Bayesian Inference and Portfolio Efficiency

Shmuel Kandel
Tel-Aviv University and University of Pennsylvania

Robert McCulloch
University of Chicago

Robert F. Stambaugh
University of Pennsylvania and National Bureau of Economic Research

A Bayesian approach is used to investigate a sample's information about a portfolio's degree of inefficiency. With standard diffuse priors, posterior distributions for measures of portfolio inefficiency can concentrate well away from values consistent with efficiency, even when the portfolio is exactly efficient in the sample. The data indicate that the NYSE-AMEX market portfolio is rather inefficient in the presence of a riskless asset, although this conclusion is justified only after an analysis using informative priors. Including a riskless asset significantly reduces any sample's ability to produce posterior distributions supporting small degrees of inefficiency.

A portfolio is *efficient* if it offers the highest expected return for a given variance of return. The concept of portfolio efficiency, pioneered by Markowitz (1952,

We are grateful for comments received from Bruce Lehmann, Andrew Lo, Craig MacKinlay, Angelo Melino, Peter Rossi, Guofu Zhou, and participants in workshop at Berkeley, Cornell, Duke, Michigan, New York University, Ohio State, the National Bureau of Economic Research, the University of Manitoba, the University of Pennsylvania, Stanford, Washington University, and the Western Finance Association. Robert McCulloch acknowledges support from the Graduate School of Business and the IBM Faculty Research Fund at the University of Chicago. Address correspondence to Robert F. Stambaugh, Finance Department, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104-6367.

The Review of Financial Studies Spring 1995 Vol. 8, No. 1, pp. 1-53

1959), offers obvious normative content as well as a framework for representing modem theories of financial asset pricing. In the latter context, for example, the capital asset pricing model (CAPM) of Sharpe (1964), Lintner (1965), and Black (1972) maintains that the market portfolio is efficient, and the consumption-based model of Breeden (1979) requires the efficiency of the portfolio whose return has the maximum correlation with consumption.

In the traditional empirical approach, the efficiency of a given portfolio p, whose composition is known to the researcher, is tested as a simple (point) hypothesis. Classical. methods of statistical inference are used to accept or reject the hypothesis at a given significance level.[1] Two departures from this traditional approach have emerged. The first of these stems from the recognition that portfolio *p* is often an imperfect "proxy" for a more theoretically interesting portfolio *q*, whose exact composition is unknown. One is thus led to entertain a composite hypothesis that allows some degree of inefficiency in portfolio *p*, even though the theory might specify a simple hypothesis of exact efficiency for portfolio *q*. Kandel and Stambaugh (1987) and Shanken (1987a) have explored approaches to investigating composite hypotheses of approximate efficiency using classic frequentist methods. A second major departure from the traditional approach relies on Bayesian inference instead of frequentist methods. Shanken (1987b), McCulloch and Rossi (1990, 1991), and Harvey and Zhou (1990) develop and apply Bayesian approaches to drawing inferences about portfolio efficiency and asset pricing models.

This study develops and analyzes a framework that combines both of the above departures from the traditional approach. The inefficiency of the observed portfolio *p* is, described by either of two univariate measures with simple intuitive appeal: Δ, equal to minus the difference in expected returns between portfolio *p* and an efficient portfolio of equal variance, and *r*, the maximum correlation between the returns on *p* and an efficient portfolio. Exact efficiency is represented by $\Delta = 0$ and $\rho = 1$ and degrees of inefficiency are rep resented by values of $\Delta > 0$ and $\rho < 1$. Both measures incorporate the restriction that an efficient portfolio must lie on the *positively* sloped portion of the minimum-variance boundary. The latter restriction-is not incorporated, for example, in the Bayesian approaches of Shanken (1987b), Harvey and Zhou (1990), and McCulloch and Rossi

[1] The earliest examples of this approach include Douglas (1969), Black, Jensen, and Scholes (1972), and Fama and MacBeth (1973). For recent developments and interpretations of such methods, see Gibbons, Ross, and Shanken (1989), Kandel and Stambaugh (1989), Zhou (1991), and Shanken (1992).

(1990, 1991) or in the likelihood-ratio tests, summarized by Kandel and Stambaugh (1989).

We apply methods of Bayesian inference to obtain posterior distributions for ρ and Δ. These distributions allow one to compute directly the posterior probability of a hypothesis about the degree of inefficiency of the observed portfolio *p*, represented by a composite hypothesis about ρ or Δ. Our focus on composite hypotheses stands in contrast to the emphasis on a point hypothesis of exact efficiency in previous Bayesian studies. Shanken (1987b), McCulloch and Rossi (1991), and Harvey and Zhou (1990) compute posterior odds ratios in order to infer the probability that the observed portfolio is exactly efficient, whereas that point hypothesis is assigned zero measure in our approach.[2]

All of the previous Bayesian studies examine portfolio efficiency in the presence of a riskless asset. The definitions of Δ and ρ apply whether or not a riskless asset is included in the asset universe. We investigate both cases and discover interesting differences between them. Our initial analysis specifies a diffuse prior distribution for the mean vector (E) and the variance-covariance matrix (V) of the joint distribution of risky asset returns. Posterior distributions of ρ and Δ are computed using 25 years of weekly returns on 12 stock portfolios, where portfolio *p* is defined as the value-weighted portfolio of the New York and American Stock Exchanges. When a riskless asset (one-week Treasury bill) is included, the posterior distributions for ρ and Δ seem to be concentrated on values well away from those representing exact efficiency. When the riskless asset is excluded, the posterior distributions for both measures, particularly the distribution for ρ, concentrate on values closer to efficiency.

Of paramount interest in any Bayesian analysis is the role played by the sample data in forming the posterior beliefs about a given parameter. When a riskless asset is included and the prior for *E* and *V* is diffuse, we show that the sample information relevant to the posterior of ρ can be summarized in terms of two sufficient statistics: $\hat{\rho}$, the sample estimate of ρ, and $\hat{\theta}$, the sample Sharpe measure of portfolio. *p*. (The Sharpe measure is the ratio of a portfolio's mean excess return to the standard deviation of its return.) This result allows us to investigate thoroughly the role played by samples of various characteristics in determining the posterior distribution of ρ. We discover that, in samples of sixes often encountered in practice, the posterior distribution of ρ concentrates around low values—even in samples where

[2] As we discuss below, Shanken (1987b) also examines a composite hypothesis, but he conditions on an unknown parameter and directly assigns prior probabilities to discrete values of r rather than specifying prior distributions for the underlying fundamental moments of returns.

portfolio p is exactly efficient ($\hat{\rho} = 1$). When prior beliefs about the distribution of returns are disperse, a posterior for ρ that concentrates around low values need not indicate that the sample information has played a significant role. In essence, disperse beliefs about the basic parameters of the return distribution can imply sharper beliefs about an inefficiency measure such as ρ, a nonlinear function of those parameters. More importantly, these prior beliefs about ρ exert a strong influence on the posterior distribution, in that an extremely large sample is required, even with $\hat{\rho} = 1$, in order to arrive at a posterior for ρ that supports only modest degrees of inefficiency (high ρ).

Excluding the riskless asset dramatically reduces the sample size necessary to infer that ρ is high, even though the prior distribution for ρ assigns much of Its mass to low values in that case as well. Thus, it seems that, with sample sixes typically encountered, a concentrated posterior can be interpreted properly as reflecting primarily sample, information. A similar conclusion is reached for the posterior distributions for Δ, both with and without a riskless asset. With that inefficiency measure, however, we find that disperse beliefs about E and V correspond to disperse beliefs about Δ.

To understand better the extent to which our actual sample does contain information about ρ, especially when a riskless asset is included, we specify a class of informative (proper) priors for the parameters of the return distribution. Priors from this class can be specified so that, unlike the diffuse prior, they imply a marginal prior distribution for ρ that is disperse between zero and one, but prior beliefs about expected returns must then be surprisingly concentrated. With such a prior, the posterior for ρ obtained with the actual 25-year sample of weekly returns is concentrated around low values, so it appears that the sample does indeed contain information indicating that the value-weighted NYSE-AMEX portfolio is not highly correlated with the efficient (tangent) portfolio. On the other hand, when the same disperse prior for p is combined with a hypothetical 25-year sample in which $\hat{\rho} = 1$, the posterior distribution is much less concentrated. Again it appears that, unless one has strong prior beliefs that a given portfolio is highly correlated with the efficient portfolio, much larger samples are required to assign high posterior probability to such an inference. Whether $\hat{\rho}$ is high or low, however, the prior plays a very important role in obtaining the posterior for ρ using typically sized samples that include a riskless asset. In the absence of a riskless asset, the information in a 25-year sample is sufficient to produce essentially identical posteriors using a wide range of priors.

Unlike previous Bayesian investigations of portfolio efficiency, the set of unknown parameters in our analysis includes the mean and variance of the return on portfolio p, denoted by μ_p and σ_p^2. While it

might seem that, to infer a portfolio's inefficiency, one surely needs to infer the mean and variance of that portfolio's return, there are in fact some measures of inefficiency for which this is not the case. When a riskless asset is included, one such measure, λ, is the difference in squared Sharpe measures between an efficient portfolio and portfolio *p*. A Bayesian posterior distribution for λ can be obtained in a regression framework, wherein the set of unknown parameters does not Include μ_p and σ_p^2 [Harvey and Zhou (1990)]. Unfortunately, numerical values of λ do not lend themselves to simple interpretations, so previous studies have attempted to develop intuition about λ by transforming it to ρ [e.g., Shanken (1987b) and Harvey and Zhou (1990)]. This transformation requires knowledge of θ, the Sharpe measure of *p*, which in turn depends on μ_p and σ_p^2. An approach followed in previous studies is to plug in a value of θ to obtain the posterior of ρ from the posterior of λ. As shown here, this "plug-in" approach yields the conditional posterior distribution of ρ given θ. We find that the problems associated with misinterpreting concentrated posterior distributions of ρ are exacerbated for this conditional posterior.

The remainder of the article proceeds as follows. Section 1 discusses several aspects of the Bayesian estimation approach that guides our attempts to draw inferences about composite hypotheses of portfolio inefficiency. Section 2 formally defines the inefficiency measures ρ and Δ, discusses their interpretations, and compares them to previously used measures. Section 3 then describes the construction of posterior distributions of ρ and Δ. Section 4 investigates the manner in which, samples of various characteristics influence the posteriors of the inefficiency measures. Section 5 considers the use of informative priors instead of the diffuse priors used in the foregoing analysis. Section 6 briefly reviews the main conclusions.

1. Bayesian Inference and Composite Hypotheses

Investigating a hypothesized restriction on the parameter space by computing the posterior distribution of a function on that space (such as ρ.) is known in the Bayesian literature as the *estimation* approach. Generally, in the estimation approach, a prior is placed on a parameter vector 6, the posterior of δ is computed given the likelihood, and then the marginal posterior of the function $g(\delta)$ is obtained. Ideally, $g(\delta)$ captures the practical importance of the departure of δ from the hypothesized restriction in a simple and easily interpretable manner. Informally, the posterior distribution of $g(\delta)$ can be used to investigate a precise, or "point," hypothesis of the form $g(\delta) = c$ simply by observing how much of the mass is placed on values of no practical difference from *c*. More formally, one can compute the expected loss

of acting as if g(δ) = *c*, as in McCulloch and Rossi (1990). The posterior distribution of g(δ) can also be used to compute the posterior probability of composite hypotheses of the form $H = \{\delta : g(\delta) \in A\}$. In our case, for example, we might want to compute the posterior probability of the composite hypotheses { ρ > .9}.

The estimation approach is popular because it is useful in capturing substantive, as opposed to just statistically significant, departures from a hypothesis [see, for example, Kadane (1987)]. In the case of a point hypothesis, however, where the set of parameters satisfying the hypothesized restriction is of lower dimension than the unrestricted parameter space, the prior placed on δ in the estimation approach typically assigns zero probability to the point hypothesis. Consequently, that hypothesis is also assigned zero posterior probability, so it is impossible to infer that the hypothesis holds exactly.

When the prior does assign positive probability to a point hypothesis, we can compute the posterior probability of that hypothesis, $P(H \mid D)$, and the corresponding *posterior odds ratio*, $P(H \mid D)/(1 - P(H \mid D))$, where *D* represents the data. In this case, the prior is usually represented as a mixture, $P = P(H)P_1 + (1 - P(H))P_2$, where $P(H)$ is the prior probability of the hypothesis, P_1 is a distribution whose support lies in the restricted parameter space corresponding to *H*, and P_2 is a distribution that gives prior support to parameters lying outside of the restricted space. The posterior odds ratio is then $[P(H) \int L(\delta) dP_1(\delta)]/[(1 - P(H)) \int L(\delta) dP_2]$ where *L* is the likelihood. This odds ratio is typically quite sensitive to the choices of the P_1 and P_2. For example, the odds can usually be made quite large by choosing P_2 to be proper but very disperse relative to P_1. Even asymptotically, the choice of priors affects the value of the odds ratio.

There is an extensive literature examining the odds ratio as a method for testing hypotheses. The well-known Jeffreys-Lindley paradox consists of a simple example in which, as the sample size goes to infinity, the p-value remains fixed at .05 (suggesting that the hypothesis may be rejected at significance level .05) but the odds ratio goes to infinity, so that the probability of the hypothesis goes to one. [For a textbook discussion, see Zellner (1971, p. 304).] More recent studies by Berger and Selke (1987) and Berger and Delampady (1987) demonstrate that wide discrepancies between the odds ratio and the standard p-value can be found even in small samples and for a wide variety of priors.

As argued at the outset, we believe that one is generally led to entertain composite hypotheses of inefficiency for an observed portfolio that is selected as a proxy for an unobserved portfolio identified by financial theory. When a composite hypothesis can be expressed in terms of the function g(δ) adopted for the estimation approach

(such as ρ), the distinction between the estimation and odds-ratio approaches can be more formal than substantive. The posterior probability of a set defined in terms of g(δ), divided by the posterior probability of its complement, is an odds ratio and can be computed directly from the posterior distribution of g(δ). For example, given the posterior distribution of ρ, we can easily compute $\Pr[\rho > .9]/\Pr[\rho \leq .9]$, which is an odds ratio. Conversely, providing the odds ratio for a variety of sets defined in terms of g(δ) is equivalent to specifying the posterior distribution of g(δ). As discussed above, the formal distinction between the two approaches is that, when computing an 'odds ratio, one specifies a prior on the set defining the null hypothesis as well as a prior on its complement. Clearly, this is just a particular method for defining the prior on the entire parameter space, which is required for the estimation approach.

2. Measures of Portfolio Inefficiency

Let the vector R_t contain returns in period *t* on *n* risky assets, where one of the *n* assets is portfolio *p,* whose efficiency is to be investigated. If a riskless asset exists, then R_t contains excess returns on these assets (i.e., returns in excess of the riskless rate). The set of efficient portfolios is defined allowing short positions in all assets. Our two inefficiency measures ρ and Δ are rather complicated nonlinear functions of the elements of *E* and *V,* the mean vector and covariance matrix of R_t. These functions will be used to obtain univariate posterior distributions for Δ and ρ from the multivariate posterior distributions for *E* and *V.* To aid in the analysis, we first provide simple characterizations of both inefficiency measures in mean-standard-deviation space.

Let *x* denote the efficient portfolio with the same standard deviation as portfolio *p,* and let *y* denote the minimum-variance portfolio with the same mean return as portfolio *p.* Let *g* denote the portfolio having minimum variance among all portfolios composed solely of risky assets. The mean and standard deviation of the return on a given portfolio *q* are denoted by μ_q and σ_q. When a riskless asset exists, the Sharpe measure of a portfolio is defined in mean-standard-deviation space as the slope of a ray connecting the portfolio to the origin. (Recall that R_t contains excess returns when a riskless asset is included.) The maximum Sharpe measure is denoted by γ, and the Sharpe measure of portfolio *p* is denoted by $\theta(=\mu_p/\sigma_p)$.

The inefficiency measure Δ is given by

$$\Delta = \mu_x - \mu_p, \tag{1}$$

or the vertical distance in mean-standarddeviation space between portfolio p and the locus of efficient portfolios. This distance is well defined whether or not a riskless asset is included. Alternatively, Δ (≥ 0) equals the average loss in rate of return associated with holding the inefficient portfolio p instead of the efficient portfolio of equal risk. If portfolio p is efficient, then $\Delta = 0$. This simple inefficiency measure obviously lends itself to economic interpretation, since it can be viewed on the same scale as expected asset returns.

The inefficiency measure ρ is defined as the maximum correlation between portfolio p and an efficient portfolio. Exact efficiency of portfolio p corresponds to $\rho = 1$, whereas inefficiency in p is characterized by $\rho < 1$. In the presence of a riskless asset, efficient portfolios lie on a ray emanating from the origin with slope γ (> 0).[3] The correlation between p and any efficient portfolio is given by

$$\rho = \theta/\gamma. \tag{2}$$

That is, ρ is equal to the ratio of portfolio *p's* Sharpe measure to the maximum Sharpe measure [see Kandel and Stambaugh (1987) and Shanken (1987a)].[4] In the absence of a riskless asset, ρ is the maximum correlation between portfolio p and any portfolio on the positively sloped portion of the (curved) minimum-standard-deviation boundary. This correlation is given by

$$\rho = \begin{cases} \sigma_y/\sigma_p & \text{if } \mu_p > \mu_g \\ \sigma_g/\sigma_p & \text{otherwise,} \end{cases} \tag{3}$$

as shown by Kandel and Stambaugh (1987). When the mean of p exceeds that of the global minimum-variance portfolio, then both (2) and (3) give $\rho = \sigma_y/\sigma_p$. [In the case of (2), this follows from the observation that γ is then the Sharpe measure of y, which has the same mean as p.] Thus, while Δ is the difference In means for portfolios having the same risk, ρ is the ratio of standard deviations for portfolios having the same mean.

To aid in economic intuition about values of ρ, it may be helpful to review some of the previous literature. Perhaps the earliest use of correlation to characterize a portfolio's inefficiency is by Roll (1977), who computes the sample correlation between a stock-market index (portfolio p) and the in-sample tangent portfolio. In other words, Roll com-

[3] In principle, the "tangent portfolio" composed solely of the n risky assets could lie on either the positively or negatively sloped portion of the minimum-standard-deviation boundary of risky assets (or a tangency could fail to exist). In all cases, however, efficient portfolios lie on a positively sloping ray emanating from the origin.

[4] Note that the correlation between p and the riskless asset (the origin) is undefined.

putes an *ex post* value of ρ. Roll argues that high values of ρ —at least 0.9 for his examples of the minimum-variance boundary—should give one serious reservations about rejecting the CAPM simply because one rejects the exact efficiency of portfolio *p*, an imperfect market proxy. Roll also observes, however, that "most reasonable proxies will be very highly correlated with each other and with the true market," and indeed Stambaugh (1982) finds that a variety of market indexes computed with rather different weights on broad asset classes exhibit returns with high mutual correlations (typically 0.8 or higher). Thus, given such an empirical prediction, one's fears about incorrectly rejecting the CAPM would lessen as ρ declines. This argument underlies the frequentist approaches of Kandel and Stambaugh (1987) and Shanken (1987a), and the same reasoning motivates our use of ρ in a Bayesian setting.

As mentioned earlier, previous studies have analyzed an alternative measure of inefficiency when a riskless asset is included. This alternative measure, denoted by λ, is the difference between the maximum squared Sharpe measure and the squared Sharpe measure of *p:*

$$\lambda = \gamma^2 - \theta^2. \tag{4}$$

Exact efficiency of portfolio *p* is equivalent to λ = 0. Although A does provide a univariate measure of portfolio inefficiency, it possesses limitations not shared by ρ and Δ, For example, if p lies close to the *negatively* sloping ray representing the lower boundary of the feasible set in mean-standarddeviation space (i.e., if θ is close to - γ), then λ will be close to zero, even though *p* is grossly inefficient. Since the sign of ρ is the same as the sign of θ, however, ρ would be close to -1 (not +1) in that case. Moreover, differences in squared Sharpe measures do not easily lend themselves to economic interpretation. The Bayesian approaches in Shanken (1987b) and Harvey and Zhou (1990) base their formal analyses on λ, but, in order to develop intuition, both studies turn to the link between λ, and ρ,

$$\rho = \text{sign}(\theta)\sqrt{\frac{1}{1+\frac{\lambda}{\theta^2}}}, \tag{5}$$

which follows from Equations (2) and (4) and the fact that the sign of p is the same as that of θ (and of μ_p). Note from (5), however, that the link between λ and ρ also involves θ, whose true value (or range) is uncertain. Later we compare inferences that ignore this uncertainty to those that include it.

3. Posterior Distributions with Diffuse Priors

3.1 The general framework

We assume that the n-vector of returns R_t is distributed multivariate normal, independently across t, with mean E and nonsingular variance-covariance matrix V. (Recall that the return on portfolio p is included as an element of vector R_t, so no linear combination of returns on the other n - 1 assets can yield the return on p.) For our initial analysis, we use the standard diffuse prior for the multivariate normal distribution [e.g., Zellner (1971)],

$$p(E, V) \propto |V|^{-(n+1)/2}. \tag{6}$$

For a sample of T observations of R_t, define the statistics

$$\hat{E} = \frac{1}{T}\sum_{t=1}^{T} R_t, \tag{7}$$

$$\hat{V} = \frac{1}{T}\sum_{t=1}^{T}(R_t - \hat{E})(R_t - \hat{E})' . \tag{8}$$

Given the sample of T observations and the prior distribution (6), the joint posterior distribution of (E, V) follows standard results. The matrix V^{-1} has a Wishart posterior distribution with parameter matrix $(T\hat{V})^{-1}$ and $(T-1)$ degrees of freedom. Given V, the posterior distribution of E is normal with mean $\hat{E}$ and covariance matrix V/T.

Note that, with the use of the diffuse prior in (6), several aspects of our Bayesian posterior distribution correspond to confidence regions obtained by standard frequentist analysis. For example, the quantity $(\hat{E} - E)'\hat{V}^{-1}(\hat{E} - E)$ has a posterior distribution (where E is random and $\hat{E}$ and $\hat{V}$ are considered fixed) exactly the same as the frequentist distribution (where E and V are considered fixed but $\hat{E}$ and $\hat{V}$ are random). The posterior and frequentist distributions are both $[n/(T-n)]\mathrm{F}_{n,T-n}$. The usual elliptical confidence region for E has posterior probability equal to the frequentist confidence level. Thus, our Bayesian results provide inferences about basic model parameters with a level of uncertainty equivalent to that obtained through the frequentist approach. This suggests that while the prior in (6) is sufficiently diffuse to represent ignorance, it does not inject an inordinate amount of uncertainty into our inference process. The Bayesian analysis allows us to compute marginal posterior distributions for the measures Δ and ρ, while it is not clear how one could compute the corresponding frequentist confidence intervals.

3.2 Obtaining the posterior distributions

Although the inefficiency measures ρ and Δ are functions of E and V, the complexity of these functions precludes a complete analytical derivation of the marginal posterior distributions of ρ and Δ from the joint posterior distribution of E and V. Instead, we follow a Monte Carlo method and repeatedly draw values for E and V from the posterior distribution given above. For each draw of E and V, we compute the corresponding values of ρ and Δ. For a large number of draws (5000 in this study), the empirical distribution obtained from the computed values of each measure will provide a good approximation to the posterior marginal distribution [see, e.g., Geweke (1989)].

When a riskless asset is included, the joint posterior distribution of λ and θ can be obtained analytically, and these parameters determine ρ [Equation (5)]. The Monte Carlo method can then be used to obtain the posterior for ρ by simply drawing repeatedly from the distribution of λ and θ. We state the analytical results here, and the derivations are provided in the Appendix. These analytical simplifications are also used in Section 4 to provide additional insights about the role of sample information in obtaining the marginal posterior distribution.

We first define a few statistics that are functions of $\hat{E}$ and $\hat{V}$. Recall that, since a riskless asset is included; returns are in excess of the riskless rate. Let $\hat{\mu}_p$ and $\hat{\sigma}_p^2$ denote the elements of $\hat{E}$ and $\hat{V}$ corresponding to the sample mean and variance of portfolio p (one of the n assets). The sample Sharpe measure of p is defined as

$$\hat{\theta} = \hat{\mu}_p / \hat{\sigma}_p. \tag{9}$$

It is easily shown that the maximum Sharpe measure is $\gamma = +\sqrt{E'V^{-1}E}$. The sample analog is defined as

$$\hat{\gamma} = +\sqrt{\hat{E}'\hat{V}^{-1}\hat{E}}, \tag{10}$$

which leads to the sample statistic

$$\hat{\rho} = \frac{\hat{\theta}}{\hat{\gamma}}, \tag{11}$$

using the expression for ρ in (2).

The statistics $\hat{\rho}$ and $\hat{\theta}$ are sufficient to determine the marginal posterior distribution for the inefficiency measure ρ. Conditional on a noncentrality parameter v, λ is proportional to a noncentral chi-square

variate with n - 1 degrees of freedom:

$$\lambda \sim \left[\frac{1+\hat{\theta}^2}{T}\right] \cdot \chi^2_{n-1}(\nu). \tag{12}$$

The noncentrality parameter ν is proportional to a central chi-square variate with T - 1 degrees of freedom:

$$\nu \sim \left(\frac{\hat{\theta}^2}{1+\hat{\theta}^2}\right)\left(\frac{1-\hat{\rho}^2}{\hat{\rho}^2}\right) \cdot \chi^2_{T-1}. \tag{13}$$

We also find that, given the data, λ is independent of θ, and the posterior distribution of θ is determined by $\hat{\theta}$. Specifically,

$$\theta \sim \frac{1}{\sqrt{T}}\left(\hat{\theta}\chi_{T-n} + z\right) \tag{14}$$

where χ_{T-n} is the square root of a central chi-square variate with T– n degrees of freedom and z is a standard normal variate, independent of χ_{T-n}. Therefore, to obtain a draw of ρ in constructing its unconditional distribution (histogram), we first draw χ_{T-n} and z, which together give a draw for θ (using (14)]. Independently, a draw for ν is obtained from (13) and then used to draw λ from the conditional distribution in (12). The value of ρ is then obtained from (5).

The ability to summarize the data with the sufficient statistics $\hat{\rho}$ and $\hat{\theta}$ easily reveals the potential effects of replacing some of the n assets with alternative choices. As long as portfolio p remains as one of the assets, then the statistic $\hat{\theta}$ is obviously unaffected. Modifying the remaining set of n - 1 assets can, thereore, affect the posterior of ρ only by changing its sample analog, $\hat{\rho}$, or, equivalently, only by changing the maximum sample Sharpe measure $\hat{\gamma}$ [using (11) and the condition that $\hat{\theta}$ does not change]. Adding assets that increase $\hat{\gamma}$ will decrease $\hat{\rho}^2$, place more mass on higher values of the noncentrality parameter used to generate λ [by (13)], and, as one would expect, place more posterior mass on lower absolute values of ρ [by (5)] Thus, a property such as near-singularity of the sample covariance matrix, the effect of which might in general be difficult to assess, is easily seen here to be fully captured through its effect on $\hat{\gamma}$, the estimated maximum risk-return tradeoff available to investors.

3.3 Sample results

Our sample consists of weekly returns for the period January 1, 1963, through December 31, 1987. The daily Center for Research in Security Prices (CRSP) master is used to compute weekly raw returns

on stocks of the New York Stock Exchange (NYSE) and the American Stock Exchange (AMEX). Weekly returns on U.S. Treasury bills (T-bill) are computed from price quotations of T-bills obtained from the *Wall Street Journal*. The T-bill returns were originally compiled by Oldfield and Rogalski (1987) and were subsequently updated by Ferson, Kandel, and Stambaugh (1987) and McCulloch and Rossi (1990).

We consider a set of 12 risky assets: 10 size-ranked portfolios and 2 indices of the stock market. The formulation and rebalancing of the size-ranked portfolios are described in a greater detail by McCulloch and Rossi (1990). Every four weeks, firms are sorted by market value and placed into decile portfolios. Portfolios' raw returns are computed by value weighting the Individual firms' returns. The two market indices are the NYSE-AMEX value-weighted and equally weighted portfolios. Portfolio p is the value-weighted market index. Excess returns are computed by subtracting the weekly returns on the T-bills from the raw returns. Results virtually identical to those reported are obtained, however, by subtracting the average T-bill return from the weekly raw returns.

The size-ranked portfolios, which are assigned stocks ranked by total market value of the firm's outstanding equity, are selected primarily because previous studies have found that firm size provides a simple device for distinguishing common stocks across a number of return characteristics. For example, Chan, Chen, and Hsieh (1985) find that estimated sensitivities of returns to a variety of economic factors exhibit monotonic associations with size, and Huberman, Kandel, and Karolyi (1987) find that differences in correlations among market-model residuals are monotonic in size. Numerous studies, beginning with Banz (1981), have also observed that average returns are inversely related to size, which has come to be known simply as the *size effect.* This last effect, if viewed as an ex post justification for the portfolio sorting criterion, heightens the possibility of "data-snooping" biases {e.g., Lo and MacKinlay (1990)]. At the same time, however, average returns in excess of a misspecified risk correction are likely to be inversely associated, *ceteris paribus,* with asset prices and related variables such as firm size [e.g., Berk (1993)]. Thus, such variables might also be viewed *ex ante* as providing appealing criteria for sorting stocks.

Figure 1A displays the marginal posterior of ρ when a riskless asset is included, and Figure 1B displays the posterior of ρ in the absence of the riskless asset. In the first case, most of the posterior mass lies between the values $\rho = -0.1$, and $\rho = .3$. The value of $\hat{\rho}$ defined in (11) is 0.1. As noted earlier, correlations of this magnitude are probably less than one would expect to find among alternative reasonable proxies for the market portfolio. A rather different result occurs in

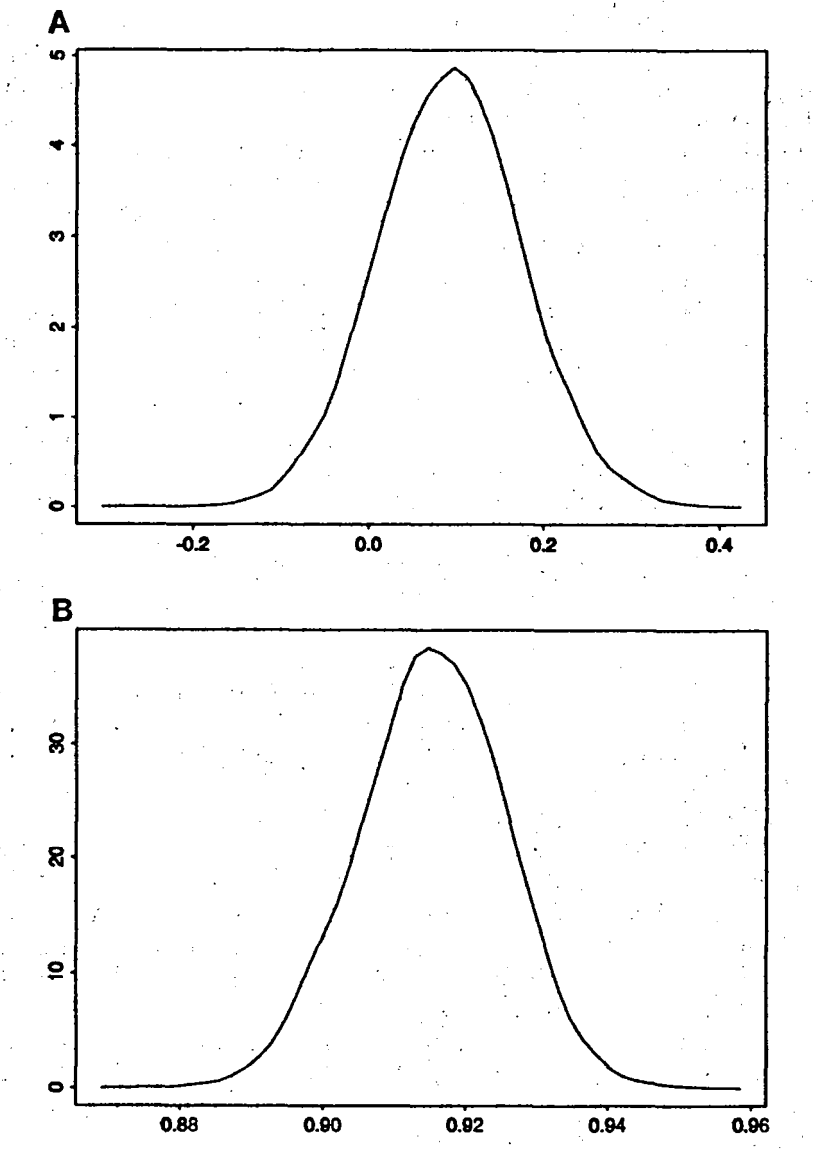

Figure 1
Posterior distributions of ρ using a diffuse prior
Each figure is based on 5000 draws from the posterior distribution of the inefficiency measure ρ, which is the maximum correlation between the value-weighted NYSE-AMEX index and an efficient portfolio. Efficient portfolios are computed using weekly returns from January 1963 through December 1987 for a set of assets consisting of 10 size-ranked portfolio and the equally and value-weighted NYSE-AMEX indexes. Figure 1A displays the posterior of ρ when a riskless asset is included, in which case returns are in excess of the return on a one-week Treasury bill. Figure 1B displays the posterior of ρ obtained when the riskless asset is excluded.

Figure 1B, where the riskless asset is excluded. In that case, most of the posterior mass lies between the values $\rho = 0.89$ and $\rho = 0.94$. (The sample estimate of ρ based on $\hat{E}$ and $\hat{V}$ is 0.92.)

Figures 2A and 2B display the posterior distributions of Δ. The values of Δ are multiplied by 52 in constructing the histograms, so Δ should be interpreted as (approximately) the expected annual return lost from holding portfolio *p* versus holding the efficient portfolio with the same risk. When a riskless asset is included, Figure 2A indicates that the posterior distribution of this expected loss centers around 30 percent per year, with most of the mass falling between 20 percent and 40 percent. Excluding the riskless asset (Figure 2B) shifts the posterior of Δ toward zero, to the degree that the posterior distributions in Figures 2A and 2B barely overlap. Nevertheless, even when the riskless asset is excluded, much of the mass for Δ falls above an expected annual loss of 10 percent. The posterior of Δ is also somewhat less disperse in that case. (The sample estimates of Δ based on $\hat{E}$ and $\hat{V}$, with and without the riskless asset, are .31 and .12).

The results obtained without the riskless asset provide an interesting contrast between the two inefficiency measures. The posterior for ρ in Figure 1B suggests that the value-weighted NYSE-AMEX index is highly correlated with the efficient portfolio having the same mean, while the posterior for Δ in Figure 2B suggests that the efficient portfolio with the same variance would be expected to outperform the index by a substantial amount. Such results need not be viewed as mutually inconsistent, since a pair of values such as $\rho = 0.9$ and $\Delta = 0.10$ is certainly a theoretical possibility. The results in Figures 1 and 2 indicate that adding a riskless asset substantially increases the degree of inefficiency of the value-weighted NYSE-AMEX index, whether that inefficiency is measured by ρ or by Δ. These results are consistent with the theoretical property that adding any asset to the universe can only, for a given portfolio *p*, lower ρ and raise Δ. [5] At the same time, however, one must be struck by the low values around which the posterior for ρ concentrates when a riskless asset is included. We discover in the next section that considerable care must be exercised in interpreting posterior distributions of these and other inefficiency measures when using the standard diffuse prior.

4. Interpreting Posteriors: The Information in a Sample

When interpreting the outcome of any Bayesian analysis, it is useful to develop some understanding of the role played by the sample in

[5] This statement **is** obvious for Δ; Kandel and Stambaugh (1987) prove this property for ρ.

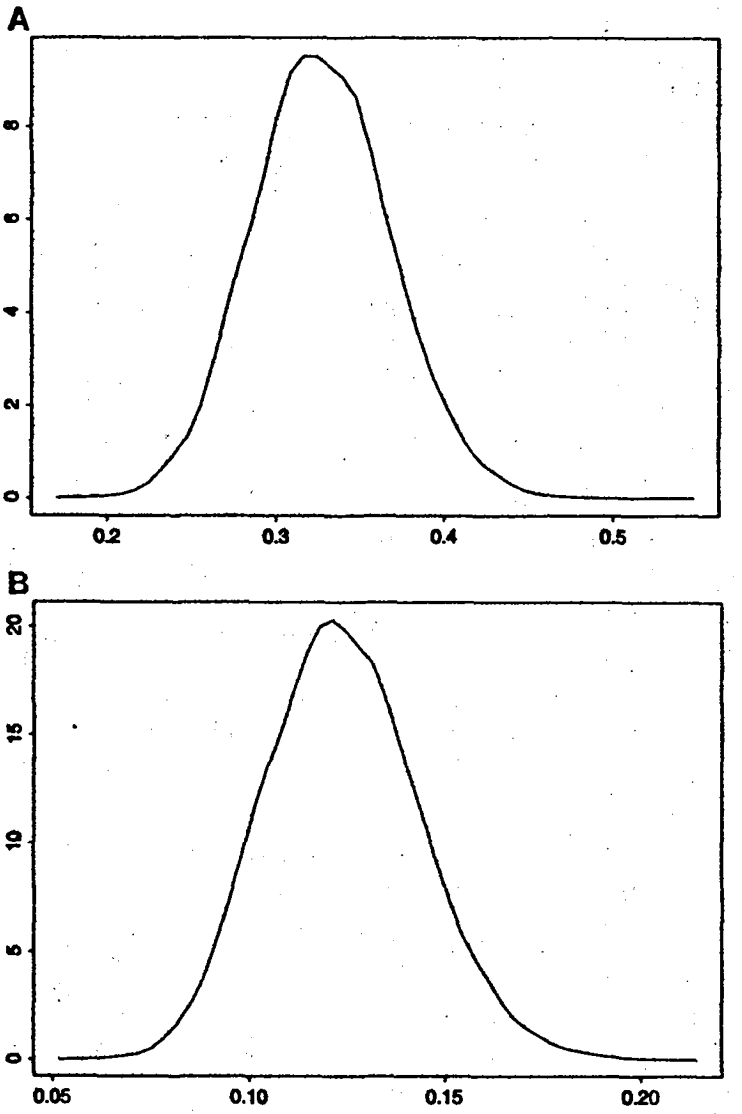

Figure 2
Posterior distributions of Δ using a diffuse prior
Each figure is based on 5000 draws from the posterior distribution of the inefficiency measure Δ, which is equal to minus the difference between the mean of the value-weighted portfolio and the efficient portfolio having the same variance. Efficient portfolios are computed using weekly returns from January 1963 through December 1987 for a set of assets consisting of 10 size-ranked portfolios and the equally and value-weighted NYSE–AMEX indexes. Figure 2*A* displays the posterior of Δ when a riskless asset is included, in which case returns are in excess of the return on a one-week Treasury bill. Figure 2*B* displays the posterior of Δ obtained when the riskless asset is excluded. All values are multiplied by 52.

determining one's posterior beliefs. The Bayesian analysis in the previous section adopts the standard specification for a diffuse prior distribution, which is intended to represent vague beliefs about the joint distribution of asset returns. With such a prior, one might be inclined to interpret disperse posterior distributions as indicating that the sample does not contain much information and to interpret concentrated posterior distributions as indicating a more informative sample. This conventional intuition is not always justified, however, particularly for arbitrary nonlinear functions of the parameters.

This section investigates the role played by sample information in determining posteriors of efficiency measures when standard diffuse priors are used to represent uninformative beliefs. The first subsection analyzes the marginal posterior distribution of ρ when a riskless asset is included. Recall from Section 3.2 that, in this case, the posterior distribution of ρ depends only on the sample size (T) and the sufficient statistics $\hat{\theta}$ and $\hat{\rho}$, defined in (9) and (11). We compute the posterior distributions of ρ for different sample sizes and values of these sufficient statistics. The second subsection considers the case where a riskless asset is included but the posterior distribution of ρ is conditioned on θ. This analysis of the conditional posterior of ρ, which we show is equivalent to the posterior developed by Harvey and Zhou (1990) in a regression framework, provides additional insights into the effects of nonlinear transformations on posterior distributions as well as the importance of incorporating uncertainty about θ.

The third subsection analyzes the posterior distribution of ρ when a riskless asset is not included, and the fourth subsection considers the posterior of Δ both with and without a riskless asset. Simplifying analytic results are not available for the computation of these posterior distributions, so the application of the Monte Carlo procedure requires that we specify, in addition to the sample size, the sufficient statistics $\hat{E}$ and $\hat{V}$, defined in (7) and (8). To keep the investigation manageable, we focus on the roles of samples of various sizes in which $\hat{E}$ and $\hat{V}$ correspond to exact sample efficiency of portfolio p ($\hat{\rho} = 1$). In essence, we investigate the maximum degree to which a given sized sample can move the posterior distribution of an inefficiency measure toward its exact-efficiency limit.

4.1 Interpreting posteriors for ρ when a riskless asset is included

As stated above, computing the posterior distribution of ρ when a riskless asset is included requires only two sufficient statistics, $\hat{\rho}$ and $\hat{\theta}$, and the sample size, T. This simplification allows us to investigate fully the role of sample information in determining the posterior of ρ.

We consider here four different values of $\hat{\rho}$ that range from exact sample efficiency to substantial inefficiency: 1.0, 0.7, 0.4, and 0.1. Two values of the sample Sharpe measure $\hat{\theta}$ are considered: 0.03 and 0.1. The first value is approximately the Sharpe measure for the value-weighted NYSE-AMEX index obtained from our 25-year sample of weekly returns (the actual estimate is 0.0323), whereas the second value would exceed historical estimates in most periods (a weekly Sharpe measure of 0.1 corresponds to an annual measure of about 0.7). Seven hypothetical sample sizes (T) are specified and numbered as follows:

Sample-size no.	Sample size, weeks
1	15
2	1.52
3	5.52
4	25.52
5	100.52
6	400.52
7	1000.52

The first sample size, 15 weeks, provides a benchmark "uninformative" case in which the sample clearly cannot provide much information about the distribution of asset returns. In that sense, the posterior in this case can be viewed roughly as a representation of prior beliefs—there is barely enough data to translate the improper diffuse prior distribution into a proper posterior distribution. The fourth sample size, 25 years, corresponds to the size of the actual sample used in the previous section. Although the last few sample sizes may seem rather large, the reasons for their inclusion will become evident as we progress.

Figures 3A to 3H display, for each of the eight combinations of $\hat{\rho}$ and $\hat{\theta}$, the posterior distribution of ρ cross the seven sample sizes. A solid line connects the 75 percent and 25 percent quantiles, and dotted lines extend out to the 1 percent and 99 percent quantiles.

In each figure, the posterior for ρ with the "uninformative" 15-week sample (no. 1) has its median around $\rho = 0.1$, and its interquartile range extends from about $\rho = -0.2$ to $\rho = 0.2$. Thus, prior beliefs about ρ are rather concentrated around low values.[6] This is our first illustration of how vague beliefs about the basic parameters of the return distribution *(E* and *V)* need not imply disperse distributions for nonlinear functions of those parameters. More will, be said on this

[6] Given the diffuse prior in (6), the marginal (improper) prior density on ρ^2 is given by

$$p(\rho^2) \propto (\rho^2)^{\frac{-(n+1)}{2}} (1-\rho^2)^{\frac{n-3}{2}},$$

which is unbounded as ρ approach zero.

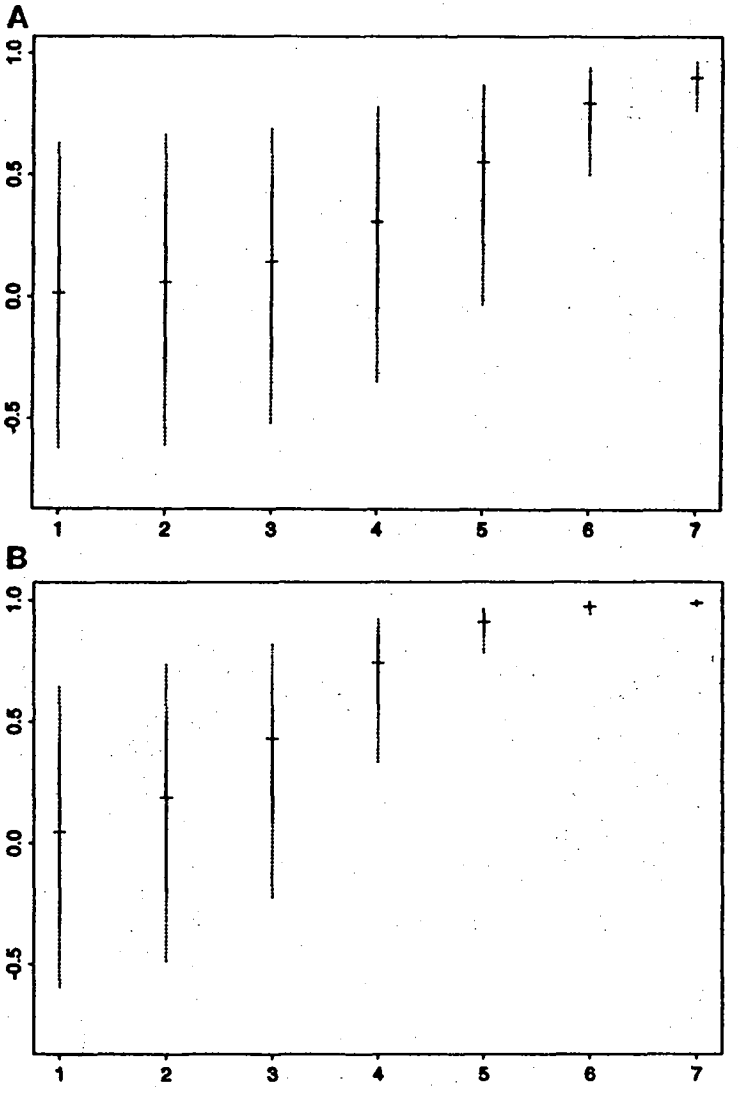

Figure 3

Posteriors distributions of ρ when a riskless asset is included, using a diffuse prior, for various sample sizes and values of the sufficient statistics

Each of Figures 3A through 3H displays, for given values of the sufficient statistics $\hat{\rho}$ and $\hat{\theta}$, the marginal posterior distributions of the inefficiency measure ρ for seven different sample sizes. The numbers 1 through 7 on the horizontal scale in each figure correspond to sample sizes of 15 weeks, 1 year, 5 years, 25 years, 100 years, 400 years, and 1000 years. For each sample size, the figure displays the median (horizontal tick) and the 1 percent, 25 percent, 75 percent, and 99 percent quantiles of 5000 draws from the posterior distribution. The solid line connects the 25 percent and 75 percent quantiles, and the dotted lines extend to the 1 percent and 99 percent quantiles. **A:** $\hat{\rho} = 1.0, \hat{\theta} = 0.03$. **B:** $\hat{\rho} = 1.0, \hat{\theta} = 0.10$.

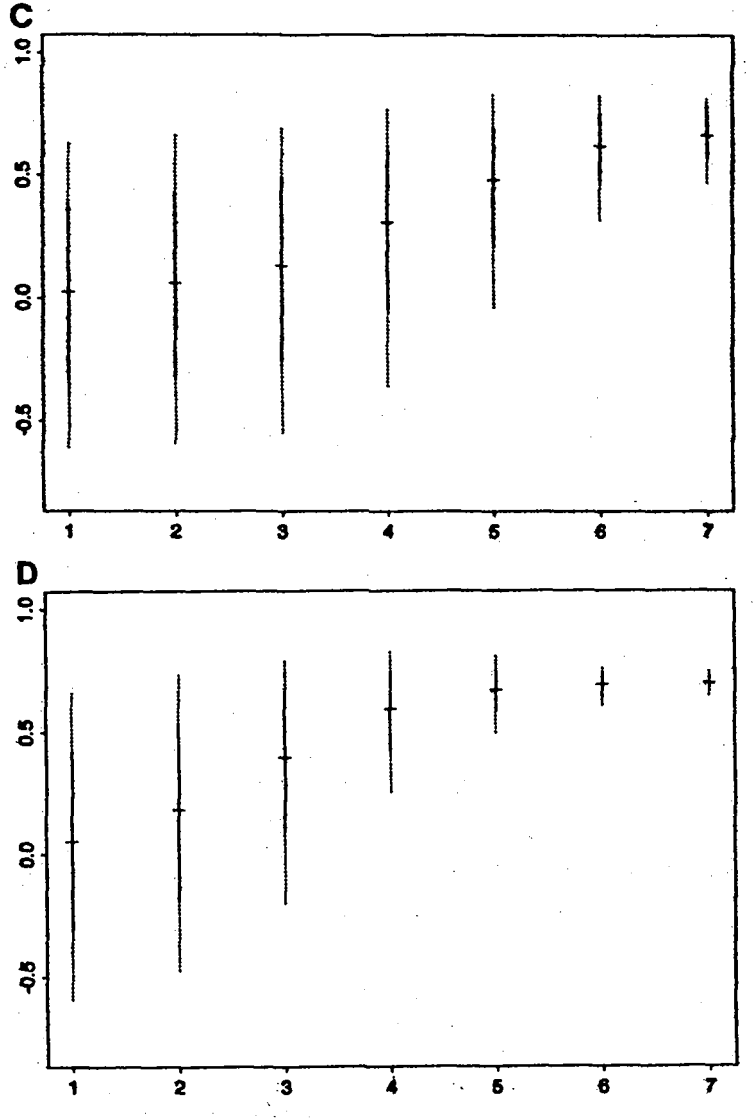

Figure 3
Continued.
C: $\hat{\rho} = 0.7, \hat{\theta} = 0.03$. D: $\hat{\rho} = 0.7, \hat{\theta} = 0.10$.

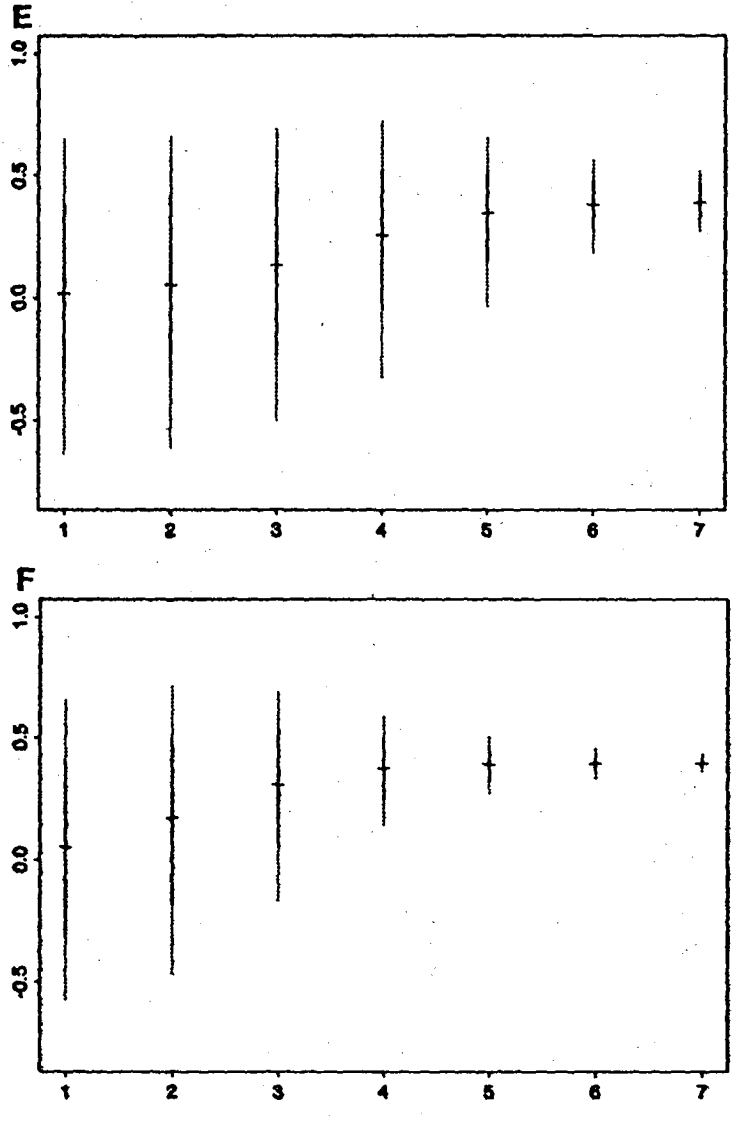

Figure 3
Continued.
E: $\hat{\rho} = 0.4$, $\hat{\theta} = 0.03$. F: $\hat{\rho} = 0.4$, $\hat{\theta} = 0.10$.

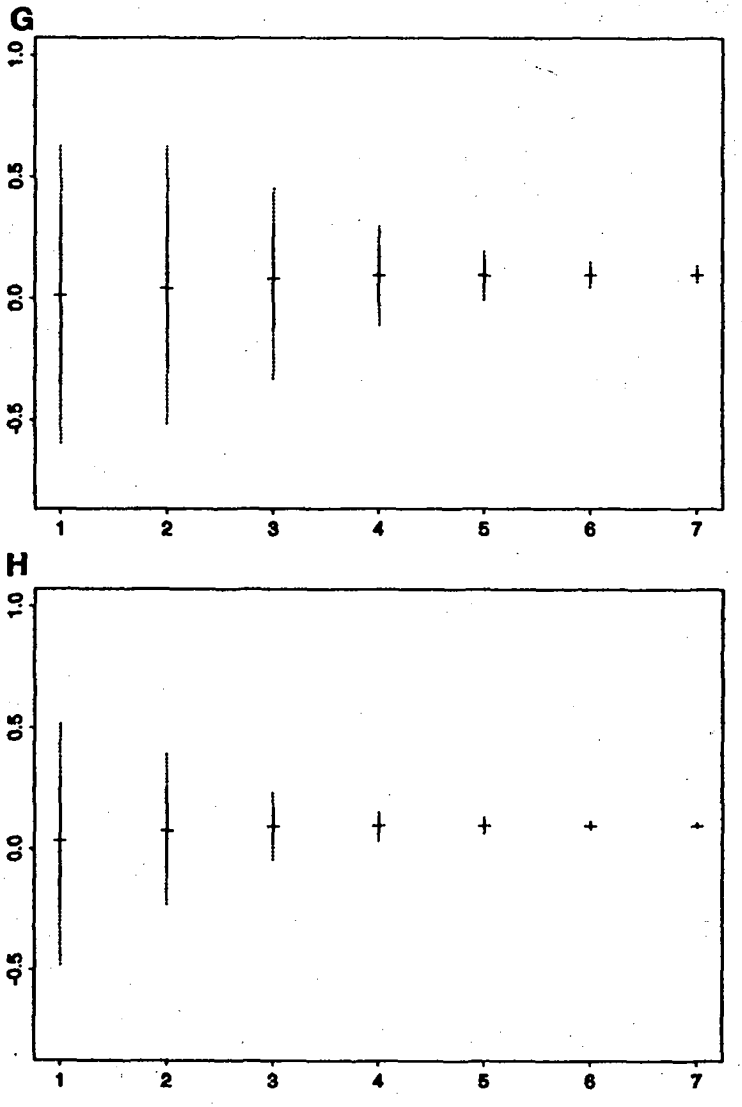

Figure 3
Continued.
G: $\hat{\rho} = 0.1, \hat{\theta} = 0.03$. H: $\hat{\rho} = 0.1, \hat{\theta} = 0.10$.

point later. This observation alone, however, does not reveal the rate at which prior beliefs become unimportant relative to the information provided by the data.

Figure 3A displays the posterior distribution for ρ when $\hat{\theta} = 0.03$ and portfolio *p* is exactly efficient in the sample (i.e., $\hat{\rho} = 1$). The amount of additional data required to influence the posterior distribution in this case is striking. A 25-year sample (no. 4) in which *p* is exactly efficient still results in a posterior distribution with about 75 percent of the mass below ρ = .5. With 400 years of weekly data (sample size no. 6), most of the posterior mass still lies below 0.9. We find that a sample of about .1000 years (no. 7) in which *p* is exactly efficient is required in order to have most of the posterior mass exceed ρ = 0.9. In a Bayesian analysis that starts with a diffuse prior about the parameters in the joint distribution of returns, a posterior for ρ that is concentrated around low values might lead one to conclude that the sample is being informative about the inefficiency of portfolio *p*. As we see, this need not be the case.

Figure 3B displays the posteriors for ρ when $\hat{\rho} = 1$ but $\hat{\theta} = .1$. As can be seen in the figure, the posterior concentrates more rapidly on values near unity as sample size increases. It can be shown analytically that, for large *T* and when $\hat{\rho} = 1$, an approximation to the posterior distribution 'depends only on the product $T \times \hat{\theta}^2$. Indeed, the posterior for the 1000-year sample size (no. 7) in Figure 3A is very similar to that for the 100-year sample size (no. 5) in Figure 3B, and $\hat{\theta}^2$ is about 11 times greater in the latter case.

Figures 3C to 3H display the posteriors for the other values of $\hat{\rho}$, where portfolio *p* possesses various degrees of inefficiency in the sample. As the "uninformative" prior for ρ is concentrated on low absolute values, one may expect that, for each sample size *T*, the concentration of the posterior around the sample statistic $\hat{\rho}$ will occur more rapidly for lower absolute values of $\hat{\rho}$. While this phenomenon is exhibited in the figures, one must nevertheless be struck by how slowly the posterior changes as a function of $\hat{\rho}$, especially for $\hat{\theta} = 0.03$. For example, the posterior obtained with a 25-year sample (no. 4) in which $\hat{\rho} = 1$ (Figure 3A) is almost identical to that obtained when $\hat{\rho} = 0.7$ (Figure 3C) and is not much different from that with $\hat{\rho} = 0.4$ (Figure 3E).

We observed in Section 3 that, when a riskless asset is included, the posterior distribution for ρ obtained with our sample concentrates around rather low values of ρ (Figure 1A). One might thereby infer that the data indicate that ρ is probably quite low for the value-weighted NYSE-AMEX portfolio. As our investigation here reveals, however, a sample of the same size in which this portfolio is exactly

efficient can still produce a posterior for ρ that has most of its mass below ρ = 0.5. It so happens that our sample produces a value of $\hat{\rho}$ (0.1) that lies well within the fairly concentrated marginal prior distribution for ρ (or, more precisely, the 15-week "uninformative" distribution). Thus, it becomes difficult in such a case to assess the role played by the data in forming our posterior beliefs. The analysis Using alternative (informative) priors in Section 5 will shed additional light on this issue.

4.2 Interpreting posteriors in the regression framework

As shown by Harvey and Zhou (1990), a marginal posterior distribution for the inefficiency measure λ, the difference in squared Sharpe measures defined in (4), can be obtained in a regression framework. Let $r_{p,t}$ denote the return on portfolio *p* in excess of the return on a riskless asset, and let r_t contain the excess returns on the other n - 1 assets. Define the multivariate regression,

$$r_t = \alpha + \beta r_{p,t} + u_t, \qquad t = 1, \ldots, T, \tag{15}$$

and assume that u_t is distributed multivariate normal, independently across *t*, with mean 0 and variance-covariance matrix Σ. The standard diffuse prior for the multivariate regression model, used by Harvey and Zhou, is given by,

$$p(\alpha, \beta, \Sigma) \propto |\Sigma|^{-\frac{n}{2}}. \tag{16}$$

The inefficiency measure λ can be expressed in terms of α and Σ [Gibbons, Ross, and Shanken (1989)],

$$\lambda = \alpha' \Sigma^{-1} \alpha, \tag{17}$$

and the joint posterior distribution of (α, Σ) can be obtained analytically. From this distribution, Harvey and Zhou draw independent realizations of (α, Σ) and compute the histogram for λ using (17), following essentially the same Monte Carlo method described in the previous section.

Shanken (1987b) and Harvey and Zhou (1990) conduct Bayesian analyses using λ, but they turn to ρ in order to develop intuition. As noted earlier, the link between λ and ρ. in Equation (5) includes θ, the Sharpe measure of portfolio *p*. Since θ is a function of μ_p and σ_p, and those, two parameters are not included in the Bayesian regression framework, that framework cannot provide the marginal posterior distribution for ρ. Harvey and Zhou (1990) obtain a posterior for ρ from the posterior for λ simply by plugging in a numerical value

for θ and then in effect, changing variables from λ to ρ. With the diffuse prior in (16), it can be shown that λ obeys the same marginal posterior distribution given in Section 3, except that the central chi-square distribution in (13) his T - 2 instead of T - 1 degrees of freedom (see Appendix). Recall from Section 3 that, when μ_p and σ_p are included as unknown parameters, λ and θ are independent in their joint posterior distribution. Thus, it immediately follows that the plug-in approach employed by Harvey and Zhou actually provides the *conditional* posterior distribution of ρ given θ [except for the previously noted difference between T - 1 and T - 2 degrees of freedom for the chi-square in (13)].

From the discussion in Section 3, it follows that the statistics $\hat{\rho}$ and $\hat{\theta}$ are sufficient to determine the conditional distribution of ρ given θ. To obtain a draw of ρ in constructing this conditional distribution, a draw for v is obtained from (14) and then used to draw λ from the distribution in (12). The value of ρ is then obtained from (5) using the given value for θ.

We compute here the conditional posterior of ρ given θ for two values of $\hat{\rho}$, 1.0 and 0.4, and the same two values for $\hat{\theta}$ considered previously, 0.03 and 0.1. In each case, the posterior for ρ is conditioned on $\theta = \hat{\theta}$. We also use the same seven sample sizes (T) as above. Figures 4A to 4D display, for each of the four combinations of $\hat{\rho}$ and $\hat{\theta}$, the conditional posterior distributions of ρ given θ across the seven sample sixes. In all four figures, the "uninformative" 15-week sample produces a posterior for ρ that is very concentrated on low values. As the sample size increases, the posterior for ρ spreads out before finally concentrating close to $\hat{\rho}$. As we also observed for the unconditional posteriors of ρ, the rate of convergence is higher for higher values of $\hat{\theta}$ and lower absolute values of $\hat{\rho}$. Comparing the conditional and unconditional posteriors in Figures 3 and 4, however, reveals that including the uncertainty in θ increases the dispersion in the posterior of ρ for all sample sixes, but the increases in dispersion are dramatic for the smaller samples. It seems that ignoring uncertainty about θ increases the potential for misinterpretating concentrated posteriors as reflecting informative samples.

The effect that a nonlinear transformation can have on the dispersion of a distribution can be illustrated by considering the conditional posterior for ρ in light of the posterior distribution of λ. It can be verified that the behavior of the posterior distribution of λ across samples of increasing size conforms to the conventional intuition. That is, the posterior mass for λ is disperse for small samples and then concentrates toward the statistic $\hat{\lambda}$ $(\equiv \hat{\gamma}^2 - \hat{\theta}^2)$ as sample size increases. Note from the expression for ρ in (5) that λ appears in the denominator

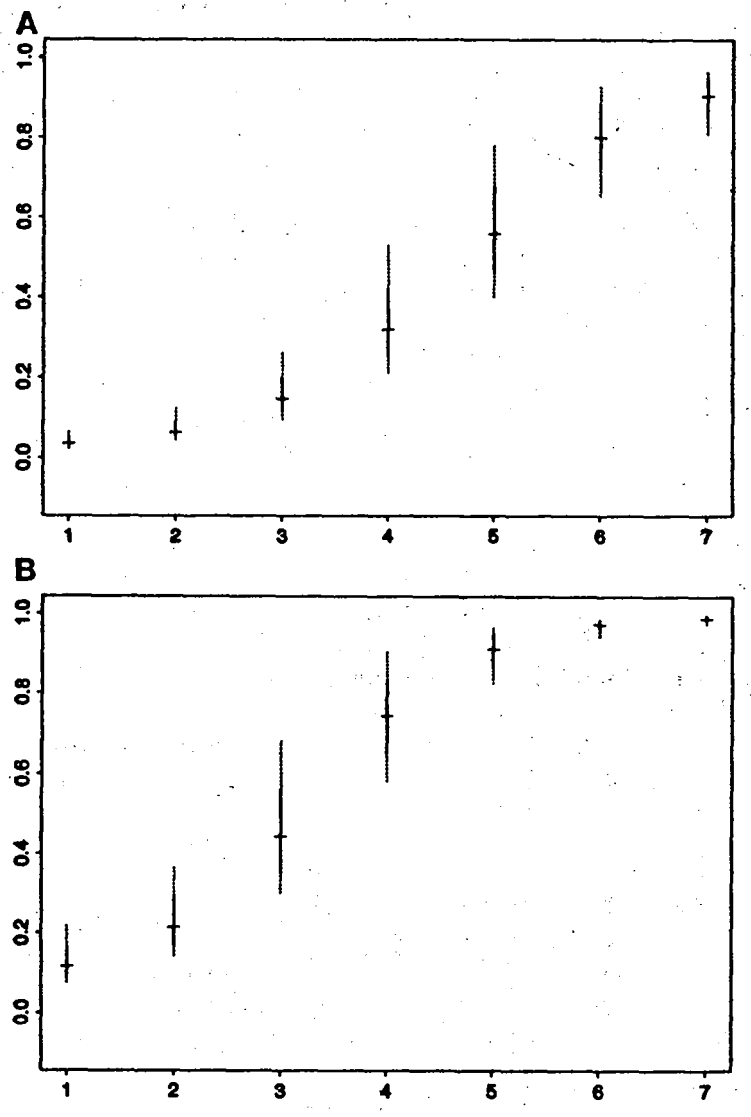

Figure 4
Conditional posterior distributions of ρ given θ, using a diffuse prior, for various sample sizes and values of the sufficient statistics
Each of Figures 4A through 4D displays, for given values of the sufficient statistics $\hat{\rho}$ and $\hat{\theta}$, the conditional posterior distributions of the inefficiency measure ρ given $\theta = \hat{\theta}$, where a riskless asset is included, for seven different sample sizes. The numbers 1 through 7 on the horizontal scale in each figure correspond to sample sizes of 15 weeks, 1 year, 5 years, 25 years, 100 years, 400 years, and 1000 years. For each sample size, the figure displays the median (horizontal tick) and the 1 percent, 25 percent, 75 percent, and 99 percent quantiles of 5000 draws from the posterior distribution. The solid line connects the 25 percent and 75 percent quantiles, and the dotted lines extend to the 1 percent and 99 percent quantiles. **A:** $\hat{\rho} = 1.0, \hat{\theta} = 0.03$. **B:** $\hat{\rho} = 1.0, \hat{\theta} = 0.10$.

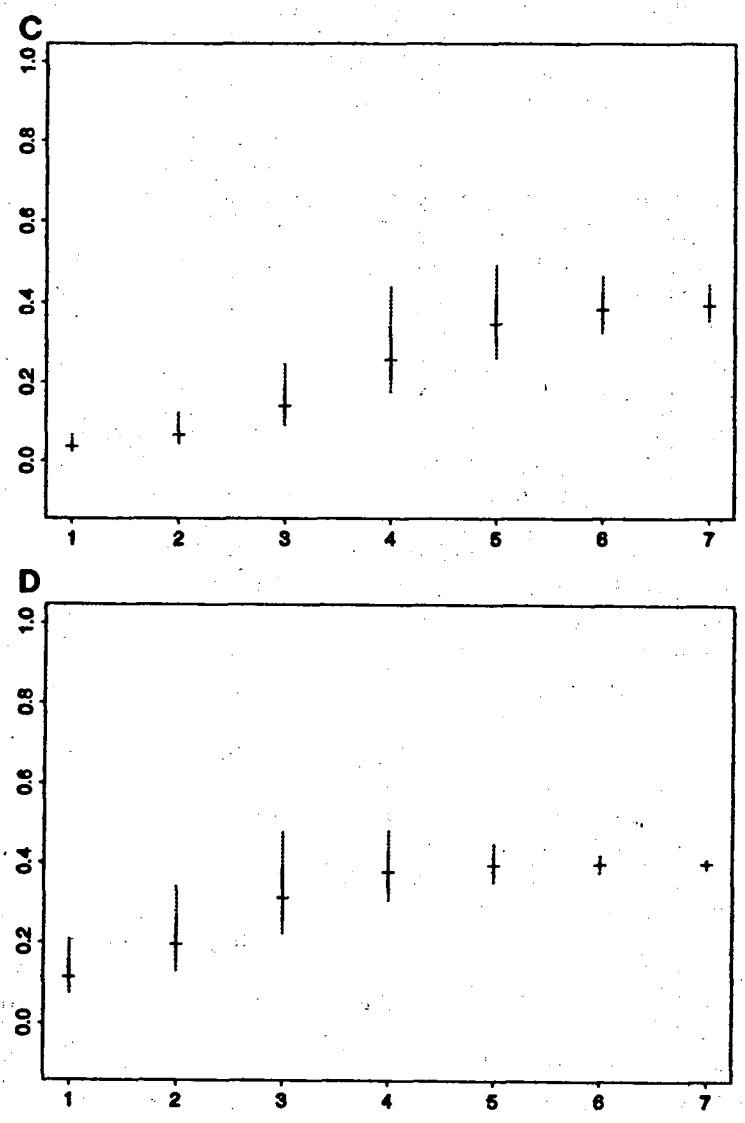

Figure 4
Continued.
C: $\hat{\rho} = 0.4$, $\hat{\theta} = 0.03$. D: $\hat{\rho} = 0.4$, $\hat{\theta} = 0.10$.

and is multiplied by $1/\theta^2$. The values assumed by the latter quantity are likely to be much larger than unity—even the largest "conceivable true value" for θ used by Harvey and Zhou (1990) implies that $1/\theta^2$

would be roughly 50.[7] Thus, in a very small sample, the transformation from λ to ρ produces, for a constant value of θ, a variable whose distribution concentrates near zero.

The results displayed in Figure 4 provide an interesting setting in which to review the conclusions reached by Harvey and Zhou (1990). They observe that their plug-in posterior distribution for ρ is concentrated away from unity, and they assign less than a 1 percent probability that ρ is greater than 0.9 for the value-weighted NYSE portfolio. As demonstrated here, such an outcome would occur even if the portfolio were exactly efficient in the sample.

4.3 The posterior for ρ in the absence of a riskless asset

When a riskless asset is not included, we must specify all of the elements of $\hat{E}$ and $\hat{V}$ in order to compute the posterior distribution of ρ. Thus, unlike the case with a riskless asset, we are unable to express the posterior of ρ in terms of a few univariate statistics and thereby present a complete characterization of the role played by the data. We instead focus here on the role of sample size (T) when $\hat{\rho} = 1$. That is, we investigate the maximum degree to which a given sized sample' can move the posterior distribution of ρ toward unity. This line of inquiry is guided by our previous observation that, when a riskless asset is included, large sample sizes are required to form posterior beliefs that ρ is close to unity.

In constructing the posterior for our actual sample, displayed in Figure 1B, the sufficient statistics $\hat{E}$ and $\hat{V}$ correspond to the unrestricted maximum likelihood estimates of E and V for the multivariate normal model. In this section, since we wish to analyze samples in which portfolio p is exactly efficient, we instead set the sufficient statistics equal to *restricted* maximum likelihood estimates for our sample. These estimates satisfy the restriction that the value-weighted NYSE-AMEX index is efficient. Otherwise, the method used here to construct the posterior for ρ is identical to that described in Section 3.

Figure 5 displays the posteriors for ρ for the seven sample sizes considered earlier. The uninformative 15-week sample produces a posterior with a median around 0.5 and an interquartile range between 0.4 and 0.6. Thus, the uninformative sample produces a posterior that is again Fairly concentrated around lower values of ρ, which might suggest problems of interpretation similar to those discussed in the case when a riskless asset is included. An examination of the other sample sixes, however, makes it readily apparent that excluding the riskless asset produces dramatic differences in the behavior of the

[7] Their monthly value for θ is halved to obtain a weekly value.

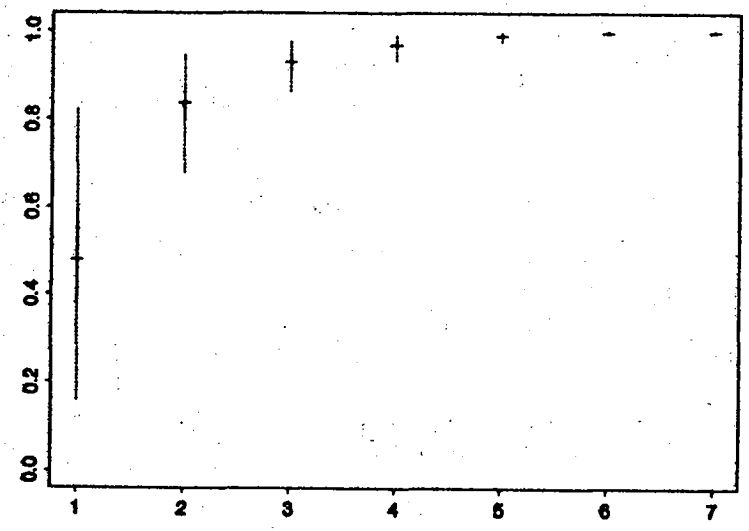

Figure 5
Posterior distributions of ρ when a riskless asset is excluded, using a diffuse prior, for various sized samples in which the portfolio is exactly efficient
The numbers 1 through 7 on the horizontal scale in each figure correspond to sample sizes for 15 weeks, 1 year, 5 years, 25 years, 100 years, 400 years, and 1000 years. For each sample size, the figure displays the median (horizontal tick) and the 1 percent, 25 percent, 75 percent, and 99 percent quantiles of 5000 draws form the posterior distribution. The solid line connects the 25 percent and 75 percent quantiles, and the dotted lines extend to the 1 percent and 99 percent quantiles.

posterior of ρ. The posterior quite rapidly concentrates on values near unity as sample size increases. In the 25-year sample, for example, virtually all of the mass for ρ lies above 0.9.

We see that, in contrast to the case where a riskless asset is included, modest sample sizes allow one to arrive at posterior beliefs supporting high values of ρ, even though the marginal prior for ρ concentrates around low values. Indeed, Figure 1B demonstrates that such posterior beliefs are possible, but that posterior, viewed by itself, does not reveal the role of the data in forming those beliefs. Given the contrast between the prior and posterior shown here, we can better assess the substantial influence of the data in this case.

4.4 The posteriors for Δ

Computing the posterior distribution of Δ, with or without a riskless asset, requires that we specify the elements of $\hat{E}$ and $\hat{V}$. Therefore, we again focus our investigation on the role of sample size when $\hat{\rho} = 1$. As in the previous subsection, the values of $\hat{E}$ and $\hat{V}$ are set to the restricted maximum likelihood estimates for our actual sample.

Figures 6A and 6B display the posteriors for Δ with and without the riskless asset. First note that the posteriors for the 15-week "uninformative" sample are not shown in either case. Those posteriors are so disperse that increasing the scale to accomodate them would remove almost all of the visible dispersion from the posteriors for sample sixes 2 through 7. Therefore, the posteriors of Δ based on our actual sample, reported earlier in Figures 2A and 2B, differ substantially from those produced by an "uninformative" sample. Recall that, when the riskless asset is included, the posteriors of ρ in both 15-week and 25-year samples in which $\hat{\rho} = 1$ are concentrated around low values, so it is more difficult to gauge the role played by our actual sample in forming posterior beliefs about that inefficiency measure.

As the sample size increases, the posteriors for Δ converge toward zero, although the convergence is faster when the riskless asset is excluded. For the 25-year sample, the 75 percent quartile falls at Δ = 8 percent with the riskless asset but at Δ = 2 percent without the riskless asset. The posterior with the riskless asset is more disperse for the 100-year sample (no. 5) than is the posterior without the riskless asset for the 25-year sample. As observed earlier for ρ, it appears that excluding the riskless asset increases's sample's ability to support an inference that portfolio *p* is nearly efficient, even when that portfolio is exactly sample efficient in both cases. We cannot supply a formal explanation for why it is that excluding the riskless asset produces this result, but we can offer some simple intuition. Suppose we add the same constant to the expected return on every risky asset (i.e., make a vertical shift in the minimum-standard-deviation boundary of the risky assets). When the riskless asset is excluded, neither ρ nor Δ is affected by such a shift. When a riskless asset is included, however, shifting the means of the risky assets relative to the observed riskless return will change both ρ and Δ. Thus, uncertainty about a common additive scale factor in expected returns on the risky assets does not contribute to uncertainty about ρ or Δ when the riskless asset is excluded, but the same is not true when the riskless asset is included.

5. Posterior Distributions with Informative Priors

The posterior distributions analyzed in Sections 3 and 4 employ the improper prior, $p(E, V) \propto |V|^{-(n+1)/2}$. The general appeal of, this prior stems from its presumed noninformative property and the close correspondence between the resulting posterior distributions and classic confidence regions for the parameters *E* and *V*. In Section 4, we observed that a posterior for the inefficiency measure ρ can be tight and yet be influenced very little by sample information, even though the prior is "diffuse." Further, the posterior can be centered

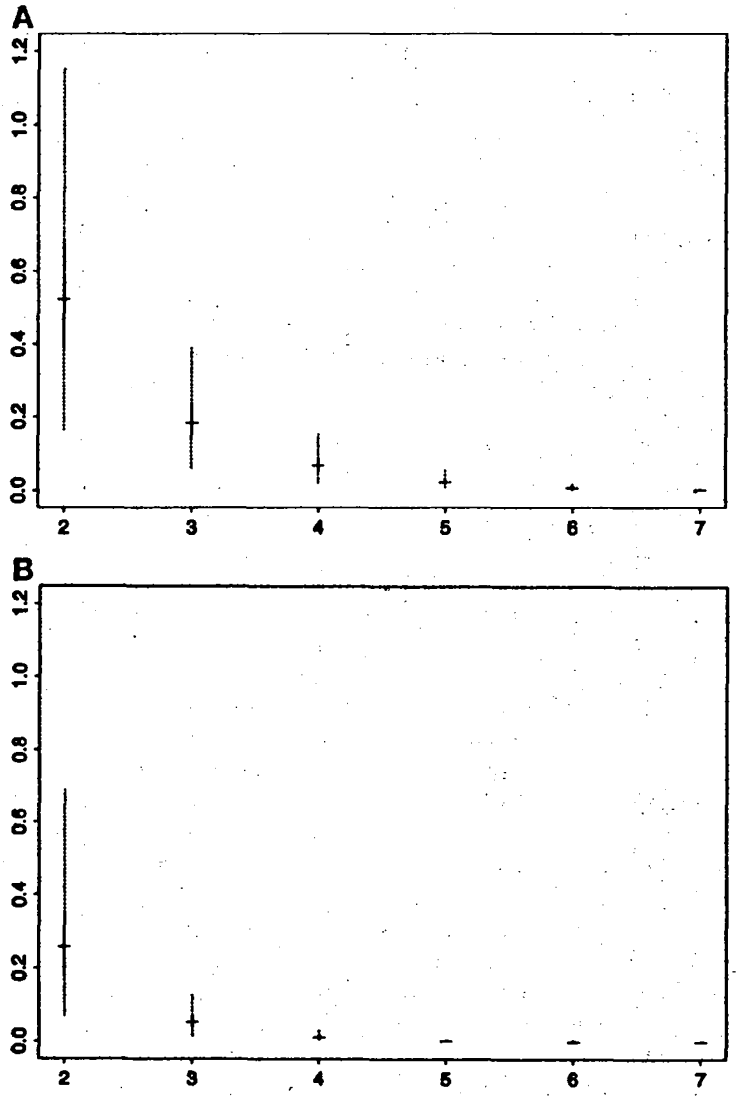

Figure 6
Posterior distributions of Δ, using a diffuse prior, for various sized samples in which the portfolio is exactly efficient
Figure 6A displays the posteriors of Δ when a riskless asset is included, and Figure 6B displays the posteriors in the absence of the riskless asset. The numbers 2 through 7 on the horizontal scale in each figure correspond to sample sizes of 1 year, 5 years, 25 years, 100 years, 400 years, and 1000 years. For each sample size, the figure displays the median (horizontal tick) and the 1 percent, 25 percent, 75 percent, and 99 percent quantiles of 5000 draws from the posterior distribution. The solid line connects the 25 percent and 75 percent quantiles, and the dotted lines extend to the 1 percent and 99 percent quantiles.

far from the maximum likelihood estimate. It is well known [see Box and Tiao (1973)] that a noninformative prior can be informative on certain marginals, and ρ provides a striking example, especially when a riskless asset is included. With a diffuse prior, therefore, it is more difficult to comprehend the effect of the data on the posterior of ρ than on the posterior of parameters such as *E*.

In this section we construct proper prior distributions for the model parameters *E* and *V*. Various priors are constructed, each one implying a marginal prior for ρ that reflects beliefs about the inefficiency of portfolio *p*. We then gauge the data's information about ρ by comparing the various priors with their corresponding posteriors. When the goal of a study is quantifying the amount of information about a hypothesis for a general audience, rather than making a particular decision, the desirability of considering a variety of priors has been noted in the Bayesian literature. [See Poirier (1992) for a lucid discussion.]

We construct prior distributions for the model parameters such that the range of marginal priors for ρ includes those (i) concentrated near zero, (ii) disperse between zero and one, and (iii) concentrated near one. In the application considered here, the last specification would correspond to beliefs that the value-weighted NYSE-AMEX index is fairly highly correlated with an efficient portfolio. Such prior beliefs might be formed, for example, by a consideration of asset pricing theory and the view that this stock market index is a reasonable market proxy. We find that, when a riskless asset is included, seemingly strong prior beliefs about the model parameters are required to obtain a marginal prior for ρ that concentrates near one. This seems to be consistent with our previous finding in Section 4 that, when the noninformative prior is used, large samples are needed to obtain posteriors for ρ with substantial mass near one.

When the riskless asset is excluded, we find that prior beliefs about the model's parameters need not be as strong in order to obtain a marginal prior for ρ that concentates near one. This observation also seems to be consistent with the results in Section 4, where we found that, with the noninformative prior, smaller samples can produce posteriors for ρ that concentrate near 1. Because the models with and without a riskless asset behave so differently, and for technical reasons that will become apparent below, we consider different types of priors for the different models. We also restrict our attention to the inefficiency measure ρ in order to keep things concise. It appears in any event that the inefficiency measure Δ is less problematic, perhaps because it is not bounded both above and below, so that large amounts of uncertainty can be represented by a disperse prior or posterior.

5.1 Informative priors when a riskless asset is included

5.1.1 Specification of the prior distribution. Our goal is to formulate a class of priors capable of reflecting various degrees of belief about ρ. Moreover, we wish to avoid priors that imply unnecessarily strong beliefs about all of the other model parameters. That is, we wish to control, for any given ρ*, the prior probability of the set $\{(E, V) : \rho > \rho^*\}$ without placing high probability on subsets of (E, V) within that set. The results of Section 4 suggest that the latter objective is easily satisfied when the desired marginal prior for ρ is concentrated near zero, since vague beliefs about (E, V) evidently imply strong beliefs that ρ is close to zero.

As one increases the desired prior probability assigned to values of ρ near 1, it becomes increasingly difficult to satisfy the above objective by specifying a prior directly for *E* and *V*. We instead transform this set of parameters to a set consisting α, β, and Σ, the parameters of the regression framework defined earlier, plus μ_p and σ_p^2, the mean and variance of portfolio *p:*

$$E \longrightarrow \begin{bmatrix} \alpha + \beta\mu_p \\ \mu_p \end{bmatrix},$$

$$V \longrightarrow \begin{bmatrix} \Sigma + \beta\beta'\sigma_p^2 & \beta\sigma_p^2 \\ \beta'\sigma_p^2 & \sigma_p^2 \end{bmatrix}. \tag{18}$$

This parameterization allows the prior mass of ρ to be placed as close to one as desired by concentrating the prior of α around zero and assigning all of the prior mass of μ_p to positive values. At the same time, the priors for the parameters other than α can remain spread out over a wide range of values, although we specify the priors to be proper. We will construct four priors that differ only in the degree to which the marginal prior for a concentrates around zero.

The joint prior distribution of $(\alpha, \beta, \Sigma, \mu_p, \sigma_p^2)$ is specified as follows. The pammeters β, Σ, μ_p, and σ_p^2 ;are independent with marginal distributions:

$$\beta \sim N(\bar{\beta}, \sigma_\beta^2 I_{n-1}), \tag{19}$$

$$\Sigma^{-1} \sim W(S_0^{-1}, \nu_\Sigma), \tag{20}$$

$$\mu_p \sim N(\bar{\mu}_p, \sigma_{\mu_p}^2)\, I[0, .2/52], \tag{21}$$

$$\sigma_p^2 \sim \frac{\nu_o \lambda_o}{\chi_{\nu_o}^2}. \tag{22}$$

Here $W(S_0^{-1}, \nu_\Sigma)$ is the Wishart distribution with parameter matrix S_0^{-1} and ν_Σ degrees of freedom such that $E(\Sigma^{-1}) = \nu_\Sigma S_0^{-1}$. The distribution of μ_p is proportional to a normal, truncated so that its support lies in the interval [0, .2/52].

The conditional distribution of α given μ_p is

$$\alpha \mid \mu_p \sim N(0, \sigma_\alpha^2 I_{n-1}), \tag{23}$$

where

$$\sigma_\alpha = \phi_0 + \phi_1 \mu_p. \tag{24}$$

We see from (5) and (17) that, if λ is, independent of θ, then smaller θ values will be associated with smaller ρ values. The specification in (24), where $\phi_1 > 0$, reduces this prior dependence between ρ and θ by allowing the prior standard deviation of the elements of α to increase with μ_p, thereby associating smaller α vectors (which are positively correlated with λ) with smaller values of μ_p (which are positively correlated with θ). One can also view (24) as associating higher absolute "mispricings" with higher overall levels of expected returns.

In constructing the priors for μ_p and σ_p^2, we set $\bar{\mu}_p$ = .07/52, σ_{μ_p} = .04/52, ν_o = 20, and $\lambda_o = .2/\sqrt{52}$. The resulting 5 percent and 95 percent quantiles are .0446/52 and .1718/52 for μ_p and .022 and .0374 for σ_p.

To aid us in the choice of S_0, we use a sample of monthly returns on NYSE size-ranked portfolios for the period 1926 through 1962 to compute $\hat{\Sigma}$, the maximum likelihood estimate of Σ subject to the restriction that $\alpha = 0$. We set $S_0 = \nu_\Sigma \hat{\Sigma}/4$, where division by four converts from monthly to weekly magnitudes. We then set ν_Σ = 20, so the prior information about Σ is comparable to 20 weeks of data.

Since the prior distribution of ρ does not depend on β, we let the prior for β be very disperse by setting σ_β = 1000. We set $\bar{\beta}$ to be a vector of ones, although the large σ_β makes the choice of $\bar{\beta}$ irrelatively unimportant.

Finally, we turn to the choices of ϕ_0 and ϕ_1 that determine the conditional prior for α in (23) and (24). We select four pairs of values for ϕ_0 and ϕ_1, where each pair is chosen to achieve a desired "average" value of σ_α as well as approximate independence between ρ and θ over a fairly wide range for 8. We report below our choices for ϕ_0 and ϕ_1 and the values of σ_α that they imply for μ_p = .1/52, which is approximately the prior mean of μ_p. (All values of ϕ_0 and ϕ_1 are multiplied by 100.)

ϕ_0	ϕ_1	σ_α for $\mu_p = .1/52$
0	25	.025/52
.0064	1.667	.005/52
.0019	0.500	.0015/52
.0013	0.333	.0010/52

The implied choices of σ_α include values that may seem quite small, implying very precise prior beliefs about α, but the reason for such choices will become apparent below.

5.1.2 Results. Figures 7A through 7D display the four marginal prior distributions for ρ as well as the posterior distributions based on our sample of 1304 weekly excess returns. The dotted line in each figure is the prior, and the solid line is the posterior. The computational algorithm used to obtain the posterior distributions is given in the Appendix.[8]

The prior in Figure 7A reflects the choices of ϕ_0 and ϕ_1 giving the largest values for σ_α, whereas the prior in Figure 7D reflects the choices giving the smallest values. As the prior on α becomes tighter around zero, the prior mass of ρ shifts to values closer to one. The four priors have been chosen to cover the possibilities of interest. The first prior is concentrated on low values of ρ and has a prior mean of .19. The second prior is spread out over the interval from zero to one with a prior mean for ρ of .47. The third prior is concentrated on ρ values greater than .8 but with a heavy left tail, and the prior mean is .83. Finally, the fourth prior is even more concentrated on large ρ values and has a prior mean of .91.

Clearly, the data shift the last three priors (Figures 7B, 7C, and 7D) to posteriors for ρ that put more mass on smaller values. For the first prior (Figure 7A), even though the posterior mode is to the right of the prior mode, the posterior is tighter in a region of quite small ρ values. This posterior, obtained with a proper but disperse prior for the elements of α, is very similar to that obtained earlier using the improper diffuse prior (Figure 1A). Perhaps the most important message conveyed by the collection of Figures 7A through 7D is that, if the four priors represent a reasonable range of prior beliefs about ρ, the data do not contain enough information to make the posteriors converge to a common opinion.

The results with the second prior (Figure 7B) are probably the most dramatic: the posterior is much more concentrated than the prior and is supported almost entirely by ρ values less than .5. After observing

[8] The plotted distributions are obtained by applying a kernel smoother to a set of draws from the true distribution. A transformation is applied to insure that the plotted distributions do not extend beyond the upper limit (1.0) of the true distribution.

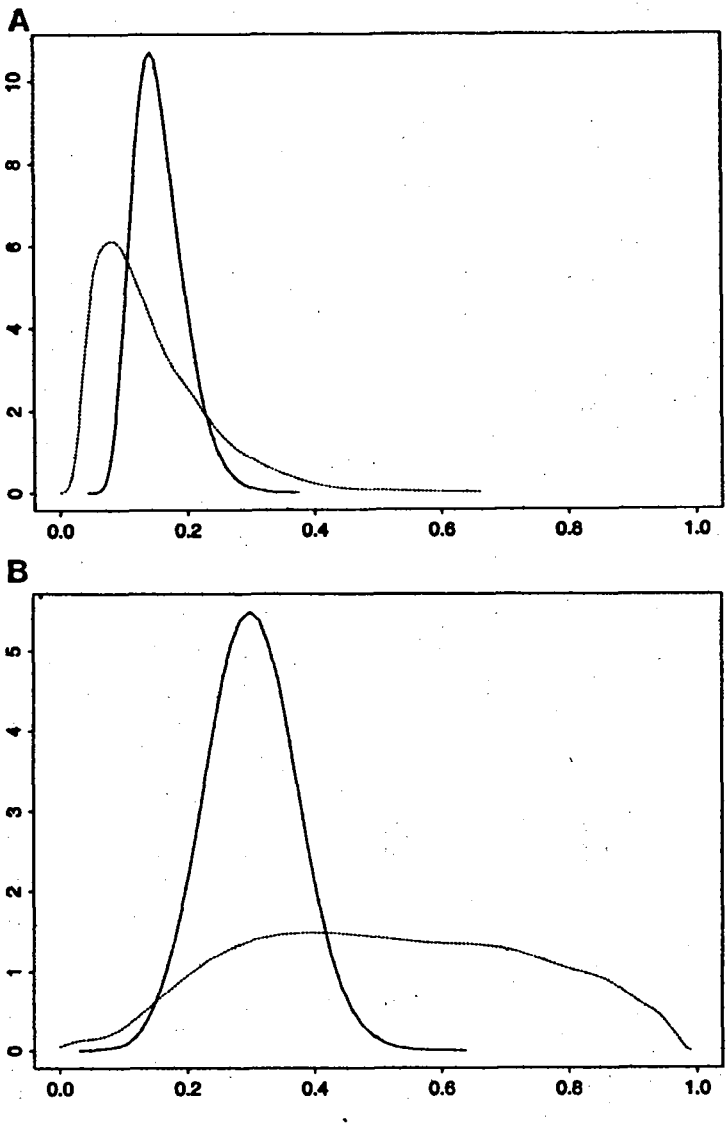

Figure 7
Posterior distributions of ρ when a riskless asset is included
Each of Figures 7A through 7D displays a marginal prior distribution for ρ (dotted line) and the corresponding posterior distribution (solid line) obtained from a sample of weekly returns from January 1963 through December 1987 for a set of assets consisting of 10 size-ranked portfolios, the equally weighted NYSE-AMEX portfolio, and the value-weighted NYSE-AMEX portfolio, whose inefficiency is characterized by ρ. All returns are in excess of the return on a one-week Treasury bill.

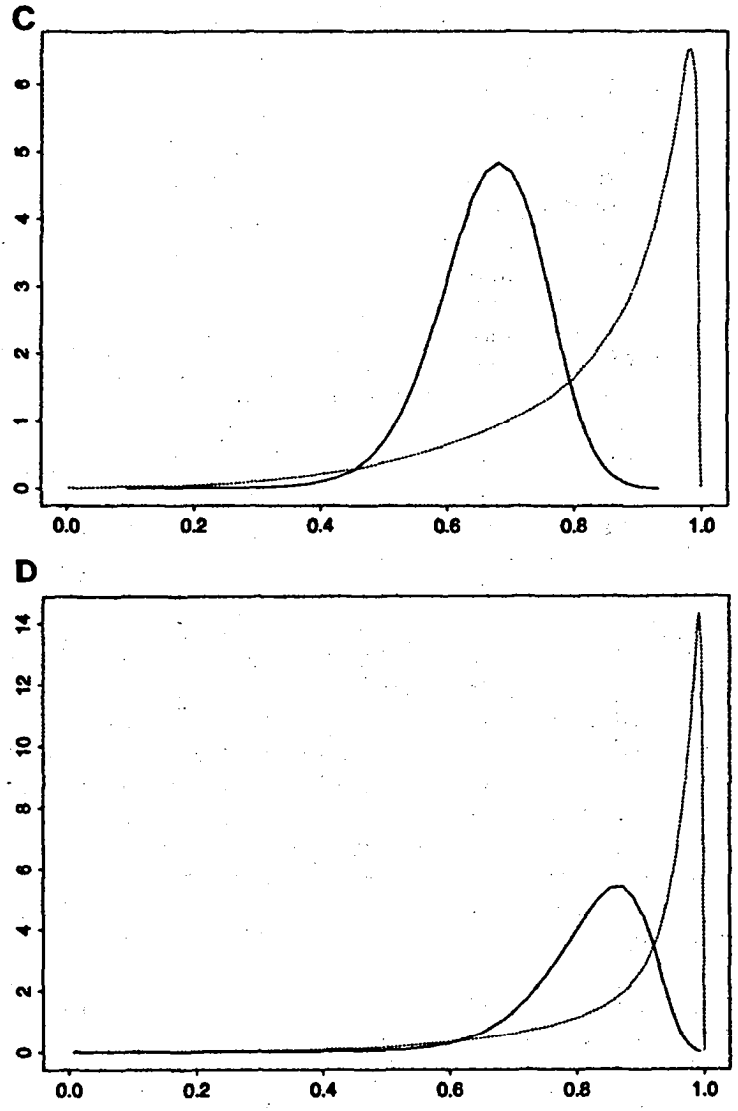

Figure 7
Continued.

this case, it seems reasonable to conclude that our data do in fact exert an important influence on the posterior distribution of ρ and thereby indicate strongly that the value-weighted NYSE-AMEX index is inefficient. Although the marginal prior for ρ in this case is disperse, it is important to realize that, to achieve this dispersion, the underlying

prior for α must be quite concentrated around zero. When μ_p is close to its prior mean, for example, the prior standard deviation of the elements of α is only 0.5 percent on an annual basis.

The third prior (Figure 7C) assigns roughly 50 percent probability to the hypothesis that $\rho > 0.9$. Shanken (1987b) specifies the same prior probability for this composite hypothesis by directly assigning prior probabilities to 13 discrete values of ρ under the null and the alternative. Conditioning on a value for θ, Shanken computes a posterior odds ratio and concludes that the posterior probability of $\rho \geq 0.9$ is 20 percent. Our distribution in Figure 7C indicates a very small posterior probability that $\rho \geq 0.9$. (As explained earlier, the posterior distribution of ρ allows one to compute easily the odds ratio $\Pr\{\rho \geq .9\}/\Pr\{\rho < .9\}$.) Unlike Shanken, we specify a prior distribution for the basic moments of the multivariate return distribution and do not condition on θ. Our data differ as well, in that Shanken uses monthly returns on NYSE stocks combined by industries while we use weekly returns on size-ranked portfolios and include the AMEX stocks. Nevertheless, we reach a similar conclusion: if one assigns substantial prior probability to the belief that the value-weighted stock portfolio is nearly efficient ($\rho \geq 0.9$), this belief is weakened considerably by the data. Note, however, that the posterior probability for, say, $\rho \geq 0.7$ is substantial. (Shanken does not report posterior probabilities for other composite hypotheses.)

As noted earlier, the choices of ϕ_0 and ϕ_1 used in constructing the priors, especially those in Figures 7C and 7D that place most of the mass on high values of p, imply marginal priors on the elements of a that are rather concentrated. An alternative approach, suggested by MacKinlay (1993), is to specify a more concentrated marginal prior for the inefficiency measure, λ, the amount by which the maximum squared Sharpe measure exceeds that of portfolio *p*. MacKinlay shows that, by incorporating the dependence between α and Σ in a manner that maintains sufficient prior mass on low values of $\lambda\ (= \alpha'\Sigma^{-1}\alpha)$, one can specify less concentrated marginal priors on the elements of a but still obtain a prior on ρ similar to the marginal prior in Figure 7C.[9] It remains for future research to investigate whether such a prior would produce a posterior for ρ substantially different from that shown in Figure 7C in samples similar to our 12-asset, 25-year data set.

Recall that Section 4 analyzed posterior distributions obtained from hypothetical samples in which $\hat{\rho} = 1$. Figures 8A through 8D display the results of a similar exercise using the four informative priors rather

[9] MacKinlay conditions on a value for θ in constructing his priors, but a similar approach could be pursued in an unconditional setting.

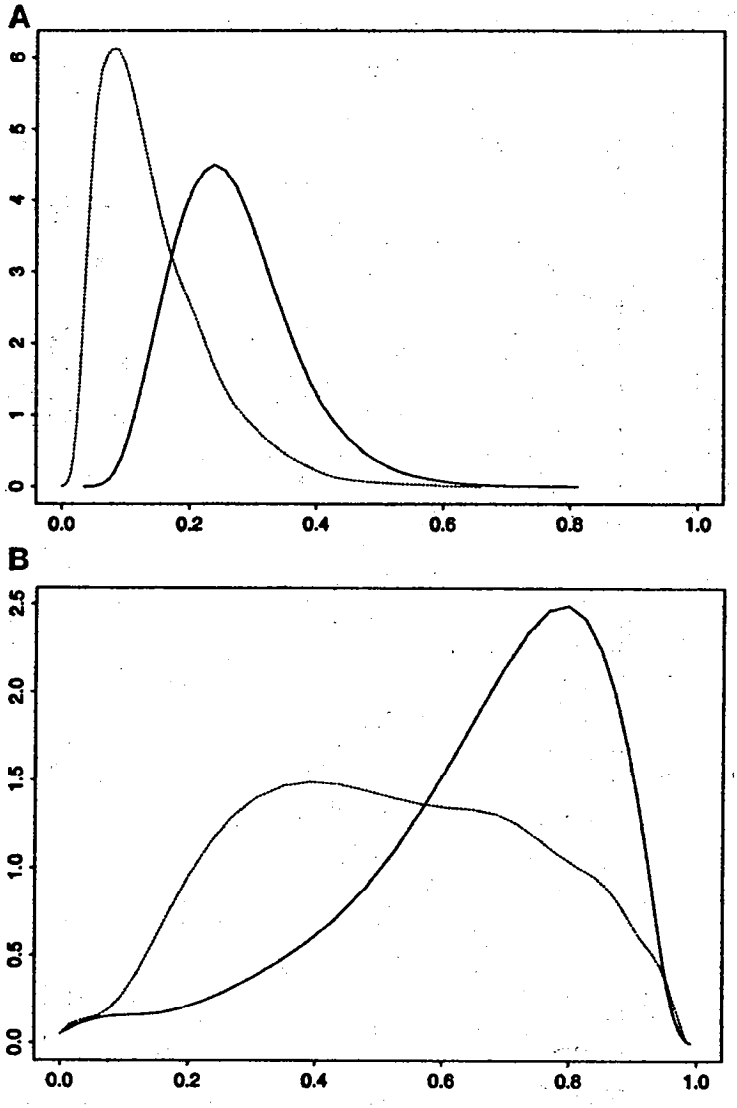

Figure 8
Prior and posterior distribution of ρ when a riskless asset is included and the portfolio is exactly efficient in the sample
Each of Figures 8A through 8D displays a marginal prior distribution for *r* (dotted line) and the corresponding posterior distribution (solid line) obtained from a hypothetical 25-year sample of weekly excess returns in which the portfolio of interest is exactly efficient.

than the improper prior used previously. For this analysis, we compute maximum likelihood estimates of the model parameters subject to the restriction that $\hat{\rho}$ =1 (or $\alpha = 0$). As explained in the Appendix,

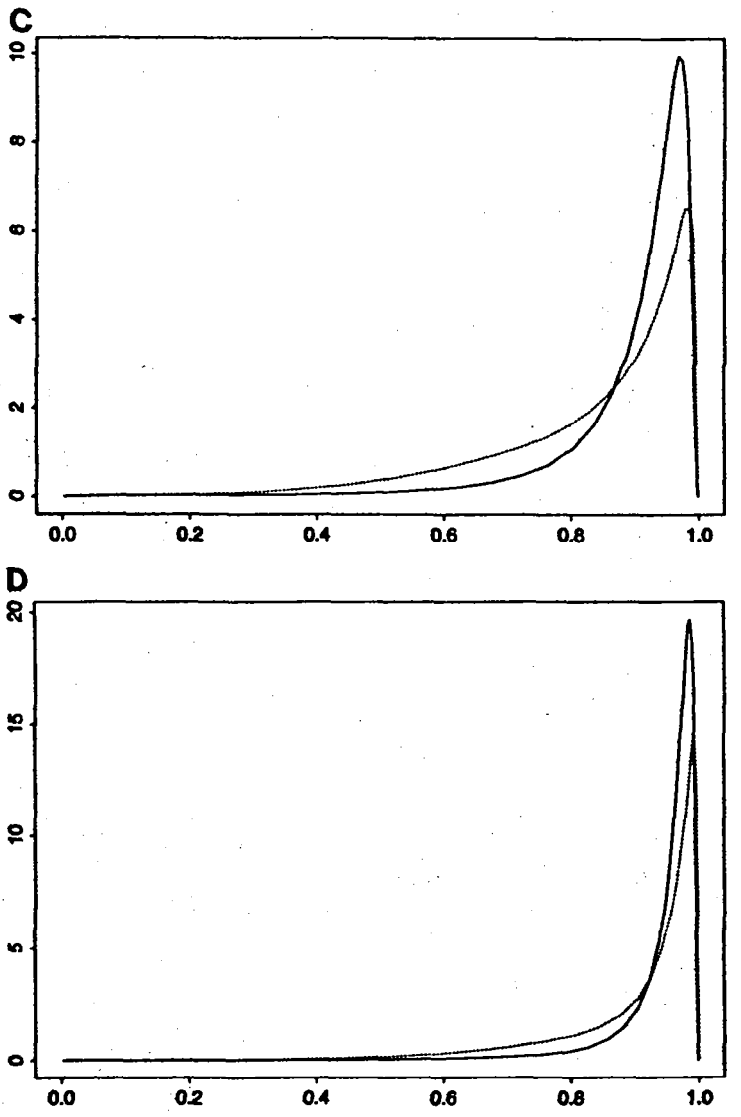

Figure 8
Continued.

the posterior distributions can be obtained using maximum likelihood estimates of the model parameters (as sufficient statistics for the data), so the *restricted* estimates will give the posterior for a hypothetical sample of the same size as our actual sample but where portfolio p is exactly efficient.

With the hypothetical data set in which $\hat{\rho} = 1$, the posteriors for ρ move toward one, as expected. Again, though, the data do not provide information that is sufficient to make the posteriors concentrate near 1 for all four priors. A comparison of Figures 7B and 8B reveals an interesting difference between the posteriors. The marginal prior for p, the same in both cases, is spread over a wide range of values. The posterior in Figure 7B, based on the actual sample in which $\hat{\rho} := 0.1$, is significantly tighter than the posterior in Figure 8B, based on the hypothetical sample in which $\hat{\rho} = 1$. It seems that, unless one believes *a priori* that a portfolio is nearly efficient, substantially more data are required to assign high probability to such a hypothesis.

5.2 Informative priors in the absence of a riskless asset

5.2.1 Specification of the prior distribution. When there is no riskless asset, it is difficult to specify a prior on *E* and *V* that allows us to control the concentration of the mass on the set $\{(E, V) : \rho > \rho^*\}$. We can tighten the prior around specific values of *E* and *V* within this set, but this approach is less desirable. On the other hand, we saw in Section 4.3 that the data seem to have a greater influence on the posterior for ρ when the riskless asset is excluded. Thus although our tools are less refined in this case, it seems easier to assess the role of the data in obtaining the posterior distribution.

We specify a prior in which *E* and *V* are independent with the following marginal distributions:

$$E \sim N(\bar{E}, \sigma_E^2 I_n), \tag{25}$$

$$V^{-1} \sim W(S_V^{-1}, \nu_V). \tag{26}$$

To specify the prior completely, we must assign values to $\bar{E}$, σ_E, ν_V, and S_V. Our approach is to obtain a reasonable pair (E, V) in the set {(E, V) : ρ = 1} and then choose σ_E and ν_V to control the prior's concentration around these values. We again use the monthly data from 1926 through 1962 and compute maximum likelihood estimates $\hat{E}$ and $\hat{V}$ subject to the restriction that $\hat{\rho} = 1$. We then set $\bar{E} = \hat{E}/4$ and $S_V = \nu_V \hat{V}/4$, where division by 4 puts the monthly quantities on a weekly basis. The marginal prior for ρ will concentrate near 1. as we increase ν_v and decrease σ_E, because we are then tightening the joint prior on (E, V) around a value such that ρ = 1.

We specify four different combinations of σ_E and ν_V:

σ_E	ν_V
.05/52	20
.05/52	2.52
.005/52	2.52
.0005/52	2.52

The first (σ_E, ν_V) combination is intended to construct a prior that, although proper, is still fairly disperse. The elements of E are independent with standard deviations of .05/52, and the information about V is comparable to that obtained with 20 weeks of data. The information about V reflected in the other three priors is the equivalent of two years of weekly data, and these priors specify progressively smaller standard deviations for the elements of E. Although two years is not large relative to our 25-year sample period, we find that such information about V is sufficient to concentrate the marginal prior for ρ about 1 as we tighten the prior for E.

5.2.2 Results. Figures 9A through 9D display the results based on our sample of 1304 weekly returns for each of the four prior distributions. As before, the dotted line is the prior, and the solid line is the posterior. The Appendix explains the computation of the posteriors.

We see that the first prior (Figure 9A) is disperse over a wide range for ρ. Thus, in contrast to the case where the riskless asset is included, a disperse prior on the parameters of the return distribution does not lead to a concentrated marginal prior on ρ. The posterior is tightly concentrated about .92 and is, in fact, quite similar to that obtained in Section 3 (Figure 1B) using the improper prior. Thus, the difference between the improper diffuse prior and the proper diffuse prior of this section is negligible relative to the information content of the data.

As we progressively tighten the marginal priors in Figures 9B through 9D, the posteriors remain essentially the same, although they are a bit tighter for the last two priors. Therefore, when the riskless asset is excluded, the data seem to be quite informative about ρ, in that the information in our 25-year sample seems to overwhelm the information supplied by the prior. The results here based on the informative prior distributions strengthen the conclusion drawn in the previous sections about the strong role played by the data in supporting the near efficiency of the value-weighted NYSE-AMEX index in the absence of a riskless asset.

6. Conclusions

The inefficiency of a given portfolio p can be investigated in a Bayesian framework by examining posterior distributions of scalar measures of portfolio inefficiency. Two such measures with simple intuitive appeal are Δ, equal to minus the difference in expected returns between portfolio p and an efficient portfolio of equal variance, and ρ, the maximum correlation between the return on p and an efficient portfolio. We obtain posterior distributions for these measures that in-

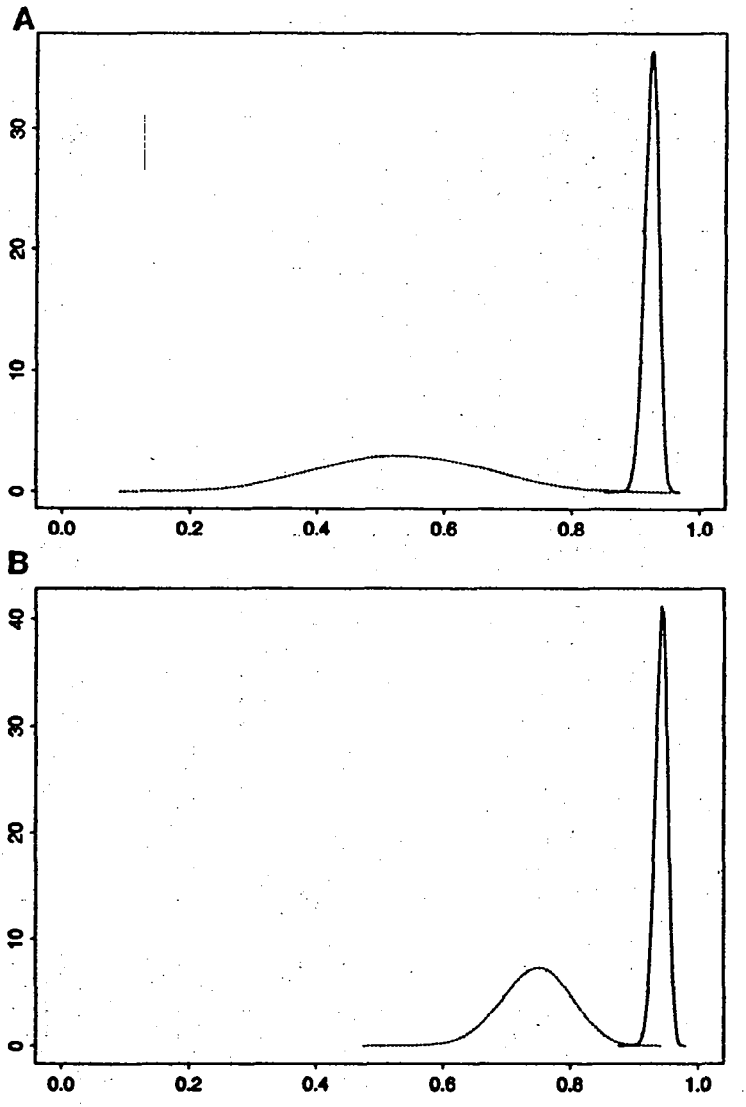

Figure 9
Prior and posterior distributions of ρ when a riskless asset is excluded
Each of Figures 9A through 9D displays a marginal prior distribution for ρ (dotted line) and the corresponding posterior distribution (solid line) obtained from a sample of weekly returns from January 1963 through December 1987 for a set of assets consisting of 10 size-ranked portfolios, the equally weighted NYSE-AMEX portfolio, and the value-weighted NYSE-AMEX portfolio, whose inefficiency is characterized by ρ.

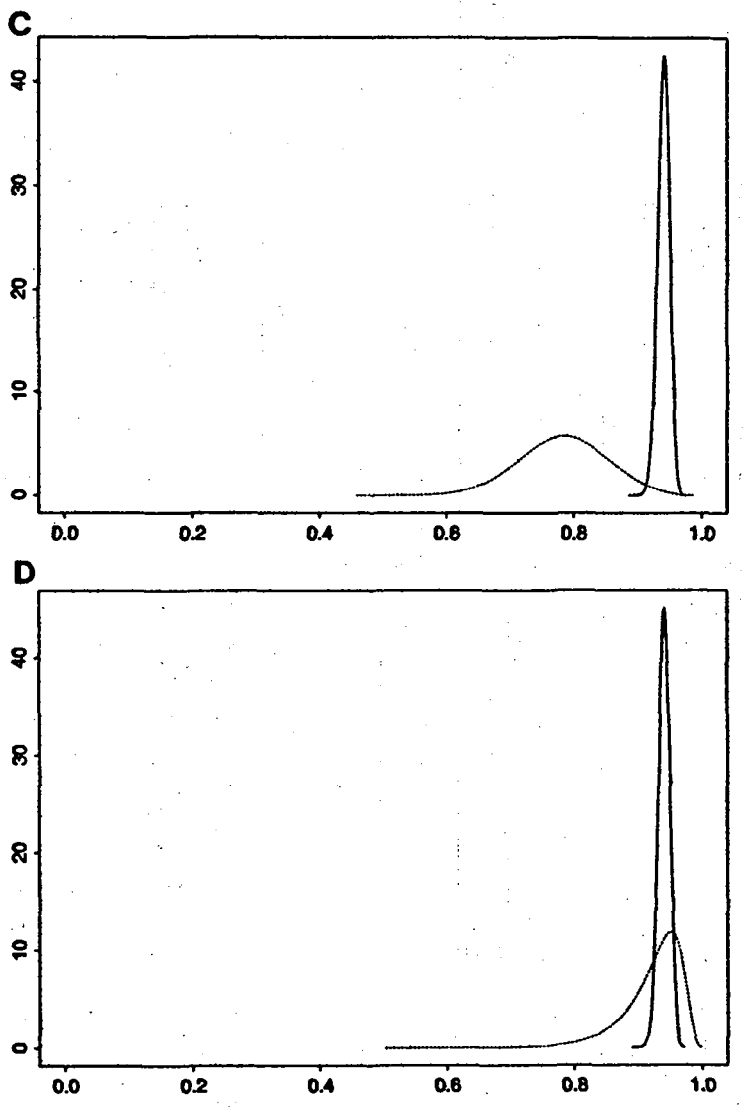

Figure 9
Continued.

corporate the inherent uncertainty about all of the parameters of the joint distribution of asset returns. We demonstrate that ignoring the uncertainty in some of the model's parameters, as is sometimes done in previous studies, can exacerbate problems inherent in interpreting posterior distributions of inefficiency measures.

The inefficiency measures Δ and ρ can be applied whether or not a riskless asset is included in the set of assets. We find, however, that including a riskless asset has significant effects on various aspects of the analysis. When prior beliefs are represented by standard uninformative (diffuse) prior distributions, the value-weighted NYSE-AMEX index seems to be quite inefficient when the riskless asset is included, in that the posteriors for both Δ and ρ place most of their mass on values quite far from $\Delta = 0$ and $\rho = 1$. A deeper analysis reveals that such posterior distributions need not reflect information in the sample. In particular, similar posteriors for ρ result when the portfolio of interest is exactly efficient in the sample. Vague prior beliefs about E and V imply a marginal prior for ρ that is concentrated around low values, due the nonlinear relation between ρ and the basic return moments.

To investigate more carefully the extent to which the data do contain information about the inefficiency of the NYSE-AMEX index when a riskless asset is included, we recompute posterior distributions of ρ using various informative prior distributions for the parameters of the return distribution. Constructing a prior with disperse beliefs about ρ requires prior distributions for other model parameters (the elements of α) that are quite concentrated. Upon constructing such a prior, we find that the data produce a marginal posterior for ρ that is concentrated around low values, so we conclude that the data do in fact contain information indicating that the value-weighted NYSE-AMEX portfolio is rather inefficient. At the same time, however, a reasonable range of informative priors for ρ produces posteriors that vary substantially. We also find that much larger samples are required to produce a posterior distribution that strongly supports only modest degrees of inefficiency (values of ρ close to 1), even if the portfolio is exactly efficient in sample. In general, with sample sizes typically encountered in practice, the prior distribution exerts an important influence on one's posterior beliefs about a portfolio's inefficiency in the presence of a riskless asset.

When a riskless asset is excluded, posterior distributions for the inefficiency measures seem to be determined primarily by the data, in that the posterior distributions are very similar for a wide range of priors. It seems likely that the value-weighted NYSE-AMEX index is highly correlated with an efficient portfolio: most of the mass in the posterior distribution for ρ is concentrated above 0.9. At the same time, the posterior for Δ places much of its mass on values between 10 percent and 15 percent, which suggests that one would expect the index to be substantially outperformed by an efficient portfolio of equal risk.

Including a riskless asset seems to decrease significantly one's ability to draw sharp inferences about a portfolio's degree of inefficiency.

With a sample of 25 years, for example, the choice of the prior is critical in the presence of a riskless asset but unimportant in its absence. This property may arise, in part, from the increased importance of inferring the levels of expected returns in first case. When a riskless asset is excluded, the inefficiency measures ρ and Δ are unaffected by a common shift in the expected return on all risky assets, whereas both inefficiency measures are affected by such a shift when a riskless asset is included.

Appendix A: Analytic Results with Diffuse Priors

We derive here the analytic results, stated in Sections 3.2 and 4.2, for the posterior distributions of ρ and λ when a riskless asset is included. Without loss of generality, we assume that the nth asset is portfolio p. That is, the $1 \times n$ vector containing the returns on all assets in period t is $R_t' = [r_t' \; r_{p,t}]$, where r_t and $r_{p,t}$ are defined in the regression in (15). Some useful functions of the data and the parameters are defined as follows:

$$
\begin{aligned}
R &\equiv [R_1 \; R_2 \; \cdots \; R_T], && n \times T \\
r_p &\equiv [r_{p,1}, \ldots, r_{p,T}]', && T \times 1 \\
Y &\equiv [r_1 \; \cdots \; r_T]', && T \times (n-1) \\
\iota_T &\equiv [1, \ldots, 1]', && T \times 1 \\
X &\equiv [\iota_T \; r_p], && T \times 2 \\
B &\equiv [\alpha \; \beta]', && 2 \times (n-1) \\
b &\equiv [\alpha' \; \beta']', && 2(n-1) \times 1 \\
\hat{B} &\equiv (X'X)^{-1}X'Y \equiv [\hat{\alpha} \; \hat{\beta}]', && 2 \times (n-1) \\
\hat{b} &\equiv [\hat{\alpha}' \; \hat{\beta}']', && 2(n-1) \times 1, \text{ and} \\
S &\equiv (Y - X\hat{B})'(Y - X\hat{B}), && (n-1) \times (n-1)
\end{aligned}
$$

The Jacobian of the transformation in (18) is $\sigma_p^{2(n-1)}$. Noting that $|V| = \sigma_p^2|\Sigma|$, the prior distribution of (E, V) in (6) is rewritten in terms of b, Σ, μ_p, and σ_p^2:

$$
p(b, \Sigma, \mu_p, \sigma_p^2) \propto \sigma_p^{n-3}|\Sigma|^{-\frac{(n+1)}{2}}. \tag{A1}
$$

The multivariate Normal sampling distribution of R, conditioned on

E and V, is

$$p(R \mid E, V) \propto |V|^{-\frac{T}{2}} \exp\left\{-\tfrac{1}{2}\sum_{t=1}^{T}(R_t - E)'V^{-1}(R_t - E)\right\}. \quad \text{(A2)}$$

Using the transformation in (18), this distribution can be factored into the product of the marginal distribution of r_p and the distribution of Y conditioned on r_p:

$$p(R \mid b, \Sigma, \mu_p, \sigma_p^2) = p(Y \mid r_p, b, \Sigma) \cdot p(r_p \mid \mu_p, \sigma_p^2) \quad \text{(A3)}$$

where

$$p(Y \mid r_p, b, \Sigma) \propto |\Sigma|^{-\frac{T}{2}} \exp\left\{-\tfrac{1}{2}(\hat{b} - b)'[(X'X)^{-1} \otimes \Sigma]^{-1}(\hat{b} - b)\right\} \cdot \exp\left\{-\tfrac{1}{2}tr(S\Sigma^{-1})\right\}, \quad \text{(A4)}$$

and

$$p(r_p \mid \mu_p, \sigma_p^2) \propto \sigma_p^{-T} \cdot \exp\left\{-\frac{T}{2\sigma_p^2}(\hat{\mu}_p - \mu_p)^2\right\} \cdot \exp\left\{-\frac{T\hat{\sigma}_p^2}{2\sigma_p^2}\right\}. \quad \text{(A5)}$$

Combining (A1) through (A5) allows us to write

$$p(b, \Sigma, \mu_p, \sigma_p^2 \mid R) = p(b \mid \Sigma, Y, r_p) \cdot p(\Sigma \mid Y, r_p) \cdot p(\mu_p \mid r_p, \sigma_p^2) \cdot p(\sigma_p^2 \mid r_p) \quad \text{(A6)}$$

where

$$p(b \mid \Sigma, Y, r_p) \propto |\Sigma|^{-1} \exp\left\{-\tfrac{1}{2}(\hat{b} - b)'[(X'X)^{-1} \otimes \Sigma]^{-1}(\hat{b} - b)\right\}, \quad \text{(A7)}$$

$$p(\Sigma \mid Y, r_p) \propto |\Sigma|^{\frac{-(T+n-1)}{2}} \exp\{-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}S\}, \quad \text{(A8)}$$

$$p(\mu_p \mid r_p, \sigma_p^2) \propto \sigma_p^{-1} \exp\left\{-\frac{T}{2\sigma_p^2}(\hat{\mu}_p - \mu_p)^2\right\}, \quad \text{(A9)}$$

and

$$p(\sigma_p^2 \mid r_p) \propto \sigma_p^{n-T-2} \exp\left\{-\frac{T\hat{\sigma}_p^2}{2\sigma_p^2}\right\}. \quad \text{(A10)}$$

We see from (A6)-(A10) that, given the data, α, β, and Σ are independent of μ_p and σ_p. When a riskless asset is included, ρ can be expressed as a function of λ and θ [Equation (5)]. As shown in (17),

λ can be expressed as a function of α and Σ. Hence, λ is independent of μ_p and σ_p ; and, therefore, is independent of θ. From (A7) we observe that

$$(\alpha \mid \Sigma, Y, r_p) \sim N(\hat{\alpha}, d\Sigma), \tag{A11}$$

where d, the (1,1) element of $(X'X)^{-1}$, can be written as

$$d = \frac{(1+\hat{\theta}^2)}{T}. \tag{A12}$$

From (A8), we see that Σ^{-1} obeys a Wishart distribution,

$$(\Sigma^{-1} \mid Y, r_p) \sim W(S^{-1}, T-1). \tag{A13}$$

In order to derive the posterior distribution of λ, define

$$v \equiv d^{-1/2}\Sigma^{-1/2}\alpha, \tag{A14}$$

$$\eta \equiv \frac{\lambda}{d} = v'v, \tag{A15}$$

and

$$v \equiv \frac{\hat{\alpha}'\Sigma^{-1}\hat{\alpha}}{d}. \tag{A16}$$

The posterior distribution of λ and v can be expressed as a product of conditional and marginal distributions:

$$p(\lambda, v \mid Y, r_p) = p(\lambda \mid v, Y, r_p) \cdot p(v \mid Y, r_p). \tag{A17}$$

From Equation (A11) and (A14) we get

$$(v \mid \Sigma, Y, r_p) \sim N(d^{-1/2}\Sigma^{-1/2}\hat{\alpha}, I). \tag{A18}$$

Combining (A15) and (A18) yields the posterior distribution of η conditioned on v:

$$(\eta \mid v, Y, r_p) \sim \chi^2_{n-1}(v). \tag{A19}$$

Using the definition of d in (A12) we get

$$(\lambda \mid v, Y, r_p) \sim \chi^2_{n-1}(v)\left[\frac{1+\hat{\theta}^2}{T}\right]. \tag{A20}$$

The marginal posterior distribution of v is derived from (A12), (A13), and (A16):

$$(v \mid Y, r_p) \sim \chi^2_{T-1} \cdot [\hat{\alpha}' S^{-1} \hat{\alpha}]/d. \tag{A21}$$

Gibbons, Ross, and Shanken (1989) show that

$$\begin{aligned} T \cdot \hat{\alpha}' S^{-1} \hat{\alpha} &= \hat{\gamma}^2 - \hat{\theta}^2 \\ &= \frac{\hat{\theta}^2 (1 - \hat{\rho}^2)}{\hat{\rho}^2} \end{aligned} \tag{A22}$$

where the second equality follows from (11). Therefore, the posterior distribution of v in (A21) can be written as

$$(v \mid Y, r_p) \sim \chi^2_{T-1} \cdot \left(\frac{\hat{\theta}^2}{1 + \hat{\theta}^2} \right) \left(\frac{1 - \hat{\rho}^2}{\hat{\rho}^2} \right). \tag{A23}$$

Based on (A17), (A20), and (A23), in order to draw a value for λ, we first draw a value of a central chi-square variate with (T - 1) degrees of freedom and calculate the corresponding value of v. We then draw a value of a noncentral chi-square variate with (n - 1) degrees of freedom and v as the noncentrality parameter and calculate the corresponding value for λ.

With the standard diffuse prior for the regression model in (16), the posterior distribution of λ is virtually identical to that given above. The exponent on $|\Sigma|$ in (A8) becomes -(T + n - 2)/2, which simply requires that one change the degrees of freedom from T - 1 to T - 2 in the Wishart distribution in (A13) and in the central chi-square distributions in (A21) and (A23).

The marginal posterior distribution of θ is easily obtained from (A9) and (A10). From (A10),

$$\sigma_p^2 = T \hat{\sigma}_p^2 \frac{1}{\chi^2_{T-n}} \tag{A24}$$

or

$$\frac{1}{\sigma_p} = \frac{1}{\sqrt{T} \hat{\sigma}_p} \chi_{T-n}, \tag{A25}$$

where χ^2_{T-n} is a chi-square variate with T - n degrees of freedom. From (A9) we can represent μ_p as

$$\mu_p = \hat{\mu}_p + \frac{\sigma_p}{\sqrt{T}} z, \tag{A26}$$

where z is a standard Normal variate that is independent of χ^2_{T-n}.

Since $\theta = \mu_p/\sigma_p$, combining (A25) and (A26) gives

$$\theta = \frac{1}{\sqrt{T}}\left[\frac{\hat{\mu}_p}{\hat{\sigma}_p}\chi_{T-n} + z\right] = \frac{1}{\sqrt{T}}\left[\hat{\theta}\chi_{T-n} + z\right]. \quad \text{(A27)}$$

Appendix B: Obtaining Posteriors Using the Informative Priors

We describe here the algorithms used to draw from the posteriors given the data and the priors of Section 5. Both algorithms use Gibbs sampling, which is described in detail in Casella and George (1992) and Tierney (1991). Briefly, the jth draw of δ_i from the posterior for $(\delta_1, \delta_2, \ldots, \delta_K)$, where each δ_i is a subvector of the parameter vector δ, is obtained from the conditional distribution:

$$\delta_i^j \mid \delta_1^j, \delta_2^j, \ldots, \delta_{i-1}^j, \delta_{i+1}^{j-1}, \ldots, \delta_K^{j-1}, D \quad \text{(A28)}$$

where D represents the data. The values are drawn sequentially, after choosing starting values δ_i^0, $i = 1, \ldots, K$, and the generated sequence $\{\delta_i,\ i = 1, \ldots, M\}$ converges in distribution to the desired posterior as $M \to \infty$. The sampled δ_i's are dependent, but we generate them in sufficient number such that two sampling runs, starting with different random-number seeds, produce virtually identical posteriors (histograms). For additional discussions of the dependence issue, see Geweke (1991), Gelman and Rubin (1992), and Geyer (1992).

In the case where the riskless asset is included, we must draw from the posterior distribution of the parameters $(\alpha, \beta, \Sigma, \mu_p, \sigma_p^2)$. We use Gibbs sampling and draw from the following conditionals:

$$(\alpha, \beta) \mid \Sigma, \mu_p, \sigma_p^2, R; \quad \text{(A29)}$$

$$\Sigma^{-1} \mid \alpha, \beta, \mu_p, \sigma_p^2, R; \quad \text{(A30)}$$

$$\mu_p \mid \alpha, \beta, \Sigma, \sigma_p^2, R; \quad \text{(A31)}$$

$$\sigma_p^2 \mid \alpha, \beta, \Sigma, \mu_p, R. \quad \text{(A32)}$$

For the conditional distribution in (A29), define

$$A \equiv \begin{bmatrix} (\phi_0 + \phi_1\mu_p)^2 & 0 \\ 0 & \sigma_\beta^2 \end{bmatrix}^{-1} \otimes I_{n-1}, \quad \text{(A33)}$$

$$\Sigma_b \equiv (X'X \otimes \Sigma^{-1} + A)^{-1}, \quad \text{(A34)}$$

and

$$\bar{b} \equiv [0\ 1]' \otimes \iota_{n-1}. \quad \text{(A35)}$$

Then the distribution of $b = [\alpha' \beta']'$ in (A29) is

$$N(\Sigma_b[\text{Vec}(\Sigma^{-1} Y'X) + A\bar{b}], \Sigma_b), \tag{A36}$$

where Vec stacks the columns of a matrix.

The conditional distribution in (A30) is

$$W([S_0 + (Y - XB)'(Y - XB)]^{-1}, \nu_\Sigma + T). \tag{A37}$$

Draws from the conditional distribution in (A31) are essentially obtained by brute force. We have

$$p(\mu_p \mid \alpha, \beta, \Sigma, \sigma_p^2, R) \propto p(\alpha \mid \mu_p) p(X \mid \mu_p, \sigma_p^2) p(\mu_p). \tag{A38}$$

We compute this for a grid of μ_p values, normalize to obtain probabilities, and then draw from the discrete distribution.

The conditional distribution in (A32) is proportional to

$$\frac{\nu_o \lambda_o + \text{rss}}{\chi^2_{\nu_o + T}} \tag{A39}$$

where rss $= \sum (r_{p,t} - \mu_p)^2$.

For simplicity, the -conditional distributions above are expressed directly in terms of the matrices containing the data (using *Y, X,* etc.). It is straightforward to show that the unconstrained maximum likelihood estimates of the parameters $\alpha, \beta, \Sigma, \mu_p, \sigma_p^2$ serve as sufficient statistics, and the conditional distributions can be rewritten using those quantities. This alternative representation is used in constructing the posteriors in Figures 8A through 8D, where, as described in Section 5.1.2, it is assumed that the hypothetical samples produce unconstrained maximum likelihood estimates yielding $\hat{\rho} = 1$.

In the case where the riskless asset is excluded, we must draw from the posterior distribution of (E, V). We use the Gibbs strategy to draw from the conditional distributions,

$$E \mid V, R; \tag{A40}$$

$$V \mid E, R. \tag{A41}$$

where *R* again denotes the returns data. Since the location model used in this specification is a special case of the regression model, these two conditionals are special cases of (A29) and (A30).

References

Banz, R. W., 1981, "The Relationship between Return and Market Value of Common Stocks," *Journal of Financial Economics*, 9, 3-18.

Berger, J. O., and M. Delampady, 1987, "Testing Precise Hypotheses," *Statistical Science,* 2, 317-352.

Berger, J. O., and T. Selke, 1987, "Testing a Point NuIl Hypothesis: The Irreconcilability of P-Values and Evidence," *Journal of the American Statistical Association,* 82, 112-122.

Berk, J. B., 1993, "A Critique of Size-Related Anomalies," working paper, University of British Columbia, Vancouver.

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 Michael C. Jensen (ed.), *Studies in the Theory of Capital Markets,* Praeger, New York.

Box, G. E. P., and G. C. Tiao, 1973, *Bayesian Inference in Statistical Analysis,* Addison-Wesley, **Reading, MA.**

Breeden, D. T., 1979, "An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities," *Journal of Financial Economics,* 7, 265-296.

Casella, G., and E. I. George, 1992, "Explaining the Gibbs Sampler," *American Statistician,* 46, 167-174.

Chan, K. C., N. Chen, and D. A. Hsieh, 1985, "An Exploratory Investigation of the Firm Size Effect," *Journal of Financial Economics,* 14, 451-471.

Douglas, G. W., 1969, "Risk in the Equity Markets An Empirical Appraisal of Market Efficiency, *Yale Economic Essays,* 9, 3-45.

Fama, E. F., and J. D. MacBeth, 1973, "Risk, Return, and Equilibrium: Empirical Tests," *Journal of Political Economy,* 81, 607-636.

Ferson, W., S. Kandel, and R. F. Stambaugh, 1987, "Tests of Asset Pricing with Time-Varying Expected Risk Premiums and Market Betas," *Journal of Finance,* 42, 201-220.

Gelman, A., and D. B. Rubin, 1992, "Inference from Interative Simulation," *Statistical Science,* 7, 457-472.

Geweke, J., 1989, Bayesian Inference in Econometric Models using Monte Carlo Integration," *Econometrica,* 57, 1317-1339.

Geweke, J., 1991, "Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments," in J. M. Bernardo et al. (eds.), *Bayesian Statistics* (4th ed.), Oxford University Press. oxford.

Geyer, C. J., 1992, "Practical Markov Chain Monte Carlo," *Statistical Science,* 7, 473-483.

Gibbons, M. R., S. A. Ross, and J. Shanken, 1989, "A Test of the Efficiency of a Given Portfolio," *Econometrica,* 57, 1121-1152.

Harvey, C R., and G. Zhou, 1990, "Bayesian Inference in Asset Pricing Tests," *Journal of Financial Economics,* 26, 221-254.

Huberman, G., S. Kandel, and G. A. Karolyi, 1987, "Size and Industry Related Covariation of Stock Returns," working paper, University of Chicago, Chicago.

Kadane, J. B., 1987, "Comment," *Statistical Science,* 2, 347-348.

Kandel, S., and R. F. Stambaugh, 1987, "On Correlations and Inferences about Mean-Variance Efficiency," *Journal of Financial Economics* 18, 61-90.

Kandel, S., and R. F. Stambaugh, 1989, "A Mean-Variance Framework for Tests of Asset Pricing Models," *Review of Financial Studies*, 2, 125-156.

Lintner, J., 1965, "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets," *Review of Economics and Statistics*, 47, 13-27.

Lo, A. W., and A. C. MacKinlay, 1990, "Data-Snooping Biases in Tests of Financial Asset Pricing Models," *Review of Financial Studies*, 3, 431-467.

MacKinlay, A. C., 1993, "Multifactor Models Do Not Explain Deviations from the CAPM," working paper, University of Pennsylvania, Philadelphia.

Markowitz, H., 1952, "Portfolio Selection," *Journal of Finance*, 7, 77-91.

Markowitz, H., 1959, *Portfolio Selection: Efficient Diversification of Investments*, John Wiley, New York.

McCulloch, R., and P. E. Rossi, 1990, "Posterior, Predictive, and Utility-Based Approaches to Testing the Arbitrage Pricing Theory," *Journal of Financial Economics*, 28, 7-38.

McCulloch, R., and P. E. Rossi, 1991, "A Bayesian Approach to Testing the Arbitrage Pricing Theory," *Journal of Econometrics*, 49, 141-168.

Oldfield, G. S., and R. J. Rogalski, 1987, "The Stochastic Properties of Term Structure Movements," *Journal of Monetary Economics*, 19, 229-254.

Poirier, D., 1992, "Window Washing: A Bayesian Perspective on Diagnostic Checking," technical report, Department of Economics, University of Toronto.

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.

Shanken, J., 1987a, "Multivariate Proxies and Asset Pricing Relations: Living with the Roll Critique," *Journal of Financial Economics*, 18, 91-110.

Shanken, J., 1987b, "A Bayesian Approach to Testing Portfolio Efficiency," *Journal of Financial Economics*, 19, 195-215.

Shanken, J., 1992, "On the Estimation of Beta-Pricing Models," *Review of Financial Studies*, 5, 1-33.

Sharpe, W. F., 1964, "Capital Asset Prices A Theory of Market Equilibrium under Conditions of Risk," *Journal of Finance*, 19, 425-442.

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

Tierney, L., 1991, "Markov Chains for Exploring Posterior Distributions," Technical Report 560, School of Statistics, University of Minnesota.

Zellner, A., 1971, *An Introduction to Bayesian Inference in Econometrics*, John Wiley, New York.

Zhou, G., 1991, "Small Sample Tests of Portfolio Efficiency," *Journal of Financial Economics*, 30, 165–191.