Doron Avramov
University of Maryland

John C. Chao
University of Maryland

An Exact Bayes Test of Asset Pricing Models with Application to International Markets*

I.   Introduction

Financial economists have derived equilibrium asset pricing models such as the Capital Asset Pricing Model (CAPM) of Sharpe (1964) and Lintner (1965) and the consumption-oriented CAPM of Breeden (1979). Subsequent work (e.g., Black, Jensen, and Scholes 1972; Fama and MacBeth 1973; Breeden, Gibbons, and Litzenberger 1989) examined the empirical performance of unconditional versions of these asset pricing models. The empirical tests met with mixed results. More recent work examined versions of pricing models that incorporate lagged variables such as the dividend yield. Studying conditional models has both theoretical and empirical appeal. Theoretically, Hansen and Richard (1987) show that, even if the unconditional CAPM fails, the conditional CAPM could be perfectly valid. In addition, Campbell (1996) shows that any instrument that forecasts future market returns or labor income

This paper develops and implements an exact finite-sample test of asset pricing models with time-varying risk premia using posterior probabilities. The strength of our approach is that it allows multiple conditional asset pricing specifications, both nested and nonnested, to be tested and compared simultaneously. We apply our procedure to international equity markets by testing and comparing the international Capital Asset Pricing Model (ICAPM) and conditional ICAPM versions of Fama and French (1998). The empirical evidence suggests that the best performing model is the ICAPM with the value premium constructed based on global earnings-to-price ratio.

growth could be a priced factor for asset returns. Empirically, Lettau and Ludvigson (2001), among others, demonstrate that asset pricing models with conditioning information explain a substantially larger fraction of the cross-sectional variation in average returns than unconditional models.

The asset pricing literature also proposed various econometric approaches to testing asset pricing models. In particular, a multivariate finite-sample test was introduced by Gibbons, Ross, and Shanken (1989). In addition, Shanken (1987), Harvey and Zhou (1990), McCulloch and Rossi (1991), and Geweke and Zhou (1996) all developed small-sample Bayesian tests. These tests, however, are designed to test unconditional models and not models that incorporate conditioning information. With conditioning information, asymptotically valid asset pricing tests are available, such as the cross-sectional approach of Fama and MacBeth (1973), the generalized method of moments (GMM) of Hansen (1982), and the distance measure of Hansen and Jagannathan (1997), but finite sample tests are yet to be developed. This paper fills a gap in developing and implementing an exact finite-sample test of asset pricing models with conditioning information. Our test allows risk premia and potential asset pricing misspecification to vary predictably in response to changing economic conditions.

To our knowledge, this work provides the first exact finite-sample test of asset pricing models with conditioning information. The method introduced here allows one to simultaneously test and compare the performance of multiple-asset pricing specifications, both nested and non-nested. There are other advantages to using our approach. For example, the test can be applied without explicit knowledge of the long-run (low-frequency) properties of the stochastic processes that generate excess stock returns, underlying risk factors, and information variables. Thus, the decision rule for evaluating model performance is not affected by the possible presence of a unit root in lagged variables such as the dividend yield.

Bayesian methods of hypothesis testing and model selection, such as the one employed here, lead to a test statistic that comprises both an in-sample goodness-of-fit measure and a penalty term for model complexity (see Kass and Raftery 1995). Since a higher-dimensional model that nests a smaller model always fits the data at least as well, if not better, as the smaller model, the presence of a penalty term in the test statistic guards against overfitting the data, which may occur if models are evaluated only on the basis of goodness of fit. Thus, our Bayesian procedure that penalizes model complexity helps prevent the inclusion of excess, useless factors in linear factor models. Indeed, implicit in the trade-off between goodness of fit and model complexity is the assertion that, since all equilibrium pricing models are mere approximations of the phenomenon under study, the use of a larger, more complex pricing model is justified over a

more parsimonious one only if the former can be shown to be significantly better at fitting the data than the latter.

To illustrate the broad applicability of our finite-sample test, we apply our procedure in an international asset pricing context. We study five competing asset pricing specifications. These are the conditional versions of the international CAPM (ICAPM) and four ICAPM's of Fama and French (1998), each of which consists of the global market portfolio and the global value premium. As in Fama and French, the global value premium is formed as the return differential between (1) high and low book-to-market stocks, (2) high and low earnings-yield stocks, (3) high and low cash-flow-to-price stocks, and (4) high and low dividend-yield stocks. Whereas, under the CAPM, the global market portfolio is conditionally mean-variance efficient, under the Fama and French model, some combination of the global market and global value premium portfolio lies on the conditional minimum-variance boundary of risky assets (e.g., Roll 1977; Hansen and Richard 1987; and Huberman and Kandel 1987).

In the empirical implementation, each of the 10 data generating models, which consist of the restricted and unrestricted versions of the 5 asset pricing specifications noted earlier, are assigned equal prior probability. When such "neutral" initial beliefs are combined with the sample, the updated beliefs strongly support the international CAPM and ICAPM. In most cases, the posterior probabilities in favor of specifications that allow for asset pricing misspecification are substantially lower than those for models where asset pricing restrictions are assumed to hold exactly. An interesting result coming out of the cross-model comparison is that the best performing asset pricing specification is the Fama-French model with value premium constructed based on earnings yield. Remarkably, this model outperforms the other Fama-French models, even in some cases where the test assets are portfolios that are not sorted on earnings yield but on other equity characteristics, such as book-to-market, cash-flow-to-price, or dividend-to-price ratios, where a priori one would expect competing models to have a decisive advantage. In contrast to the earnings yield specification, the book-to-market-based ICAPM does not perform nearly as well in our series of horse races, recording zero or near zero posterior probabilities for all but a few of the test assets studied.

The remainder of the paper proceeds as follows. Section II introduces a generic form of conditional asset pricing models, discusses its testable implications, and describes our Bayesian approach to hypothesis testing, model selection, and model combination. Section III presents exact formulae for both restricted and unrestricted factor model specifications and states our decision rules for selecting among competing models. Section IV presents empirical results, and we offer some concluding remarks and ideas for future research in Section V. Unless otherwise noted all derivations are presented in the appendix.

II. Evaluating Conditional Asset Pricing Models

A. Deriving Testable Restrictions

To evaluate asset pricing models with conditioning information, we first specify the dynamics of stock returns, underlying factors, and information variables. In particular, let r_t denote an N-vector of returns on test assets in excess of the risk-free rate; let f_t denote a K-vector of returns on portfolio-based factors, such as a claim to total wealth; and let z_t denote an M-vector of conditioning information, such as the dividend yield. We describe the multivariate form of the data-generating process for excess returns, factors, and predictive variables as

$$Y = XA + U, \tag{1}$$

where

$$Y = \begin{bmatrix} r_1', f_1', z_1' \\ \vdots \\ r_T', f_T', z_T' \end{bmatrix}, \tag{2}$$

where $X = [\iota_T, Z_{-1}], Z_{-1} = [z_0, \ldots, z_{T-1}]', \iota_T$ is a $T \times 1$ vector of ones, T is the sample size, and U is a $T \times (N + K + M)$ matrix of the regression disturbances. It is assumed that the variance covariance matrix of the disturbances (denoted by Σ) is constant over time. For the analysis that follows, we make the following partitions $Y = [R, F, Z], A = [A_R, A_F, A_Z],$ and $U = [U_R, U_F, U_Z],$ where $R = [r_1, \ldots, r_t]', F = [f_1, \ldots, f_T]',$ and $Z = [z_1, \ldots, z_T]'.$ Asset pricing models impose testable restrictions on the multivariate system in equation (1). Next, we derive those asset pricing restrictions and develop a novel procedure for testing the restrictions.

Consider now a beta factor model obeying the form

$$E(r_t|z_{t-1}) = \alpha(z_{t-1}) + \beta E(f_t|z_{t-1}), \tag{3}$$

where $\alpha(z_{t-1})$ is an N-vector of mispricing across assets with respect to benchmark portfolios under consideration, and β is an $N \times K$ matrix of factor loadings. It is assumed that $\alpha(z_{t-1})$ linearly depends on information variables observed at time $t - 1$, thereby taking the form $\alpha(z_{t-1}) = G'x_{t-1},$ where G is an $(M + 1) \times N$ asset mispricing matrix and $x_{t-1} = [1, z_{t-1}']'.$ The matrix of factor loadings is $\beta = [\text{Cov}\{r_t, f_t'|z_{t-1}\}] \times [\text{Cov}\{f_t, f_t'|z_{t-1}\}]^{-1}.$ As noted earlier, the matrix Σ is assumed constant over time, thereby the conditional beta does not vary with time and is equal to the fixed quantity $\Sigma_{RF}\Sigma_{FF}^{-1},$ where $\Sigma_{RF} = \text{Cov}\{r_t, f_t'|z_{t-1}\}$ and $\Sigma_{FF} = \text{Cov}\{f_t, f_t'|z_{t-1}\}.$ That is, the conditional beta can be directly computed from the distinct partitions of the covariance matrix Σ. In adopting

a constant beta framework, we follow the work of Campbell (1987) and others. We note that, although treating the conditional beta as fixed has potential costs, such a treatment has the advantage that it leads to a more parsimonious specification that avoids the need to estimate additional set of parameters characterizing the evolution of beta variation. Moreover, Ghysels (1998) shows that, if the dynamics of beta are misspecified, serious pricing errors may result. Those pricing errors could be larger than those with a constant beta model.

It should also be noted that, while this paper adopts a beta pricing representation in conducting empirical analysis of asset pricing models, an alternative approach, the so-called stochastic discount factor (SDF) approach, instead takes a pricing kernel representation of pricing models. An example of an interesting recent paper that takes this approach is Lettau and Ludvigson (2001), which documents the importance of scaling the pricing kernel parameters by information variables. As has been argued forcibly by Kan and Zhou (1999, 2003), a main advantage of the beta pricing approach is that, because it incorporates a fully specified model, it is at least as efficient as the SDF approach, if not more so. Indeed, within a conditional framework, full specification of the dynamics of asset returns and their dependence on information variables is likely to lead to rather substantial efficiency gains. In addition, we note that the constant beta paradigm is not at odds with time-varying pricing kernel parameters. To illustrate this point, consider a pricing kernel obeying the linear form $\xi_{t+1} = a(z_t) + b(z_t)' f_{t+1}$. Following Cochrane (1996), the equivalence between a beta pricing model and a pricing kernel specification implies that

$$a(z_t) = 1/r_{ft}[1 + \lambda(z_t)'\Sigma_{FF}^{-1}\lambda(z_t)], \tag{4}$$

$$b(z_t) = -1/r_{ft}\Sigma_{FF}^{-1}\lambda(z_t), \tag{5}$$

where r_{ft} is the 1-period conditionally riskless T-bill and $\lambda(z_t) = \alpha_F + a_F z_t$, and where the $K \times 1$ vector α_F and the $K \times M$ matrix a_F are the corresponding partitions of A_F. Observe that time-varying risk premia $\lambda(z_t)$ and time-varying riskless rate imply time-varying pricing kernel parameters, even when beta is fixed.

When factors are portfolio based, one can obtain the following relationship between several blocks in the regression coefficients A and Σ and the mispricing matrix G:

$$A_R = G + A_F\Sigma_{FF}^{-1}\Sigma_{FR}. \tag{6}$$

This relation is similar to that derived by Campbell (1987), except that, here, factors are prespecified as opposed to being latent, as in the framework of Campbell. Observe from equation (6) that, if expected return

variation is due to common risk factors ($G = 0$), the predictable component of returns is a linear transformation of the predictable component of returns on factor-mimicking portfolios, with the transformation matrix being equal to beta. This relationship has an intuitive appeal. If all the assets in the economy are priced by a lower dimensional set of K benchmark positions, then the problem of exploring the predictability of returns on N various securities having random payoffs boils down to exploring the predictability of benchmark asset returns. Differences in predictability across assets, reflected through distinct columns in the matrix A_R, are attributable to different loadings on the benchmark positions. The presence of equity mispricing breaks this relationship, in that cross-sectional differences in predictability are attributable not only to different factor loadings but also to model mispricing, whose magnitude can differ across the test assets.

A hypothesis that favors the prevalence of an asset pricing model, such as the CAPM or the Fama and French (1993, 1996, 1998) multifactor models, restricts all elements of $\alpha(z_{t-1})$ to be equal to 0 at every time period. In that case, the error terms obtained by regressing excess returns on factor-mimicking portfolio returns are not priced. To test the zero-intercept restrictions, we formulate asset pricing restrictions as a sharp null hypothesis:

$$H_0 : G = 0 \text{ if and only if } \alpha(z_{t-1}) = 0 \text{ for all } t,$$

$$H_1 : G \neq 0 \text{ if and only if there is a } t \text{ such that } \alpha(z_{t-1}) \neq 0. \quad (7)$$

Under the null hypothesis, the time variation in expected returns is driven by economywide risk factors only. Under the alternative, time-varying expected returns can be explained, among others, by under- or overpricing. For example, DeBondt and Thaler (1987), Lakonishok, Shleifer, and Vishny (1994), and Haugen (1995) attribute the value premium to overreaction to corporate performance. The notion is that financial market participants undervalue distressed stocks and overvalue growth stocks. When mispricing is corrected, high-value stocks have high returns relative to growth stocks. This correction governs stock return predictability.

B. A Bayesian Method for Hypothesis Testing and Model Selection

We evaluated the validity of the multivariate restrictions formulated in equation (7) using a hypothesis testing approach based on the Bayesian posterior odds (or, equivalently, the Bayes factor when prior model probabilities are taken to be equal, as explained later). To describe this approach, we consider the case where there are L competing models (or hypotheses), denoted by

$$\mathcal{M}_i : \text{data has density } f(D|\theta_i, \mathcal{M}_i), \quad i = 1, \dots, L, \quad (8)$$

where D stands for the data and $\theta_i \in \Theta_i$ is the unknown parameter vector of model $\mathcal{M}_i$. Hypothesis testing from a Bayesian perspective proceeds by computing the posterior probabilities with each of the alternative models via Bayes's theorem to yield the formula

$$P(\mathcal{M}_i|D) = \frac{\pi(\mathcal{M}_i)m(D|\mathcal{M}_i)}{\sum_{i=1}^{L} \pi(\mathcal{M}_l)m(D|\mathcal{M}_l)}, \tag{9}$$

where $\pi(\mathcal{M}_l)$ gives the prior model probability of $\mathcal{M}_l$ and $m(D|\mathcal{M}_l)$ denotes the marginal density (or the marginal likelihood) of the data under $\mathcal{M}_l$. The marginal likelihood can be represented by the integral expression

$$m(D|\mathcal{M}_l) = \int_{\Theta_l} f(D|\theta_l, \mathcal{M}_l)\pi(\theta_l|\mathcal{M}_l)d\theta_l, \tag{10}$$

with $\pi(\theta_l|\mathcal{M}_l)$ denoting the (proper) prior density of the parameter vector θ_l under model $\mathcal{M}_l$. Pairwise comparison between two models (say, $\mathcal{M}_i$ and $\mathcal{M}_j$ for $i \neq j$) then can be conducted using the posterior odds ratio:

$$\frac{P(\mathcal{M}_i|D)}{P(\mathcal{M}_j|D)} = \frac{\pi(\mathcal{M}_i)m(D|\mathcal{M}_i)}{\pi(\mathcal{M}_j)m(D|\mathcal{M}_j)}. \tag{11}$$

Looking at expression (11), we note that the factor $\pi(\mathcal{M}_i)/\pi(\mathcal{M}_j)$ is simply the prior odds ratio of $\mathcal{M}_i$ to $\mathcal{M}_j$. We can also define B_{ij}, the Bayes factor for testing $\mathcal{M}_i$ against $\mathcal{M}_j$, as the posterior odds ratio over the prior odds ratio, so that in light of equation (11), the Bayes factor is simply the ratio of the respective marginal likelihoods. Hence, the Bayes factor can be interpreted as the odds for $\mathcal{M}_i$ relative to $\mathcal{M}_j$ after the prior odds ratio has been updated by the information in the data. Note further that in the typical empirical situation where the prior probability is taken to be the same for each model, that is, where $\pi(\mathcal{M}_1) = \pi(\mathcal{M}_2) = \ldots = \pi(\mathcal{M}_l)$, the Bayes factor is equivalent to the posterior odds ratio.

A difficulty with computing the Bayesian factor is that, in the case where the parameter space Θ_l is unbounded, one cannot take the prior density $\pi(\theta_l|\mathcal{M}_l)$ in the marginal likelihood expression (10) to be that of a (improper) noninformative prior, such as the (improper) uniform prior (see Kass and Raftery 1995). To see the problem that the use of an improper prior creates, suppose we specify the uniform prior densities $\pi(\theta_i|\mathcal{M}_i) = \pi_i$ and $\pi(\theta_j|\mathcal{M}_j) = \pi_j$ for models $\mathcal{M}_i$ and $\mathcal{M}_j$, respectively. Then, the Bayes factor is

$$B_{ij} = \frac{\pi_i \int_{\Theta_i} f(D|\theta_i, \mathcal{M}_i)d\theta_i}{\pi_j \int_{\Theta_j} f(D|\theta_j, \mathcal{M}_j)d\theta_j}, \tag{12}$$

given that π_i and π_j are constants not depending on θ_i and θ_j. However, since the priors are improper (i.e., their densities integrate to infinity over the parameter spaces Θ_i and Θ_j), there are no unique normalization constants for these densities. Hence, we can just as well take as our prior densities the alternative uniform densities $\pi(\theta_i|\mathcal{M}_i) = c_i\pi_i$ and $\pi(\theta_j|\mathcal{M}_j) = c_j\pi_j$, for constants c_i and c_j with $c_i \neq c_j$. The last prior specifications lead to

$$B_{ij}^* = \frac{c_i\pi_i \int_{\Theta_i} f(D|\theta_i, \mathcal{M}_i)d\theta_i}{c_j\pi_j \int_{\Theta_i} f(D|\theta_j, \mathcal{M}_j)d\theta_j} \neq B_{ij}, \qquad (13)$$

for $c_i \neq c_j$. Thus, there is an indeterminacy with respect to the Bayes factor specification.

A common solution to the problem described previously, in the case where the researcher does not wish to specify (proper) subjective prior densities for the model parameters, is to split the total sample of data D into two subsamples: a training sample, denoted $D(\tau)$, and a primary sample, denoted $D(-\tau)$. In the time series context we study here, there is a natural ordering of the data, so we can think of the training sample as comprising the first τ observations of the data, or equivalently, comprising data up to time $t = \tau$; whereas the primary sample comprises observations from $t = \tau + 1$ to T. The strategy commonly employed is to combine a (possibly improper) noninformative prior density, say, $\pi^{\mathcal{N}}(\theta_i|\mathcal{M}_i)$, with data from the training sample to obtain a (proper) posterior density:

$$\pi[\theta_i|D(\tau), \mathcal{M}_i] = \frac{f[D(\tau)|\theta_i, \mathcal{M}_i]\pi^{\mathcal{N}}(\theta_i|\mathcal{M}_i)}{\int_{\Theta_i} f[D(\tau)|\theta_i, \mathcal{M}_i]\pi^{\mathcal{N}}(\theta_i|\mathcal{M}_i)d\theta_i}, \qquad (14)$$

where the size of the training sample must be chosen such that the density $\pi[\theta_i|D(\tau), \tau, \mathcal{M}_i]$ is indeed proper. This proper (posterior) density then is used as a prior density and combined with data from the primary sample to compute the posterior model probabilities for model comparison. Thus, analogous to expressions (9) and (10), we obtain

$$P_\tau(\mathcal{M}_i|D) = \frac{\pi(\mathcal{M}_i)m[D(-\tau)|D(\tau), \mathcal{M}_i]}{\sum_{l=1}^{L}\pi(\mathcal{M}_l)m[D(-\tau)|D(\tau), \mathcal{M}_l]}, \qquad (15)$$

where

$$m[D(-\tau)|D(\tau), \mathcal{M}_l] = \int_{\Theta_l} f[D(-\tau)|\theta_l, D(\tau), \mathcal{M}_l]\pi[\theta_l|D(\tau), \mathcal{M}_l]d\theta_l,$$

$$(16)$$

and where the subscript τ on the posterior model probability $P_\tau(\mathcal{M}_i|D)$ emphasizes that this probability depends on the training sample size, as denoted by τ.

Implementation of this training sample approach to Bayesian model selection requires a specification of τ. Rather than specifying a particular value of τ, such as the minimal size of the sample needed to make (14) a proper density, our approach in this paper is to average the posterior model probabilities expressed in equation (15) across a range of τ values, say, $\tau = T_*, T_* + 1, \ldots, T^*$, where T_* is the value of the split point that yields the minimal training sample and $T^* = [T/2]$ is the midpoint of the total sample, where the notation $[x]$ denotes the integer part of x. This leads to the average posterior model probability for model $\mathcal{M}_i$ given by the formula

$$
P_{\mathcal{M}_i}^{\text{AVG}} = \frac{1}{T^* - T_* + 1} \sum_{\tau=T_*}^{T^*} P_\tau(\mathcal{M}_i|D)
$$

$$
= \frac{1}{T^* - T_* + 1} \sum_{\tau=T_*}^{T^*} \left\{ \frac{\pi(\mathcal{M}_i) m[D(-\tau)|D(\tau), \mathcal{M}_i]}{\sum_{l=1}^{L} \pi(\mathcal{M}_l) m[D(-\tau)|D(\tau), \mathcal{M}_l]} \right\}, \quad (17)
$$

where $P_\tau(\mathcal{M}_i|D)$ and $m[D(-\tau)|D(\tau), \mathcal{M}_i]$ are as defined in equations (15) and (16). The advantage of averaging across training samples is that it leads to a more stable procedure relative to that based on a particular choice of τ, since empirical results may be sensitive to the choice of τ. Note also that the odds ratio for making pairwise comparisons of models can be constructed by simply taking the ratio of expression (17) for two competing models.

It should be pointed out that our approach of averaging across posterior model probabilities with different training sample size τ follows the spirit of the intrinsic Bayes factor introduced by Berger and Pericchi (1996). In that influential paper, procedures were proposed for averaging Bayes factors across different training samples of minimal size. However, an important distinction between our paper and that of Berger and Pericchi (1996) is that whereas Berger and Pericchi focus only on the case of cross-sectional data, our work is set explicitly within a time-series framework. Hence, our procedure here can be viewed as extending the work of Berger and Pericchi to the time series context.

Another crucial task in the implementation of Bayesian procedures for hypothesis testing, model selection, and model comparison is the computation of the marginal likelihood for each of the rival models. With regard to our procedure, we need to calculate the conditional marginal likelihood of the data, as given by expression (16). It turns out that an attractive feature of our approach is that an exact analytical formula for expression (16) can be obtained both in the case where the model is subject to restrictions

of an asset pricing model, that is, $G = 0$ in equation (7), and where model restrictions are disregarded.

The availability of exact formula for the marginal likelihoods is an important advantage of our approach, since this allows exact posterior model probabilities to be computed without asymptotic approximation. In fact, growing evidence suggests that the asymptotically justified classical hypotheses test performs poorly in finite samples. Ferson and Forester (1994), for example, provided Monte Carlo evidence suggesting that, with respect to testing CAPMs, GMM-based tests may overreject for large systems with many assets and suffer from low power for null models with a small number of risk factors. In contrast, finite sample tests of restrictions of the form given in equation (7) within a multivariate linear system cannot be readily implemented within a classical framework, even if the residuals in the regression of excess returns on a constant intercept and a set of K factors are assumed to be Gaussian. In particular, note that the exact finite sample distribution of the classical likelihood ratio test for testing the restriction (7) is complicated; and as far as we know, no critical values for such a test has ever been tabulated.

Note also that, when the inference is conducted conditional on available data, one need not impose further conditions on the low-frequency (or long-run) behavior of the system given by equation (1). See Sims (1988), Sims and Uhlig (1991), and Phillips and Ploberger (1996) for further discussion on this point. In particular, we make no further assumptions about the order of integration of the time series variables r_t, f_t, and z_t, so they may be either stationary processes or possibly unit root processes. We believe this is an advantage of the approach taken here, as we can proceed with inference about equilibrium asset pricing theories without worrying about the difficult issues involved in pretesting the data for unit roots and cointegration or possibly having to transform the data to "induce" stationarity. Indeed, information variables such as dividend yield, term spread, and default spread are highly persistent and may involve a unit root.

Furthermore, our approach could be attractive even from a sample theoretic or frequentist standpoint. It is well known that, given a fixed significance level, the classical approach to hypothesis testing (say, one based on the Wald or the likelihood ratio) is consistent only in the sense that the probability of committing a Type II error vanishes as the sample size approaches infinity; however, the probability of a Type I error for such a test does not approach zero, even in large samples. On the other hand, Bayesian tests based on the posterior odds or the Bayes factor can be shown under general regularity conditions to be completely consistent, so that the probability of both Type I and Type II errors vanishes asymptotically.

Before leaving this section to discuss the explicit implementation of our procedure, we want to note that, while many papers have tested asset pricing models, none pursues a framework similar to that developed here.

Shanken (1987), Harvey and Zhou (1990), and McCulloch and Rossi (1991) all developed posterior odds ratios to test the intercept restriction as implied by factor based models. In addition, Geweke and Zhou (1996) entertained a finite-sample Bayesian approach for testing asset pricing models when factors are unobserved. However, these studies assumed that expected returns are constant. In contrast, here, the intercept in the regression of excess returns on asset pricing factors and the price of beta risk could vary in response to changing economic conditions. Incorporating time-varying expected returns is motivated on both theoretical and empirical grounds. Related to this work, Avramov (2002) derives posterior probability in a predictive regression framework but disregarding pricing model restrictions. In addition, Avramov (2004) exploits the asset pricing restrictions presented in equation (7) to form informative prior beliefs about the extent of return predictability in an investment context.

III.   An Exact Test of Asset Pricing Models with Conditioning Information

Let $\mathcal{M}_R$ denote the multivariate model (1) subject to the asset pricing restrictions (7) and let $\mathcal{M}_U$ denote the unrestricted version of (1). Then, in the case where $\pi(\mathcal{M}_R) = \pi(\mathcal{M}_U) = 1/2$, we obtain directly from expression (17) the following formulas for $P_{\mathcal{M}_R}^{\text{AVG}}$ and $P_{\mathcal{M}_U}^{\text{AVG}}$:

$$P_{\mathcal{M}_R}^{\text{AVG}} = \left(\frac{1}{q}\right) \sum_{\tau=T_*}^{[T/2]} \left\{ \frac{m[D(-\tau)|D(\tau), \mathcal{M}_R]}{m[D(-\tau)|D(\tau), \mathcal{M}_R] + m[D(-\tau)|D(\tau), \mathcal{M}_U]} \right\}$$

(18)

and

$$P_{\mathcal{M}_U}^{\text{AVG}} = \left(\frac{1}{q}\right) \sum_{\tau=T_*}^{[T/2]} \left\{ \frac{m[D(-\tau)|D(\tau), \mathcal{M}_R]}{m[D(-\tau)|D(\tau), \mathcal{M}_R] + m[D(-\tau)|D(\tau), \mathcal{M}_U]} \right\},$$

(19)

where, in these expressions, $q = [T/2] - T_* + 1$ and where $m[D(-\tau)|D(\tau), \mathcal{M}_R]$ and $m[D(-\tau)|D(\tau), \mathcal{M}_U]$ denote the marginal likelihoods under $\mathcal{M}_R$ and $\mathcal{M}_U$, respectively.

In the appendix, we give detailed derivations of exact expressions for $m[D(-\tau)|D(\tau), \mathcal{M}_R]$ and $m[D(-\tau)|D(\tau), \mathcal{M}_U]$. Specifically, the appendix contains four propositions. Proposition 1 (3) derives informative prior beliefs about unknown parameters pertaining to the restricted (unrestricted) specification. Proposition 2 (4) derives the marginal likelihood of the restricted (unrestricted) specification. Following Harvey and Zhou (1990), McColluch and Rossi (1991), Kandel, McCulloch, and Stambaugh (1995), and Geweke and Zhou (1996), we make the standard Gaussian assumption on the errors of the multivariate regression

model (1). Of course, in light of the recent work by Tu and Zhou (2004), it would be of interest to explore the robustness of the Gaussian assumption here by extending our approach to the case where the error distribution belongs to the multivariate t family.

Given analytic expressions for the marginal likelihoods, the average posterior model probabilities $P_{\mathcal{M}_R}^{\text{AVG}}$ and $P_{\mathcal{M}_U}^{\text{AVG}}$ can be computed easily using expressions (18) and (19). It follows that a Bayesian test of asset pricing restrictions can be implemented using either the odds ratio $P_{\mathcal{M}_R}^{\text{AVG}}/P_{\mathcal{M}_U}^{\text{AVG}}$ or its natural log transform $\ln(P_{\mathcal{M}_R}^{\text{AVG}}) - \ln(P_{\mathcal{M}_U}^{\text{AVG}})$. Taking a loss function that is symmetric with respect to Type I and II errors, the decision rule for choosing the restricted model $\mathcal{M}_R$ over the unrestricted model $\mathcal{M}_U$ can be stated as either

$$\text{Choose } \mathcal{M}_R \text{ over } \mathcal{M}_U \text{ if } P_{\mathcal{M}_R}^{\text{AVG}}/P_{\mathcal{M}_U}^{\text{AVG}} > 1 \qquad (20)$$

or

$$\text{Choose } \mathcal{M}_R \text{ over } \mathcal{M}_U \text{ if } \ln\left(P_{\mathcal{M}_R}^{\text{AVG}}\right)\Big/\ln\left(P_{\mathcal{M}_U}^{\text{AVG}}\right) > 0. \qquad (21)$$

Moreover, note that since $P_{\mathcal{M}_R}^{\text{AVG}}$ and $P_{\mathcal{M}_U}^{\text{AVG}}$ are (discrete) probabilities, satisfying the conditions $P_{\mathcal{M}_R}^{\text{AVG}} \geq 0, P_{\mathcal{M}_U}^{\text{AVG}} \geq 0$, and $P_{\mathcal{M}_R}^{\text{AVG}} + P_{\mathcal{M}_U}^{\text{AVG}} = 1$, we also have the decision rule:

$$\text{Choose } \mathcal{M}_R \text{ over } \mathcal{M}_U \text{ if } P_{\mathcal{M}_R}^{\text{AVG}} > \frac{1}{2}. \qquad (22)$$

(See Zellner 1971, pp. 294–97, for further discussion of Bayesian decision rules under symmetric loss function.)

While expressions for the marginal likelihoods as given by equations (A.15) and (A.25) in the appendix may appear to be cumbersome our method, in fact, is easily implementable using a simple Matlab code, which is available from us on request.

IV. An Empirical Study: The Case of International Markets

To illustrate the broad applicability of our approach, we apply our procedure in an international asset pricing context. We study five competing models, the international CAPM and four ICAPMs, which are the conditional versions of the unconditional models studied by Fama and French (1998). Inherent in the two-factor model is the assertion that the value premium is compensation for a global risk factor missed by the global market portfolio. Under the CAPM restriction, the global market portfolio is conditionally mean-variance efficient. Under the two-factor model restriction, some combination of the global market portfolio and the global value premium lies on the conditional mean-variance frontier. In addition to testing each restricted model against its unrestricted counterpart, we

also conduct direct comparisons of all the models together on the basis of their posterior model probabilities. Note that such a comparison cannot be done easily within a classical statistical framework, since the different ICAPM versions are mutually nonnested. In contrast, our procedure is especially suitable for comparing both nested and nonnested models.

In particular, the empirical analysis compares and evaluates asset pricing models on the basis of their posterior probabilities. Any candidate model is represented by two competing return generating models. One, the pricing model, implies zero intercepts in a multivariate regression of excess returns on portfolio based factors. The other allows for the presence of asset pricing misspecification, the extent of which may vary in response to changing economic conditions. The appendix describes the derivation of the posterior probabilities for both the restricted and unrestricted specifications. The posterior probability indicates the odds that stock returns are generated either by the restricted model under $G = 0$ in equation (7) or by its unrestricted counterpart. An interesting question is how to interpret the magnitude of the posterior odds ratio or the Bayes factor if the prior odds ratio is assumed to be unity. Jeffreys (1961) suggests a qualitative interpretation. According to him, the evidence against the alternative or in favor of the null is as follows. If the Bayes factor is between 1 and 3.2, the evidence does not justify more than a bare mention; if it is between 3.2 and 10, the evidence is substantial; if it is between 10 to 100, the evidence is strong; if it is greater than 100, the evidence is decisive.

For the empirical implementation, we make three assumptions often used in an international framework. First, world equity markets are perfectly integrated, as in Korajczyk and Viallet (1989), Harvey (1991), and Dumas and Solnik (1995). Second, investors are not concerned about deviations from purchasing power parity (PPP). Otherwise, they would hedge against foreign exchange risk, and as a result, additional risk premia corresponding to the covariances of returns with exchange rates would emerge. For example, Solnik (1974) shows that, when investors in international markets face exchange risk, such risk should be priced, even in a world otherwise similar to the unconditional CAPM. In our setting, foreign exchange risk is not priced separately from the market risk, since the PPP assumption implies that investors do not perceive real changes in relative prices. Third, as in Korajczyk and Viallet (1989), Harvey (1991), and Fama and French (1998), our empirical implementation takes the view of a global investor who cares about U.S. dollar returns.

A. Data

We study returns on market, value, and growth portfolios for the United States and 12 major EAFE (Europe, Australia, and the Far East) countries over the period 1975–2000. We adhere to the Fama and French's (1998) portfolio definitions. HB/M and LB/M denote high and low book-to-market portfolios, respectively; HE/P and LE/P denote high and low

TABLE 1 **Monthly Dollar Returns in Excess of U.S. T-Bill Rate for Market, Value, and Growth Portfolios: 1975–2000**

	Market	HB/M	LB/M	HE/P	LE/P	HC/P	LC/P	HD/P	LD/P
United States	.80	1.08	.74	1.11	.69	1.03	.72	.90	.82
	(4.45)	(4.32)	(4.90)	(4.66)	(4.94)	(4.28)	(4.99)	(3.82)	(5.19)
Japan	.61	1.07	.27	.96	.19	1.05	.20	.85	.36
	(6.67)	(7.15)	(7.09)	(6.49)	(7.27)	(7.03)	(6.76)	(7.17)	(7.01)
United Kingdom	1.09	1.27	.97	1.26	1.04	1.37	1.02	1.13	.96
	(6.79)	(7.28)	(7.04)	(7.00)	(6.98)	(7.17)	(6.99)	(6.71)	(7.07)
France	.87	1.19	.77	1.09	.71	1.28	.76	1.15	.54
	(6.66)	(7.56)	(6.76)	(7.64)	(6.93)	(7.73)	(6.88)	(6.91)	(7.23)
Germany	.68	1.09	.65	.73	.74	1.00	.37	.79	.64
	(5.73)	(6.08)	(6.15)	(6.00)	(6.19)	(5.79)	(5.90)	(5.67)	(6.27)
Italy	.64	.61	.72	.60	.76	.88	.21	.90	.43
	(7.80)	(8.62)	(7.79)	(8.88)	(7.99)	(8.69)	(8.05)	(8.47)	(8.02)
Netherlands	1.01	1.07	.87	1.07	.79	.74	.87	1.10	.67
	(5.13)	(6.80)	(5.40)	(5.84)	(5.79)	(7.17)	(5.60)	(5.66)	(6.18)
Belgium	.84	.97	.70	1.06	.79	1.12	.79	1.02	.70
	(5.52)	(6.42)	(5.75)	(5.69)	(5.84)	(6.21)	(6.15)	(5.91)	(5.96)
Switzerland	.76	1.09	.68	.83	.73	.71	.69	.91	.67
	(5.34)	(5.87)	(5.47)	(5.78)	(5.79)	(6.05)	(5.96)	(5.81)	(5.67)
Sweden	.98	1.20	1.00	1.32	1.04	1.08	1.09	1.26	1.00
	(6.57)	(7.93)	(6.73)	(7.29)	(7.05)	(7.73)	(7.06)	(7.27)	(6.91)
Australia	.63	1.07	.39	1.10	.40	1.19	.27	.97	.46
	(6.88)	(6.91)	(7.68)	(6.42)	(8.00)	(7.11)	(7.82)	(6.36)	(7.96)
Hong Kong	1.44	1.52	1.25	1.71	1.36	1.76	1.35	1.76	1.52
	(9.46)	(11.46)	(8.88)	(10.36)	(9.47)	(9.72)	(9.95)	(9.60)	(9.95)
Singapore	.84	1.34	.70	.98	.81	.93	.54	.67	.74
	(8.22)	(10.72)	(7.93)	(8.62)	(9.00)	(8.37)	(8.22)	(8.47)	(9.25)

NOTE.—Value and growth portfolios are formed on book-to-market equity (B/M), earnings to price (E/P), cash flow to price (C/P), and dividend to price (D/P) in the same manner as described in Fama and French (1998). Also following Fama and French (1998), we denote value (high) and growth (low) portfolios using the letters H and L, respectively. The first number reported in each cell of table 1 is the annual average return of the given portfolio for each country, while the numbers in parentheses are the t-statistics for testing the null that the average return is no different from 0.

earnings yield portfolios, respectively; HC/P and LC/P denote high and low cash-flow-to-price portfolios, respectively; and HD/P and LD/P denote high and low dividend yield portfolios, respectively. All portfolio returns are obtained from Ken French's Web site.

Following previous studies (e.g., Harvey 1991; Dumas and Solnik 1995), our list of instruments includes five variables: the excess rate of return on the world index lagged 1 month, the January dummy, the U.S. term structure slope, the dividend yield on the U.S. value-weighted index, and the 1-month rate of interest on a Eurodollar deposit. Indeed, some of the instruments used are local. This is due to Harvey (1991) and others, who show that U.S. instruments have some power in predicting equity returns in foreign markets.

Table 1 displays summary statistics across the 13 equity markets. We report monthly excess returns on country-specific market, value, and growth portfolios. Excess returns are monthly in percent. Figures in

parentheses are the *t*-statistics for testing the null that the average return is no different from 0. Table 1 draws on table 3 in Fama and French (1998), except that the time series of returns spans the longer period 1975–2000. Table 1 shows some evidence in favor of the value premium in international returns. Firms with high ratios of book-to-market equity, earnings to price, cash flow to price, or dividend to price display higher average returns than firms with low ratios.

B. Results

Hypothesis testing using the Bayesian procedure. Here, we report the results of our Bayes test of the conditional asset pricing restriction $\alpha(z_{t-1}) = 0$ using the international data set described earlier. In all, five asset pricing models are examined: the conditional version of the international CAPM plus conditional versions of the four Fama-French two-factor models. Like the international CAPM, returns on the international market portfolio enter into each of the four two-factor models as one of the two risk factors. The other risk factor in these models is taken to be one of four proxies for relative financial distress (HB/M − LB/M, HE/P − LE/P, HC/P − LC/P, or HD/P − LD/P), as proposed originally by Fama and French (1998).

Results from testing each restricted model against its unrestricted counterpart are reported in tables 2 and 3. The entries in these tables are the average posterior probabilities of the restricted models, that is, $P_{\mathcal{M}_R}^{\text{AVG}}$, as calculated from expression (18). To provide a brief description of these two tables, we note first that tables 2 and 3 provide results on different groups of test assets. In particular, table 2 reports results for the case where the test assets are country-specific market portfolios, while table 3

TABLE 2   International CAPM and Two-Factor Models: Market Portfolios

$R - F \ P_{\mathcal{M}_R}^{\text{AVG}} \times 100$	CAPM	FF(B/M)	FF(E/P)	FF(C/P)	FF(D/P)
United States	99.69	99.61	99.84	99.63	96.94
Japan	99.37	99.15	98.53	98.02	98.13
United Kingdom	97.66	98.70	98.37	97.29	96.99
France	98.29	97.61	98.20	97.84	97.39
Germany	95.40	96.36	95.71	95.62	94.15
Italy	99.12	99.15	99.09	99.18	99.07
Netherlands	95.59	97.89	98.30	96.83	94.37
Belgium	96.74	97.47	96.68	97.46	98.10
Switzerland	91.14	94.24	92.90	92.02	92.33
Sweden	55.95	56.29	66.41	47.26	42.56
Australia	98.08	98.54	99.07	98.50	97.98
Hong Kong	97.30	98.61	98.72	98.10	97.44
Singapore	97.69	97.67	98.30	97.75	97.87
Multivariate	100.00	100.00	100.00	100.00	100.00

NOTE.—Entries in the table are posterior probability of the restricted model stated in percentage terms. The test assets are the market portfolios of the United States and 12 major EAFE countries. The last row of table 2 gives the posterior probability for testing a multivariate system.

TABLE 3   International CAPM and Two-Factor Models: Global Characteristics

	CAPM	FF(B/M)	FF(E/P)	FF(C/P)	FF(D/P)
HB/M	71.64		99.45	97.25	95.97
LB/M	83.72		96.96	96.85	97.85
HE/P	80.00	98.98		95.03	95.60
LE/P	49.72	90.40		91.84	64.84
HC/P	58.06	91.06	91.56		78.01
LC/P	97.66	94.02	97.22		96.98
HD/P	86.92	97.98	98.22	94.77	
LD/P	71.18	97.80	93.55	90.32	

NOTE.—Entries in this table are (average) posterior probabilities of the restricted models stated in percentage terms. Test assets are global characteristics-sorted portfolios.

provides results where the test assets are global, as opposed to country-specific, characteristic-sorted portfolios. We also tested each restricted model against its unrestricted counterpart for the cases where the test assets are taken to be various country-specific characteristic-sorted portfolios. However, we choose not to report these results here because they are qualitatively very similar to the results reported in table 2, where the test assets are country-specific market portfolios. These results, however, are available from us on request. Note also that, for both tables 2 and 3, the first column of each table gives results for the conditional version of the single-factor international CAPM; whereas the other four columns provide results on the various two-factor models, so that the numbers reported in columns 2, 3, 4, and 5 correspond to results for the Fama-French model whose second factor is HB/M − LB/M, HE/P − LE/P, HC/P − LC/P, and HD/P − LD/P, respectively.

Looking first at table 2, we see that, when the test assets employed are country-specific market portfolios, our test results uniformly provide strong and unambiguous evidence in favor of conditional versions of both the international CAPM and the four Fama-French models relative to models that allow for asset mispricing. Excluding Sweden, the smallest value of $P_{\mathcal{M}_R}^{\mathrm{AVG}}$ recorded across all countries and for the different asset pricing models exceeds 91%, indicating strong evidence for the null hypothesis. Indeed, based on the decision rule (22), our Bayes procedure chooses the CAPM and the two-factor models over their unrestricted counterparts for each of the countries in table 2, with the exception of Sweden.

In the case of Sweden, the posterior model probability $P_{\mathcal{M}_R}^{\mathrm{AVG}}$ recorded for the Fama-French model whose second factor is HC/P − LC/P and the Fama-French model whose second factor is HD/P − LD/P are both under 0.5 or 50%, indicating the possible inadequacy of these models in explaining the market returns of Sweden. Moreover, $P_{\mathcal{M}_R}^{\mathrm{AVG}}$ recorded for the CAPM, the Fama-French model whose second factor is HB/M − LB/M, and the Fama-French model whose second factor is HE/P − LE/P

are, respectively, 55.95%, 56.29%, and 66.41%; all of which are substantially lower than the posterior model probabilities for the various restricted models recorded for other countries, which, as noted earlier, exceed 90%. The last row of table 2 reports results for the case where all the country-specific market returns are tested simultaneously in a multivariate regression framework. Here, the posterior probability in favor of any pricing model studied is 100%. Hence, simultaneous testing yields stronger evidence in favor of the five asset pricing models over their unrestricted alternatives.

Our results somewhat differ from those of Fama and French (1998). Whereas they reject the null hypothesis that the international CAPM is an appropriate model for the excess returns on country-specific HB/M portfolios using the Gibbons, Ross, and Shanken (1989) GRS test; we find the international CAPM to be an adequate model for these returns when judged against an alternative, unrestricted specification that allows for time-varying asset mispricing. It should be noted, of course, that any comparison of our results here with those of Fama and French (1998) should be done with care, not only because of the difference in the methodology employed (i.e., Fama and French 1998 used classical hypothesis testing methods, while we use Bayesian testing procedures here) but also because Fama and French (1998) do not account for potential stock return predictability and our sample is 5 years longer.

Turning our attention now to table 3, where the test assets are various global characteristic-sorted portfolios, we note that, while the results presented in this table still find in favor of the various restricted specifications over their unrestricted counterparts, the evidence in favor of the CAPM vis-à-vis an unrestricted alternative appears to be weaker here, as judged by the smaller value of its posterior model probabilities. Indeed, comparing the results from this table with those of the previous tables, we see that, in contrast with the cases where the test assets are country-specific portfolios, the posterior model probabilities recorded for the CAPM are less than 0.8 or 80% for half of the eight categories of test assets examined.

Model comparison.   Tables 4 and 5 exhibit posterior probabilities for 10 data-generating models under consideration, the restricted and unrestricted versions of the 5 pricing models studied here. We assigned a prior probability of 10% to each specification. Figures reported in the tables are useful for both comparing the pricing abilities of the CAPM and four ICAPMs and incorporating model uncertainty in investment-based experiments. In particular, consider an investor who must allocate funds among multiple securities. The investor is uncertain about which specification, if any, is useful in pricing. That investor can use model posterior probabilities for averaging across the 10 data-generating specifications. Investment decisions can then be made based on a general model that optimally nests the individual specifications using posterior

TABLE 4   Model Comparison with Country-Specific Test Assets: Multivariate Case

	$\mathcal{M}_R^C$	$\mathcal{M}_U^C$	$\mathcal{M}_R^{B/M}$	$\mathcal{M}_U^{B/M}$	$\mathcal{M}_R^{E/P}$	$\mathcal{M}_U^{E/P}$	$\mathcal{M}_R^{C/P}$	$\mathcal{M}_U^{C/P}$	$\mathcal{M}_R^{D/P}$	$\mathcal{M}_U^{D/P}$
Market	.00	.00	.02	.00	97.12	.00	.01	.00	2.86	.00
HB/M	.00	.00	100.00	.00	.00	.00	.00	.00	.00	.00
LB/M	.00	.00	.77	.00	98.37	.00	.00	.00	.87	.00
HE/P	.00	.00	.00	.00	100.00	.00	.00	.00	.00	.00
LE/P	.00	.00	.00	.00	99.75	.00	.00	.00	.25	.00
HC/P	.00	.00	.02	.00	51.26	.00	48.72	.00	.00	.00
LC/P	.00	.00	.00	.00	.02	.00	99.98	.00	.00	.00
HD/P	.00	.00	.00	.00	.00	.00	.00	.00	100.00	.00
LD/P	.00	.00	.00	.00	16.68	.00	.00	.00	83.32	.00

NOTE.—Entries are (average) posterior model probabilities stated in percentage terms. Test assets are market portfolios and country-specific characteristics.

probabilities as weights. In this paper, we attempt to compare the empirical performance of various asset pricing specifications. Analyzing investment decisions in the presence of uncertainty about the correct asset-pricing specification and the set of predictive variables provides a direction for future work.

Observe from tables 4 and 5 that, when neutral initial beliefs about the pricing abilities of the five factor models are combined with the data, the best performing model is the Fama-French whose second factor is (global) HE/P − LE/P (denoted by the symbol $\mathcal{M}_R^{E/P}$). Focusing first on table 4, where the test assets are country-specific market and characteristic-sorted portfolios, we note that $\mathcal{M}_R^{E/P}$ is found to be the best model for five of the nine categories of test assets examined. In fact, it is a bit surprising that, as judged by posterior model probability, $\mathcal{M}_R^{E/P}$ does a better job of explaining the returns on country-specific LB/M portfolios than $\mathcal{M}_R^{B/M}$, the Fama-French model whose second factor is (global) HB/M − LB/M. Moreover, $\mathcal{M}_R^{E/P}$ also has slightly higher posterior model probability than M_R^{CP}, the Fama-French model whose second factor is (global) HC/P − LC/P when the test assets are

TABLE 5   Model Comparison with Global Characteristic-Sorted Portfolios as Test Assets

	$\mathcal{M}_R^C$	$\mathcal{M}_U^C$	$\mathcal{M}_R^{B/M}$	$\mathcal{M}_U^{B/M}$	$\mathcal{M}_R^{E/P}$	$\mathcal{M}_U^{E/P}$	$\mathcal{M}_R^{C/P}$	$\mathcal{M}_U^{C/P}$	$\mathcal{M}_R^{D/P}$	$\mathcal{M}_U^{D/P}$
HB/M	.00	.00			100.00	.00	.00	.00	.00	.00
LB/M	.00	.00			100.00	.00	.00	.00	.00	.00
HE/P	.00	.00	100.00	.00			.00	.00	.00	.00
LE/P	.00	.00	1.15	.00			79.97	.00	18.88	.00
HC/P	.00	.00	.00	.00	100.00	.00			.00	.00
LC/P	.00	.00	.00	.00	88.44	.00			11.56	.00
HD/P	.00	.00	2.24	.00	10.49	.00	87.27	.00		
LD/P	.00	.00	2.51	.00	97.49	.00	.00	.00		

NOTE.—Entries are (average) posterior model probabilities stated in percentage terms.

country-specific HC/P portfolios. The CAPM, on the other hand, does not compare favorably with the Fama-French ICAPM's as a model for any of the country-specific test assets examined. Note also that return-generating specifications that permit asset mispricing are not at all supported by the data. Applying the intuition of Jeffreys (1961) to our results, we record decisive evidence in favor of the restricted specifications.

Next, in table 5, the results are based on global characteristic-sorted portfolios. Here, for each category of test assets, we exclude from the comparison that Fama-French model whose second factor is constructed using returns from that category of test assets. For example, $\mathcal{M}_R^{E/P}$ is excluded from the comparison when the test assets are either (global) HE/P or (global) LE/P. Again, we see that, overall, $\mathcal{M}_R^{E/P}$ is the best performer, as it is the best model for all but one of the categories of test assets in which it is compared with other models. Moreover, note that the CAPM also does not perform well vis-à-vis the Fama-French models with respect to returns on these global characteristic-sorted portfolios.

In summary, both the CAPM and the four Fama-French models have been found to perform well relative to their unrestricted counterparts for most test assets examined in this paper. The lone exception to this overall statement is the case of Sweden, where there is some evidence of asset mispricing especially with regard to returns on growth portfolios for that country. In addition, when all the models are compared directly in terms of their posterior model probability, the Fama-French model with value premium based on earnings yield is found to be the best model for explaining returns on most test assets examined.

V.   Conclusion

This paper develops and implements an exact multivariate procedure for testing and selecting among alternative asset-pricing specifications with conditioning information using posterior probability. The finite-sample multivariate test introduced by the seminal Gibbons et al. (1989) paper, while exact, is suitable for testing unconditional asset pricing models. It cannot be readily extended for testing models where alpha is allowed, under the alternative hypothesis, to vary with the state of the economy. Moreover, the availability of an exact test in this context seems desirable, since the alternative GMM-based test, whose justification is based on asymptotic analysis, has been shown by various Monte Carlo studies (see Ahn and Gadarowski 1999) to suffer from poor finite sample properties. In addition, our approach allows us to simultaneously compare the performance of multiple asset-pricing specifications, both nested and non-nested, and optimally combine those models into a one general weighted model that could be useful for making investment decisions under model uncertainty.

We applied our procedure to testing international asset pricing models using the Fama and French (1998) data set extended to the year 2000. Not only did we test each asset pricing model individually against its unrestricted counterpart, we ran a comprehensive horse race as well. That horse race simultaneously evaluated the relative performance of all asset pricing specifications considered in this paper in terms of their ability to explain the excess returns of various test assets. The most interesting result that arises from our horse race comparison is that, for most test assets, the conditional Fama-French model with a value premium constructed based on earnings yield appears to be the best model. Overall, this model outperforms an alternative Fama-French specification whose value premium is constructed from portfolios sorted on the basis of book-to-market ratio, even though this latter specification, thus far, has received greater attention in the literature.

The research presented here can be extended in a number of directions. First, although beta is assumed to be time invariant, beta variation can also be incorporated within this testing framework. In this case, an exact analytical expression for the marginal likelihood under the unrestricted specification can still be obtained. Gibbs sampling methods will be needed to compute the marginal likelihood for the specification that conforms to conditional pricing restrictions. Second, as explained earlier, the Bayesian framework of posterior odds ratios has important advantages both in testing nonnested hypotheses and the simultaneous comparison of multiple models. In particular, in comparing nonnested models, the method used here will not lead to problems of intransitivity, which can occur when classical tests of nonnested hypotheses are implemented. Thus, it seems worthwhile to extend our framework to implement an even broader comparison of the wide array of asset pricing models proposed in the finance literature, including specifications whose factors are not portfolio based. Finally, analyzing portfolio selection in the presence of uncertainty about the pricing specification could be a worthy topic for future research.

Appendix

Exact Marginal Likelihood for the Pricing Restrictions

We first partition equation (1) into its components and obtain the three multivariate regressions:

$$R = XA_R + U_R, \qquad (A.1)$$

$$F = XA_F + U_F, \qquad (A.2)$$

$$Z = XA_Z + U_Z. \qquad (A.3)$$

Under the pricing restriction, we can rewrite the multivariate predictive regression given by expression (A.1) in the alternative form:

$$R = XA_F\Sigma_{FF}^{-1}\Sigma_{FR} + U_R = FB_{FR} + V_R, \tag{A.4}$$

where the second equality is obtained by setting $B_{FR} = \Sigma_{FF}^{-1}\Sigma_{FR}$ (note $B_{FR} = \beta'$) and $V_R = (U_R - U_F\Sigma_{FF}^{-1}\Sigma_{FR})$. Next, we write $V = [V_R V_F V_Z]$ and $U = [U_R U_F U_Z]$ and note that $V = U_H$, where

$$H = \begin{pmatrix} I_N & 0 & 0 \\ -\Sigma_{FF}^{-1}\Sigma_{FR} & I_K & 0 \\ 0 & 0 & I_M \end{pmatrix}. \tag{A.5}$$

Since $|H| = 1$, it follows that, under the Gaussian error assumption, the likelihood function under the restricted model $\mathcal{M}_R$ can be written as

$$\mathcal{L}(A_F, A_Z, \Sigma | Y, X, \mathcal{M}_R)$$
$$= (2\pi)^{-\frac{1}{2}T(M+N+K)}|\Omega(\Sigma)|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}Tr\left[\Omega(\Sigma)^{-1}V'V\right]\right\}, \tag{A.6}$$

where

$$\Omega(\Sigma) = \begin{pmatrix} \Sigma_{RR.F} & 0 & \Sigma_{RZ.F} \\ 0 & \Sigma_{FF} & \Sigma_{FZ} \\ \Sigma'_{RZ.F} & \Sigma'_{FZ} & \Sigma_{ZZ} \end{pmatrix}, \tag{A.7}$$

$\Sigma_{RR.F} = \Sigma_{RR} - \Sigma_{RF}\Sigma_{RF}^{-1}\Sigma_{FR}$, and $\Sigma_{RZ.F} = \Sigma_{RZ} - \Sigma_{RF}\Sigma_{RF}^{-1}\Sigma_{FZ}$. Hence, if we start out with the diffuse prior $\pi_0(A_F, A_Z, \Sigma) \propto |\Sigma|^{-1/2h} = |\Omega(\Sigma)|^{1/2h}$ and use the first τ observations as the training sample; then, the diffuse-prior posterior distribution for the training sample takes the form

$$\tau(A_F, A_Z, \Sigma | D(\tau), \mathcal{M}_R)$$
$$\propto (2\pi)^{-\frac{1}{2}\tau(M+N+K)}|\Omega(\Sigma)|^{-\frac{1}{2}(\tau+h)}\exp\left\{-\frac{1}{2}Tr\left[\Omega(\Sigma)^{-1}V(\tau)'V(\tau)\right]\right\}, \tag{A.8}$$

where $D(\tau) = [Y(\tau), X(\tau)]$ and $Y(\tau), X(\tau)$, and $V(\tau)$ denote matrices comprising the first τ rows of $Y = [R, F, Z], X$, and V, respectively. Observe that (A.8) yields a proper (posterior) distribution on the elements of the parameter matrices A_F, A_Z, Σ when τ is sufficiently large. As we use an appropriately normalized version of this posterior to serve as the prior distribution in our construction of the marginal likelihood under the restricted model; in subsequent discussion, we often refer to expression (A.8) as the prior density under the restricted model $\mathcal{M}_R$.

To fix some additional notations, let $L = N + K + M$ and $M_* = M + 1$. Next, we define $W = [X, F, R]$ and partition $R = [R(\tau)', R(-\tau)']', F = [F(\tau)', F(-\tau)']', Z = [Z(\tau)', Z(-\tau)']', V_R = [V_R(\tau)', V_R(-\tau)']', U_F = [U_F(\tau)', U_F(-\tau)']',$ and $W = [W(\tau)', W(-\tau)']'$ and where $R(\tau), F(\tau), Z(\tau), V_R(\tau), U_F(\tau),$ and $W(\tau)$ are matrices

of dimensions $\tau \times N, \tau \times K, \tau \times M, \tau \times N, \tau \times K$, and $\tau \times L$, respectively, while $R(-\tau), F(-\tau), Z(-\tau), V_R(-\tau), U_F(-\tau)$, and $W(-\tau)$ are matrices of dimensions $T_1 \times N, T_1 \times K, T_1 \times M, T_1 \times N, T_1 \times K$, and $T_1 \times L$, respectively, where $T_1 = T - \tau$. Also, we define $B_{FZ} = \Sigma_{FF}^{-1}\Sigma_{FZ}, B_{RZ.F} = \Sigma_{RR.F}^{-1}\Sigma_{RZ.F}$, and $\Sigma_{ZZ.R.F} = \Sigma_{ZZ} - \Sigma_{RZ.F}'\Sigma_{RR.F}^{-1}\Sigma_{RZ.F} - \Sigma_{FZ}'\Sigma_{FF}^{-1}\Sigma_{FZ}$; and let $\theta = [\mathrm{vec}(B_{FR})', \mathrm{vec}(B_{FZ})', \mathrm{vec}(B_{RZ.F})',$ $\mathrm{vech}(\Sigma_{FF})', \mathrm{vech}(\Sigma_{RR.F})', \mathrm{vech}(\Sigma_{ZZ.R.F})']'$, where $\mathrm{vec}(\cdot)$ is the usual column stacking operator and $\mathrm{vech}(\cdot)$ is the column stacking operator, stacking elements on and below the main diagonal for a symmetric matrix. Now, to write the training sample posterior density, given by expression (A.8), in a more convenient form, we make the one-to-one transformation $\mathrm{vech}(\Sigma) \to \theta$, noting that the Jacobian for this transformation can be calculated to be $(|\Sigma_{FF}|^{(M+N)}|\Sigma_{RR.F}|^M)$. Under this alternative parameterization, the training sample posterior density given in expression (A.8) can be factored into the product of conditional and marginal densities, each of which is either multivariate normal or inverted Wishart. We summarize this result in the proposition that follows.

PROPOSITION 1.   Let $\pi_0(A_F, A_Z, \Sigma) \propto |\Sigma|^{-1/2h} = |\Omega(\Sigma)|^{-1/2h}$, then under the restriction $A_R = A_F \Sigma_{FF}^{-1}\Sigma_{FR}$, the training sample posterior density takes the form

$$\pi(\Phi, \Sigma_{ZZ.R.F}, B_{FR}, \Sigma_{RR.F}, A_F, \Sigma_{FF}|D(\tau), \mathcal{M}_R),$$

$$\propto \pi(\Phi|\Sigma_{ZZ.R.F}, \Sigma_{RR.F}, \Sigma_{FF}, B_{FR}, A_F, D(\tau), \mathcal{M}_R),$$

$$\pi(\Sigma_{ZZ.R.F}|\Sigma_{RR.F}, \Sigma_{FF}, B_{FR}, A_F, D(\tau), \mathcal{M}_R),$$

$$\pi(B_{FR}|\Sigma_{RR.F}, \Sigma_{FF}, A_F, D(\tau), \mathcal{M}_R),$$

$$\pi(\Sigma_{RR.F}|\Sigma_{FF}, A_F, D(\tau), \mathcal{M}_R),$$

$$\pi(A_F|\Sigma_{FF}, D(\tau), \mathcal{M}_R),$$

$$\pi(\Sigma_{FF}|D(\tau), \mathcal{M}_R), \tag{A.9}$$

where

$$\pi[\Phi|\Sigma_{ZZ.R.F}, \Sigma_{RR.F}, \Sigma_{FF}, B_{FR}, A_F, D(\tau), \mathcal{M}_R]$$
$$= f_{MN}\left\{\Phi|\hat{\Phi}_{T_0}, \Sigma_{ZZ.R.F} \otimes \left[\tilde{W}(\tau)'\tilde{W}(\tau)\right]^{-1}\right\},$$

$$\pi(\Sigma_{ZZ.R.F}|\Sigma_{RR.F}, \Sigma_{FF}, B_{FR}, A_F, D(\tau), \mathcal{M}_R)$$
$$= f_{IW}\left[\Sigma_{ZZ.R.F}|Z(\tau)'Q_{\tilde{W}(\tau)}Z(\tau), \nu_{\tau,h} - 1\right],$$

$$\pi(B_{FR}|\Sigma_{RR.F}, \Sigma_{FF}, A_F, D(\tau), \mathcal{M}_R) = f_{MN}\left\{B_{FR}|\hat{B}_{FR,\tau}, \Sigma_{RR.F} \otimes \left[F(\tau)'F(\tau)\right]^{-1}\right\},$$

$$\pi(\Sigma_{RR.F}|\Sigma_{FF}, A_F, D(\tau), \mathcal{M}_R) = f_{IW}\left[\Sigma_{RR.F}|R(\tau)'Q_{F(\tau)}R(\tau), \nu_{\tau,h}\right],$$

$$\pi(A_F|\Sigma_{FF}, D(\tau), \mathcal{M}_R) = f_{MN}\left\{A_F|\hat{A}_{F,\tau}, \Sigma_{FF} \otimes \left[X(\tau)'X(\tau)\right]^{-1}\right\},$$

$$\pi(\Sigma_{FF}|D(\tau), \mathcal{M}_R) = f_{IW}\left[\Sigma_{FF}|F(\tau)'Q_{X(\tau)}F(\tau), \nu_{\tau,h}^*\right];$$

where $\Phi = [A_Z', B_{FZ}', B_{RZ.F}']'$, $\nu_{\tau,h} = \tau + h - 2M - N - K - 1$, $\nu_{\tau,h}^* = \nu_{\tau,h} - M - N - 1$, $\tilde{W}(\tau) = [X(\tau)U_F(\tau)V_R(\tau)]$, $\hat{\Phi}_\tau = [\tilde{W}(\tau)'\tilde{W}(\tau)]^{-1}\tilde{W}(\tau)'Z(\tau)$, $\hat{B}_{FR,\tau} = [F(\tau)'F(\tau)]^{-1}F(\tau)'R(\tau)$, and $\hat{A}_{F,\tau} = [X(\tau)'X(\tau)]^{-1}X(\tau)'F(\tau)$; and where $f_{MN}(\cdot|\cdot)c$ and $f_{IW}(\cdot|\cdot)$

denote, respectively, the probability density function of a matrix-normal distribution and that of an inverted Wishart distribution.

Proof. To begin, note that on making the transformation $\text{vech}(\Sigma) \to \theta$, where $\theta = [\text{vec}(B_{FR})', \text{vec}(B_{FZ})', \text{vec}(B_{RZ.F})', \text{vech}(\Sigma_{FF})', \text{vech}(\Sigma_{RR.F})', \text{vech}(\Sigma_{ZZ.R.F})']'$, the density of the diffuse-prior posterior distribution based on the training sample can be written as

$$\pi(\Phi, \Sigma_{ZZ.R.F}, B_{FR}, \Sigma_{RR.F}, A_F, \Sigma_{FF} | D(\tau), \mathcal{M}_R) \propto (2\pi)^{-\frac{1}{2}\tau(M+N+K)} |\Sigma_{FF}|^{-\frac{1}{2}(\tau+h-2M-2N)}$$

$$|\Sigma_{RR.F}|^{-\frac{1}{2}(r+h-2M)} |\Sigma_{ZZ.R.F}|^{-\frac{1}{2}t} \exp\left\{-\frac{1}{2}\text{Tr}\left[\Omega(\theta)^{-1}V(\tau)'V(\tau)\right]\right\}, \quad \text{(A.10)}$$

where $V(\tau) = [V_R(\tau)U_F(\tau)U_Z(\tau)]$ and $V_R(\tau)$, $U_F(\tau)$, and $U_Z(\tau)$ are matrices made up of the first τ rows of the partitioned error matrices $V_R = [V_R(\tau)', V_R(-\tau)']'$, $U_F = [U_F(\tau)', U_F(-\tau)']'$, and $U_Z = [U_Z(\tau)', U_Z(-\tau)']'$, and where $\Omega(\theta)^{-1}$ denotes the inverse of the error covariance matrix $\Omega(\Sigma)$ but expressed as a function of the new parameterization θ. Moreover, applying the formula for the inverse of a 3×3 matrix partition, we can write $\Omega(\theta)^{-1}$ in explicit form as follows:

$$\Omega(\theta)^{-1} = \begin{pmatrix} \Omega^{11} & \Omega^{12} & \Omega^{13} \\ \Omega^{21} & \Omega^{22} & \Omega^{23} \\ \Omega^{31} & \Omega^{32} & \Omega^{33} \end{pmatrix}, \quad \text{(A.11)}$$

where $\Omega^{11} = \Sigma_{RR.F}^{-1} + B_{RZ.F}\Sigma_{ZZ.R.F}^{-1}B_{RZ.F}'$, $\Omega^{12} = B_{RZ.F}\Sigma_{ZZ.R.F}^{-1}B_{FZ}$, $\Omega^{22} = \Sigma_{FF}^{-1} + B_{FZ}\Sigma_{ZZ.R.F}^{-1}B_{FZ}$, $\Omega^{13} = -B_{RZ.F}\Sigma_{ZZ.R.F}^{-1}$, $\Omega^{23} = -B_{FZ}\Sigma_{ZZ.R.F}^{-1}$, and $\Omega^{33} = \Sigma_{ZZ.R.F}^{-1}$ and where $\Omega^{21} = \Omega^{12'}$, $\Omega^{31} = \Omega^{13'}$, and $\Omega^{32} = \Omega^{23'}$. It follows from (A.10) that the training sample posterior density depends on the parameters A_F, A_Z, B_{FR}, B_{FZ}, and $B_{RZ.F}$ through the relationships $V_R(\tau) = R(\tau) - F(\tau)B_{FR}$, $U_F(\tau) = F(\tau) - X(\tau)A_F$, and $U_Z(\tau) = Z(\tau) - X(\tau)A_Z$ and through the matrix function $\Omega(\theta)^{-1}$.

Next, making use of expression (A.11) and some straightforward algebra, it is easy to show that the trace expression in the exponential term of equation (A.10) can be rewritten as follows:

$$\text{Tr}[\Omega(\theta)^{-1}V(\tau)'V(\tau)] = \text{Tr}\{\Sigma_{FF}^{-1}[F(\tau)'Q_{X(\tau)}F(\tau)$$

$$+ (A_F - \hat{A}_{F,\tau})'X(\tau)'X(\tau)(A_F - \hat{A}_{F,\tau})]\}$$

$$+ \text{Tr}[\Sigma_{RR.F}^{-1}R(\tau)'Q_{F(\tau)}R(\tau) + (B_{FR} - \hat{B}_{FR,\tau})'F(\tau)'F(\tau)(B_{FR} - \hat{B}_{FR,\tau})]$$

$$+ \text{Tr}\{\Sigma_{ZZ.R.F}^{-1}[Z(\tau)'Q_{\tilde{W}(\tau)}Z(\tau) + (\Phi - \hat{\Phi}_\tau)'\tilde{W}(\tau)'\tilde{W}(\tau)(\Phi - \hat{\Phi}_\tau)]\}, \quad \text{(A.12)}$$

where $\hat{A}_{F,\tau}, \hat{B}_{F,\tau}, \hat{\Phi}_\tau, F(\tau), R(\tau)$, and $Z(\tau)$ are as previously defined.

Now, note that $\tilde{W}(\tau) = W(\tau)\tilde{H}$, where $W(\tau) = [X(\tau)F(\tau)R(\tau)]$ and

$$\tilde{H} = \begin{bmatrix} I_{M+1} & -A_F & 0 \\ 0 & I_K & -B_{FR} \\ 0 & 0 & I_N \end{bmatrix}, \quad \text{(A.13)}$$

and it is obvious that $|\tilde{H}| = 1$, so $\tilde{H}$ is nonsingular. It follows that, for each value of B_{FR} and A_F, we have that $Z(\tau)'Q_{\tilde{W}(\tau)}Z(\tau) = Z(\tau)'Q_{W(\tau)}Z(\tau)$ and $|\tilde{W}(\tau)'\tilde{W}(\tau)| = |W(\tau)'W(\tau)|$; so that, in particular, neither $Z(\tau)'Q_{\tilde{W}(\tau)}Z(\tau)$ nor $|\tilde{W}(\tau)'\tilde{W}(\tau)|$ depends on the unknown parameters A_F and B_R. Using these facts, we can factor (A.10) into a product of (unnormalized) conditional and marginal probability density functions, each of which is either the density of a matrix-variate normal distribution or that of an inverted Wishart distribution. Hence, on appropriate renormalization, we obtain the training sample posterior density given by expression (A.9). Q.E.D.

Next, we derive a closed form expression for the marginal likelihood under the restricted model, $\mathcal{M}_R$. To proceed, we update the prior information supplied by the training sample with data information from the second subsample, as provided by the likelihood function:

$$
\mathcal{L}(\Phi, \Sigma_{ZZ.R.F}, B_{FR}, \Sigma_{RR.F}, A_F, \Sigma_{FF} | D(-\tau), \mathcal{M}_R) = (2\pi)^{-\frac{1}{2}T_1(M+N+K)} |\Omega(\Sigma)|^{-\frac{T_1}{2}}
$$

$$
\times \exp\left\{-\frac{1}{2}\mathrm{Tr}[\Omega(\Sigma)^{-1}V(-\tau)'V(-\tau)]\right\}, \tag{A.14}
$$

where $D(-\tau) = [Y(-\tau), X(-\tau)]$ and where $Y(-\tau), X(-\tau)$, and $V(-\tau)$ denote matrices comprising the last $T_1 = T - \tau$ rows of the matrices Y, X, and V, respectively. Thus, the marginal likelihood under asset-pricing restrictions can be computed by combining the prior density (A.10) with the likelihood function (A.14) then integrating with respect to the parameters $\Phi, \Sigma_{ZZ.R.F}, B_{FR}, \Sigma_{RR.F}, A_F$, and Σ_{FF}. Now, define $S_{YY.X}$ generically as the sum of squared residual of the regression of Y on X, so that $S_{YY.X} = (Y - X\hat{\beta})'(Y - X\hat{\beta})$, where $\hat{\beta} = (X'X)^{-1}X'Y$, then the following proposition gives an exact analytical formula for this restricted marginal likelihood.

PROPOSITION 2.   Let the prior density be as given by expression (31). Then, the marginal likelihood for the restricted model, $\mathcal{M}_R$, computed on the basis of observations at $t = \tau + 1, \ldots, T$ has the form

$$
m[D(-\tau)|D(\tau), \mathcal{M}_R] = (\pi)^{-\frac{1}{2}T_1(M+N+K)}
$$

$$
\times \left\{\frac{\Gamma_M[0.5(v_{T,h}-1)]\Gamma_N[0.5(v_{T,h})]\Gamma_K[0.5(v_{T,h}-M-N-1)]}{\Gamma_M[0.5(v_{\tau,h}-1)]\Gamma_N[0.5v_{\tau,h}]\Gamma_K[0.5(v_{\tau,h}-M-N-1)]}\right\}
$$

$$
\times \left[\frac{|S_{Z(\tau)Z(\tau).W(\tau)}|^{\frac{1}{2}(v_{\tau,h}-1)}}{|S_{ZZ.W}|^{\frac{1}{2}(v_{T,h}-1)}}\right] \left[\frac{|S_{R(\tau)R(\tau).F(\tau)}|^{\frac{1}{2}v_{\tau,h}}}{|S_{RR.F}|^{\frac{1}{2}v_{T,h}}}\right]
$$

$$
\times \left[\frac{|S_{F(\tau)F(\tau).X(\tau)}|^{\frac{1}{2}(v_{\tau,h}-M-N-1)}}{|S_{FF.X}|^{\frac{1}{2}(v_{T,h}-M-N-1)}}\right]
$$

$$
\times \left[\frac{|W(\tau)'W(\tau)|}{|W'W|}\right]^{\frac{1}{2}M} \left[\frac{|F(\tau)'F(\tau)|}{|F'F|}\right]^{\frac{1}{2}N} \left[\frac{|X(\tau)'X(\tau)|}{|X'X|}\right]^{\frac{1}{2}K}, \tag{A.15}
$$

where $v_{T,h} = T + h - 2M - N - K - 1$ and $v_{\tau,h}$ is as defined in proposition 1.

Proof. Set $T_1 = \mathrm{T} - \tau$ and note that the likelihood function for the second part of the sample, as given by expression (A.14), can be factored as follows:

$$\mathcal{L}(\Phi, \Sigma_{ZZ.R.F}, B_{FR}, \Sigma_{RR.F}, A_F, \Sigma_{FF} | D(-\tau), \mathcal{M}_R) = (2\pi)^{-\frac{1}{2}T_1(N+M+K)} |\Sigma_{RR.F}|^{-\frac{1}{2}T_1}$$

$$\exp\left(-\frac{1}{2}\mathrm{Tr}\left\{\Sigma_{RR.F}^{-1}\left[S_{R(-\tau)R(-\tau).F(-\tau)} + (B_{FR} - \hat{B}_{FR,\tau})'F(-\tau)'F(-\tau)(B_{FR} - \hat{B}_{FR,\tau})\right]\right\}\right)$$

$$|\Sigma_{ZZ.R.F}|^{-\frac{1}{2}T_1}\ \exp\left(-\frac{1}{2}\mathrm{Tr}\left\{\Sigma_{ZZ.R.F}^{-1}\left[S_{Z(-\tau)Z(-\tau).\tilde{W}(-\tau)}\right.\right.\right.$$

$$\left.\left.\left. + (\Phi - \hat{\Phi}_{T_1})'\tilde{W}(-\tau)'\tilde{W}(-\tau)(\Phi - \hat{\Phi}_{T_1})\right]\right\}\right)$$

$$|\Sigma_{FF}|^{-\frac{1}{2}T_1}\ \exp\left(-\frac{1}{2}\mathrm{Tr}\left\{\Sigma_{FF}^{-1}\left[S_{F(-\tau)F(-e).X(-\tau)}\right.\right.\right.$$

$$\left.\left.\left. + (A_F - \hat{A}_{F,\tau})'X(-\tau)'X(-\tau)(A_F - \hat{A}_{F,\tau})\right]\right\}\right), \tag{A.16}$$

and where $\hat{\Phi}_{T_1} = [\tilde{W}(-\tau)'\tilde{W}(-\tau)]^{-1}\tilde{W}(-\tau)'Z(-\tau), \hat{B}_{FR,T_1} = [F(-\tau)'F(-\tau)]^{-1} \times F(-\tau)'R(-\tau), A_{F,T_1} = [X(-\tau)'X(-\tau)]^{-1}X(-\tau)'F(-\tau),$ and $\tilde{W}(-\tau) = [X(-\tau), V_R(-\tau), U_F(-\tau)]$.

Next, observe that, for each value of B_{FR} and A_F, the following relationships hold: $Z'Q_{\tilde{W}}Z = Z'Q_W Z$ and $|\tilde{W}(\tau)'\tilde{W}(\tau) + \tilde{W}(-\tau)'\tilde{W}(-\tau)| = |W'W|$, where the matrix $\tilde{H}$ is as defined in expression (A.13). Thus, neither $Z'Q_{\tilde{W}}Z$ nor $|\tilde{W}'\tilde{W}|$ depends on the unknown parameters A_F and B_{FR}. Making use of these facts, we can construct the joint posterior density of $\Phi, \Sigma_{ZZ.R.F}, B_{FR}, \Sigma_{RR.F}, A_F$, and Σ_{FF} given the data by combining the likelihood function (A.16) with the prior density given in expression (A.9). The marginal likelihood given in expression (A.15) then can be obtained straightforwardly by integrating this joint posterior density with respect to the parameters $\Phi, \Sigma_{ZZ.R.F}, B_{FR}, \Sigma_{RR.F}, A_F$, and Σ_{FF} over their respective support.

Computing the Marginal Likelihood for the Unrestricted Model

Our point of departure is the linear system given by (A.1)–(A.3). Observe that, in the unrestricted case, we can rewrite equation (A.1) as

$$R = X(A_R - A_F B_{FR}) + F B_{FR} + \tilde{V}_R, \tag{A.17}$$

where $B_{FR} = \Sigma_{FF}^{-1}\Sigma_{FR}$ and $\tilde{V}_R = U_R - U_F\Sigma_{FF}^{-1}\Sigma_{FR}$. Note, of course, that equation (A.4) of the last subsection is nested within the more general equation given by (A.17) here. Now, to specify a prior based on the training sample posterior distribution, we start again with the diffuse prior $\pi_0(A_R, A_F, A_Z, \Sigma) \propto |\Sigma|^{-1/2h} = |\Omega(\Sigma)|^{-1/2h}$, with $\Omega(\Sigma)$ as defined in expression (A.7), and use the first τ observations as a training sample.

The resulting posterior distribution for the training sample then takes the form

$$\pi[A_R, A_F, A_Z, \Sigma \mid D(\tau), \mathcal{M}_U]$$

$$\propto (2\pi)^{-\frac{1}{2}\tau(M+N+K)} \mid \Omega(\Sigma) \mid^{-\frac{1}{2}(\tau+h)} \exp\left\{-\frac{1}{2}\text{Tr}\left[\Omega(\Sigma)^{-1}\tilde{V}(\tau)'\tilde{V}(\tau)\right]\right\},$$

$$(A.18)$$

where $D(\tau) = [Y(\tau), X(\tau)]$ and where $Y(\tau)(\tau \times L), X(\tau)(\tau \times M_*)$, and $\tilde{V}(\tau) = [\tilde{V}_R(\tau), U_F(\tau), U_Z(\tau)](\tau \times L)$ here denote matrices comprising the first T_0 rows of $Y = [R, F, Z], X$, and $\tilde{V} = [\tilde{V}_R, U_F, U_Z]$, respectively. It is worthwhile to emphasize that given equations (A.17), (A.2), and (A.3), the posterior density (A.18) depends on the parameters A_R, A_F, and A_Z through the relationships $\tilde{V}_R(\tau) = R(\tau) - X(\tau)(A_R - A_F\Sigma_{FF}^{-1}\Sigma_{FR}) - F(\tau)\Sigma_{FF}^{-1}\Sigma_{FR}, U_F(\tau) = F(\tau) - X(\tau)A_F$, and $U_Z(\tau) = Z(\tau) - X(\tau)A_Z$; where as before, $R(\tau), X(\tau), F(\tau)$, and $Z(\tau)$ denote matrices comprising the first τ rows of R, X, F, and Z, respectively. Thus, (A.18) differs from the training sample posterior density for the restricted model, as given by expression (A.8), only with respect to the fact that, in the (more general) unrestricted model here, the restriction $A_R - A_F\Sigma_{FF}^{-1}\Sigma_{FR} = 0$ is not assumed to hold necessarily.

As in the restricted case, we use a properly normalized version of (A.18) as the prior density in constructing the marginal likelihood for the unrestricted model. To proceed, we once more make the change in parameterization $\text{vech}(\Sigma) \rightarrow \theta$, where $\theta = [\text{vec}(B_{FR})', \text{vec}(B_{FZ})', \text{vec}(B_{RZ.F})', \text{vech}(\Sigma_{FF})', \text{vech}(\Sigma_{RR.F})', \text{vech}(\Sigma_{ZZ.R.F})']'$ and where, as noted in the last subsection, the Jacobian of this one-to-one transformation is $|\Sigma_{FF}|^{(M+N)}|\Sigma_{RR.F}|^M$. Note that, since the unrestricted model is a linear, multivariate regression model in the conventional sense, we could have proceeded under the original parameterization in computing the marginal likelihood. However, we choose to work with the parameters $A_R, A_F, A_Z, B_{FR}, B_{FZ}, B_{RZ.F}, \Sigma_{FF}, \Sigma_{RR.F}$, and $\Sigma_{ZZ.R.F}$ instead because, this way, the marginal likelihood calculations given for the unrestricted case here can be compared easily with that given for the restricted case reported in the previous subsection, although the final result will be the same regardless of which parameterization we use. Hence, analogous to proposition 1, the training sample posterior density for the unrestricted case can be factored as follows.

PROPOSITION 3.   Given $\pi_0(A_R, A_F, A_Z, \Sigma) \propto |\Sigma|^{-1/2h} = |\Omega(\Sigma)|^{-1/2h}$, the training sample posterior density for the unrestricted model takes the form:

$$\pi[\Phi, \Sigma_{ZZ.R.F}, \Psi, \Sigma_{RR.F}, A_F, \Sigma_{FF}|D(\tau), \mathcal{M}_U]$$

$$\propto \pi[\Phi|\Sigma_{ZZ.R.F}, \Sigma_{RR.F}, \Sigma_{FF}, \Psi, A_F, D(\tau), \mathcal{M}_U]$$

$$\pi[\Sigma_{ZZ.R.F}|\Sigma_{RR.F}, \Sigma_{FF}, \Psi, A_F, D(\tau), \mathcal{M}_U]$$

$$\pi[\Psi|\Sigma_{RR.F}, \Sigma_{FF}, A_F, D(\tau), \mathcal{M}_U]$$

$$\pi[\Sigma_{RR.F}|\Sigma_{FF}, A_F, D(\tau), \mathcal{M}_U]$$

$$\pi[A_F|\Sigma_{FF}, D(\tau), \mathcal{M}_U]$$

$$\pi[\Sigma_{FF}|D(\tau), \mathcal{M}_U], \qquad\qquad (A.19)$$

where $\Psi = [A_R' B_{FR}']'$, $\Phi = [A_Z' B_{FZ}' B_{RZ.F}']'$, and

$$\pi[\Phi | \Sigma_{ZZ.R.F}, \Sigma_{RR.F}, \Sigma_{FF}, \Psi, A_F, D(\tau), \Sigma_{FF} | D(\tau), \mathcal{M}_U]$$

$$= f_{MN}\left\{ \Phi | \hat{\Phi}_\tau, \Sigma_{ZZ.R.F} \otimes \left[\tilde{W}(\tau)' \tilde{W}(\tau) \right]^{-1} \right\}$$

$$\pi[\Sigma_{ZZ.R.F} | \Sigma_{RR.F}, \Sigma_{FF}, \Psi, A_F, D(\tau), \Sigma_{FF} | D(\tau), \mathcal{M}_U]$$

$$= f_{IW}\left[\Sigma_{ZZ.R.F} | Z(\tau)' Q_{\tilde{W}(\tau)} Z(\tau), \nu_{\tau,h} - 1 \right],$$

$$\pi[\Psi | \Sigma_{RR.F}, \Sigma_{FF}, A_F, D(\tau), \Sigma_{FF} | D(\tau), \mathcal{M}_U]$$

$$= f_{MN}\left\{ \Psi | \hat{\Psi}_\tau, \Sigma_{RR.F} \otimes \left[\tilde{W}(\tau)^{(1)'} \tilde{W}(\tau)^{(1)} \right]^{-1} \right\}$$

$$\pi[\Sigma_{RR.F} | \Sigma_{FF}, A_F, D(\tau), \Sigma_{FF} | D(\tau), \mathcal{M}_U]$$

$$= f_{IW}\left[\Sigma_{RR.F} | R_0' Q_{\tilde{W}^{(1)}(\tau)} R_0, \nu_{\tau,h} - M - 1 \right],$$

$$\pi(A_F | \Sigma_{FF}, D(\tau), \Sigma_{FF} | D(\tau), \mathcal{M}_U) = f_{MN}\left\{ A_F | \hat{A}_{F,\tau}, \Sigma_{FF} \otimes \left[X(\tau)' X(\tau) \right]^{-1} \right\},$$

$$\pi(\Sigma_{FF} | D(\tau), \Sigma_{FF} | D(\tau), \mathcal{M}_U) = f_{IW}\left[\Sigma_{FF} | F(\tau)' Q_{X(\tau)} F(\tau), \nu_{\tau,h}^* \right],$$

where $\tilde{W}(\tau) = [X(\tau), U_F(\tau), \tilde{V}_R(\tau)]$, $\hat{\Phi}_\tau = [\tilde{W}(\tau)' \tilde{W}(\tau)]^{-1} \tilde{W}(\tau)' Z(\tau)$, $\nu_{\tau,h}^* = \nu_{\tau,h} - M - N - 1$, $\hat{\Psi}_\tau = [\tilde{W}(\tau)^{(1)'} \tilde{W}(\tau)^{(1)}]^{-1} \tilde{W}(\tau)^{(1)'} R(\tau)$, and $\hat{A}_{F,\tau} = [X(\tau)' X(\tau)]^{-1} \times [X(\tau)' F(\tau)$.

Proof. Note that, on making the transformation $\mathrm{vech}(\Sigma) \rightarrow \theta$, where $\theta = [\mathrm{vec}(B_{FR})', \mathrm{vec}(B_{FZ})', \mathrm{vec}(B_{RZ.F})', \mathrm{vech}(\Sigma_{FF})', \mathrm{vech}(\Sigma_{RR.F})', \mathrm{vech}(\Sigma_{ZZ.R.F})']'$, the density of the diffuse-prior posterior distribution based on the training sample can be written as

$$\pi(\Phi, \Sigma_{ZZ.R.F}, \Psi, \Sigma_{RR.F}, A_F, \Sigma_{FF} | D(\tau), \mathcal{M}_U) \propto (2\pi)^{-\frac{1}{2}\tau(M+N+K)}$$

$$\times |\Sigma_{FF}|^{-\frac{1}{2}(\tau+h-2M-2N)} |\Sigma_{RR.F}|^{-\frac{1}{2}(\tau+h-2M)} |\Sigma_{ZZ.R.F}|^{-\frac{1}{2}T_0}$$

$$\times \exp\left\{ -\frac{1}{2} \mathrm{Tr}\left[\Omega(\theta)^{-1} \tilde{V}(\tau)' \tilde{V}(\tau) \right] \right\}, \tag{A.20}$$

where $\Omega(\theta)^{-1}$ is as defined in expression (A.11) and where $\tilde{V}(\tau) = [\tilde{V}_R(\tau), U_F(\tau), U_Z(\tau)]$. The posterior density (A.20), thus, depends on the parameters A_R, A_F, A_Z B_{FR}, B_{FZ}, and $B_{RZ.F}$ through the relationships $\tilde{V}_R(\tau) = R(\tau) - X(\tau)(A_R - A_F B_{FR}) - F(\tau)$, $U_F(\tau) = F(\tau) - X(\tau)A_F$, and $U_Z(\tau) = Z(\tau) - X(\tau)A_Z$ and through the matrix function $\Omega(\theta)^{-1}$. Now, similar to the proof of proposition 1, one can show by straightforward calculation that the trace expression in the exponential component of (A.20) can be written in the form

$$\mathrm{Tr}\left[\Omega(\theta)^{-1} \tilde{V}(\tau)' \tilde{V}(\tau) \right] = \mathrm{Tr}\Big\{ \Sigma_{FF}^{-1}\Big[F(\tau)' Q_{X(\tau)} F(\tau)$$

$$+ \left(A_F - \hat{A}_{F,\tau} \right)' X(\tau)' X(\tau) \left(A_F - \hat{A}_{F,\tau} \right) \Big] \Big\}$$

$$+ \mathrm{Tr}\Big\{ \Sigma_{RR.F}^{-1}\Big[R(\tau)' Q_{\tilde{W}(\tau)^{(1)}} R(\tau) + \left(\Psi - \hat{\Psi}_\tau \right)' \tilde{W}(\tau)^{(1)'} \tilde{W}(\tau)^{(1)} \left(\Psi - \hat{\Psi}_\tau \right) \Big] \Big\}$$

$$+ \mathrm{Tr}\Big\{ \Sigma_{ZZ.R.F}^{-1}\Big[Z(\tau)' Q_{\tilde{W}(\tau)} Z(\tau) + (\Phi - \hat{\Phi}_\tau)' \tilde{W}(\tau)' \tilde{W}(\tau)(\Phi - \hat{\Phi}_\tau) \Big] \Big\}, \tag{A.21}$$

where $\hat{A}_{F,\tau}, \hat{\Psi}_\tau, \hat{\Phi}_\tau, F(\tau), R(\tau)$, and $Z(\tau)$ are as defined in Section II of the paper, and where $\tilde{V}_R(\tau), U_F(\tau)$, and $U_Z(\tau)$ are matrices made up of the first τ rows of the partitioned error matrices $\tilde{V}_R = [\tilde{V}_R(\tau)', \tilde{V}_R(-\tau)']'$, $U_F = [U_F(\tau)', U_F(-\tau)']$, and $U_Z = [U_Z(\tau)', U_Z(-\tau)']$.

Now, write $\tilde{W}(\tau) = W(\tau)\tilde{H}_U$ and $\tilde{W}(\tau)^{(1)} = W(\tau)^{(1)}\tilde{H}_U^{1,1}$, where $W(\tau) = [X(\tau), F(\tau), R(\tau)]$, $W(\tau)^{(1)} = [X(\tau), F(\tau)]$,

$$\tilde{H}_U = \begin{pmatrix} I_{M+1} & -A_F & -(A_R - A_F B_{FR}) \\ 0 & I_K & -B_F R \\ 0 & 0 & I_N \end{pmatrix}, \quad (A.22)$$

and

$$\tilde{H}_U^{(1,1)} = \begin{pmatrix} I_{M+1} & -A_F \\ 0 & I_K \end{pmatrix}. \quad (A.23)$$

Obviously, $|\tilde{H}_U| = |\tilde{H}_U^{1,1}| = 1$, so that both $\tilde{H}_U$ and $\tilde{H}_U^{(1,1)}$ are nonsingular. It follows that $|\tilde{W}(\tau)'\tilde{W}(\tau)| = |W(\tau)'W(\tau)|$ and $|\tilde{W}(\tau)^{(1)\prime}\tilde{W}(\tau)^{(1)\prime}| = |W(\tau)^{(1)\prime}W(\tau)^{(1)\prime}|$, so that neither $|\tilde{W}(\tau)'\tilde{W}(\tau)|$ nor $|\tilde{W}(\tau)^{(1)\prime}\tilde{W}(\tau)^{(1)}|$ depends on the unknown parameters A_R, A_F, and B_{FR}. Moreover, observe that $Z(\tau)'Q_{\tilde{W}(\tau)}Z(\tau) = Z(\tau)'Q_{W(\tau)}Z(\tau)$ and $Z(\tau)'Q_{\tilde{W}(\tau)}Z(\tau) = Z(\tau)'Q_{W(\tau)^{(1)}}Z(\tau)$. Using these facts, we can factor (A.20) into a product of (unnormalized) conditional and marginal probability density functions, each of which is either the density of a matrix-variate normal distribution or that of an inverted Wishart distribution. Hence, on appropriate renormalization, we obtain the training sample posterior density given by expression (A.19). Q.E.D.

We now proceed to derive the marginal likelihood for the unrestricted model, M_U. Analogous to the restricted case, we start by updating the prior information from the training sample with data information from the second subsample, as provided by the unrestricted likelihood function:

$$\mathsf{L}(\Phi, \Sigma_{ZZ.R.F}, \Psi, \Sigma_{RR.F}, A_F, \Sigma_{FF} | D(-\tau), \mathsf{M}_U)$$
$$= (2\pi)^{-\frac{1}{2}T_1(M+N+K)} |\Omega(\Sigma)|^{-\frac{T_1}{2}} \exp\left\{ -\frac{1}{2} Tr[\Omega(\Sigma)^{-1} \tilde{V}(-\tau)' \tilde{V}(-\tau)] \right\}, \quad (A.24)$$

where $D(-\tau) = [Y(-\tau), X(-\tau)]$, and $Y(-\tau), X(-\tau)$ and $\tilde{V}(-\tau)$ denote matrices making up the last $T_1 = T - \tau$ rows of the matrices Y, X, and $\tilde{V}$, respectively. Thus, the marginal likelihood under asset-pricing restrictions can be computed by combining the prior density (A.19) with the likelihood function (A.24) then integrating with respect to the parameters $\Phi, \Sigma_{ZZ.R.F}, \Psi, \Sigma_{RR.F}, A_F$, and Σ_{FF}. Again, $S_{YY.X}$ denotes generically the sum of squared residual from the regression of Y on X; the following proposition gives an exact analytical formula for the marginal likelihood of the unrestricted model.

Proposition 4. Let the prior density be as given by expression (A.19). Then, the marginal likelihood for the unrestricted model, M_U, computed on the basis of observations at $t = \tau + 1, \ldots, T$ has the form

$$
m[D(-\tau)|D(\tau), M_U] = (\pi)^{-\frac{1}{2}T_1(M+N+K)}
$$

$$
\times \left\{ \frac{\Gamma_M[0.5(v_{T,h} - 1)]\Gamma_N[0.5(v_{T,h} - M - 1)]\Gamma_K[0.5(v_{T,h} - M - N - 1)]}{\Gamma_M[0.5(v_{\tau,h} - 1)]\Gamma_N[0.5(v_{\tau,h} - M - 1)]\Gamma_K[0.5(v_{\tau,h} - M - N - 1)]} \right\}
$$

$$
\times \left[\frac{|S_{Z(\tau)Z(\tau).W(\tau)}|^{\frac{1}{2}(v_{\tau,h}-1)}}{|S_{ZZ.W}|^{\frac{1}{2}(v_{T,h}-1)}} \right] \left[\frac{|S_{R(\tau)R(\tau).W(\tau)^{(1)}}|^{\frac{1}{2}(v_{\tau,h}-M-1)}}{|S_{RR.W^{(1)}}|^{\frac{1}{2}(v_{T,h}-M-1)}} \right]
$$

$$
\times \left[\frac{|S_{F(\tau)F(\tau).X(\tau)}|^{\frac{1}{2}(v_{\tau,h}-M-N-1)}}{|S_{FF.X}|^{\frac{1}{2}(v_{T,h}-M-N-1)}} \right]
$$

$$
\times \left[\frac{|W(\tau)'W(\tau)|}{|W'W|} \right]^{\frac{1}{2}M} \left[\frac{|W(\tau)^{(1)'}W(\tau)^{(1)}|}{|W^{(1)'}W^{(1)}|} \right]^{\frac{1}{2}N} \left[\frac{|X(\tau)'X(\tau)|}{|X'X|} \right]^{\frac{1}{2}K}, \qquad (A.25)
$$

where, as before, $v_{\tau,h} = \tau + h - 2M - N - K - 1$ and $v_{T,h} = T + h - 2M - N - K - 1$.

Proof. Note first that the unrestricted likelihood for the primary sample can be factored as follows:

$$
\mathcal{L}[\Phi, \Sigma_{ZZ.R.F}, \Psi, \Sigma_{RR.F}, A_F, \Sigma_{FF} | D(-\tau), M_U] = (2\pi)^{-\frac{1}{2}T_1(M+N+K)}|\Sigma_{ZZ.R.F}|^{-\frac{1}{2}T_1}
$$

$$(A.26)$$

$$
\exp\left(-\frac{1}{2}\text{Tr}\left\{\Sigma_{ZZ.R.F}^{-1}\left[Z(-\tau)'Q_{\tilde{W}(-\tau)}Z(-\tau)\right.\right.\right.
$$

$$
\left.\left.\left. + \left(\Phi - \hat{\Phi}_{T_1}\right)'\tilde{W}(-\tau)'\tilde{W}(-\tau)\left(\Phi - \hat{\Phi}_{T_1}\right)\right]\right\}\right) \times |\Sigma_{RR.F}|^{-\frac{1}{2}T_1}
$$

$$
\times \exp\left(-\frac{1}{2}\text{Tr}\left\{\Sigma_{RR.F}^{-1}\left[R(-\tau)'Q_{\tilde{W}(-\tau)^{(1)}}R(-\tau)\right.\right.\right.
$$

$$
\left.\left.\left. + \left(\Psi - \hat{\Psi}_{T_1}\right)'\tilde{W}(-\tau)^{(1)'}\tilde{W}(-\tau)^{(1)}\left(\Psi - \hat{\Psi}_{T_1}\right)\right]\right\}\right) \times |\Sigma_{FF}|^{-\frac{1}{2}T_1}
$$

$$
\times \exp\left(-\frac{1}{2}\text{Tr}\left\{\Sigma_{FF}^{-1}\left[F(-\tau)'Q_{X(-\tau)}F(-\tau)\right.\right.\right.
$$

$$
\left.\left.\left. + \left(A_F - \hat{A}_{F,T_1}\right)'X(-\tau)'X(-\tau)\left(A_F - \hat{A}_{F,T_1}\right)\right]\right\}\right),
$$

where $\hat{\Phi}_{T_1} = [\tilde{W}(-\tau)'\tilde{W}(-\tau)]^{-1}\tilde{W}(-\tau)'Z(-\tau)$, $\hat{\Psi}_{T_1} = [\tilde{W}(-\tau)^{(1)'}\tilde{W}(-\tau)^{(1)}]^{-1}$ $\tilde{W}(-\tau)^{(1)'}R(-\tau), \hat{A}_{F,T_1} = [X(-\tau)'X(-\tau)]^{-1}X(-\tau)'F(-\tau)$, and $\tilde{W}(-\tau) = [X(-\tau),$ $U_F(-\tau), \tilde{V}_R(-\tau)]$.

Next, observe that, for each value of A_R, A_F, and B_{FR}, the following relationships hold: $Z'Q_{\tilde{W}}Z = Z'Q_W Z, R'Q_{\tilde{W}^{(1)}}R = R'Q_{W^{(1)}}R, |\tilde{W}'\tilde{W}| = |W'W|$, and $|\tilde{W}^{(1)'}\tilde{W}^{(1)}| = |W^{(1)'}W^{(1)}|$, where the matrix $\tilde{H}_U$ and $\tilde{H}_U^{(1,1)}$ are as defined in expressions (A.22)

and (A.23). It follows that, conditioned on the data, the expressions $Z'Q_{\tilde{W}}Z$, $R'Q_{\tilde{W}^{(1)}}R$, $|W'W|$, and $|W^{(1)'}W^{(1)}|$ do not depend on the unknown parameters A_R, A_F, and B_{FR}. Making use of these facts, we can construct the joint posterior density of $\Phi, \Sigma_{ZZ.R.F}, B_{FR}, \Sigma_{RR.F}, A_F$, and Σ_{FF} given the data by combining the likelihood function for the unrestricted model, expression (A.26) with the prior density given in expression (A.19). The marginal likelihood for the unrestricted model as given by expression (A.25) can then be obtained in a straightforward manner by integrating this joint posterior density with respect to the parameters $\Phi, \Sigma_{ZZ.R.F}, B_{FR}, \Sigma_{RR.F}, A_F$, and Σ_{FF} over their respective support. Q.E.D.

References

Ahn, S. C., and C. Gadarowski. 1999. Small sample properties of the model specification test based on the Hansen-Jagannathan distance. Working paper, Arizona State University.

Avramov, D. 2002. Stock return predictability and model uncertainty. *Journal of Financial Economics* 64:423–58.

———. 2004. Stock return predictability and asset pricing models. *Review of Financial Studies* 17:699–738.

Berger, J. O., and L. R. Pericchi. 1996. The intrinsic Bayes factor for model selection and prediction. *Journal of the American Statistical Association* 91:109–22.

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

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

Breeden, D. T., M. Gibbons, and R. Litzenberger. 1989. Empirical tests of the consumption oriented CAPM. *Journal of Finance* 44:231–62.

Campbell, J. 1987. Stock returns and the term structure. *Journal of Financial Economics* 18:373–99.

———. 1996. Understanding risk and return. *Journal of Political Economy* 104:572–621.

Cochrane, J. H. 1996. A cross-sectional test of an investment-based asset pricing model. *Journal of Political Economy* 104:572–621.

DeBondt, W., and R. Thaler. 1985. Does the stock market overreact? *Journal of Finance* 40:793–805.

Dumas, B., and B. Solnik. 1995. The world price of foreign exchange risk. *Journal of Finance* 50:445–79.

Fama, E. F., and K. R. French. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33:3–56.

———. 1996. Multifactor explanations of asset pricing anomalies. *Journal of Finance* 51:55–84.

———. 1998. Value versus growth: The international evidence. *Journal of Finance* 53:1975–99.

Fama, E. F., and J. D. MacBeth. 1973. Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy* 81:607–36.

Ferson, W., and S. R. Forester. 1994. Finite sample properties of the generalized method of moments in tests of conditional asset pricing models. *Journal of Financial Economics* 36:29–56.

Geweke, J., and G. Zhou. 1996. Measuring the pricing errors of the arbitrage pricing theory. *Review of Financial Studies* 9:557–87.

Gibbons, M. R., S. A. Ross, and J. Shanken. 1989. A test of the efficiency of a given portfolio. *Econometrica* 57:1121–52.

Ghysels, E. 1998. On stable factor structures in the pricing of risk: Do time-varying betas help or hurt? *Journal of Finance* 53:549–73.

Hansen, L. P. 1982. Large sample properties of generalized method of moments estimators. *Econometrica* 50:1029–054.

Hansen, L. P., and R. Jagannathan. 1997. Assessing specification errors in stochastic discount factor models. *Journal of Finance* 52:557–90.

Hansen, L. P., and S. F. Richard. 1987. The role of conditioning information in deducing testable. *Econometrica* 55:587–613.

Harvey, C. 1991. The world price of covariance risk. *Journal of Finance* 46:111–57.

Harvey, C., and G. Zhou. 1990. Bayesian inference in asset pricing tests. *Journal of Financial Economics* 26:221–54.

Haugen, R. 1995. *The new finance: The case against efficient markets*. Englewood Cliffs, N.: Prentice-Hall.

Huberman, G., and S. Kandel. 1987. Mean variance spanning. *Journal of Finance* 42:873–88.

Jeffreys, H. 1961. *Theory of probability*. New York: Oxford University Press.

Kan, R., and G. Zhou. 1999. A critique of the stochastic discount factor methodology. *Journal of Finance* 54:1221–48.

———. 2003. Modeling non-normality using multivariate *t*: Implications for asset pricing. Working paper, Washington University.

Kass, R. E., and A. E. Raftery. 1995. Bayes factors. *Journal of the American Statistical Association* 90:773–95.

Kandel, S., R. McCulloch, and R. Stambaugh. 1995. Bayesian inference and portfolio efficiency. *Review of Financial Studies* 8:1–53.

Korajczyk, R. A., and C. J. Viallet. 1989. An empirical investigation of international asset pricing. *Review of Financial Studies* 4:553–85.

Lakonishok, J., A. Shleifer, and R. Vishny. 1994. Contrarian investment, extrapolation, and risk. *Journal of Finance* 49:1541–78.

Lettau, M., and S. Ludvigson. 2001. Resurrecting the (C)CAPM: A cross sectional test when risk premia are time-varying. *Journal of Political Economy* 109:1238–87.

Lintner, J. 1965. Security prices, risk, and maximal gains from diversification. *Journal of Finance* 20:587–615.

McCulloch, R., and P. E. Rossi. 1991. A Bayesian approach to testing the arbitrage pricing theory. *Journal of Econometrics* 49: 141–168.

Phillips, P. C. B., and W. Ploberger. 1996. An asymptotic theory of Bayesian inference for time series. *Econometrica* 64:381–412.

Roll, R. 1977. A critique of the asset pricing theory's tests. Part I. On past and potential testability of the theory. *Journal of Financial Economics* 4:129–76.

Shanken, J. 1987. A Bayesian approach to testing portfolio efficiency. *Journal of Financial Economics* 19:217–44.

Sharpe, W. F. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance* 19:425–42.

Sims, C. A. 1988. Bayesian skepticism on unit root econometrics. *Journal of Economic Dynamics and Control* 12:463–74.

Sims, C. A., and H. Uhlig. 1991. Understanding unit root: A helicopter tour. *Econometrica* 59:1591–99.

Solnik, B. 1974. An equilibrium model of the international capital market. *Journal of Economic Theory* 8:500–24.

Tu, J., and G. Zhou. 2004. Data-generating process uncertainty: What difference does it make in portfolio decisions? *Journal of Financial Economics* 72:385–421.

Zellner, A. 1971. *An introduction to Bayesian inference in econometrics*. New York: John Wiley and Sons.