# Measuring the Pricing Error of the Arbitrage Pricing Theory

**John Geweke**
University of Minnesota and
Federal Reserve Bank of Minneapolis

**Guofu Zhou**
Washington University in St. Louis

***This article provides an exact Bayesian framework for analyzing the arbitrage pricing theory (APT). Based on the Gibbs sampler, we show how to obtain the exact posterior distributions for functions of interest in the factor model. In particular, we propose a measure of the APT pricing deviations and obtain its exact posterior distribution. Using monthly portfolio returns grouped by industry and market capitalization, we find that there is little improvement in reducing the pricing errors by including more factors beyond the first one.***

As an important extension of the asset pricing model of Sharpe (1964) and Lintner (1965), Ross (1976, 1977) derived the arbitrage pricing theory (APT) which addresses a fundamental problem in finance: to characterize the expected return on a security. The APT implies that the expected return is approximately a

We are grateful to Franklin Allen (the Executive Editor), Siddhartha Chib, Philip Dybvig, Mark Grinblatt, Raymond Kan, Campbell Harvey, Ravi Jagannathan, Sunil Panikkath, Richard Roll, Jay Shanken, Jack Strauss, Sheridan Titman, participants of the Second Asian-Pacific Finance Conference in Hong Kong, and especially to Robert Korajczyk (the editor) and an anonymous referee for many helpful comments that substantially improved this article. Geweke's work is supported in part by NSF grant no. SES-9210070. Part of this article was written while the second author was visiting the Federal Reserve Bank of Minneapolis. Views expressed here are not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve system. Address correspondence to Guofu Zhou, John M. Olin School of Business, Washington University, Campus Box 1133, St. Louis, MO 63130.

*The Review of Financial Studies* Summer 1996 Vol. 9, No. 2, pp. 557–587

linear function of the risk premiums on systematic factors in the economy. Subsequently, there have been both a large theoretical literature extending the APT and a large empirical literature testing its implications.[1]

There are mainly two testing approaches that have been applied to the empirical study of the APT. Traditional factor analysis is the first approach. Burmeister and McElroy (1991), among others, tested nonlinear restrictions of the APT in the factor model. In most studies, a likelihood ratio test is used. In order to obtain it, one has to estimate the parameters under nonlinear restrictions, which is difficult to accomplish in practice. As a result, it is difficult to obtain the asymptotic distribution of the likelihood ratio test.[2] Given that the test has an asymptotic $\chi^2$-distribution, it remains unclear whether or not the asymptotic inference is reliable in the sample size commonly used. The second approach is a two-pass procedure. Chen (1983), Connor and Korajczyk (1988), Lehmann and Modest (1988), Roll and Ross (1980), and many others developed this procedure. In the first pass, either the factor loadings or the factors are estimated. Then, in the second pass, the regression of the returns on the estimated loadings or the factors is estimated. Treating the estimates as the true variables, the APT restrictions become linear constraints (implying zero-intercepts) on the regression coefficients in the multivariate regression and hence can be tested by using standard methods. However, this procedure suffers an errors-in-variables problem, because the estimated rather than the actual factor loadings or factors are used in the second pass tests. As known in the errors-in-variables literature, ignoring the uncertainty of the estimates can potentially lead to incorrect inference.

This article provides an exact statistical framework for analyzing the APT. There are at least two interesting aspects of our approach. First, our approach is a one-step procedure that is consistent with the return generating process. Given the fact that there are unobservable factors in the return generating process, our procedure implicitly incorporates this uncertainty into inference. As a result, there is no need to estimate separately either the factors or the factor loadings to infer the validity of the APT. Second, our approach makes it possible to examine virtually any function of the parameters that possesses important economic interpretations. In particular, it provides the exact posterior density for a proposed measure of the APT pricing errors, which indicates how far the data deviates from the APT pricing equations.

[1] Connor and Korajczyk (1992) provide an excellent survey of the literature.

[2] As shown by Anderson and Amemiya (1988), asymptotic distributions of the parameter estimates are very complex in the factor model, and the constrained estimates should be even more complex, making it difficult to analyze the likelihood ratio test and related asymptotic tests.

Exact inference is an important advantage of our approach over the existing approaches because the latter often do not have asymptotic distributions for functions of interest, and the asymptotic distributions may not be reliable in finite sample even if they become available.

Our approach is Bayesian. With the problem in hand, it is very difficult to apply classical statistical analysis, and Bayesian inference becomes a natural choice. McCulloch and Rossi (1990, 1991) developed a Bayesian analysis of the APT, whereas Harvey and Zhou (1990) and Shanken (1987a) proposed Bayesian tests for efficiency of a given portfolio. However, McCulloch and Rossi's approach remains a two-pass procedure in which the factors are extracted, before the Bayesian analysis starts, by using Connor and Korajczyk's (1988) asymptotic principal components (APC) approach. In contrast, our approach is a one-step procedure and is based on the Gibbs sampler. The Gibbs sampler permits us to obtain the exact posterior distributions for functions of interest in the factor model. In particular, this method makes it possible to provide exact posterior distributions for both the measure of the APT pricing errors and measures of the systematic and idiosyncratic risks.

This article is organized as follows. In the first section, the exact Bayesian framework is proposed. In the second section, the proposed approach is applied to portfolio returns grouped by both industry and size. The empirical results show that a one-factor model has a modest APT pricing error and there is little improvement in reducing the pricing errors by including more factors beyond the first one. Concluding remarks are offered in the final section.

## 1. Methodology

In this section, we examine first the APT restrictions and propose a measure quantifying the pricing deviations. Then, ignoring the identification issue for simplicity, we show how to obtain the exact posterior moments for this measure of pricing errors and other functions of interest. Next, coming back to the identification problem, we show how to identify the factor model and how the earlier analysis can incorporate the identification conditions imposed on the factor model. Then, we discuss how prior information may be utilized in the posterior analysis. Finally, we compare the proposed approach with the usual two-pass procedure.

### 1.1 APT restrictions

The basic APT model assumes that the returns on a vector of $N$ assets are related to $K$ pervasive and unknown factors by a $K$-factor model:

$$r_{it} = \alpha_i + \beta_{i1} f_{1t} + \cdots + \beta_{iK} f_{Kt} + \epsilon_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T, \quad (1)$$

where

$r_{it}$ = the return on asset $i$ at time $t$,
$\alpha_i = E[r_{it}]$, the expected return on asset $i$,
$f_{kt}$ = the $k$-th pervasive factor at time $t$,
$\epsilon_{it}$ = the idiosyncratic factor of asset $i$ at time $t$,
$\beta_{ik}$ = the beta or factor loading of the $k$-th factor for asset $i$,
$N$ = the number of assets, and
$T$ = the number of periods.

In what follows, it will be convenient for us to work with the vector form of the model:

$$\mathbf{r}_t = \boldsymbol{\alpha} + \beta \mathbf{f}_t + \boldsymbol{\epsilon}_t, \tag{2}$$

where $\mathbf{r}_t$ is an $N \times 1$ vector of returns, $\boldsymbol{\alpha}$, $N \times 1$, $\boldsymbol{\beta}$, $N \times K$, $\mathbf{f}_t$ and $\epsilon_t$ are defined accordingly. The standard assumptions on the factor model are

$$E\left[\mathbf{f}_t\right] = \mathbf{0}, \quad E\left[\mathbf{f}_t \mathbf{f}_t'\right] = \mathbf{I}, \quad E\left[\boldsymbol{\epsilon}_t | \mathbf{f}_t\right] = \mathbf{0}, \quad E\left[\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t' | \mathbf{f}_t\right] = \boldsymbol{\Sigma}, \tag{3}$$

where $\boldsymbol{\Sigma} = diag(\sigma_1^2, \ldots, \sigma_N^2)$. In this article, as in most studies, we make the standard assumptions that $\boldsymbol{\epsilon}_t$ and $\mathbf{f}_t$ are independent and both follow multivariate normal distributions.

Ross (1976, 1977) and many subsequent authors [e.g., Chamberlain and Rothschild (1983)] have shown that the absence of riskless arbitrage opportunities implies an approximate linear relationship between the expected asset returns and their risk exposures:

$$\alpha_i \approx \lambda_0 + \beta_{i1}\lambda_1 + \cdots + \beta_{iK}\lambda_K, \quad i = 1, \ldots, N, \tag{4}$$

as the number of assets satisfying Equation (1) tends toward infinity where $\lambda_0$ is the intercept of the pricing relationship (zero-beta rate) and $\lambda_k$ is the risk premium on the $k$-th factor ($k = 1, \ldots, K$). Since unknown parameters $\beta_{iK}$ and $\lambda_k$ enter into the constraints [Equation (4)] by multiplication with one another, the constraints are nonlinear. Equation (4) is the implication of no asymptotic arbitrage, and similar approximate pricing relations can be obtained under much weaker conditions [Shanken (1992)]. In contrast, with the much stronger assumption of competitive equilibrium, Connor's (1984) equilibrium version APT replaces the approximation with an equality.

Consider a measure of the pricing errors:

$$Q^2 = \frac{1}{N} \sum_{i=1}^{N} (\alpha_i - \lambda_0 - \beta_{i1}\lambda_1 - \cdots - \beta_{iK}\lambda_K)^2. \tag{5}$$

This is an average of the squared pricing errors across the assets. For the equilibrium version APT, Equation (4) is valid exactly, implying $Q$

is zero. For the asymptotic APT, $Q$ converges to zero as the number of assets approaches infinity. However, for a given $N$, $Q$ will not necessarily be small [Shanken (1992)]. Nevertheless, there are at least two theoretical reasons to examine the pricing errors in this case. First, conditional on an assumption about the multiple correlation between factors (proxies) and an equilibrium benchmark portfolio, Shanken (1987b) derived testable restrictions on the pricing deviaton for each individual asset, implying that $Q$ should be small if the correlation is close to one. Second, $Q$ provides information about the slope of the efficient frontier [Shanken (1992)]. Conditional on $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, the minimized average pricing error is

$$Q^2 = \frac{1}{N}\boldsymbol{\alpha}'\left[\mathbf{I}_N - \boldsymbol{\beta}^*(\boldsymbol{\beta}^{*\prime}\boldsymbol{\beta}^*)^{-1}\boldsymbol{\beta}^{*\prime}\right]\boldsymbol{\alpha}, \tag{6}$$

where $\boldsymbol{\beta}^* = (\mathbf{1}_N, \boldsymbol{\beta})$ and $\mathbf{1}_N$ is an $N \times 1$ vector of ones.[3] The sampling distribution of $Q$ or $Q^2$ is difficult to determine, whereas its exact posterior distribution can be easily constructed by using our proposed approach.[4]

### 1.2 Bayesian inference

To simplify the presentation, we ignore for the time being identification conditions for the factor model, but will incorporate them into the analysis in the next subsection. In a Bayesian framework, the parameters are treated as random variables. In particular, the pricing error $Q$ is a random variable. To characterize it, it is sufficient to find its posterior distribution. This distribution is analytically intractable, but the approach taken here allows one to determine the exact posterior distribution of $Q$ numerically. It is then straightforward to assess the economic importance of the pricing errors. For example, if $Q$ is found to have its posterior mass concentrated at 5 percent for monthly data, this implies the average pricing error is likely to be about 5 percent for monthly returns. Because, on the average, the asset returns are only about 1 percent for monthly data, we would regard the 5 percent average pricing error as too high, and so we would reject the APT restrictions. But if $Q$ is found to have a concentration at 0.01 percent,

---

[3] Interestingly, $Q$, a measure of pricing deviations, is very similar in mathematical form to the noncentrality parameter of the Gibbons, Ross, and Shanken (1989) test. The term $[\mathbf{I}_N - \boldsymbol{\beta}^*(\boldsymbol{\beta}^{*\prime}\boldsymbol{\beta}^*)^{-1}\boldsymbol{\beta}^{*\prime}]$ plays the role of their $\boldsymbol{\Sigma}^{-1}$.

[4] The approach also applies to the case where a riskless asset exists. In this case, $\lambda_0$ must be equal to the (observable) riskless rate of return, and the minimized average pricing error is

$$Q^2 = \frac{1}{N}(\boldsymbol{\alpha} - \lambda_0\mathbf{1}_N)'[\mathbf{I}_N - \boldsymbol{\beta}(\boldsymbol{\beta}'\boldsymbol{\beta})^{-1}\boldsymbol{\beta}'](\boldsymbol{\alpha} - \lambda_0\mathbf{1}_N),$$

where $\mathbf{1}_N$ is an $N$-vector of ones.

for example, the pricing errors would be regarded as negligible from an economic perspective. As a result, we would then regard the APT as an adequate pricing model for the assets.

Bayesian analysis transforms our prior belief into a posterior belief in light of the data. For simplicity, we consider a standard diffuse prior first and defer discussion of informative priors to Section 1.4. The standard diffuse prior has the following form:

$$P_0(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-1/2} = (\sigma_1 \cdots \sigma_N)^{-1}, \tag{7}$$

where $|\cdot|$ represents the determinant of the variance-covariance matrix $\boldsymbol{\Sigma}$. Let $\sigma_i^2$ be the $i$-th diagonal element of $\boldsymbol{\Sigma}$, then $|\boldsymbol{\Sigma}| = \sigma_1^2 \cdots \sigma_N^2$.

Let $\mathbf{R}$ be a $T \times N$ matrix of asset returns observed over the $T$ periods. Based on Bayes' rule, the joint posterior density function of the parameters, $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\Sigma}$, is

$$P(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-1/2} f(\mathbf{R}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Sigma}), \tag{8}$$

where $f(\mathbf{R}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Sigma})$ is the density of the data conditional on the parameters, or the likelihood function for the factor model [Equation (1)].

Denote all the parameters by $\boldsymbol{\theta}$ and let $g(\boldsymbol{\theta})$ be a function of interest. The general task of Bayesian inference is to obtain the expected value of $g(\boldsymbol{\theta})$ under the posterior density,

$$E[g(\boldsymbol{\theta})] = \int_{\boldsymbol{\Theta}} g(\boldsymbol{\theta}) P(\boldsymbol{\theta})\, d\boldsymbol{\theta}, \tag{9}$$

where $\boldsymbol{\Theta}$ is the domain of $\boldsymbol{\theta}$. This poses at least two difficulties. First, an analytical evaluation of Equation (9) is intractable if not impossible. This is seen by noting that the inverse of the complex covariance matrix $\boldsymbol{\beta}\boldsymbol{\beta}' + \boldsymbol{\Sigma}$ enters into the posterior density. Second, although Kloek and van Dijk (1978) show that the standard Monte Carlo approach can be a solution to such a high dimensional integration problem, it is not an easy matter to implement it because the posterior density function is of unknown form, and hence it is difficult to draw samples from this density. Monte Carlo integration with importance sampling [Geweke (1989)] may be an alternative if an adequate importance density (a density function that approximates the posterior density well) can be found. However, it is not clear how the importance density may be constructed given the complexity of the posterior density in the factor model. Fortunately the Gibbs sampling-data augmentation approach can be used to sample from the posterior distribution.

To explain the Gibbs sampling-data augmentation approach (Appendix A provides another explanation in a simpler model), we ob-

serve first that

$$f(\mathbf{R}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Sigma}) = \int f^*(\mathbf{R}, \mathbf{f}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Sigma})\, d\mathbf{f}, \tag{10}$$

where $\mathbf{f}$ denotes all the factors, and $f^*(\mathbf{R}, \mathbf{f}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Sigma})$ is the joint probability density of $\mathbf{R}$ and $\mathbf{f}$ conditional on the parameters $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$. We wish to approximate:

$$E[g(\alpha, \beta, \Sigma)] = \frac{\iiiint g(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Sigma})|\boldsymbol{\Sigma}|^{-1/2} f^*(\mathbf{R}, \mathbf{f}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Sigma})\, d\mathbf{f}\, d\boldsymbol{\alpha}\, d\boldsymbol{\beta}\, d\boldsymbol{\Sigma}}{\iiiint |\boldsymbol{\Sigma}|^{-1/2} f^*(\mathbf{R}, \mathbf{f}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Sigma})\, d\mathbf{f}\, d\boldsymbol{\alpha}\, d\boldsymbol{\beta}\, d\boldsymbol{\Sigma}}. \tag{11}$$

Suppose that we are able to draw samples from the conditional posterior density function for $\boldsymbol{\alpha}$,

$$P(\boldsymbol{\alpha}|\boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{f}, \mathbf{R}) = f^*(\mathbf{R}, \mathbf{f}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Sigma}) \Big/ \int f^*(\mathbf{R}, \mathbf{f}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Sigma})\, d\boldsymbol{\alpha},$$

and similarly from the other three conditional posterior density functions, $P(\boldsymbol{\beta}|\boldsymbol{\alpha}, \boldsymbol{\Sigma}, \mathbf{f}, \mathbf{R})$ and $P(\boldsymbol{\Sigma}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{f}, \mathbf{R})$, as well as from the conditional density $P(\mathbf{f}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{R})$. It turns out that it is in fact easy to do this, as will soon be shown. Now suppose further that we were given a *single* draw from the full posterior density:

$$P(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{f}) = \frac{|\boldsymbol{\Sigma}|^{-1/2} f^*(\mathbf{R}, \mathbf{f}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Sigma})}{\int\int\int\int |\boldsymbol{\Sigma}|^{-1/2} f^*(\mathbf{R}, \mathbf{f}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Sigma})\, d\mathbf{f}\, d\boldsymbol{\alpha}\, d\boldsymbol{\beta}\, d\boldsymbol{\Sigma}}.$$

If we replace the value of $\boldsymbol{\alpha}$ in this draw with a new value drawn from $P(\boldsymbol{\alpha}|\boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{f}, \mathbf{R})$, the new $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{f})$ must still be a draw from the full posterior distribution. If the value of $\boldsymbol{\beta}$ is then replaced with a draw from $P(\boldsymbol{\beta}|\boldsymbol{\alpha}, \boldsymbol{\Sigma}, \mathbf{f}, \mathbf{R})$, the new draw still comes from the full posterior distribution. Similarly, replacing $\boldsymbol{\Sigma}$ and $\mathbf{f}$ in succession with draws from their conditional posterior distributions, we are left with a value for $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{f})$ that is completely different from the original draw, but still comes from the full posterior distribution. The process may then be repeated, starting with $\boldsymbol{\alpha}$ and proceeding through $\mathbf{f}$. At the end of each repetition, the process yields a new draw from the full posterior distribution.

This algorithm is unrealistic in assuming an initial draw of $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{f})$ from the full posterior distribution. Under fairly general conditions [Gelfand and Smith (1990) and Geman and Geman (1984)] the initial draw may be replaced with any legitimate value for $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{f})$, and the sequence of draw just described will then converge in distribution to the posterior distribution [Roberts and Smith (1992) and Tierney (1991)]. One such condition is that it be possible to move from any point in the support of $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{f})$ to any other point in exactly

one full iteration of the Gibbs sampling-data augmentation algorithm just described. That condition is satisfied here, and so the sequence of draw will converge to the posterior distribution. The numerical accuracy can be assessed based on Geweke (1991b).

Thus we need only consider how to draw from the conditional distributions. The parameter vector formed by the $i$-th row of $\mathbf{B} \equiv (\boldsymbol{\alpha}, \boldsymbol{\beta})$ has a multivariate normal distribution conditional on $\sigma_i$:

$$f(\mathbf{b}_i|\mathbf{f}, \sigma_i) \propto exp\left(-\frac{1}{2\sigma_i^2}(\mathbf{b}_i - \hat{\mathbf{b}}_i)'\mathbf{F}'\mathbf{F}(\mathbf{b}_i - \hat{\mathbf{b}}_i)\right), \tag{12}$$

where $\mathbf{F} = (\mathbf{1}_T, \mathbf{f})$ is a $T \times (K+1)$ matrix formed by a vector of ones and the factors, and $\hat{\mathbf{b}}_i$ is the classical OLS estimator of the regression coefficients. Because the regressions given $\mathbf{f}$ in the factor model are mutually independent, the conditional distribution of $\mathbf{b}_i$ is independent of $\mathbf{b}_i (j \neq i)$. Each diagonal element of $\boldsymbol{\Sigma}$ has an inverted gamma distribution conditional on $\mathbf{b}_i$:

$$f(\sigma_i|\mathbf{f}, \mathbf{b}_i) \propto \frac{1}{\sigma_i^{\nu+1}} exp\left(-\frac{\nu s_i^2}{2\sigma_i^2}\right), \tag{13}$$

where

$$s_i^2 = \frac{1}{T}\sum_{t=1}^{T}(\mathbf{r}_{it} - \mathbf{F}_t\mathbf{b}_i)'(\mathbf{r}_{it} - \mathbf{F}_t\mathbf{b}_i), \tag{14}$$

and $\nu = T$ is the degrees of freedom. Equivalently, $\nu s_i^2/\sigma_i^2 \sim \chi^2(T)$.

Consider now how to draw $\mathbf{f}$ conditional on $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\Sigma}$, and the data. To do so, the probability distribution of $\mathbf{f}$ has to be specified. Consistent with Equations (2) and (3), $\mathbf{f}_t$ and $\mathbf{r}_t$ are jointly normally distributed:[5]

$$\begin{pmatrix} \mathbf{f}_t \\ \mathbf{r}_t \end{pmatrix} \sim N\left[\begin{pmatrix} \mathbf{0} \\ \boldsymbol{\alpha} \end{pmatrix}, \begin{pmatrix} \mathbf{I} & \boldsymbol{\beta}' \\ \boldsymbol{\beta} & \boldsymbol{\beta}\boldsymbol{\beta}' + \boldsymbol{\Sigma} \end{pmatrix}\right]. \tag{15}$$

Hence, the conditional samples of $\mathbf{f}$ at time $t$ can be drawn from a multivariate normal distribution with mean

$$E(\mathbf{f}_t|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{r}_t) = \boldsymbol{\beta}'(\boldsymbol{\beta}\boldsymbol{\beta}' + \boldsymbol{\Sigma})^{-1}(\mathbf{r}_t - \boldsymbol{\alpha}), \tag{16}$$

and covariance matrix

$$\text{Cov}(\mathbf{f}_t|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{r}_t) = \mathbf{I} - \boldsymbol{\beta}'(\boldsymbol{\beta}\boldsymbol{\beta}' + \boldsymbol{\Sigma})^{-1}\boldsymbol{\beta}. \tag{17}$$

[5] In the classical framework where $\alpha$, $\beta$, and $\Sigma$ are treated as constant parameters, Equation (15) is the standard assumption necessary to facilitate the maximum likelihood estimation. This assumption is also used in applying the EM algorithm to factor analysis [see, e.g., Lehmann and Modest (1988)].

Because $N$ is often far greater than $K$, it is computationally simpler to obtain the inversion of the $N \times N$ matrix $\boldsymbol{\beta}\boldsymbol{\beta}' + \boldsymbol{\Sigma}$ by using Woodbury's identity [see, e.g., Seber (1984, p. 520)]:

$$(\boldsymbol{\beta}\boldsymbol{\beta}' + \boldsymbol{\Sigma})^{-1} = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}(\mathbf{I} + \boldsymbol{\beta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta})^{-1}\boldsymbol{\beta}'\boldsymbol{\Sigma}^{-1}. \tag{18}$$

Since $\boldsymbol{\Sigma}$ is diagonal, its inversion is trivial to compute. So, only the inversion of $\mathbf{I} + \boldsymbol{\beta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}$, a $K \times K$ matrix, is needed to invert the $N \times N$ matrix $\boldsymbol{\beta}\boldsymbol{\beta}' + \boldsymbol{\Sigma}$.

### 1.3 Identification

For pedagogical reasons, we have so far ignored the well-known identification problem in the factor model. There are in fact two indeterminacies of the parameters. First, the information is not enough to determine all of the parameters if the number of factors is greater than or equal to half the number of assets [Seber (1984, p. 214)]. This is because the observable returns can determine only its mean and covariance matrix $\mathbf{V}$, which are related to the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ by

$$\mathbf{V} = \boldsymbol{\beta}\boldsymbol{\beta}' + \boldsymbol{\Sigma} \tag{19}$$

There are only $N(N+1)/2$ distinct elements of $\mathbf{V}$, but there are $NK+N$ elements on the right-hand side. To determine those parameters, we must have $N(N+1)/2 \geq NK + N$, or $N \geq 2K + 1$. For example, if there are $N = 10$ assets, and if no other restrictions are imposed on the parameters, we can only estimate a factor model up to four factors.

Second, there is an indeterminacy of the factor rotation. For any $K \times K$ orthogonal matrix $\mathbf{P}$, there is an equivalent factor model,

$$\mathbf{r}_t = \boldsymbol{\alpha} + \boldsymbol{\beta}^*\mathbf{f}_t^* + \boldsymbol{\epsilon}_t, \tag{20}$$

in which the new factors $\mathbf{f}_t^* = \mathbf{P}\mathbf{f}_t$ is a rotation of the old factors $\mathbf{f}_t$. The same moment conditions valid for the old factors are also valid for the new factors, that is, $E[\mathbf{f}_t^*] = 0$ and $E[\mathbf{f}_t^*\mathbf{f}_t^{*\prime}] = \mathbf{I}$. Moreover, the factor loadings are also rotated. The new loadings are linked to the old ones through $\boldsymbol{\beta}^* = \boldsymbol{\beta}\mathbf{P}'$. Because these new factor loadings and factors give rise to the same distribution for the returns, they cannot be identified from the observed returns unless further restrictions are imposed.

Because $\boldsymbol{\beta}$ has rank $K$, we assume, without loss of generality, that the first $K$ rows of $\boldsymbol{\beta}$ are independent. Let $\boldsymbol{\beta}^K$ be the $K \times K$ matrix composed by the first $K$ rows, then $\boldsymbol{\beta}^K$ is nonsingular. By a theorem in matrix theory [see, e.g., Muirhead (1982, p. 592), Theorem A9.8], there exists a unique orthogonal matrix $\mathbf{P}$ such that $\boldsymbol{\beta}^K\mathbf{P}'$ is a lower triangular

matrix with positive diagonal elements.[6] Therefore, to identify the factor model, we assume in what follows that $\boldsymbol{\beta}^K$ is of the form

$$\boldsymbol{\beta}^K = \begin{pmatrix} \beta_{11} & 0 & \cdots & 0 \\ \beta_{21} & \beta_{22} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ \beta_{K1} & \beta_{K2} & \cdots & \beta_{KK} \end{pmatrix}, \tag{21}$$

where $\beta_{ii} > 0$, $i = 1, \ldots, K$. This condition uniquely identifies the loadings and the associated factors. For example, we cannot identify the betas and factors by the returns data alone in a one-factor model, because both

$$r_{it} = \alpha_i + \beta_{i1} f_{1t} + \epsilon_{it}$$

and

$$r_{it} = \alpha_i + (-\beta_{i1})(-f_{1t}) + \epsilon_{it}$$

imply the same data generating process, but they have different betas and factors. However, by restricting $\beta_{11} > 0$ in the one-factor model, we uniquely identify both the betas and the associated factors.

Under the identification condition, all parameters, except for $\boldsymbol{\beta}^K$ and $\alpha_1, \ldots, \alpha_K$, have exactly the same posterior distributions as before. To draw $\boldsymbol{\beta}^K$ and $\alpha_1, \ldots, \alpha_K$ from their new posterior distributions, let $\mathbf{b}_i^* = (\alpha_i, \beta_{i1}, \ldots, \beta_{ii})'$, $i = 1, \ldots, K$. Simple algebra shows that $\mathbf{b}_1^*, \ldots, \mathbf{b}_K^*$ are independently multivariate normally distributed:

$$f(\mathbf{b}_i^* | \mathbf{f}, \sigma_i) \propto exp\left( -\frac{1}{2\sigma_i^2} (\mathbf{b}_i^* - \hat{\mathbf{b}}_i^*)' \mathbf{F}_i' \mathbf{F}_i (\mathbf{b}_i^* - \hat{\mathbf{b}}_i^*) \right), \quad i = 1, \ldots, K, \tag{22}$$

where $\mathbf{F}_i$ is a $T \times i$ matrix consisting of the first $i$ columns of $\mathbf{F}$, and $\hat{\mathbf{b}}_i^*$ is the OLS estimator of the regression of $r_i$ on $(1, f_1, \ldots, f_i)$. Because of the identification condition, draws from Equation (22) should be rejected[7] if they violate $\beta_{ii} > 0$. Combining these conditional distributions with those in Section 1.2, it is straightforward to evaluate the posterior means of functions of interest.

### 1.4 Informative priors

Only the diffuse prior has been used thus far in our Bayesian analysis. This prior represents no prior information or "ignorance" on $\boldsymbol{\alpha}, \boldsymbol{\beta}$,

[6] The orthogonal matrix $\mathbf{P}$ can be explicitly constructed as $\mathbf{P} = \mathbf{L}^{-1}\boldsymbol{\beta}^K$, where $\mathbf{L}$ is the $L$ matrix in the LU decomposition of the positive definite matrix $\boldsymbol{\beta}^K \boldsymbol{\beta}'^K$ ($\mathbf{L}$ is the lower triangular matrix such that $\mathbf{L}\mathbf{L}' = \boldsymbol{\beta}^K \boldsymbol{\beta}'^K$).

[7] Our applications show that the fraction of samples being rejected is often less than 30 percent. A more effective, but more complex, approach is provided in Appendix B.

and $\mathbf{\Sigma}$. As a function of these parameters, the pricing error $Q$ will also have a diffuse prior density. To induce an informative prior on $Q$, informative priors on $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\mathbf{\Sigma}$ have to be utilized. Consider the following class of informative priors:

$$\boldsymbol{\alpha}_i|\boldsymbol{\beta}_i \sim N(\alpha_{0i}, \eta_{0i}), \quad i = 1, \ldots, N, \tag{23}$$

$$\boldsymbol{\beta}_i \sim N(\hat{\boldsymbol{\beta}}_{0i}, \zeta_{0i}\mathbf{I}), \quad \beta_{ii} > 0, \quad i = 1, \ldots, K, \tag{24}$$

$$\boldsymbol{\beta}_i \sim N(\hat{\boldsymbol{\beta}}_{0i}, \xi_{0i}\mathbf{I}), \quad i = (K+1), \ldots, N, \tag{25}$$

$$\nu_{0i}s_{0i}^2/\sigma_i^2 \sim \chi^2(\nu_{0i}), \quad i = 1, \ldots, N, \tag{26}$$

where all variables with subscripts 0 (except $\alpha_{0i}$) are constants, chosen to reflect our prior degrees of belief on the distributions of the parameters. For example, $\hat{\beta}_{0i}$ represents our prior mean value for $\boldsymbol{\beta}_i$, and $\zeta_{0i}$ measures how close the mass of $\beta_i$ is to its mean. In Equation (23), $\alpha_{0i}$ is defined by $\alpha_{0i} = \lambda_o + \sum_{k=1}^{K} \beta_{ik}\lambda_k$, where $\lambda_0, \lambda_1, \ldots, \lambda_K$ are chosen constants. This says that the prior distribution of $\boldsymbol{\alpha}_i$ is dependent on $\boldsymbol{\beta}_i$. In other words, the prior distribution of $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}_i$ is specified jointly as a product of the marginal distribution of $\boldsymbol{\beta}_i$ and the distribution of $\boldsymbol{\alpha}_i$ conditional on $\boldsymbol{\beta}_i$. In addition, the priors are assumed to be independent across $i$.

Given the above priors on the model parameters, the prior distribution of $Q$ is readily computed. By varying the constants such as $\eta_{0i}$, we obtain different prior distributions of $Q$ which in turn reflect our varying degrees of prior beliefs on $Q$. The posterior distribution of $Q$ will show how our priors are changed in light of the data. Clearly, this posterior distribution is straightforward to obtain if samples of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\mathbf{\Sigma}$ can be drawn from their posterior distributions.

To draw $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\mathbf{\Sigma}$, we use again the Gibbs sampler by drawing them from their conditional distributions. Based on our earlier analysis (Section 1.3), it is seen that the alphas can be drawn from a normal distribution:

$$\boldsymbol{\alpha}_i|\boldsymbol{\beta}_i \sim N(\tilde{\alpha}_i, \eta_{1i}), \quad i = 1, \ldots, N, \tag{27}$$

where $\tilde{\alpha}_i = (\eta_i\alpha_{0i}+\eta_{0i}\hat{\alpha}_i)/(\eta_i+\eta_{0i})$, $\eta_{1i} = \eta_i\eta_{0i}/(\eta_i+\eta_{0i})$, $\hat{\alpha}_i = \sum(r_{it} - \beta_{i1}f_{1t} - \cdots - \beta_{iK}f_{Kt})/T$ and $\eta_i = \sigma_i^2/T$. The remaining parameters can be drawn as follows:

$$\boldsymbol{\beta}_i \sim N(\tilde{\boldsymbol{\beta}}_i, \mathrm{Var}(\boldsymbol{\beta}_i)), \quad \beta_{ii} > 0, \quad i = 1, \ldots, K, \tag{28}$$

$$\boldsymbol{\beta}_i \sim N(\tilde{\boldsymbol{\beta}}_i, \mathrm{Var}(\boldsymbol{\beta}_i)), \quad i = (K+1), \ldots, N, \tag{29}$$

$$\nu_{1i}s_{1i}^2/\sigma_i^2 \sim \chi^2(\nu_{1i}), \quad i = 1, \ldots, N, \tag{30}$$

where for $i = 1, \ldots, K$, $\text{Var}(\boldsymbol{\beta}_i) = (\mathbf{I}/\zeta_{0i} + \mathbf{F}_i^{*\prime}\mathbf{F}_i^*/\sigma_i^2)^{-1}$, $\tilde{\boldsymbol{\beta}}_i = \text{Var}(\boldsymbol{\beta}_i)(\hat{\boldsymbol{\beta}}_{0i}/\zeta_{0i} + \mathbf{F}_i^{*\prime}\mathbf{F}_i^*\hat{\boldsymbol{\beta}}_i^*/\sigma_i^2)$, $\mathbf{F}_i^*$ is the $\mathbf{F}_i$ matrix without the first column, and $\hat{\boldsymbol{\beta}}_i^*$ is the OLS estimator of the regression of $(r_i - \alpha_i)$ on $(f_1, \ldots, f_{i-1})$; and for $i = (K+1), \ldots, N$, $\text{Var}(\boldsymbol{\beta}_i) = (\mathbf{I}/\xi_{0i} + \mathbf{F}^{*\prime}\mathbf{F}^*/\sigma_i^2)^{-1}$, $\tilde{\boldsymbol{\beta}}_i^* = \text{Var}(\boldsymbol{\beta}_i)(\hat{\boldsymbol{\beta}}_{0i}/\xi_{0i} + \mathbf{F}^{*\prime}\mathbf{F}^*\hat{\boldsymbol{\beta}}_i^*/\sigma_i^2)$, $\mathbf{F}^*$ is the $\mathbf{F}$ matrix without the first column, and $\hat{\boldsymbol{\beta}}_i^*$ is the OLS estimator of the regression of $(r_i - \alpha_i)$ on $(f_1, \ldots, f_K)$. Finally, for $i = 1, \ldots, N$, $\nu_{1i} = \nu_{0i} + T$ and $s_{1i}^2 = (\nu_{0i}s_{0i}^2 + \nu s_i^2)/\nu_{1i}$.

### 1.5 A comparison with the two-pass procedure

The two-pass procedure (reviewed briefly in the introduction) usually works as follows. In the first pass, either the factor loadings or the factors are estimated from the factor model [Equation (2)]. Then, in the second pass, a multivariate regression is run of the returns on the estimated loadings or the factors. The equilibrium version APT implies zero-intercepts of the multivariate regression, and this implication is often tested by using Gibbons, Ross, and Shanken's (1989) (GRS) test.

The most flexible two-pass procedure is the one developed by Connor and Korajczyk (1986, 1988), which is a cross-sectional approach that can be applied to a large number of assets to extract the factors. In contrast, our approach is a time series one that can only be applied to a relatively small number of assets. Specifically, $N$ can be any large number on Connor and Korajczyk's framework, but it has to be less than or equal to $T - K$ in our setting in order to estimate the (nonsingular) covariance matrix of the returns. However, most multivariate tests of the APT are carried out in the second step of the two-pass procedure, and it is also necessary to estimate the covariance matrix of the returns. As a result, most of the tests are eventually applied to a small number of assets (usually about 10). The errors-in-variables problem is often ignored, but this can potentially yield incorrect inference as known in the errors-in-variables literature.

In analyzing a small number of assets, our approach suggests that it is possible to obtain exact inference that automatically recognizes the errors-in-variables problem. Furthermore, with as many as 100 assets, our simulations show that the proposed approach is still feasible, and capable of providing reliable inference.[8] Therefore, the proposed approach should be a useful complement to Connor and Korajczyk's (1986, 1988) in the case where a relatively small number of assets (portfolios) are used to test the APT.

[8] With $N = 100$, $T = 731$, and $K = 2$, the simulation took about three days' CPU on a SUN SPARCstation 10. For a year after the publication of this article, a Fortran program of the simulation will be available from the second author through e-mail (zhou@zhoufin.wustl.edu) upon request.

## 2. Empirical Results

In this section, we provide first the summary statistics of the data and then apply our methodology to obtain the pricing error of the APT in the U.S. equity market. To get additional insight, measures of the systematic and idiosyncratic risks are also provided. As a diagnostic for model fitting, we compare the regression of the returns on the market index with that on the factor extracted from the one-factor model, and we find that there is a gain in $R^2$ by using the extracted factor. However, the diagnostic alone does not mean that the pricing error is small. It simply says that the extracted factor fits the returns data better than the market index. Because the pricing error is of primary interest, we examine it further by showing how its exact posterior density may vary under a class of informative priors.

### 2.1 The data and summary statistics

There are two sets of data. The first is the returns on the industry portfolios grouped by following Breeden, Gibbons and Litzenberger (1989), Ferson and Harvey (1991), Gibbons, Ross, and Shanken (1989), and Sharpe (1964) with raw data available from the Center for Research in Security Prices (CRSP) at the University of Chicago. There are 12 industries: petroleum, finance/real estate, consumer durables, basic industries, food/tobacco, construction, capital goods, transportation, utilities, textiles/trade, services, and leisure. The returns are monthly from February 1926 to December 1986, a total of 61 years data ($T = 731$). For a comparison of results, we also use decile portfolios from the CRSP. This is our second data set, which is the monthly returns on market value sorted NYSE portfolio deciles varying from size 1 to size 10.

Means, standard deviations, and autocorrelations of the data are presented in Table 1. The means range from 0.849 percent per month for the utilities industry (industry 9) to 1.118 percent per month for the consumer durables industry (industry 3). The lowest standard deviation, a measure of the total industry risk, is found in the utilities industry, and the highest is found in the consumer durables industry. Although for both of these industries the high or low average returns are associated with their total industry risks, the petroleum industry (industry 1) has lower risk and higher return than the capital goods industry (industry 7). However, this is not in contradiction with financial theories. For example, the equilibrium version of APT asserts only that the high returns should be associated with their high systematic risks, and the systematic risks are determined by the asset's exposure to the factors. As shown in Table 1, there is some evidence of first-order autocorrelation in the returns. In the factor model, both

**Table 1**
**Means, standard deviations, and autocorrelations of asset returns**

| Variable | Mean (percent) | Std. dev. (percent) | Autocorrelation $\rho_1$ | $\rho_2$ | $\rho_3$ | $\rho_4$ | $\rho_{12}$ | $\rho_{24}$ |
|---|---|---|---|---|---|---|---|---|
| Industry portfolio returns[1] | | | | | | | | |
| Industry 1 | 1.040 | 6.300 | 0.009 | −0.020 | −0.054 | 0.091 | 0.017 | −0.010 |
| Industry 2 | 0.976 | 6.078 | 0.107 | −0.049 | −0.140 | 0.014 | 0.053 | 0.028 |
| Industry 3 | 1.118 | 7.610 | 0.145 | 0.000 | −0.114 | 0.010 | −0.019 | −0.014 |
| Industry 4 | 0.993 | 6.430 | 0.111 | 0.010 | −0.125 | 0.037 | −0.018 | 0.031 |
| Industry 5 | 0.939 | 4.869 | 0.096 | −0.027 | −0.088 | 0.011 | 0.025 | −0.023 |
| Industry 6 | 0.892 | 7.117 | 0.161 | 0.045 | −0.097 | −0.019 | −0.022 | −0.002 |
| Industry 7 | 1.031 | 6.550 | 0.117 | 0.001 | −0.101 | 0.015 | 0.001 | 0.014 |
| Industry 8 | 0.868 | 7.785 | 0.144 | −0.005 | −0.158 | −0.014 | 0.000 | 0.023 |
| Industry 9 | 0.849 | 4.837 | 0.149 | −0.036 | −0.134 | 0.000 | −0.013 | 0.039 |
| Industry 10 | 0.931 | 6.178 | 0.132 | −0.005 | −0.071 | 0.008 | −0.006 | −0.016 |
| Industry 11 | 0.968 | 7.441 | 0.013 | 0.041 | −0.003 | 0.055 | 0.047 | −0.030 |
| Industry 12 | 0.994 | 7.556 | 0.200 | 0.034 | −0.075 | −0.047 | 0.026 | 0.030 |
| Decile portfolio returns[2] | | | | | | | | |
| Decile 1 | 1.720 | 11.440 | 0.158 | −0.012 | −0.079 | −0.062 | 0.082 | 0.030 |
| Decile 2 | 1.489 | 9.786 | 0.156 | 0.003 | −0.090 | −0.102 | 0.049 | 0.029 |
| Decile 3 | 1.289 | 8.752 | 0.195 | −0.003 | −0.097 | −0.085 | 0.012 | 0.024 |
| Decile 4 | 1.281 | 8.068 | 0.176 | 0.014 | −0.106 | −0.058 | 0.016 | −0.014 |
| Decile 5 | 1.207 | 7.590 | 0.146 | 0.005 | −0.104 | −0.049 | 0.009 | 0.008 |
| Decile 6 | 1.200 | 7.331 | 0.163 | 0.002 | −0.121 | −0.032 | 0.000 | 0.011 |
| Decile 7 | 1.185 | 6.977 | 0.137 | 0.026 | −0.106 | −0.009 | −0.023 | −0.009 |
| Decile 8 | 1.008 | 6.516 | 0.123 | 0.011 | −0.112 | 0.006 | −0.004 | −0.004 |
| Decile 9 | 1.044 | 6.234 | 0.098 | −0.002 | −0.133 | 0.021 | 0.007 | 0.005 |
| Decile 10 | 0.871 | 5.364 | 0.085 | −0.015 | −0.121 | 0.041 | 0.008 | 0.028 |

[1]The industry groups are 1 = petroleum, 2 = finance/real estate, 3 = consumer durables, 4 = basic industries, 5 = food/tobacco, 6 = construction, 7 = capital goods, 8 = transportation, 9 = utilities, 10 = textiles/trade, 11 = services, and 12 = leisure.

[2]These are returns on market value sorted NYSE monthly portfolio deciles compiled by the Center for Research in Security Prices (CRSP) at the University of Chicago. For both the industry and decile portfolios, the data is monthly from February 1926 to December 1986 (731 observations).

the residuals and factors are assumed to be serially independent, and so are the returns. Nevertheless, the autocorrelation does not seem to be severe. Therefore, as is the case for most empirical studies, we adopt the working assumption that the returns are independent and identically distributed, and the $K$-factor model is well specified.

For the decile portfolios, there is the well-known pattern that lower deciles tend to have large mean returns that are accompanied by large standard deviations. Generally speaking, small firms tend to have higher returns and, at the same time, are subject to more economic risks. In contrast to the industry returns, there are greater first-order autocorrelations which are concentrated largely in the low deciles. Nevertheless, this pattern does not seem to be severe. Similar to the industry returns, higher order autocorrelations die out very fast.

To understand more about the data, Table 2 provides both the eigenvalues and sample correlation matrix from principal components analysis. The largest eigenvalue dominates other eigenvalues and the

difference between the largest eigenvalue, 0.0431, and the second largest one, 0.0022, is substantial. Moreover, the first eigenvalue explains 81.81 percent of the total variation of returns, and the first four eigenvalues explain about 91.96 percent. Based on the asymptotic principal components analysis of Connor and Korajczyk (1986, 1993), the eigenvalues can be interpreted (asymptotically) as explaining the portions of the systematic risk in the factor model. If the number of factors is $K$, the eigenvalues excluding the $K$ largest ones should be equal. However, it is difficult to determine whether a subset of the sample eigenvalues are significantly different from one another. As a result, we examine values of $K$ from 1 to 4 in our Bayesian factor analysis of the APT restriction. This may be a reasonable choice given that the first four sample eigenvalues explain about 92 percent of the systematic risk.

The decile portfolio returns have in general greater correlations than the industry returns. In addition, the first eigenvalue explains more than 92.84 percent of the variations, and the first four explain 98.95 percent. There is relatively stronger evidence that a factor model with $K$ from 1 to 4 should describe the returns.

### 2.2 The APT pricing error

Under the standard diffuse priors on all the parameters in the factor model, the posterior distribution of any function of interest is readily evaluated by using the methods discussed in Section 1.3. The posterior mean of the pricing error is provided in Table 3. Panel A reports the results for the whole sample period, from February 1926 to December 1986, whereas panels B and C report the results for the subperiods, from February 1926 to June 1956 and from July 1956 to December 1986.

Consider first the results for the whole sample period. For purposes of comparison, we examine the APT constraints starting from the case where there are no factors. In this case, $K = 0$ and the minimum pricing error $Q$ is the average of the squared pricing errors across assets, where the pricing error for each individual asset is the deviation of its expected return from the average of all the expected returns. Recall that $Q$ is a random variable in a Bayesian framework. Both the posterior mean and standard deviation of $Q$ are of interest, and they are, as reported in Table 3, 0.2408 percent and 0.0536 percent, respectively. The mean seems small as compared with the magnitude of the expected returns, showing that there are small deviations in the expected returns across the industries, a fact reflected from the summary statistics in Table 1. To further assess the pricing error, we provide the 90 percent Bayesian confidence interval [0.1564 percent, 0.3233 percent], which states that there is 90 percent probability that

**Table 2**
**Principal components analysis of the data**

Industry portfolio returns

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Eigenvalues | | | | | | |
| 4.312 | 0.217 | 0.177 | 0.141 | 0.098 | 0.083 | 0.064 | 0.057 | 0.043 | 0.031 | 0.026 | 0.023 |
| | | | | | Correlation matrix | | | | | | |
| 1.000 | 0.810 | 0.732 | 0.781 | 0.695 | 0.719 | 0.754 | 0.696 | 0.671 | 0.630 | 0.603 | 0.670 |
| 0.810 | 1.000 | 0.880 | 0.895 | 0.883 | 0.865 | 0.879 | 0.835 | 0.859 | 0.830 | 0.740 | 0.848 |
| 0.732 | 0.880 | 1.000 | 0.923 | 0.845 | 0.883 | 0.920 | 0.833 | 0.789 | 0.853 | 0.706 | 0.850 |
| 0.781 | 0.895 | 0.923 | 1.000 | 0.869 | 0.894 | 0.934 | 0.839 | 0.795 | 0.832 | 0.713 | 0.840 |
| 0.695 | 0.883 | 0.845 | 0.869 | 1.000 | 0.833 | 0.860 | 0.766 | 0.813 | 0.875 | 0.713 | 0.853 |
| 0.719 | 0.865 | 0.883 | 0.894 | 0.833 | 1.000 | 0.893 | 0.806 | 0.761 | 0.824 | 0.724 | 0.846 |
| 0.754 | 0.879 | 0.920 | 0.934 | 0.860 | 0.893 | 1.000 | 0.836 | 0.781 | 0.838 | 0.731 | 0.860 |
| 0.696 | 0.835 | 0.833 | 0.839 | 0.766 | 0.806 | 0.836 | 1.000 | 0.731 | 0.740 | 0.699 | 0.809 |
| 0.671 | 0.859 | 0.789 | 0.795 | 0.813 | 0.761 | 0.781 | 0.731 | 1.000 | 0.753 | 0.695 | 0.763 |
| 0.630 | 0.830 | 0.853 | 0.832 | 0.875 | 0.824 | 0.838 | 0.740 | 0.753 | 1.000 | 0.674 | 0.850 |
| 0.603 | 0.740 | 0.706 | 0.713 | 0.713 | 0.724 | 0.731 | 0.699 | 0.695 | 0.674 | 1.000 | 0.741 |
| 0.670 | 0.848 | 0.850 | 0.840 | 0.853 | 0.846 | 0.860 | 0.809 | 0.763 | 0.850 | 0.741 | 1.000 |

Decile portfolio returns

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Eigenvalues | | | | | |
| 5.919 | 0.313 | 0.052 | 0.025 | 0.021 | 0.013 | 0.011 | 0.009 | 0.008 | 0.006 |
| | | | | Correlation mix | | | | | |
| 1.000 | 0.956 | 0.931 | 0.908 | 0.888 | 0.870 | 0.846 | 0.811 | 0.796 | 0.716 |
| 0.956 | 1.000 | 0.970 | 0.960 | 0.948 | 0.934 | 0.913 | 0.886 | 0.870 | 0.801 |
| 0.931 | 0.970 | 1.000 | 0.973 | 0.962 | 0.950 | 0.935 | 0.910 | 0.890 | 0.828 |
| 0.908 | 0.960 | 0.973 | 1.000 | 0.979 | 0.972 | 0.964 | 0.946 | 0.925 | 0.867 |
| 0.888 | 0.948 | 0.962 | 0.979 | 1.000 | 0.982 | 0.974 | 0.961 | 0.948 | 0.894 |
| 0.870 | 0.934 | 0.950 | 0.972 | 0.982 | 1.000 | 0.980 | 0.973 | 0.963 | 0.912 |
| 0.846 | 0.913 | 0.935 | 0.964 | 0.974 | 0.980 | 1.000 | 0.981 | 0.969 | 0.927 |
| 0.811 | 0.886 | 0.910 | 0.946 | 0.961 | 0.973 | 0.981 | 1.000 | 0.981 | 0.945 |
| 0.796 | 0.870 | 0.890 | 0.925 | 0.948 | 0.963 | 0.969 | 0.981 | 1.000 | 0.960 |
| 0.716 | 0.801 | 0.828 | 0.867 | 0.894 | 0.912 | 0.927 | 0.945 | 0.960 | 1.000 |

The table provides the eigenvalues (multiplied by 100) of the sample correlation matrix for the monthly industry and decile returns, respectively.

**Table 3**
**Average pricing errors**

| | Industry returns ($N = 12$) | | | Decile returns ($N = 10$) | | |
|---|---|---|---|---|---|---|
| $K$ | $Q$ | Std. error | 90% interval | $Q$ | Std. error | 90% interval |
| | | | Panel A: February 1926 to December 1986 (whole period) | | | |
| 0 | 0.2408 | 0.0536 | [0.1564, 0.3323] | 0.3520 | 0.0859 | [0.2184, 0.5009] |
| 1 | 0.1184 | 0.0286 | [0.0746, 0.1679] | 0.0845 | 0.0292 | [0.0448, 0.1388] |
| 2 | 0.1169 | 0.0299 | [0.0710, 0.1690] | 0.0649 | 0.0228 | [0.0347, 0.1077] |
| 3 | 0.1070 | 0.0283 | [0.0624, 0.1558] | 0.0437 | 0.0127 | [0.0244, 0.0660] |
| 4 | 0.0982 | 0.0273 | [0.0557, 0.1458] | 0.0422 | 0.0130 | [0.0234, 0.0658] |
| | | | Panel B: February 1926 to June 1956 (subperiod) | | | |
| 0 | 0.4237 | 0.0941 | [0.2750, 0.5836] | 0.5499 | 0.1443 | [0.3336, 0.8072] |
| 1 | 0.2083 | 0.0485 | [0.1331, 0.2916] | 0.1647 | 0.0573 | [0.0865, 0.2715] |
| 2 | 0.1767 | 0.0443 | [0.1100, 0.2540] | 0.1350 | 0.0387 | [0.0769, 0.2040] |
| 3 | 0.1681 | 0.0426 | [0.1019, 0.2421] | 0.1095 | 0.0283 | [0.0672, 0.1594] |
| 4 | 0.1504 | 0.0445 | [0.0812, 0.2275] | 0.0854 | 0.0241 | [0.0391, 0.1436] |
| | | | Panel C: July 1956 to December 1986 (subperiod) | | | |
| 0 | 0.2807 | 0.6212 | [0.1836, 0.3871] | 0.3268 | 0.0731 | [0.2122, 0.4528] |
| 1 | 0.1693 | 0.0317 | [0.1185, 0.2226] | 0.0854 | 0.0236 | [0.0498, 0.1275] |
| 2 | 0.1524 | 0.0334 | [0.0910, 0.2086] | 0.0601 | 0.0183 | [0.0325, 0.1004] |
| 3 | 0.1302 | 0.0294 | [0.0831, 0.1793] | 0.0467 | 0.0145 | [0.0257, 0.0824] |
| 4 | 0.1273 | 0.0480 | [0.0751, 0.2051] | 0.0416 | 0.0159 | [0.0214, 0.0811] |

Let $r_{it}$ be the return on asset $i$ at time $t$. Assume the $K$-factor model for the returns:

$$r_{it} = \alpha_i + \beta_{i1} f_{1t} + \cdots + \beta_{iK} f_{Kt} + \epsilon_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T,$$

where $\alpha_i = E[r_{it}]$ is the expected return on asset $i$, $f_{kt}$ the $k$-th pervasive factor at time $t$, $\epsilon_{it}$ the idiosyncratic factor of asset $i$ at time $t$, $\beta_{ik}$ the beta or factor loading of the $k$-th factor for asset $i$. The pricing error from the APT is measured by $Q \geq 0$, where $Q^2 = \frac{1}{N}\sum_{i=1}^{N}(\alpha_i - \lambda_0 - \beta_{i1}\lambda_1 - \cdots - \beta_{iK}\lambda_K)^2 = \frac{1}{N}\boldsymbol{\alpha}'[\mathbf{I}_N - \boldsymbol{\beta}^*(\boldsymbol{\beta}^{*\prime}\boldsymbol{\beta}^*)^{-1}\boldsymbol{\beta}^{*\prime}]\boldsymbol{\alpha}$, $\lambda_0$ is the intercept of the APT pricing relationship, $\lambda_k$ is the risk premium on the $k$-th factor ($k = 1, \ldots, K$), $\boldsymbol{\beta}^* = (1_N, \boldsymbol{\beta})$, $\boldsymbol{\beta}$ is an $N \times K$ matrix of the factor loadings, and $1_N$ is an $N \times 1$ vector of ones. The data are monthly industry and decile returns from February 1926 to December 1986. With alternative assumptions on the number of factors, the table provides the posterior means, standard deviations, and the 90 percent Bayesian confidence intervals for $Q$ (the results are multiplied by 100) over the whole sample period and its subperiods.

the pricing error is in the interval. As there is not much difference between the mean and the values in the confidence interval, the posterior density of the average pricing error is concentrated heavily near its mean. This may be interpreted as evidence of the informativeness of the data on the pricing errors.

In a one-factor model, the mean pricing error is 0.1184 percent, the standard deviation is 0.0286 percent, and the 90 percent Bayesian confidence interval is [0.0746 percent, 0.1679 percent]. The mean pricing error is much smaller than the average sample mean returns of the assets, 0.9666 percent. This indicates that the deviation of the expected returns from the risk premiums multiplied by the factor load-

ings (including the constant) is just about 10 percent of the magnitude of the expected returns. However, in comparison with the previous $Q = 0.2408$ percent in a zero-factor model, a value of 0.1184 percent reduces the pricing error by about 50 percent. To examine the sensitivity of the pricing error to the number of factors, we allow $K$ to vary from 1 to 4. The mean pricing error goes down to 0.1169 percent, 0.1070 percent, and 0.0982 percent in two-, three- and four-factor models, respectively. This suggests that there are no substantial reductions in the pricing errors when more factors are allowed beyond the first one.

The empirical results are more striking for the decile returns. First, there are relatively large variations in the expected returns of the assets. When $K = 0$, the mean pricing error is $Q = 0.3520$ percent, about 46 percent larger than the mean pricing error for the industry returns. This large value reflects greater variation in the cross-sectional expected returns. This is perhaps more evident in the summary statistics, where the maximum difference of the sample mean returns between the decile portfolios is 0.8490 percent, as compared to only 0.2686 percent for the industry returns. Second, there is relatively large reduction in the zero-factor pricing error when at least one factor is included. Without factors, the mean pricing error is 0.3520 percent, but that reduces to 0.0845 percent in a one-factor model. This is a far greater reduction than that for the industry portfolios. In the case of multiple factors, although there are additional reductions in the pricing error, the percentage of the reduction is small compared to the one-factor case. For example, one additional factor in a three-factor model barely reduces the pricing error, but a one-factor model shrinks the pricing error of a zero-factor model by 76 percent.

Consider now the results for the subperiods reported in panels B and C of Table 3. The sample size becomes half as large as before. As a result, the uncertainty in the estimation goes up. For example, in the zero-factor case for the industry returns, the standard error of the mean pricing error increases from 0.0535 to 0.0941 percent in panel B and to 0.6212 percent in panel C. Similar increases also occur for the decile returns in the first subperiod. Because of the increased uncertainty, the mean pricing errors are less accurately estimated, and they are in general larger than before. However, the uncertainty in the estimation for the decile returns in the second subperiod is about the same as before, which is due to the behavior of the data, as seen from the zero-factor model where both the cross-sectional mean differences and the standard deviations are almost the same as those for the whole sample period. This explains the relatively unchanged mean pricing errors in the one- to four-factor models of the decile returns in the second subperiod. Overall, in combination with the

results of the whole sample period, we still find that there are no substantial reductions in the pricing errors when more factors are allowed beyond the first one.

### 2.3 Risk measures in the APT model

It is of interest to examine the expected returns, the systematic risks, and the unsystematic or idiosyncratic risks in the factor model. The results are provided in Table 4. In the Bayesian framework, the point estimate of the expected asset returns are the posterior means of the alphas. In comparison with the sample means of the asset returns as reported in Table 1, there are virtually no differences. So, as far as the expected asset returns are concerned, both the classical and the Bayesian approaches yield similar results. However, the advantage of the Bayesian approach is that it can yield small sample results about the idiosyncratic risks and other functions of interest.

The Bayesian idiosyncratic risks are the posterior means of the sigmas. In the case of $K = 0$, the idiosyncratic risks are point estimates of the unconditional standard deviations, matching those sampling estimates in Table 1. In the case of $K = 1$, the idiosyncratic risks across the industry assets are only about half of those in a zero-factor model. The reduction for the decile returns is much more substantial. For example, the idiosyncratic risk of the seventh decile asset it 0.912 percent, much smaller than the 7.341 percent level in a zero-factor model. For both the industry and decile returns, the idiosyncratic risks generally decrease as $K$ varies from 1 to 4. This is expected because some of the variations in the returns can be explained by the variations in the factors. However, the decrease is much less substantial than from a zero-factor model to a one-factor model. When the idiosyncratic risks are compared with the pricing errors, it is interesting that the $Q$s are much within the variations of the idiosyncratic risk of each asset.

The systematic risks are not reported. Instead, we report the proportions of the idiosyncratic risks to the total risk. The proportions measure the importance of the idiosyncratic risks. The higher the proportions, the greater the idiosyncratic risks relative to the systematic risks. The proportions are computed as the ratios of the diagonal elements of $\mathbf{\Sigma}$ to those of $(\boldsymbol{\beta}'\boldsymbol{\beta} + \mathbf{\Sigma})$, from which the systematic risks can be backed out (as the ratios of the idiosyncratic risks multiplied by one minus the proportions, to the proportions). As shown in Table 4, the average proportions are below 50 percent in a one-factor model, and become smaller as $K$ increases. However, the idiosyncratic risks remain a major proportion of the total risk even with up to four factors. Interestingly, the idiosyncratic risks are more disperse for the decile returns, and the largest one is found for the first decile (the

**Table 4**
**Expected returns and risks**

Industry portfolio returns

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Expected returns** | | | | | | | | | | | | |
| K = 0 | 1.037 | 0.976 | 1.119 | 0.992 | 0.939 | 0.886 | 1.028 | 0.868 | 0.847 | 0.936 | 0.968 | 0.994 |
| K = 1 | 1.037 | 0.972 | 1.114 | 0.991 | 0.936 | 0.888 | 1.027 | 0.863 | 0.849 | 0.928 | 0.965 | 0.992 |
| K = 2 | 1.034 | 0.966 | 1.103 | 0.980 | 0.931 | 0.880 | 1.017 | 0.854 | 0.841 | 0.921 | 0.958 | 0.980 |
| K = 3 | 1.038 | 0.989 | 1.136 | 1.006 | 0.951 | 0.908 | 1.045 | 0.881 | 0.859 | 0.947 | 0.982 | 1.009 |
| K = 4 | 1.048 | 0.989 | 1.136 | 1.010 | 0.949 | 0.907 | 1.048 | 0.887 | 0.859 | 0.943 | 0.981 | 1.010 |
| **Idiosyncratic risks** | | | | | | | | | | | | |
| K = 0 | 6.308 | 6.804 | 7.618 | 6.437 | 4.873 | 7.127 | 6.557 | 7.798 | 4.841 | 6.185 | 7.453 | 7.564 |
| K = 1 | 3.878 | 2.055 | 2.414 | 1.828 | 1.976 | 2.680 | 1.936 | 3.804 | 2.582 | 2.846 | 4.750 | 3.236 |
| K = 2 | 3.702 | 0.676 | 2.298 | 1.755 | 1.959 | 2.656 | 1.730 | 3.804 | 2.404 | 2.856 | 4.762 | 3.275 |
| K = 3 | 5.802 | 0.924 | 2.289 | 1.681 | 1.730 | 2.670 | 1.709 | 3.736 | 2.370 | 1.749 | 4.767 | 3.106 |
| K = 4 | 3.653 | 0.917 | 2.251 | 0.102 | 1.305 | 2.665 | 1.870 | 3.657 | 2.362 | 2.340 | 4.697 | 2.923 |
| **Proportions of idiosyncratic risks** | | | | | | | | | | | | |
| K = 0 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| K = 1 | 0.612 | 0.336 | 0.315 | 0.282 | 0.403 | 0.374 | 0.294 | 0.486 | 0.531 | 0.458 | 0.635 | 0.426 |
| K = 2 | 0.585 | 0.110 | 0.300 | 0.271 | 0.400 | 0.371 | 0.262 | 0.486 | 0.494 | 0.460 | 0.636 | 0.431 |
| K = 3 | 0.880 | 0.100 | 0.199 | 0.174 | 0.240 | 0.253 | 0.173 | 0.338 | 0.349 | 0.189 | 0.477 | 0.279 |
| K = 4 | 0.550 | 0.146 | 0.280 | 0.016 | 0.263 | 0.368 | 0.282 | 0.460 | 0.475 | 0.368 | 0.612 | 0.373 |

**Table 4 (continued)**

| | Decile portfolio returns | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Expected returns | | | | | |
| K = 0 | 1.718 | 1.481 | 1.284 | 1.281 | 1.202 | 1.201 | 1.186 | 1.002 | 1.046 | 0.872 |
| K = 1 | 1.863 | 1.623 | 1.411 | 1.395 | 1.317 | 1.305 | 1.285 | 1.100 | 1.132 | 0.942 |
| K = 2 | 1.710 | 1.473 | 1.273 | 1.265 | 1.191 | 1.183 | 1.168 | 0.991 | 1.029 | 0.856 |
| K = 3 | 1.834 | 1.582 | 1.365 | 1.342 | 1.257 | 1.244 | 1.222 | 1.034 | 1.066 | 0.882 |
| K = 4 | 1.646 | 1.423 | 1.229 | 1.223 | 1.154 | 1.149 | 1.137 | 0.964 | 1.005 | 0.840 |
| | | | | | Idiosyncratic risks | | | | | |
| K = 0 | 11.451 | 9.799 | 8.760 | 8.080 | 7.597 | 7.341 | 6.987 | 6.522 | 6.239 | 5.368 |
| K = 1 | 5.370 | 3.213 | 2.425 | 1.531 | 1.066 | 0.912 | 1.122 | 1.377 | 1.622 | 2.150 |
| K = 2 | 3.173 | 1.648 | 1.444 | 1.183 | 1.072 | 0.991 | 0.964 | 0.773 | 0.965 | 1.443 |
| K = 3 | 3.760 | 2.313 | 1.378 | 1.084 | 1.049 | 0.987 | 0.895 | 0.803 | 0.178 | 1.408 |
| k = 4 | 3.239 | 0.877 | 0.714 | 1.073 | 1.051 | 0.981 | 0.934 | 0.826 | 0.665 | 1.313 |
| | | | | | Proportions of idiosyncratic risks | | | | | |
| K = 0 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| K = 1 | 0.464 | 0.324 | 0.273 | 0.187 | 0.138 | 0.123 | 0.159 | 0.208 | 0.257 | 0.396 |
| K = 2 | 0.278 | 0.169 | 0.165 | 0.147 | 0.142 | 0.136 | 0.139 | 0.119 | 0.155 | 0.270 |
| K = 3 | 0.318 | 0.233 | 0.156 | 0.133 | 0.137 | 0.133 | 0.127 | 0.122 | 0.028 | 0.258 |
| K = 4 | 0.280 | 0.087 | 0.079 | 0.131 | 0.137 | 0.132 | 0.132 | 0.124 | 0.105 | 0.239 |

Let $r_{it}$ be the return on asset $i$ at time $t$. Assume the $K$-factor model for the returns:

$$r_{it} = \alpha_i + \beta_{i1} f_{1t} + \cdots + \beta_{iK} f_{Kt} + \epsilon_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T,$$

where $\alpha_i = E[r_{it}]$ is the expected return on asset $i$, $f_{kt}$ the $k$-th pervasive factor at time $t$, $\epsilon_{it}$ the idiosyncratic factor of asset $i$ at time $t$, $\beta_{ik}$ the beta or factor loading of the $k$-th factor for asset $i$. The data are monthly industry and decile returns February 1926 to December 1986 ($T = 731$, and $N = 12$ and 10 for the industry and decile returns, respectively). The table provides the posterior means of the expected returns, of the idiosyncratic risks, and of the proportions of the idiosyncratic risks relative to the total risks, respectively. The results are multiplied by 100.

smallest capitalization). In contrast, the idiosyncratic risks are fairly even across assets for the industry returns.

### 2.4 A comparison of the market index with the APT factor

In applications of the one-factor APT model, the factor is frequently prespecified as a market index, say, the CRSP value-weighted index. Table 5 provides the results of regressing the industry returns on the index and the APT factor, where the APT factor is extracted from the one-factor model by using our exact Bayesian procedure (the extracted factor is the posterior mean of the factor draws). The index is seen to have substantial explanatory power. The adjusted $\bar{R}^2$ is at least 58.33 percent and as high as 93.34 percent. But the average (across the industry portfolios) is 80.59 percent. In contrast, the APT factor has a minimum $\bar{R}^2$ of 60.25 percent, a maximum of 93.31 percent, and an average of 81.28 percent.

It should be noted that, while there is a time-series $\bar{R}^2$ gain from using the extracted APT factor, this does not mean the factor will necessarily yield smaller pricing deviations. Indeed, the regressions are only a diagnostic for model fitting, showing only that the extracted factor fits the returns data better than the market index. The reverse side is perhaps more interesting. Although the market index is prespecified (computed from a simple value-weighting scheme), its performance is comparable with the extracted factor which is estimated to best explain the variations across the industry returns, a conclusion similar to Brown's (1989). In addition, as shown in Table 5, it is remarkable that there is more than 99 percent correlation between the extracted APT factor and the prespecified market index. To further assess the performance of the CAPM versus the one-factor APT model, Table 5 also reports the largest absolute value and the average of the absolute values of the regression intercepts. For the APT factor, these values are 0.1294 and 0.0518 percent, respectively. They are smaller than those for the index, 0.1980 and 0.0756 percent. Because the intercepts measure pricing deviations of each equation of the model, the additional diagnostic confirms the $R^2$ analysis which suggests that the one-factor APT model fits the returns better than the single index model.

Table 6 provides the results of regressing the decile returns on the index and on the factor extracted from the decile returns. In the regression on the index, the lowest $\bar{R}^2$ is 49.31 percent, the maximum 97.05 percent, and the average is 85.37 percent. The explanatory power increases monotonously as the decile portfolio increases its size. This is expected because the index is value weighted, which gives more weight to large companies, and hence it fits larger decile portfolios better. In contrast, the APT factor explains the decile returns fairly evenly well across portfolio size, but explains relatively the worst for

**Table 5**
**Regression of industry returns on the market index and APT factor**

| Variable | $\alpha_{mi}$ (% per month) | $\beta_{mi}$ | $\bar{R}^2_{mi}$ | $\alpha_{fi}$ (% per month) | $\beta_{fi}$ | $\bar{R}^2_{fi}$ |
|---|---|---|---|---|---|---|
| Industry 1 | 0.1506 | 0.9334 | 0.7033 | 0.0966 | 0.8468 | 0.6310 |
| | (0.1279) | (0.0223) | | (0.1414) | (0.0239) | |
| Industry 2 | 0.0303 | 1.0204 | 0.9043 | 0.0104 | 0.9750 | 0.8992 |
| | (0.0700) | (0.0122) | | (0.0713) | (0.0121) | |
| Industry 3 | 0.0072 | 1.2750 | 0.8997 | 0.1080 | 1.2300 | 0.9132 |
| | (0.0898) | (0.0156) | | (0.0828) | (0.0140) | |
| Industry 4 | −0.0022 | 1.0976 | 0.9334 | 0.0149 | 1.0505 | 0.9331 |
| | (0.0618) | (0.0108) | | (0.0615) | (0.0104) | |
| Industry 5 | 0.1496 | 0.7804 | 0.8261 | 0.0118 | 0.7585 | 0.8481 |
| | (0.0755) | (0.0131) | | (0.0701) | (0.0119) | |
| Industry 6 | −0.1353 | 1.1465 | 0.8313 | −0.0993 | 1.1238 | 0.8713 |
| | (0.1089) | (0.0190) | | (0.0943) | (0.0160) | |
| Industry 7 | 0.0305 | 1.1052 | 0.9111 | 0.0497 | 1.0663 | 0.9264 |
| | (0.0728) | (0.0127) | | (0.0656) | (0.0111) | |
| Industry 8 | −0.1980 | 1.2059 | 0.7700 | −0.1294 | 1.1582 | 0.7733 |
| | (0.1390) | (0.0242) | | (0.1370) | (0.0232) | |
| Industry 9 | 0.0852 | 0.7405 | 0.7541 | −0.0679 | 0.6975 | 0.7266 |
| | (0.0892) | (0.0155) | | (0.0934) | (0.0158) | |
| Industry 10 | 0.0278 | 0.9559 | 0.7681 | −0.0269 | 0.9349 | 0.8002 |
| | (0.1108) | (0.0193) | | (0.1020) | (0.0173) | |
| Industry 11 | 0.0345 | 1.0028 | 0.5833 | 0.0028 | 0.9774 | 0.6025 |
| | (0.1787) | (0.0311) | | (0.1734) | (0.0293) | |
| Industry 12 | −0.0557 | 1.1815 | 0.7857 | −0.0040 | 1.1638 | 0.8291 |
| | (0.1301) | (0.0227) | | (0.1154) | (0.0195) | |
| | Market index | APT factor | | | | |
| $\frac{1}{N}\sum \lvert\alpha_i\rvert$ | 0.0756 | 0.0518 | | | | |
| $100 \times \max \lvert\alpha_i\rvert$ | 0.1980 | 0.1294 | | | | |
| $\frac{1}{N}\sum \bar{R}^2_i$ | 0.8059 | 0.8128 | | | | |
| $\rho^1$ | 0.9922 | | | | | |

Let $r_m$ be the return on the CRSP value-weighted index in excess of the 1-month Treasury bill rate and $r_i$ be the excess return on the $i$-th industry sorted portfolio. In the case of the APT factor, $r_f$ is the factor estimates plus the associated risk premium and $r_i$ is the return in excess of the zero-beta rate. The regression is

$$r_{it} = \alpha_{vi} + \beta_{vi} r_{vt} + \epsilon_{it}, \quad t = 1, \ldots, T, \quad i = 1, \ldots, N,$$

where $v = m$ or $f$. The data are monthly returns from February 1926 to December 1986 (731 observations) and there are $N = 12$ industries.

[1] $\rho$ is the correlation between $r_m$ and $r_f$, the market index and the APT factor.

both the smallest and the largest deciles. However, the averaged $\bar{R}^2$ is 92.65 percent, much better than the index's performance. Interestingly, the correlation between the index and the extracted factor is still as high as 95.65 percent.

## 2.5 The APT pricing error under informative priors

The use of the class of informative priors proposed in Section 1.4 requires specifying the constants that determine the prior densities. To aid this task, we use the principal factor analysis approach [Seber (1984, pp. 219–221)] to get a rough estimate of the loadings based on

**Table 6**
**Regression of decile returns on the market index and APT factor**

| Variable | $\alpha_{mi}$ (% per month) | $\beta_{mi}$ | $\bar{R}^2_{mi}$ | $\alpha_{fi}$ (% per month) | $\beta_{fi}$ | $\bar{R}^2_{fi}$ |
|---|---|---|---|---|---|---|
| Decile 1 | 0.5577 | 1.2863 | 0.4931 | 0.0568 | 1.6122 | 0.7412 |
| | (0.2303) | (0.0552) | | (0.1671) | (0.0395) | |
| Decile 2 | 0.3722 | 1.2299 | 0.6669 | −0.0089 | 1.4639 | 0.9041 |
| | (0.1536) | (0.0368) | | (0.0837) | (0.0198) | |
| Decile 3 | 0.3074 | 1.2014 | 0.7265 | −0.0249 | 1.3999 | 0.9446 |
| | (0.1303) | (0.0312) | | (0.0595) | (0.0141) | |
| Decile 4 | 0.2628 | 1.1630 | 0.7884 | −0.0080 | 1.3176 | 0.9693 |
| | (0.1065) | (0.0255) | | (0.0411) | (0.0097) | |
| Decile 5 | 0.1733 | 1.1359 | 0.8242 | −0.0593 | 1.2643 | 0.9790 |
| | (0.0928) | (0.0222) | | (0.0325) | (0.0077) | |
| Decile 6 | 0.1738 | 1.1269 | 0.8542 | −0.0347 | 1.2359 | 0.9852 |
| | (0.0823) | (0.0197) | | (0.0266) | (0.0063) | |
| Decile 7 | 0.1392 | 1.1151 | 0.8994 | −0.0203 | 1.1828 | 0.9719 |
| | (0.0659) | (0.0158) | | (0.0353) | (0.0083) | |
| Decile 8 | 0.0634 | 1.0836 | 0.9239 | −0.0542 | 1.1231 | 0.9520 |
| | (0.0550) | (0.0132) | | (0.0443) | (0.0105) | |
| Decile 9 | 0.0813 | 1.0470 | 0.9458 | 0.0217 | 1.0452 | 0.9071 |
| | (0.0443) | (0.0106) | | (0.0587) | (0.0139) | |
| Decile 10 | −0.0466 | 0.9470 | 0.9705 | 0.0432 | 0.8407 | 0.7384 |
| | (0.0292) | (0.0070) | | (0.0878) | (0.0207) | |
| | Market index | APT factor | | | | |
| $\frac{1}{N}\sum \lvert\alpha_i\rvert$ | 0.1402 | 0.0410 | | | | |
| $100 \times \max \lvert\alpha_i\rvert$ | 0.5577 | 0.0593 | | | | |
| $\frac{1}{N}\sum \bar{R}^2_i$ | 0.8537 | 0.9265 | | | | |
| $\rho^1$ | 0.9565 | | | | | |

Let $r_m$ be the return on the CRSP value-weighted index in excess of the 1-month Treasury bill rate and $r_i$ be the excess return on the $i$-th industry sorted portfolio. In the case of the APT factor, $r_f$ is the factor estimates plus the associated risk premium and $r_i$ is the return in excess of the zero-beta rate. The regression is

$$r_{it} = \alpha_{vi} + \beta_{vi} r_{vt} + \epsilon_{it}, \quad t = 1, \ldots, T, \quad i = 1, \ldots, N,$$

where $v = m$ or $f$. The data are monthly returns from February 1926 to December 1986 (731 observations) and there are $N = 10$ decile returns.

[1] $\rho$ is the correlation between $r_m$ and $r_f$, the market index and the APT factor.

the first 10 years of data and use the estimates as the prior means for the entire sample. In addition, we use the associated standard errors as a benchmark for the standard errors of the prior densities. To reflect various degrees of belief about $Q$, we provide two specifications, priors $A$ and $B$, for the prior standard errors. Prior $A$ is "large" in which the prior standard errors are five times the benchmark, and prior $B$ is "small" in which the prior standard errors are one-fifth of the benchmark. Given either prior $A$ or $B$, the prior density of the pricing error $Q$ is straightforward to compute. Table 7 reports the prior means, standard deviations, and the 90 percent Bayesian confidence intervals for $Q$. Panels $A$ and $B$ are obtained by using priors $A$ and $B$, respectively. Under prior $A$, $Q$ has a prior mean of 1.7340 percent, and its 90 percent Bayesian confidence interval is

**Table 7**
**Average pricing errors under informative priors**

| | Industry returns | | | Decile returns | | |
|---|---|---|---|---|---|---|
| $K$ | $Q$ | Std. error | 90% interval | $Q$ | Std. error | 90% interval |
| | | | Panel A: A large prior error | | | |
| 1 | 1.7340 | 0.3919 | [1.1127, 2.4012] | 0.3129 | 0.0791 | [0.1879, 0.4469] |
| | 0.1214 | 0.0282 | [0.0777, 0.1694] | 0.0983 | 0.0343 | [0.0534, 0.1633] |
| 2 | 1.6231 | 0.3877 | [1.0113, 2.2866] | 0.1463 | 0.0397 | [0.0849, 0.2151] |
| | 0.1169 | 0.0277 | [0.0734, 0.1642] | 0.0596 | 0.0204 | [0.0327, 0.0979] |
| 3 | 2.1268 | 0.5373 | [1.2856, 3.0608] | 0.1492 | 0.0442 | [0.0805, 0.2253] |
| | 0.1079 | 0.0298 | [0.0260. 0.0673] | 0.0456 | 0.0125 | [0.0260, 0.0673] |
| 4 | 2.1276 | 0.5852 | [1.2122, 3.1403] | 0.1468 | 0.0477 | [0.0734, 0.2292] |
| | 0.0913 | 0.0262 | [0.0521, 0.1376] | 0.0394 | 0.0113 | [0.0214. 0.0586] |
| | | | Panel B: A small prior error | | | |
| 1 | 0.0693 | 0.0156 | [0.0564, 0.1342] | 0.0313 | 0.0079 | [0.0188, 0.0447] |
| | 0.0357 | 0.0060 | [0.0258, 0.0458] | 0.0295 | 0.0075 | [0.0177, 0.0424] |
| 2 | 0.0649 | 0.0155 | [0.0405, 0.0215] | 0.0146 | 0.0040 | [0.0085, 0.0215] |
| | 0.0591 | 0.0136 | [0.0377, 0.0822] | 0.0145 | 0.0039 | [0.0084, 0.0214] |
| 3 | 0.0854 | 0.0215 | [0.0520, 0.1224] | 0.0149 | 0.0044 | [0.0080, 0.0225] |
| | 0.0692 | 0.0168 | [0.0428, 0.0981] | 0.0146 | 0.0043 | [0.0079, 0.0220] |
| 4 | 0.0846 | 0.0232 | [0.0488, 0.1247] | 0.0146 | 0.0048 | [0.0073, 0.0230] |
| | 0.0675 | 0.0175 | [0.0398, 0.0976] | 0.0145 | 0.0046 | [0.0074, 0.0227] |

Let $r_{it}$ be the return on asset $i$ at time $t$. Assume the $K$-factor model for the returns:

$$r_{it} = \alpha_i + \beta_{i1} f_{1t} + \cdots + \beta_{iK} f_{Kt} + \epsilon_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T,$$

where $\alpha_i = E[r_{it}]$ is the expected return on asset $i$, $f_{kt}$ the $k$-th pervasive factor at time $t$, $\epsilon_{it}$ the idiosyncratic factor of asset $i$ at time $t$, $\beta_{ik}$ the beta or factor loading of the $k$-th factor for asset $i$. The pricing error from the APT is measured by $Q \geq 0$, where $Q^2 = \boldsymbol{\alpha}'[\mathbf{I}_N - \boldsymbol{\beta}^*(\boldsymbol{\beta}^{*\prime}\boldsymbol{\beta}^*)^{-1}\boldsymbol{\beta}^{*\prime}]\boldsymbol{\alpha}/N$, $\boldsymbol{\beta}^* = (\mathbf{1}_N, \boldsymbol{\beta})$, $\boldsymbol{\beta}$ is an $N \times K$ matrix of the factor loadings, and $\mathbf{1}_N$ is an $N \times 1$ vector of ones. The data are monthly industry and decile returns from February 1926 to December 1986 ($T = 731$). The priors in panel A have larger pricing errors than those in panel B. In the table, the first row next to a given number of $K$ summarizes the prior distribution of $Q$, and the second reports the posterior means, standard deviations, and the 90 percent Bayesian confidence intervals (the results are multiplied by 100).

[1.1127 percent, 2.4012 percent] in the $K = 1$ case. This prior is rather large relative to the cross-sectional difference between the expected returns of the industries. In contrast, the prior mean under prior $B$ is only 0.0693 percent, relatively small as compared with the cross-sectional difference.

Table 7 also reports for $Q$ the posterior means, standard deviations, and the 90 percent Bayesian confidence intervals. The results are in the second row next to a given number of $K$. Under prior $A$ and $K = 1$, the posterior mean for the industries is 0.1213 percent, a sharp reduction from the prior mean level of 1.7340 percent. Interestingly, this posterior mean is also very close to 0.1184 percent, the posterior mean under the diffuse prior. However, the posterior mean under prior $B$ is in general smaller than those obtained under the diffuse

prior. For example, the posterior mean for the industries in the $K = 1$ case is 0.0357 percent, smaller than 0.1184 percent. Overall, under any of the informative or diffuse priors, there is little progress in reducing the pricing error by including more factors beyond the first one.

## 3. Conclusions

In this article we propose an exact Bayesian framework for examining the APT pricing restrictions. First, our approach is a one-step approach. In contrast to existing studies, no preestimates of either the factors or the factor loadings are required. Second, we propose a simple measure of pricing errors and obtain its exact posterior distribution. Unlike the likelihood ratio test in the classical framework, our measure indicates the extent to which the APT restrictions deviate from the data. As an application of our approach, we study the APT pricing restrictions by using monthly portfolio returns grouped by industry and market capitalization. We find that there is little improvement in reducing the pricing errors by including more factors beyond the first one. Furthermore, our approach can also be applied to study a variety of other asset pricing models, and similar measures of pricing errors can be proposed. Although it is difficult to obtain the exact sampling distributions of these measures in many applications, it is easy to evaluate the exact posterior distributions in the Bayesian framework.

## Appendix A

### An introduction to the Gibbs sampler

The Gibbs sampler is a path-breaking technique for generating random samples from a multivariate distribution by using its conditional distributions without having to compute the full joint density. In many problems, such as the ARCH, GARCH, regime-switching, and latent variables models, the full joint density is extremely difficult to calculate, but the conditional distributions are easy to evaluate. Hence, the Gibbs sampler can be used to make difficult Bayesian analysis tractable. In addition, it is also useful in classical statistics such as in the evaluation of likelihood functions [see Casella and George (1992) and references therein].

The idea of the Gibbs sampling technique is simple. To get a sample from a complex density function $f(\theta_1, \theta_2)$, it starts from an arbitrary initial value $(\theta_1, \theta_2) = (\theta_1^0, \theta_2^0)$ in the support of $f(\theta_1, \theta_2)$ and obtains a new value $(\theta_1^1, \theta_2^1)$ with $\theta_1^1$ drawn from $f(\theta_1|\theta_2^0)$ and $\theta_2^1$ from $f(\theta_2|\theta_1^1)$. Iterating this process gives rise to a sequence $\{(\theta_1^n, \theta_2^n)\}$. Under fairly general conditions, $(\theta_1^n, \theta_2^n)$ approximates well a random sample from the joint density $f(\theta_1, \theta_2)$.

To illustrate an application of the Gibbs sampler, consider an AR(1) model:

$$x_t = \rho x_{t-1} + \epsilon_t, \quad t = 1, 2, \ldots, T, \tag{A.1}$$

where $|\rho| < 1$, and $\epsilon_t$ is i.d.d. and normally distributed with $E(\epsilon_t) = 0$ and $\text{Var}(\epsilon_t) = \sigma^2$. Let $I_\rho$ be an indicator function, $I_\rho = 1$ if $|\rho| < 1$ and 0 otherwise. Then $p_0(\rho, \sigma^2) \propto \frac{1}{\sigma} I_\rho$ is a diffuse prior imposing only the stationarity condition. The posterior density is

$$p(\rho, \sigma^2) \propto p_0(\rho, \sigma^2) L(\rho, \sigma^2), \tag{A.2}$$

where $L(\rho, \sigma^2)$ is the (exact) likelihood function, which is very complex [Amemiya (1985, p. 162)], and hence it is difficult to draw samples from it. Treating $x_0$ as a parameter, the joint posterior density of $\rho, \sigma^2$, and $x_0$ is

$$p(\rho, \sigma^2, x_0) \propto p_0(\rho, \sigma^2) L(\rho, \sigma^2, x_0), \tag{A.3}$$

where $L(\rho, \sigma^2, x_0)$ is the likelihood function conditional on $x_0$ and is trivially obtained as

$$L(\rho, \sigma^2, x_0) = (2\pi)^{-T/2} \sigma^{-T} \exp\left[ -\sum_{t=1}^{T} (x_t - \rho x_{t-1})^2 / 2\sigma^2 \right]. \tag{A.4}$$

Clearly, the density $p(\rho, \sigma^2)$ in Equation (A.2) is given by $p(\rho, \sigma^2, x_0)$ after integrating $x_0$ out, implying that $p(\rho, \sigma^2, x_0)$ should provide all information about $p(\rho, \sigma^2)$. For example, the posterior mean of $\rho$ will be given by $\int \int \int \rho p(\rho, \sigma^2, x_0) d\sigma^2 dx_0 d\rho$. Hence, we need only be concerned about drawing samples from $p(\rho, \sigma^2, x_0)$.

By the Gibbs sampler, the samples are obtained from the conditional distributions: $\rho$ and $\sigma^2$ are drawn from a normal (truncated at $|\rho| < 1$) and a $\chi^2$-distribution conditional on $x_0$, and $x_0$ is drawn, conditional on $\rho$ and $\sigma$, from a normal distribution, $x_0 \sim N(\rho x_1, \sigma^2)$. This procedure generates a sequence of $(\rho^n, \sigma^n, x_0^n)$ which can then be used to approximate the expected value of a function of interest, $E[g(\rho, \sigma)]$, by Monte Carlo integration:

$$\bar{E}[g(\rho, \sigma)] = \frac{1}{N} \sum_{n=1}^{N} g(\rho^n, \sigma^n). \tag{A.5}$$

For example, if $g(\rho, \sigma) = \rho$, Equation (A.5) delivers a numerical approximation to the posterior mean of $\rho$. The accuracy increases as $N$ goes up.

Based on the draws $(\rho^n, \sigma^n, x_0^n)$, the Gibbs approximation to the

likelihood function $L(\rho, \sigma^2)$ is

$$\overline{L}(\rho, \sigma^2) = \frac{1}{N}\sum_{n=1}^{N} L(\rho, \sigma^2, x_0^n). \tag{A.6}$$

This is useful in applications where the exact likelihood function is difficult to compute, whereas the conditional likelihood function is easy to obtain.

## Appendix B

### An alternative Gibbs sampling method for B

The key for drawing $\mathbf{B}$ lies in drawing its first $K$ rows with positive $\beta_{ii}(i = 1, \ldots, K)$. Let $\mathbf{b}_1^* = \alpha_1$ and $\mathbf{b}_i^* = (\alpha_i, \beta_{i1}, \ldots, \beta_{i(i-1)})'$ for $i = 2, \ldots, K$. Then, $\mathbf{b}_1^*, \ldots, \mathbf{b}_K^*$ are mutually independent and multivariate normally distributed:

$$f(\mathbf{b}_i^* | \mathbf{f}, \sigma_i, \beta_{ii}) \propto \exp\left(-\frac{1}{2\sigma_i^2}(\mathbf{b}_i^* - \hat{\mathbf{b}}_i^*)'\mathbf{F}_i^{*\prime}\mathbf{F}_i^*(\mathbf{b}_i^* - \hat{\mathbf{b}}_i^*)\right), \quad i \le K, \tag{B.1}$$

where $\mathbf{F}_i^*$ is a $T \times i$ matrix consisting of the first $i$ columns of $\mathbf{F}$ and $\hat{\mathbf{b}}_i^*$ is the OLS estimator of the regression of $(r_i - \beta_{ii} f_i)$ on $(1, f_1, \ldots, f_{i-1})$. The conditional distributions of $\beta_{11}, \ldots, \beta_{KK}$ are also mutually independent, and each is truncated normal:

$$\beta_{ii} | \mathbf{f}, \mathbf{b}_i^* \sim N(\hat{\beta}_{ii}, \kappa_i), \quad \beta_{ii} > 0, \quad i = 1, \ldots, K, \tag{B.2}$$

where $\hat{\beta}_{ii} = \sum(r_{it} - \alpha_i - \beta_{i1} f_{1t} - \cdots - \beta_{i(i-1)} f_{(i-1)t}) f_{it} / \sum f_{it}^2$ and $\kappa_i = \sigma_i^2 / \sum f_{it}^2$. There is also an efficient method for implementing Equation (B.2). To draw $x > c$ from $x \sim N(a, b)$, a normal random variate truncated above $c$, let $x = a + \sqrt{b}y$; then $y \sim N(0, 1)$ and $y > d \equiv (c - a)/\sqrt{b}$. Following Geweke (1991a), $y$ is drawn efficiently by using (i) a simple normal rejection method if $a \le 0.5$; and (ii) an exponential rejection method if $a > 0.5$ (the exponential rejection works in two steps: draw $z$ and $u$ independently from the uniform distribution on $[0, 1]$ and compute both $y = a - 2\log z$ and $h = e^{-(y^2 - y - a^2 + a)}$; if $h < u$, reject and redo; otherwise, accept $y$ as the sample).

**References**

Amemiya, Y., 1985, *Advanced Econometrics*, Harvard University Press, Cambridge.

Anderson, T. W., and Y. Amemiya, 1988, "The Asymptotic Normal Distribution of Estimators in Factor Analysis under General Conditions," *Annals of Statistics*, 16, 759–771.

Black, F., 1972, "Capital Market Equilibrium with Restricted Borrowing," *Journal of Business*, 45, 444–454.

Breeden, D. T., M. R. Gibbons, and R. H. Litzenberger, 1989, "Empirical Tests of the Consumption Based CAPM," *Journal of Finance*, 44, 231–262.

Brown, S. J., 1989, "The Number of Factors in Security Returns," *Journal of Finance*, 46, 1247–1262.

Burmeister, E., and M. B. McElroy, 1981, "The Residual Market Factor, the APT, and the Mean-variance Efficiency," *Review of Quantitative Finance and Accounting*, 1, 27–49.

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

Chamberlain, G., and M. Rothschild, 1983, "Arbitrage, Factor Structure, and Mean Variance Analysis on Large Asset Markets," *Econometrica*, 52, 1281–1304.

Chen, N., 1983, "Some Empirical Tests of the Theory of Arbitrage Pricing," *Journal of Finance*, 38, 1393–1414.

Connor, G., 1984, "A United Beta Pricing Theory," *Journal of Economic Theory*, 34,13–31.

Connor, G., and R. A. Korajczyk, 1986, "Performance Measurement with the Arbitrage Pricing Theory: A New Framework for Analysis," *Journal of Financial Economics*, 15, 373–394.

Connor, G., and R. A. Korajczyk, 1988, "Risk and Return in an Equilibrium APT: An Application of a New Methodology," *Journal of Financial Economics*, 21, 255–289.

Connor, G., and R. A. Korajczyk, 1992, "The Arbitrage Pricing Theory and Multifactor Models of Asset Returns," working paper, Northwestern University.

Connor, G., and R. A. Korajczyk, 1993, "A Test for the Number of Factors in an Approximate Factor Model," *Journal of Finance*, 48, 1263–1291.

Dybvig, P. H., 1983, "An Explicit Bound on Individual Assets' Deviations from APT Pricing in a Finite Economy," *Journal of Financial Economics*, 12, 483–496.

Dybvig, P. H., and S. A. Ross, 1982, "Yes, APT is Testable," *Journal of Finance*, 40, 1173–1188.

Ferson, W. E., and C. R. Harvey, 1991, "The Variation of Economic Risk Premiums," *Journal of Political Economy*, 99, 385–415.

Gelfand, A. E., and A. F. M. Smith, 1990, "Sampling-based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85, 398–409.

Geman, S., and D. Geman, 1984, "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721–741.

Geweke, J. F., and K. J. Singleton, 1980, "Interpreting the Likelihood Ratio Statistic in Factor Models When Sample Size Is Small," *Journal of the American Statistical Association*, 75, 133–137.

Geweke, J. F., 1989, "Bayesian Inference in Econometric Models Using Monte Carlo Integration," *Econometrica*, 57, 1317–1339.

Geweke, J. F., 1991a, "Efficient Simulation from the Multivariate Normal and Student-$t$ distributions Subject to Linear Constraints," in E. M. Keramidas and S. M. Kaufman (eds.), *Proceedings of the Twenty-Third Symposium on the Interface*.

Geweke, J. F., 1991b, "Evaluating the Accuracy of Sampling-based Approaches to the Calculation of Posterior Moments," in J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (eds.), *Bayesian Statistics*, vol. 4, Oxford University Press, Oxford.

Gibbons, M. R., 1982, "Multivariate Tests of Financial Models: A New Approach," *Journal of Financial Economics*, 10, 3–27.

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

Grinblatt, M., and S. Titman, 1983, "Factor Pricing in a Finite Economy," *Journal of Financial Economics*, 12, 497–507.

Harvey, C. R., 1991, "World Price of Covariance Risk," *Journal of Finance*, 46, 111–157.

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

Kloek, T., and H. K. van Dijk, 1978, "Bayesian Estimates of Equation System Parameters: An Application of Integration by Monte Carlo," *Econometrica*, 46, 1–20.

Lehmann, B. N., and D. M. Modest, 1988, "The Empirical Foundations of the Arbitrage Pricing Theory," *Journal of Financial Economics*, 21, 213–254.

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, 13–37.

McCulloch, R., and P. E. Rossi, 1990, "Posterior, Predictive and Utility Based Approaches to Testing 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.

Muirhead, R. J., 1982, *Aspects of Multivariate Statistical Theory*, John Wiley, New York.

Roberts, G. O., and A. F. M. Smith, 1992, "Simple Conditions for the Convergence of the Gibbs Sampler and Metropolis-Hastings Algorithms," University of Cambridge Statistical Laboratory Research Report No. 92-30.

Roll, R. W., and S. A. Ross, 1980, "An Empirical Investigation of the Arbitrage Pricing Theory," *Journal of Finance*, 35, 1073–1103.

Ross, S. A., 1976, "The Arbitrage Theory of Capital Asset Pricing," *Journal of Economic Theory*, 13, 341–360.

Ross, S. A., 1977, "Risk, Return, and Arbitrage," in I. Friend and J. L. Bicksler (eds.), *Risk and Return in Finance*, vol. 1, Ballinger, Cambridge.

Seber, G. A. F., 1984, *Multivariate Observations*, John Wiley, New York.

Shanken, J., 1982, "The Arbitrage Pricing Theory: Is It Testable?" *Journal of Finance*, 37, 1129–1140.

Shanken, J., 1985, "Multi-beta CAPM or Equilibrium APT?: A Reply," *Journal of Finance*, 40, 1189–1196.

Shanken, J., 1987a, "A Bayesian Approach to Testing Portfolio Efficiency," *Journal of Financial Economics*, 19, 195–216.

Shanken, J., 1987b, "Multivariate Proxies and Asset Pricing Relations," *Journal of Financial Economics*, 19, 91–110.

Shanken, J., 1992, "The Current State of the Arbitrage Pricing Theory," *Journal of Finance*, 47, 1569–1574.

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

Tanner, M. A., and W. H. Wong, 1987, "The Calculation of Posterior Distributions by Data Augmentation," *Journal of the American Statistical Association*, 82, 528–540.

Tierney, L., 1991, "Markov Chains for Exploring Posterior Distributions," Technical Report No. 560, University of Minnesota School of Statistics. Forthcoming in *Annals of Statistics*.

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