding. Our approach has advantages, however. First, we are able to use prior beliefs on all the parameters. Second, we can conduct posterior analysis of the parameters as well as functions of the parameters. Third, the odds in favor of efficiency can be calculated for several forms of the prior beliefs. In the Savage density case, it is even possible to assess the relative efficiency implicit in the prior beliefs.

How strong is the evidence against the model's restrictions? The classical F statistic proposed by Gibbons, Ross, and Shanken (1989) suggests a p-value of 2.5%, which is strong evidence against the null hypothesis. The Bayesian posterior probability that the market portfolio is mean–variance efficient ranges from 27% to less than 1%, depending on the form of the prior. Both points of view suggest departures from the null hypothesis in the 1926–1987 period. Likely explanations for these departures are nonstationarity of the parameters over this sample, incorrect specification of the market portfolio, nonnormality of the joint distribution of returns, or a more fundamental misspecification of the asset pricing model. The framework presented in this paper does not attribute causes to the failure of the model. The Bayesian approach does provide a different perspective, however, by attempting to answer the question: 'Given the data (and my prior beliefs), what is the probability that the model's restrictions are valid?'

Appendix A: An introduction to Monte Carlo integration

Monte Carlo integration is a general approach to evaluating of high-dimensional integrals. The traditional grid method of numerical integration is impossible to implement when the dimension is large. For example, the use of a 10-point grid in the evaluation of the λ parameter would require 10^{90} calculations of the integrand. Not only is this infeasible even with today's

computer technology, it is not clear with this coarse grid that the final result will be accurate.

The Monte Carlo approach was drawn to the attention of econometricians by Kloek and van Dijk (1978) and further developed by Geweke (1988, 1989).[7] In finance, Monte Carlo methods have been applied in option pricing problems by Boyle (1977) and Hull and White (1987). This appendix describes how the Monte Carlo integration is used in the Bayesian approach. Two examples and their FORTRAN programs also included.

Most Bayesian inference problems can be expressed as the evaluation of the expectation of the function of interest, $g(\theta)$ under the posterior $P(\theta)$,

$$\bar{g} = \mathrm{E}[g(\theta)] = \frac{\int g(\theta) P(\theta)\, \mathrm{d}\theta}{\int P(\theta)\, \mathrm{d}\theta} \tag{A.1}$$

where θ is the parameters of the model. Let $\{\theta_i\}$ be a sequence of i.i.d. random samples drawn from the density $I(\theta)$. This is often referred to as the importance sampling approach. $I(\theta)$ is the importance density. $\{\theta_i\}$ is the importance sample. Under very weak assumptions [see Geweke (1989, sect. 2)],

$$\bar{g}_n \equiv \frac{\sum_{i=1}^{n} [g(\theta_i) P(\theta_i)/I(\theta_i)]}{\sum_{i=1}^{n} [P(\theta_i)/I(\theta_i)]}, \tag{A.2}$$

almost surely converges to $\bar{g}$ in (A.1), where n is the random sample size or replication number. Furthermore, under stronger conditions, the rate of convergence can be characterized by

$$\sqrt{n}\,(g_n - \bar{g}) \to \mathrm{N}(0, \sigma^2), \tag{A.3}$$

and σ^2 may be estimated consistently.

With samples generated from $P(\theta)$, (A.3) says that the random error $(g_n - \bar{g})$ caused by the numerical integration is asymptotically (in the number of replications) normal with mean zero and standard deviation $\sigma/\sqrt{n}$, where σ^2 is the variance of $g(\theta)$. In practice, an approximation of σ, σ_n, is used to obtain the numerical standard error of the Monte Carlo integration, $\sigma_n/\sqrt{n}$. The numerical standard error is often a good indication of the accuracy. For example, a replication number of 10,000 often has an error of less than 2% of the standard deviation of $g(\theta)$.

[7]Other applications can be found in Gallant and Monahan (1985), Zellner, Bauwens, and van Dijk (1988), Richard and Steel (1988), and Barnett, Geweke, and Yue (1988).

To demonstrate how the approach is used, consider two examples. First, suppose we wanted to evaluate the integral:

$$I = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 \, e^{-x^2/2} \, dx.$$

It is obvious that the analytical solution is 1. The Monte Carlo integration is

$$\frac{1}{n} \sum_{i=1}^{n} x_i^2,$$

where $\{x_i\}$ are drawn from a standard normal. The FORTRAN code is:

```
      DOUBLE PRECISION DSEED
      DSEED = 37581.0D00
      DO 10 I = 1,10000
      CALL GGNML (DSEED, 1, X)
      S = S + X**2
   10 CONTINUE
      S = S/10000
      WRITE (*,20) S
   20 FORMAT (F12.8)
      STOP
      END
```

where GGNML is the standard normal random number generator in IMSL. With 10,000 replications, the numerical value is 0.999.

Second, suppose we wanted to evaluate the following integral:

$$I = \int_0^1 \frac{\sin[\ln(1+x)]}{x} \, dx.$$

The analytical solution to this example is not obvious. However, the Monte Carlo integration is straightforward. Consider the FORTRAN code:

```
      DOUBLE PRECISION DSEED
      DSEED = 37581.0D00
      DO 10 I = 1, 10000
      CALL GGUBS (DSEED, 1, X)
      S = S + SIN(LOG(1 + X))/X
   10 CONTINUE
      S = S/10000
      WRITE (*,20) S
   20 FORMAT (F12.8)
      STOP
      END
```

where GGUBS is the standard uniform random number generator in IMSL. With 10,000 replications, the numerical value is 0.7977.

Appendix B: Proofs

B.1. Proof of (8)

Let I be the identity matrix, a and b column vectors, then a well-known formula in matrix theory states that:

$$|I + ab'| = 1 + a'b. \tag{B.1}$$

Applying (B.1) to (8.43) of Zellner (1971), we immediately obtain (8).

B.2. Proof of (12)

As in Zellner (1971, p. 232), we have

$$(B - \hat{B})'X'X(B - \hat{B}) = \frac{1}{a}(\alpha - \hat{\alpha})(\alpha - \hat{\alpha})' + \frac{1}{b}(\beta - \bar{\beta})(\beta - \bar{\beta})', \tag{B.2}$$

where $b = (r_p'r_p)^{-1}$ and

$$\bar{\beta} = \hat{\beta} - br_p'I(\alpha - \hat{\alpha}).$$

So,

$$\operatorname{tr}\left[(B - \hat{B})'X'X(B - \hat{B})\Sigma^{-1}\right]$$

$$= \frac{1}{a}(\alpha - \hat{\alpha})'\Sigma^{-1}(\alpha - \hat{\alpha}) + \frac{1}{b}(\beta - \bar{\beta})'\Sigma^{-1}(\beta - \bar{\beta}).$$

We know that β has a multivariate normal distribution with mean $\bar{\beta}$ and covariance matrix $b\Sigma$, conditional on α, Σ, and the data. Therefore, we can readily integrate out the β parameters in (7) and establish (12).

B.3. Evaluation of the mean of λ

Results of (7) and (12) imply the mean is given by

$$\bar{\lambda} = \int\left[\int(\alpha'\Sigma^{-1}\alpha)P(\alpha|\Sigma)\,d\alpha\right]P(\Sigma)\,d\Sigma, \tag{B.3}$$

where $P(\boldsymbol{\alpha}|\boldsymbol{\Sigma})$ is the multivariate normal density with mean $\hat{\boldsymbol{\alpha}}$ and covariance matrix $a\boldsymbol{\Sigma}$, $P(\boldsymbol{\Sigma})$ is the inverted Wishart density, i.e., the one obtained from the right-hand side of (7) by adding a constant. To evaluate (B.3), we make the following transformation of the parameters:

$$\boldsymbol{\alpha} = (a\boldsymbol{\Sigma})^{1/2}\boldsymbol{\theta} + \hat{\boldsymbol{\alpha}}. \tag{B.4}$$

Then $\boldsymbol{\theta}$ will follow the standard multivariate normal distribution, and

$$\boldsymbol{\alpha}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\alpha} = a\boldsymbol{\theta}'\boldsymbol{\theta} + 2a^{1/2}\hat{\boldsymbol{\alpha}}'\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\theta} + \hat{\boldsymbol{\alpha}}'\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\alpha}}.$$

Therefore, after performing the integration in the $\boldsymbol{\alpha}$ or $\boldsymbol{\theta}$ space in (B.3), we have the result

$$\hat{\boldsymbol{\alpha}}'\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\alpha}} + Na. \tag{B.5}$$

Now let $\boldsymbol{A} = \boldsymbol{\Sigma}^{-1}$, then $\boldsymbol{A}$ has a Wishart density $W(S^{-1}, T-2, N)$ [see Zellner (1971, app. B)].

Apply the moments results of a Wishart distribution to

$$\hat{\boldsymbol{\alpha}}'\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\alpha}} = \sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\alpha}_{ip}\alpha_{ij}\hat{\alpha}_{jp}.$$

We immediately obtain (13), that is:

$$\bar{\lambda} = Na + \hat{\boldsymbol{\alpha}}'\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\alpha}}.$$

References

Anderson, Theodore W., 1984, An introduction to multivariate statistical analysis, 2nd ed. (Wiley, New York, NY).

Barnett, William A., John Geweke, and Piyu Yue, 1988, Semiparametric Bayesian estimation of the asymptotically ideal model: The AIM demand system, Unpublished working paper (Duke University, Durham, NC).

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

Box, George E.P. and George C. Tiao, 1965, Multiparameter problems from a Bayesian point of view, Annals of Mathematical Statistics 36, 1468–1482.

Boyle, Phelim P., 1977, Options: A Monte Carlo approach, Journal of Financial Economics 4, 323–338.

Breeden, Douglas T., 1979, An intertemporal asset pricing model with stochastic consumption and investment opportunities, Journal of Financial Economics 7, 265–296.

Breeden, Douglas T., Michael R. Gibbons, and Robert H. Litzenberger, 1989, Empirical tests of the consumption based CAPM, Journal of Finance 44, 231–262.

Dickey, James M., 1971, The weighted likelihood ratio: Linear hypotheses on normal location parameters, Annals of Mathematical Statistics 42, 204–223.

Dybvig, Philip H., 1983, An explicit bound on individual assets' deviations from APT pricing in a finite economy, Journal of Financial Economics 12, 483–496.
Ferson, Wayne E. and Campbell R. Harvey, 1991, The variation of economic risk premiums, Unpublished working paper (Duke University, Durham, NC), forthcoming in the Journal of Political Economy.
Gallant, A. Ronald and John F. Monahan, 1985, Explicit infinite dimensional Bayesian analysis of production technologies, Journal of Econometrics 30, 171–202.
Geisser, Seymour and J. Cornfield, 1963, Posterior distributions for multivariate normal parameters, Journal of the Royal Statistical Society B 25, 368–376.
Geweke, John, 1986, Exact inference in the inequality constrained normal linear regression model, Journal of Applied Econometrics 1, 127–141.
Geweke, John, 1988, Antithetic acceleration of Monte Carlo integration in Bayesian inference, Journal of Econometrics 38, 73–90.
Geweke, John, 1989, Bayesian inference in econometric models using Monte Carlo integration, Econometrica 57, 1311–1339.
Gibbons, Michael R., 1982, Multivariate tests of financial models: A new approach, Journal of Financial Economics 10, 3–27.
Gibbons, Michael R., Stephen A. Ross, and Jay Shanken, 1989, A test of the efficiency of a given portfolio, Econometrica 57, 1121–1152.
Grinblatt, Mark and Sheridan Titman, 1983, Factor pricing in a finite economy, Journal of Financial Economics 12, 497–508.
Hull, John and Alan White, 1987, The pricing of options on assets with stochastic volatilities, Journal of Finance 42, 281–300.
Jeffreys, Harold, 1961, The theory of probability (Oxford University Press, Oxford).
Kloek, Teun and Herman K. van Dijk, 1978, Bayesian estimates of equation system parameters: An application of integration by Monte Carlo, Econometrica 46, 1–20.
Krishnaiah, Paruchuri R., 1980, Computations of some multivariate distributions, in: P.R. Krishnaiah, ed., Handbook of statistics (North-Holland, Amsterdam).
Lintner John, 1965, The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets, Review of Economics and Statistics 47, 13–37.
Long, John B. Jr., 1974, Stock prices, inflation, and the term structure of interest rates, Journal of Financial Economics 1, 131–170.
MacKinlay, A. Craig, 1987, On multivariate tests of the capital asset pricing model, Journal of Financial Economics 18, 341–372.
McCulloch, Robert and Peter E. Rossi, 1988, Testing the arbitrage pricing theory: Predictive distributions, posterior distributions and odds ratios, Unpublished working paper (University of Chicago, Chicago, IL).
McCulloch, Robert and Peter E. Rossi, 1989a, A Bayesian approach to testing arbitrage pricing theory, Unpublished working paper (University of Chicago, Chicago, IL), forthcoming in the Journal of Econometrics.
McCulloch, Robert and Peter E. Rossi, 1989b, Posterior, predictive, and utility-based approaches to testing the arbitrage pricing theory, Unpublished working paper (University of Chicago, Chicago, IL).
Merton, Robert C., 1973, An intertemporal capital asset pricing model, Econometrica 41, 867–887.
Press, S. James, 1982, Applied multivariate analysis: Using Bayesian and frequentist methods of inference, 2nd ed. (Krieger, Malibar, FL).
Richard, Jean-Francois and Mark F.J. Steel, 1988, Bayesian analysis of systems of seemingly unrelated regression equations under a recursive extended natural conjugate prior density, Journal of Econometrics 38, 7–37.
Ross, Stephen A., 1976, The arbitrage theory of capital asset pricing, Journal of Economic Theory 13, 341–360.
Rossi, Peter E., 1980, Testing hypotheses in multivariate regression: Bayes vs. non-Bayes procedures, Unpublished working paper (University of Chicago, Chicago, IL).
Rubinstein, Mark, 1976, The valuation of uncertain income streams and the pricing of options, Bell Journal of Economics 7, 407–425.
Scheffé, Henry, 1977, A note on the reformulation of the S-method of multiple comparison, Journal of the American Statistical Association 72, 143–144.

Shanken, Jay, 1985, Multivariate tests of the zero-beta CAPM, Journal of Financial Economics 14, 327–348.
Shanken, Jay, 1986, Testing portfolio efficiency when the zero-beta rate is unknown: A note, Journal of Finance 41, 269–276.
Shanken, Jay, 1987a, Multivariate proxies and asset pricing relations: Living with the Roll critique, Journal of Financial Economics 18, 91–110.
Shanken, Jay, 1987b, A Bayesian approach to testing portfolio efficiency, Journal of Financial Economics 19, 195–216.
Shanken, Jay, 1990, Intertemporal asset pricing: An empirical investigation, forthcoming in the Journal of Econometrics.
Sharpe, William, 1964, Capital asset prices: A theory of market equilibrium under conditions of risk, Journal of Finance 19, 425–442.
Sharpe, William, 1982, Factors in New York Stock Exchange security returns, 1931–1979, Journal of Portfolio Management 8, 5–19.
Stambaugh, Robert F., 1982, On the exclusion of assets from tests of the two parameter model: A sensitivity analysis, Journal of Financial Economics 10, 237–268.
Zellner, Arnold, 1971, An introduction to Bayesian inference in econometrics (Wiley, New York, NY).
Zellner, Arnold, A. Luc Bauwens and Herman K. van Dijk, 1988, Bayesian specification analysis and estimation of simultaneous equation models using Monte Carlo methods, Journal of Econometrics 38, 39–72.
Zellner, Arnold and Aloysius Siow, 1980, Posterior odds ratios for selected regression hypotheses, in: J.M. Bernardo, D.V. Lindley, and A.F.M. Smith, eds., Bayesian analysis: Proceedings of the first international meeting (University Press, Spain).