

FINANCE RESEARCH SEMINAR SUPPORTED BY UNIGESTION

“Comparing Asset Pricing Models”

Prof. Jay SHANKEN

Emory University, Goizueta Business School

Abstract

A Bayesian asset-pricing test is developed that is easily computed in closed-form from the standard F-statistic. Given a set of candidate traded factors, we show how this test can be adapted to permit an analysis of Bayesian model comparison, i.e., the computation of model probabilities for the collection of all possible pricing models that are based on subsets of the given factors. We find that the recent q-factor model is superior to the Fama-French three-factor model augmented by profitability and net investment factors, but both models are dominated by five or six-factor models that include a momentum factor and value and profitability factors that are updated monthly. Thus, although the standard value factor is redundant, our tests show that a version that incorporates more timely price information is not.

Monday, May 4, 2015, 14:00-15:30

Room 126, Extranef building at the University of Lausanne

Comparing Asset Pricing Models

Francisco Barillas and Jay Shanken¹

First Draft: August 31st 2014

Current Draft: April 24th 2015

Abstract

A Bayesian asset-pricing test is developed that is easily computed in closed-form from the standard F-statistic. Given a set of candidate traded factors, we show how this test can be adapted to permit an analysis of Bayesian model comparison, i.e., the computation of model probabilities for the collection of all possible pricing models that are based on subsets of the given factors. We find that the recent models of Hou, Xue and Zhang (2015a,b) and Fama and French (2015a,b) are both dominated by five and six-factor models that include a momentum factor along with value and profitability factors that are updated monthly. Thus, although the standard value factor is redundant, a version that incorporates more timely price information is not.

¹ Barillas and Shanken are from Goizueta Business School at Emory University. Shanken is also from the National Bureau of Economic Research. Corresponding author: Jay Shanken, Goizueta Business School, Emory University, 1300 Clifton Road, Atlanta, Georgia, 30322, USA; telephone: (404)727-4772; fax: (404)727-5238. E-mail: jay.shanken@emory.edu. Thanks to seminar participants at Southern Methodist University.

Beginning with the capital asset pricing model (CAPM) of Sharpe (1964) and Lintner (1965), the asset pricing literature in finance has attempted to understand the determination of risk premia on financial securities. The central theme of this literature is that the risk premium should depend on a security's market beta or other measure(s) of systematic risk. In a classic test of the CAPM, Black, Jensen and Scholes (1972), building on the earlier insight of Jensen (1968), examine the intercepts in time-series regressions of excess test-portfolio returns on market excess returns. Given the CAPM implication that the market portfolio is efficient, these intercepts or "alphas" should be zero. A joint F-test of this hypothesis is later developed by Gibbons, Ross and Shanken (1989), who also explore the relation of the test statistic to standard portfolio geometry.² In contrast, recalling the adage, "it takes a model to beat a model," our main goal in this paper is to develop a statistical procedure for determining which of several models is the best.

Like other asset pricing analyses based on alphas, we require that the benchmark factors in an asset pricing model are traded portfolio excess returns or return spreads. For example, in addition to the market excess return, Mkt, the influential three-factor model of Fama and French (1993), hereafter, FF3, includes a book-to-market or "value" factor HML (high-low), which is the difference between the returns on a portfolio of stocks with high book-to-market ratios and a portfolio with low ratios. This model also includes a size factor, SMB (small-big), based on stock-market capitalizations. Over the years, other traded factors have been considered as well. Although consumption growth and intertemporal hedge factors are not traded, one can always substitute mimicking (maximally correlated) portfolios for the non-traded factors.³ While this introduces additional estimation issues, simple spread-portfolio factors are often viewed as proxies for the relevant mimicking portfolios, e.g., Fama and French (1996).

Barillas and Shanken (2015) show that model comparison with traded factors only requires an examination of each model's ability to price the factors in the other model(s). It was not clear to us how to combine statistics from various factor regressions in a classical framework, but this turned out to be feasible using a Bayesian approach. We start by analyzing the joint alpha restriction in a factor model for a set of test assets. Interestingly, we show that the Bayesian summary of the evidence for this hypothesis, the *Bayes factor*, is a function of the GRS F-statistic. For the purpose of comparing models, we specialize this result to the case of a single left-hand-side asset – a factor. The heart of

² See related work by Jobson and Korkie (1982)

³ See Merton (1973) and Breeden (1979), especially footnote 8.

our approach is then the method that we use to aggregate Bayes factors for the different factor regressions.

The joint alpha restriction has been analyzed previously in the Bayesian setting by Shanken (1987b), Harvey and Zhou (1990) and McCulloch and Rossi (1991). The test that we develop here builds on one of the prior specifications considered by Harvey and Zhou. This approach is appealing in that standard “diffuse” priors are used for the betas and residual covariance parameters. Thus, the data dominate estimation of these parameters and the researcher is freed from the obligation to think about “reasonable” values. However, an informative prior *must* be specified for the alphas, which represent deviations from the pricing model restrictions - the main focus of the analysis. In this context, Harvey and Zhou observe that the Bayes factor is “extremely complicated to evaluate” and so numerical methods are utilized. However, the function of the conventional F-statistic mentioned above is easy to calculate and so computational issues need not be an impediment to applications. Henceforth, we refer to this procedure as the Bayesian GRS test or simply B-GRS.

Asset-pricing model comparison is an area of testing that has received very little attention. In fact, we know of no alpha-based statistical analysis of this problem in the finance literature.⁴ We start by considering the nested case in which the factors in one model are a subset of those in a larger model. For example, suppose we wish to test CAPM versus FF3. At first glance, this would appear to require a comparison of the one-factor alphas of the test assets with their three-factor counterparts. Barillas and Shanken (2015) show, however, that if the one-factor alphas of the factors, HML and SMB, are both zero then the usual CAPM zero-alpha restriction for the test assets is equivalent to the FF3 zero-alpha restriction. Viewed in this way, the CAPM is seen to be a restricted version of the FF3 model.

Since both CAPM and FF3 require that the three-factor alphas of the test assets equal zero, examining those alphas cannot help in distinguishing between the models, though they are, of course, relevant to evaluating how good the models are. The bottom line is that for the purpose of *model comparison*, a test of CAPM versus the less restrictive FF3 model only requires an evaluation of the hypothesis that the alphas of HML and SMB on Mkt are both zero. If they are, CAPM is favored, otherwise FF3 is judged superior. In other words, we need only test whether the pricing of the

⁴ A recent paper by Kan, Robotti and Shanken (2013) provides asymptotic results for comparing model R^2 s in a cross-sectional regression framework that does not impose traded-factor constraints and allows the zero-beta rate to differ from the riskfree rate. See Malatesta and Thompson (1993) for an application of Bayesian methods in comparing multiple hypotheses in a corporate finance event study context.

factors is consistent with CAPM. Similarly, in testing non-nested models, only the pricing of factors excluded from each model is relevant for model comparison.

As an illustration of our econometric approach, consider models based on the FF3 factors and assume the market is included in the models. Comparing the two-factor model {Mkt HML} to FF3, for example, amounts to testing whether the intercept in the regression of SMB on Mkt and HML is zero. Calculating the associated Bayes factor is easy enough, in that it is a special case of the B-GRS formula. We will also need to compare CAPM to {Mkt HML}, which involves the HML intercept on Mkt. The comparison of CAPM and FF3 then follows by standard formulas. For the other nested-model comparisons, we follow the same procedure with SMB in place of HML. In general, we end up considering regressions corresponding to all orderings of the non-market factors (two orderings here, depending on whether HML comes before or after SMB). But what about comparing the two-factor models {Mkt HML} and {Mkt SMB} to each other? The key to aggregating all of the nested-model evidence in a comparison of such non-nested models is simply to assign equal prior weights to the various orderings of the non-market factors. This amounts to allocating prior probabilities not only to individual models, but also to groups of models, in a way that permits us to calculate posterior probabilities for all of the models (nested or non-nested).

It is sometimes observed that all models are necessarily simplifications of reality and hence must be false in a literal sense. This provides some motivation for considering whether the models hold approximately, rather than as sharp null hypotheses. Additional motivation comes from recognizing that the factors used in asset-pricing tests are generally *proxies* for the relevant theoretical factors.⁵ With these considerations in mind, we extend our results to obtain simple formulas for testing whether asset-pricing models hold approximately. Implementation of this approach allows us to go beyond the simple test of an exact model and to obtain insight into a model's goodness of fit.

In our main empirical application, we compare models that combine many prominent factors from the literature. In addition to the FF3 factors, we consider the momentum factor, UMD (up minus down), introduced by Carhart (1997) and motivated by the work of Jegadeesh and Titman (1993). We also include the new factors in the recently proposed five-factor model of Fama and French (2015a), hereafter FF5. The additional factors are RMW (robust minus weak) based on the

⁵ Kandel and Stambaugh (1987) and Shanken (1987a) analyze pricing restrictions based on proxies for the market portfolio or other equilibrium benchmark.

profitability of firms, and CMA (conservative minus aggressive) related to firms' new net investments. Hou, Xue and Zhang (2015a, 2015b), henceforth HXZ, have also proposed their own versions of size (ME), investment (IA) and profitability (ROE) factors. In particular, ROE incorporates the most recent earnings information from quarterly data. Finally, we consider the value factor HML^m from Asness and Frazzini (2013), which is based on book-to-market rankings that use the most recent monthly stock price in the denominator. This is in contrast to Fama and French (1993), who use annually updated lagged prices in constructing HML. In total, we have ten factors in our analysis.

Rather than mechanically apply our methodology with all nine of the non-market factors treated symmetrically, we structure the prior so as to recognize that several of the factors are just different versions of the same underlying construct. Therefore, we only consider models that contain one version of the factors in each *category*: size (SMB or ME), profitability (RMW or ROE), value (HML or HML^m) and investment (CMA or IA). Using data from 1972 to 2013 we find that the individual model with highest posterior probability is {Mkt IA ROE SMB HML^m UMD} with six factors. Thus, in contrast to previous findings by HXZ and FF5, value is no longer a redundant factor when the more timely version HML^m is considered; and whereas HXZ also found momentum redundant, this is no longer true with inclusion of HML^m .

The other top models are closely related to our best model, replacing SMB with ME, IA with CMA, or excluding size factors entirely. We also conduct direct tests that compare the best six-factor model either to the HXZ four-factor model or FF5. There is overwhelming support for the six-factor model (or the five-factor model that excludes SMB) in these tests. These and our other model-comparison results are qualitatively similar for different prior specifications, ranging from priors motivated by a market-efficiency perspective to others that allow for large departures from market efficiency.

The model comparison results assess the relative performance of competing models. We also look at absolute performance for the top-ranked model and for the HXZ model. These tests examine the extent to which the models do a good job of pricing a set of test assets, with and without any excluded factors. Although various test assets were examined, results are presented for two sets: 25 portfolios based on independent rankings for either size and momentum or for book-to-market and investments. For the most part, this evidence casts strong doubt on the validity of both models. The "rejection" of the six-factor model is less overwhelming, however, when an approximate version is considered that allows for relatively small departures (average absolute value 0.8% per annum) from

exact pricing. With an average absolute alpha of 1.2%, the approximate model is favored for a range of reasonable priors.

The rest of the paper is organized as follows. Section 1 considers the classic case of testing a pricing model against a general alternative. Section 2 then considers the comparison of nested pricing models. Bayesian model comparison is analyzed in Section 3 and Section 4 provides empirical results for various pricing models. Section 5 concludes. Several proofs of key results are provided in an appendix.

1. Testing a Pricing Model Against a General Alternative

Traditional tests of factor-pricing models compare a single restricted null asset-pricing model to an unrestricted alternative return-generating process that nests the null model. In the classic case considered by GRS, the model's factors are traded zero-investment portfolios and the pricing restrictions constrain the factor model alphas to equal zero. Bayesian tests of these restrictions have been developed by Shanken (1987), Harvey and Zhou (1990) and McCulloch and Rossi (1991).

Statistical Assumptions and Portfolio Algebra

In this section, we focus on a factor model for test asset returns in which the residual returns have a multivariate normal distribution and standard Jeffreys priors are assumed for the factor model betas and the residual covariance matrix. The key alpha parameters have a multivariate normal conditional prior with mean zero and covariance matrix proportional to the residual covariance matrix. This specification was considered earlier by Harvey and Zhou (1990). Here, we derive a formula for the Bayes factor as a function of the standard Wald test statistic (equivalently, the GRS F-statistic) for testing the restricted model in a classical framework. This allows us to avoid the computational difficulties that were emphasized by Harvey and Zhou.

First, we lay out the factor model notation and assumptions. The factor model is a multivariate linear regression with N asset excess returns, r_t , and K factors, for each of T months:

$$r_t = \alpha + \beta f_t + \varepsilon_t \quad \varepsilon_t \sim N(0, \Sigma),$$

where r_t , ε_t and α are $N \times 1$, β is $N \times K$ and f_t is $K \times 1$. The normal distribution of the ε_t is assumed to hold conditional on the factors and the ε_t are independent over time. In matrix form,

$$R = XB + E$$

where

(1.1)

$$R = \begin{bmatrix} r_1' \\ r_2' \\ \vdots \\ r_T' \end{bmatrix}, \quad X = \begin{bmatrix} 1 & f_1' \\ 1 & f_2' \\ \vdots & \vdots \\ 1 & f_T' \end{bmatrix}, \quad B = \begin{bmatrix} \alpha' \\ \beta' \end{bmatrix} \quad \text{and} \quad E = \begin{bmatrix} \varepsilon_1' \\ \varepsilon_2' \\ \vdots \\ \varepsilon_T' \end{bmatrix}.$$

Here, R is $T \times N$, X is $T \times (K+1)$, B is $(K+1) \times N$ and E is $T \times N$. The $T \times K$ matrix of factor data is denoted by F , with F_j referring to the time series for the j th factor.

We assume that the factors are zero-investment returns like the excess return on the market or the spread between two portfolios, like the Fama-French value-growth factor. Under the null hypothesis, $H_0 : \alpha = 0$, we have the usual simple linear relation between expected returns and betas:

$$E(r_t) = \beta E(f_t), \quad (1.2)$$

where $E(f_t)$ is the $K \times 1$ vector of factor premia.

The GRS test of this null hypothesis is based on the F-statistic with degrees of freedom N and $T-N-K$, which equals $(T-N-K)/N$ times the Wald statistic:

$$W = T \frac{\hat{\alpha}' \hat{\Sigma}^{-1} \hat{\alpha}}{1 + \text{sh}(F)^2} = T \frac{(\text{sh}(F, R)^2 - \text{sh}(F)^2)}{1 + \text{sh}(F)^2}. \quad (1.3)$$

Here, $\text{sh}(F)^2 = \bar{F}' \hat{\Sigma}_F^{-1} \bar{F}$ is the maximum squared sample Sharpe ratio over portfolios of the factors, while $\text{sh}(F, R)^2$ is the corresponding measure based on both factor and asset returns. One can also show that W is $T/(T-K-1)$ times the maximum squared t-statistic for the regression intercept, taken over all possible portfolios of the test assets. Note that $\hat{\Sigma}$ and $\hat{\Sigma}_F$ are MLE's for the covariance matrices Σ and Σ_F . The population Sharpe ratios, $\text{Sh}(f)^2$ and $\text{Sh}(f, r)^2$, are based on the true means and covariance matrices.

Under the alternative hypothesis, $H_1 : \alpha \neq 0$, the F-statistic has a noncentral F distribution with noncentrality parameter λ such that

$$\lambda (1 + \text{sh}(F)^2) / T = \alpha' \Sigma^{-1} \alpha = \text{Sh}(f, r)^2 - \text{Sh}(f)^2 = \text{Sh}(f)^2 (\rho^2 - 1) \quad (1.4)$$

where the relative efficiency measure $\rho = \text{Sh}(f) / \text{Sh}(f, r)$ is the correlation between the respective tangency portfolios and indicates the fraction of the maximum Sharpe ratio obtainable with the factors. The first two equalities are established in Gibbons, Ross and Shanken (1989), while the last is

derived in Shanken (1987b) and Kandel and Stambaugh (1987). Under the null hypothesis, $\lambda = 0$ and $\rho = Sh(f)/Sh(f, r) = 1$, so the tangency portfolio corresponding to the factor and asset returns $\tau(f, r)$ equals that based on the factors alone, $\tau(f)$. Thus, the expected return relation in (1.2) is equivalent to this equality of tangency portfolios and their associated squared Sharpe ratios. The larger are the alphas, the lower is the relative efficiency ρ of the factors.

A Bayesian GRS Test

Shanken (1987b) develops a Bayesian approach to testing portfolio efficiency based on (1.4), utilizing the likelihood function for the single parameter λ (equivalently, ρ with $Sh(f)$ treated as known). Here we start with a more conventional Bayesian analysis of the multivariate linear regression system. Since our primary focus is on the economically-important alpha restriction, we posit a standard diffuse prior for β and Σ as in Jeffreys (1961):

$$P(\beta, \Sigma) = |\Sigma|^{-(N+1)/2} \quad (1.5)$$

Asset-pricing theory provides some motivation for linking beliefs about the magnitude of alpha to residual variance. For example, Dybvig (1983) and Grinblatt and Titman (1983) derive bounds on an individual asset's deviation from a multifactor pricing model that are proportional to the asset's residual variance. Pastor and Stambaugh (2000) also adopt a prior for α with covariance matrix proportional to the residual covariance matrix. Building on ideas in McKinlay (1995), they stress the desirability of a positive association between α and Σ in the prior in that this makes extreme increases in $Sh(f, r)^2 - Sh(f)^2$ less likely, as implied by (1.4).⁶ The prior for α is concentrated at 0 under the null hypothesis. Under the alternative, we assume a multivariate normal informative prior for α conditional on β and Σ :

$$P(\alpha | \beta, \Sigma) = \text{MVN}(0, k\Sigma), \quad (1.6)$$

where the parameter $k > 0$ reflects our belief about the potential magnitude of deviations from the expected return relation.

For an individual asset, (1.6) implies that k is the prior expectation of the squared alpha divided by residual variance, or the square of the asset's *information ratio*. By (1.4), this is the

⁶ Also see related work by Pastor and Stambaugh (1999) and Pastor (2000).

expected increment to the maximum squared Sharpe ratio from adding the asset to the given factors. In general, with a vector of N returns, the quadratic form $\alpha'(k\Sigma)^{-1}\alpha'$ is distributed as chi-square with N degrees of freedom, so the prior expected value of $\alpha'\Sigma^{-1}\alpha'$ is k times N . Therefore, given a target value Sh_{\max} for the square root of the expected maximum, the required k is

$$k = (\text{Sh}_{\max}^2 - \text{Sh}(f)^2) / N \quad (1.7)$$

Alternatively, we can just think about the expected return relation and our assessment of plausible deviations from that relation. This is similar to the approach in Pastor (2000). If our subjective view is, say, that alphas should be less than 6% (annualized) with probability 95%, then we would want to choose k such that σ_α is about 3% (annualized). Given a residual standard deviation of 10% per annum, for example, the implied k would be $0.03^2/0.10^2 = 0.09$.

Formally, the Bayes factor BF is the ratio of the marginal likelihoods: $ML(H_0)/ML(H_1)$, where each ML is a weighted-average of the likelihoods for various parameter values.⁷ The weighting is done by the prior densities associated with the different hypotheses. Since the parameters are integrated out, the ML can be viewed as a function of the data (factor and test-asset returns):

$$ML = P(F, R) = \int \int P(F, R | \alpha, \beta, \Sigma) P(\alpha | \beta, \Sigma) P(\beta, \Sigma) d\alpha d\beta d\Sigma \quad (1.8)$$

Here, the likelihood function is the joint density $P(F, R | \alpha, \beta, \Sigma)$ viewed as a function of the parameters. The (unrestricted) $ML(H_1)$ is computed using the priors given in (1.5) and (1.6); the (restricted) $ML(H_0)$ also uses (1.5), but substitutes the zero vector for α .

We can also view the test of $H_0: \alpha=0$ vs $H_1: \alpha \neq 0$ in terms of the proportionality constant in the prior covariance matrix for α ; a test of the value 0 vs the value k . More generally, the null hypothesis can be modified to accommodate an approximate null that allows for (small) deviations from the exact model, as captured by the prior parameter $k_0 < k$. The usual exact null is obtained with $k_0 = 0$. We can now state our main result.

⁷ Use of improper (diffuse) priors for “nuisance parameters” like β and Σ , that appear in both null and alternative models, but proper (informative) priors for parameters like α that are of interest in testing, is in keeping with Jeffreys (1961) and others. See the discussion in Kass and Raftery (1995). Also see our discussion of Bartlett’s paradox in section 4.2.

Proposition 1. Given the factor model in (1.1) and the prior in (1.5)-(1.6), the Bayes factor for $H_0 : \alpha = 0$ vs $H_1 : \alpha \neq 0$ equals

$$BF = \frac{1}{Q} \left(\frac{|S|}{|S_R|} \right)^{(T-K)/2} \quad (1.9)$$

where S and S_R are the $N \times N$ cross-product matrices of the OLS residuals with α unconstrained or constrained to equal zero, respectively. $P(\Sigma | F, R)$, the density for Σ , is given in the appendix, and

$$\begin{aligned} Q &= \int \int \exp \left(\frac{-1}{2a} (\alpha - \hat{\alpha})' \Sigma^{-1} (\alpha - \hat{\alpha}) P(\alpha | \Sigma) P(\Sigma | F, R) \right) d\alpha d\Sigma \\ &= \left(1 + \frac{a}{(a+k)} (W/T) \right)^{-(T-K)/2} \left(1 + \frac{k}{a} \right)^{-N/2}, \end{aligned} \quad (1.10)$$

where $a = (1 + sh(F)^2) / T$, W is given in (1.3) and equals the GRS F-statistic times $NT/(T-N-K)$.

Letting Q_{k_0} be the value of Q obtained with prior value k_0 , the BF for k_0 vs k is

$$BF_{k_0, k} = Q_{k_0} / Q \quad (1.11)$$

Proof. See the Appendix.^{8,9}

It is easy to verify that BF is a decreasing function of W ; the larger the test statistic, the stronger is the evidence against the null that α is (approximately) zero. When $N = 1$, W equals $T/(T-K-1)$ times the squared t-statistic for the intercept in the factor model. Other things equal, the greater the magnitude and precision of the intercept estimate, the bigger is that statistic, the lower is BF and the weaker is the support for the null. For $N > 1$, the same conclusion applies to the maximum squared t-statistic over all portfolios of the test assets.

In terms of the representation in (1.9), the BF decreases as the determinant of the matrix of restricted OLS sums of squared residuals increases relative to that for unrestricted OLS, suggesting that the zero-alpha restriction does not fit the data. BF also decreases as Q increases, where a large Q indicates a relatively small distance between the alpha estimate and the values of alpha anticipated under the prior for the unrestricted model. Q is always less than one since the exponent in (1.10) is

⁸ Harvey and Zhou derive (1.9) and the integral expression for Q . The function of W in (1.10) is our simplification, while (1.11) is both a simplification and generalization of (1.9).

⁹ The formula is identical, apart from minor differences in notation, to the Bayes factor that Shanken (1987b) derives by conditioning directly on the F-statistic, rather than all the data. Thus, it turns out that this simplification entails no loss of information under the diffuse prior assumptions made here.

uniformly negative. As the ratio of determinants is likewise less than 1, a BF favoring the null ($BF > 1$) occurs when Q is sufficiently low, i.e., the prior for α is “inconsistent” with the estimate.

2. Comparing Nested Pricing Models

In the previous section, we considered a test of a factor-pricing model against a general alternative. In this section, we address the testing of one factor-pricing model against another model that includes additional factors. For example, CAPM is nested in FF3 in this sense. This scenario, though clearly one of great interest for asset pricing has not, to our knowledge, been treated formally in the traditional time-series alpha-based framework.¹⁰

Barillas and Shanken (2015) establish a basic equivalence that greatly simplifies the task of comparing nested models. Consider the CAPM as nested in the Fama-French (1993) three-factor model (FF3), for example. In this case, the usual alpha restriction of the single-factor CAPM can be reformulated in terms of the one-factor intercept restriction for the excluded-factor returns (HML and SMB) and the FF3 intercept restriction for the test-asset returns. Since the models differ only with respect to the excluded-factor restrictions, this yields the surprising conclusion that the test-asset returns do not play any role in comparing CAPM and FF3.

More generally, as in the previous section, let f denote a K -vector of traded factors (we omit the time subscript here) and r an N -vector of test-asset excess returns. M is the corresponding pricing model. Assume that $f = (f_1, f_2)$, where f_1 consists of the first J factors in f , with $J < K$. The pricing model with factors f_1 is denoted M_1 . In this context, Barillas and Shanken (2015) show that the nested model M_1 holds for both the test asset returns r and the excluded-factor returns f_2 if and only if those excluded-factor returns satisfy M_1 and the larger model M holds for the test asset returns. Although a purely statistical proof is given, they note that this has a parallel interpretation in portfolio analysis: if the factors f_1 already span the tangency portfolio for the investment universe that includes both sets of factors and the test-asset returns, then adding the f_2 factors will not improve on this tangency portfolio, nor will adding test assets to the factors (and conversely).

Thus, a model M_1 that is nested in a larger model M , in the sense that its factors are all included in M , is nested in the statistical sense that M_1 may be obtained by imposing restrictions on the parameterization of M . As noted above, it follows that the only condition relevant in

¹⁰ In a cross-sectional regression framework, Chen, Roll and Ross (1986) nest the CAPM in a multifactor model with betas on macro-related factors included as well.

distinguishing between M_1 and M is the requirement that M_1 hold for the excluded-factor returns f_2 . The test-asset restriction $\alpha = 0$ is common to both models and, therefore, cannot help in deciding which model performs better. Hence, the test asset returns are not relevant in *comparing* M_1 and M though they are, of course, important for assessing model validity. Barillas and Shanken (2015) further show that the irrelevance of test assets extends to the comparison of a pair of non-nested models, M_{1a} and M_{1b} , since both models can be nested in the model M (their union). As a result, we need only consider the pricing of the factors in each model relative to the factors in the other model (the common factors will automatically be priced). In Section 3, we go further and show how to aggregate this regression evidence to obtain posterior probabilities for each model.

Three asset-pricing tests (classical or Bayesian) naturally present themselves in connection with the nested model equivalence. We can conduct a test of M with factors f and test-asset returns r . Or, we can test M_1 with factors f_1 and test-asset returns r plus the excluded factors f_2 . Finally, we can test M_1 vs M with factors f_1 and “test-asset returns” f_2 . The first two *absolute* tests pit the models (M or M_1) against unrestricted alternatives for the distribution of the test-asset returns. The third *relative* test compares M_1 to M . There is an interesting relation between the B-GRS versions of these tests. We denote the Bayes factor for M in the first test as BF_M^{abs} , for M_1 in the second test as $BF_{M_1}^{abs}$, and for M_1 in the third test as BF^{rel} . Given some additional assumptions similar to those made earlier, we then have

Proposition 2. In addition to the assumptions of Section 1, suppose the regression of f_2 on f_1 has normally distributed errors independent over time. The prior for these regression parameters is of the form in (1.6), but independent of that prior. Then the BFs are related as follows:

$$BF_{M_1}^{abs} = BF^{rel} \times BF_M^{abs} \quad (2.1)$$

Proof. The ML is the expectation under the prior of the likelihood function. Factor the joint density (likelihood function) of factor and test-asset returns into the marginal density for f_1 , times the conditional density for f_2 given f_1 , times the conditional density for R given f . By the prior independence assumptions, the prior expectation of the product is the product of the expectations. Under M , the density for f_2 is unrestricted while that for R is restricted (zero intercepts). Under M_1 , both densities are restricted while under the alternative, both are unrestricted. Equation (2.1) now

follows from the fact that $BF_{M_1}^{abs}$ is the ratio of MLs, with restrictions imposed in the numerator, but not in the denominator. \square

The proposition tells us that the support for the nested model M_1 equals that for the larger model M times the relative support for M_1 in comparison with M . Equivalently, the relative support for M_1 vs M can be backed out from the absolute B-GRS tests, as $BF_{M_1}^{abs} / BF_M^{abs}$. Thus, whether we compare the models directly or relate the B-GRS outcomes for each model, the result is the same. This reflects the fact that, as shown by Barillas-Shanken (2015), the impact of the original test-asset returns r on the absolute tests is the same for each model (equal to BF_M^{abs}) and so cancels out in the model comparison (the ratio).

A Three-Factor Nested-Models Example

To illustrate these ideas, suppose we want to test whether the FF3 model is superior to the CAPM. In this case, $f = \{\text{Mkt}, \text{HML}, \text{SMB}\}$, $f_1 = \{\text{MKT}\}$ and $f_2 = \{\text{HML}, \text{SMB}\}$. The relevant restriction is that the CAPM alphas of SMB and HML are both 0. We evaluate this hypothesis from both the classical and Bayesian perspectives, using factor data for the period 1927 to 2013 obtained from Ken French's website. The GRS statistic is 4.56 with associated p-value 0.01, statistically significant in the conventional sense.

To implement the Bayesian approach, we need to specify the value of k in the prior. Assuming the full model M includes K factors and the nested model M_1 consists of the market factor only, adaptation of the earlier formula for k in (1.7) gives

$$k = (\text{Sh}_{\max}^2 - \text{Sh}(\text{Mkt})^2) / (K - 1). \quad (2.2)$$

The divisor is $K-1$ here since $K-1$ factors are added to the market factor. In practice, since $\text{Sh}(\text{Mkt})$ is unknown, in its place we use the posterior expected value based on the Mkt time series. The time series is assumed to be independent and identically normally distributed in this context, with the usual diffuse priors on the Mkt mean and variance parameters. The resulting posterior expectation is slightly smaller than the sample estimate.¹¹

¹¹ The dependence of the prior for alpha on the *posterior* expected value of $\text{Sh}(\text{Mkt})$ may at first seem odd, but it makes sense when we recall that the entire analysis, including the prior is conditioned on the Mkt returns. This is possible

In our current example, the question is, how big do we think the Sharpe ratio increase might be as a result of adding HML and SMB to the market portfolio? We allow for a 25% increase in this illustration, i.e., $Sh_{\max} = 1.25 \times Sh(\text{Mkt})$. More precisely, the square root of the prior expected squared Sharpe ratio is 1.25 time the market's squared ratio. With $K = 3$ and an estimated value of 0.115 for $Sh(\text{Mkt})$, (2.5) gives $k = 0.0037$. This value of k translates into prior standard deviations of 2.51% and a 2.23% for the HML and SMB CAPM alphas, respectively. The latter is smaller since the SMB residual variance in the Mkt model is lower than that of HML over the full period.

Given this prior specification, the BF for the null (CAPM) vs the alternative (FF3) is 0.13. Thus the data (viewed through the lens of the prior), strongly favor the conclusion that the two alphas are not both zero by odds of more than 7 to 1. Using the fact that the probability for the alternative is one minus the probability for the null, it follows that the posterior probability that the null is true is $BF/(1 + BF)$ when the prior probabilities for both models are 0.5. This gives a posterior probability of 11.6% for CAPM with the BF of 0.13. As the p-value calculation does not even consider the alternative hypothesis, the 1% p-value cannot meaningfully be compared to this posterior probability. Shanken (1987b) discusses this issue in detail. Later, we provide an example in which the posterior probability favors the null, even though the p-value is low by conventional standards.

3. Bayesian Model Comparison

Orderings and Joint Densities

In the previous section, we saw how to compare two nested factor-pricing models. Now suppose we wish to simultaneously compare a sequence of nested models like $\{\text{Mkt}\}$, $\{\text{Mkt HML}\}$ and $\{\text{Mkt HML SMB}\}$. We use braces to denote models, which correspond to subsets of the given factors. This particular sequence of nested models is associated with a factor *ordering* in which Mkt comes first, HML next and SMB last. We use parentheses to refer to an ordering: (Mkt HML SMB).

As in this example, we will assume from now on that Mkt is the first factor in each ordering, which amounts to a prior belief that the market factor is necessarily in the true model. This is motivated by the fact that the market portfolio represents the aggregate supply of securities and, therefore, holds a unique place in portfolio analysis and the equilibrium pricing of assets, e.g., the Sharpe-Lintner CAPM and the Merton (1973) intertemporal CAPM.¹² Beyond this economic

because Mkt, as the first factor in each ordering, always plays the role of an explanatory variable in the time-series regressions.

¹² See Fama (1996) for an analysis of the role of the market portfolio in the ICAPM.

motivation, the assumption that Mkt is first will simplify the specification of our prior for the alphas, which can be conditioned on the observed Mkt returns and the corresponding Mkt Sharpe ratio (see below). In particular, the fact that the average market excess return has historically been around 6-8% per year can be used as a reference point in thinking about plausible alphas.

We will soon see that different orderings of the factors allow for different ways of parameterizing the joint density of the factor returns. This will permit a simple representation of all the models that can be formed from a given set of factors and will facilitate their comparison. Anticipating that, we introduce notation that allows for probabilities that are conditional on an ordering. Given a single ordering and the corresponding sequence of nested models, the analysis reduces to a standard problem in Bayesian model comparison. We need only introduce prior probabilities $P(M|w)$ for each model in a given ordering w . The posterior model probabilities are then given in terms of the prior probabilities and pairwise BFs as:

$$P(M_i | w, F) = \left\{ \sum_j B_{ji}(w) P(M_j | w) / P(M_i | w) \right\}^{-1} \quad (3.1)$$

where

$$B_{ji}(w) = P(F | M_j, w) / P(F | M_i, w).$$

We see that the factor data F influence the *conditional* posterior probabilities through the conditional marginal likelihoods $P(F | M, w)$. These MLs average the likelihood function for a given model over the various parameter values according to the prior for that model, conditional on the given ordering. The higher is $P(F | M_i, w)$ or the lower is $P(F | M_j, w)$, the lower is $B_{ji}(w)$ and the higher is $P(M_i | w, F)$.

Next, we discuss the computation of the conditional MLs, $P(F | M, w)$. The formal details are given in the appendix. The idea is the same as that in (1.8) except that the data now consist of factor returns only. Therefore, we work with the joint density of the factors, assumed to be multivariate normal, and priors for the parameters in the joint density. It turns out to be useful to factor the joint density in accordance with the ordering. For example, if we consider the factors in FF3, there are two orderings with Mkt first: (Mkt HML SMB) and (Mkt SMB HML).

For the ordering (Mkt HML SMB), we factor the joint density by starting with the marginal density for Mkt, multiplying by the conditional density of HML given Mkt, and then by the

conditional density of SMB given Mkt and HML.¹³ Given joint normality of the factors, the conditional densities correspond to linear regressions with normally distributed disturbances.

$$\text{HML} = a + b \cdot \text{Mkt} + e \quad (3.2a)$$

and

$$\text{SMB} = c + d \cdot \text{Mkt} + g \cdot \text{HML} + u. \quad (3.2b)$$

Thus, the parameter space for this characterization of the joint density consists of the Mkt mean and variance, the “alphas” a and c , the “betas” b , d and g , and the residual variances of e and u .

Each regression is a special case of the multivariate regression in Proposition 1, with the dependent variable now a single factor return, rather than a vector of test-asset returns. We have a single residual variance parameter for each regression and the diffuse prior for the betas and residual variance in each is given by (1.6) with $N=1$. As earlier, the regression intercept is either constrained to equal 0 or assumed, under the prior, to be a draw from a normal distribution with mean 0 and variance equal to k times the residual variance. Under $\text{CAPM} = \{\text{Mkt}\}$, the intercepts a and c are both 0. For the model $\{\text{Mkt HML}\}$, c is 0 but a is unconstrained. Finally, under the “full” model $\text{FF3} = \{\text{Mkt HML SMB}\}$, both intercepts are unrestricted.

Test-asset constraints on the full model can be evaluated in the B-GRS framework, but as discussed in the previous section, the test assets are not relevant for model comparison. Therefore, FF3 is simply treated as an unconstrained model in this context. From a portfolio perspective, it is clear that the squared Sharpe ratio will always be maximized with all factors included. The question that is being addressed when we consider the various nested models is whether that maximum can still be attained with a proper subset of the factors. In other words, are some of the original factors unneeded? Fama (1998) considers a related hypothesis in identifying the number of priced state variables in an intertemporal CAPM setting.

The value of k in the prior for the intercepts is determined as in (2.2). It follows from the discussion below (1.6) that, in the context of a factor ordering, k is the expected increment (under the alternative) to the squared Sharpe ratio at each step from the addition of one more factor. By concavity, therefore, the expected increment to the Sharpe ratio declines as more factors are included in the model. This seems like a reasonable property to impose *a priori* - a kind of diminishing returns to “active” investment condition. Likewise, for a given dependent factor, since the residual variance

¹³ Assuming independence of the factor returns over time, the density for each factor is a product of the densities for each month of the sample period.

declines as we include more factors, the expected squared values of the intercepts decline with constant k . It bears noting that, while we think this is a reasonable way of specifying the prior, the results are not sensitive to alternative methods for distributing the total increase in the squared Sharpe ratio.

The prior specification is completed by assuming sequential independence, i.e., the Mkt parameters Θ_1 are independent of the regression parameters Θ_2 in (3.2a) and Θ_3 in (3.2b), which are independent of each other. This implies factorization of the joint prior as $P(\Theta) = P(\Theta_1)P(\Theta_2)P(\Theta_3)$. The factorization of the joint density and the prior are shown in the appendix to imply factorization of the MLs and BFs as well. Thus, the BF comparing CAPM vs FF3 equals the BF comparing {Mkt} vs {Mkt HML}, times the BF comparing {Mkt HML} vs {Mkt HML SMB}.

More generally, assume there are K factors and consider the standard ordering, $w = (F_1 F_2 \dots F_K)$. Each model M_J associated with this ordering consists of the factors $\{F_1 F_2 \dots F_J\}$, for some J between 1 and K . The density under model M_J is a product of densities for each factor j from 1 to K , with the j th conditional density (regression of F_j on $F_1, F_2, \dots F_{j-1}$) equal to

$$P(F_j | F_1, F_2, \dots F_{j-1}; M_J, w) = \begin{cases} \text{restricted if } j > J \\ \text{unrestricted if } j \leq J \end{cases} \quad (3.3)$$

Here, restricted means that the intercept is 0 in that regression. The corresponding ML is then¹⁴

$$P(F | M_J, w) = P(F_1 | M_J, w)P(F_2 | F_1, M_J, w) \dots P(F_K | F_1, F_2, \dots F_{K-1}; M_J, w). \quad (3.4)$$

As in (3.3), the j th term in this product involves restricted densities (prior and data) for $j > J$ and unrestricted densities for $j \leq J$.

We've described the procedure for the standard ordering. Any other ordering, for example $w^* = (F_1 F_3 F_2)$ with $K=3$, will correspond to a permutation of the factor indices. Here, the indices ordering for w^* is (1 3 2). The implied permutation function Φ satisfies $\Phi(1)=1, \Phi(2)=3$

¹⁴ The first (Mkt) term in each ML is always the same and, therefore, cancels out in the resulting BFs and posterior model probabilities.

and $\Phi(3) = 2$. To determine the associated marginal likelihood $P(F | M_J, w^*)$, we proceed as above, everywhere substituting $F_{\Phi(1)}$ for F_1 , $F_{\Phi(2)}$ for F_2 and $F_{\Phi(3)}$ for F_3 .¹⁵

Model Priors and Aggregation of Probabilities over Orderings

Given a prior over the models in each ordering, the conditional BFs can be combined, as in (3.1), to obtain conditional model probabilities. Beyond that, there is also the issue of how to derive a single overall model probability, given that models are generally associated with multiple orderings. Toward this end, suppose the decision as to which non-market factor comes first and which follows is considered a matter of indifference, a priori. Formally, then, we can treat the various possible orderings as indexed by a hyperparameter, over which there is a uniform prior distribution. Not only will this technical device provide a means of aggregating the probabilities for a given model over the different orderings, but it will also permit us to include non-nested models in the model comparison!

As discussed above, with the three Fama-French factors (two non-market factors) there two orderings: $w_1 = (\text{Mkt HML SMB})$ and $w_2 = (\text{Mkt SMB HML})$. There are four models (sets of factors) included in these orderings: Mkt = CAPM in both orderings, $\{\text{Mkt HML}\}$ in ordering 1, $\{\text{Mkt SMB}\}$ in ordering 2 and $\{\text{Mkt HML SMB}\} = \{\text{Mkt SMB HML}\} = \text{FF3}$ in both orderings. Therefore, we would like each model to have prior probability $1/4$. The two orderings have prior probability 0.5 each. Since CAPM and FF3 are associated with both orderings, their conditional probabilities need to be $1/4$ as well ($0.5 \cdot 1/4 + 0.5 \cdot 1/4 = 1/4$). The remaining probability of $1/2$ in each ordering goes to the 2-factor model, either $\{\text{Mkt SMB}\}$ or $\{\text{Mkt HML}\}$, yielding unconditional probabilities equal to $0.5 \cdot 1/2 + 0.5 \cdot 0 = 1/4$. The general case is treated in the appendix.

To summarize, our prior specification starts with equal probabilities for each ordering, i.e., all parameterizations of the joint density of factor returns receive the same weight, apart from the special role of the market. The number of orderings that contain a given model depends on the number of factors in the model, however. Therefore, the conditional model probabilities for each ordering are selected in such a way that all models have the same *unconditional* prior probability. Consequently, other than the market, all factors are treated symmetrically and all models start with the same chance of being chosen, regardless of the number of factors in the model.

¹⁵ Note that the model M_J depends on the ordering. An additional subscript for w is implicit and should be clear from the context.

In the general case of K factors, there are $(K-1)!$ orderings of the non-market factors, so the prior probability is $1/(K-1)!$ for each ordering. Consider a model M with L non-market factors, $0 \leq L \leq K-1$. Given that these factors can be arranged in $L!$ ways after the market factor, and the remaining $K-1-L$ factors can be arranged in $(K-1-L)!$ ways after that, the number of orderings that contain M is $L! \times (K-1-L)!$. Hence, the fraction of orderings that contain M is one over the binomial coefficient

$$\binom{K-1}{L} = (K-1)! / \{L! \times (K-1-L)!\}.$$

Since the total number of models is 2^{K-1} , the unconditional prior probability for each model should be $1/2^{K-1}$. Therefore, assuming the prior probability for M is the same in each ordering w that contains it (zero otherwise), the product of this probability and the fraction of all orderings that contain M must equal $1/2^{K-1}$ or, equivalently,

$$P(M|w) = \binom{K-1}{L} / 2^{K-1}. \quad (3.5)$$

Note that these numbers can indeed be viewed as conditional probabilities since the sum over all models contained in a given ordering w , i.e., the sum over L , is one (the binomial coefficients sum to 2^{K-1}). Having fully specified the prior over orderings and for models in each ordering, we can now state a key result.

Proposition 3. The unconditional (not conditional on an ordering) posterior model probabilities are given by

$$P(M_i | F) = E_{w|F} \{P(M_i | w, F)\}, \quad (3.6)$$

where the expectation is taken with respect to the posterior over the orderings w :

$$P(w|F) = P(F|w)P(w)/P(F)$$

with

$$P(F|w) = E_{M|w} \{P(F | M, w)\} \quad (3.7)$$

and

$$P(F) = E_w \{P(F|w)\},$$

where the first expectation in (3.7) is taken with respect to the conditional prior over the models included in the given ordering and the second expectation is with respect to the prior over the orderings.

Proof. A general principle that we use repeatedly is $P(Y) = E_X \{P(Y|X)\}$. In (3.6), we condition on F throughout in the “background.” Hence the expectation is taken with respect to the distribution of w ,

conditioned on the data. This is the posterior for the hyperparameter w , given in (3.7). The first part of (3.7) is just Bayes theorem. In the second part, we again apply the general principle, but now condition on w in the background. Since we do not condition on the data here, the relevant distribution is the conditional *prior* for the models given the ordering w . The expression for $P(F)$ is a direct application of the principle.

Corollary 1. Test assets are irrelevant in our Bayesian model comparison procedure in the sense that the posterior model probabilities are unchanged if we condition on the test asset returns R , as well as the factor returns F .

Proof. The BF between a pair of models in the same ordering is the ratio of the corresponding absolute BFs in (2.1). The term that reflects the impact of R is BF_M^{abs} , which cancels out in each ratio. Similarly, the prior (conditional on w) for the parameters in the restricted conditional distribution of r given f is the same for all models and orderings and cancels out in (3.1). Therefore, the conditional posterior model probabilities $P(F | M, w)$ in (3.1) and (3.7) do not depend on R either. Similarly, the R -related term in the conditional model priors cancels out in the ratio of $P(F|w)$ to $P(F)$. It follows that the ordering posterior in (3.7) and the unconditional posterior model probabilities in (3.6) are independent of R . \square

In our standard model-comparison procedure, the prior over the orderings will be uniform and $P(F)$ is then a simple average of the conditional MLs, $P(F|w)$. As in the three-factor example, the conditional prior probabilities will be specified so as to induce a uniform (unconditional) prior over the models as well. The uniform approach ensures that differences in posterior probabilities are driven by the data, which seems desirable in this sort of research setting. Nonetheless, Proposition 3 is more general and accommodates situations in which the orderings are not weighted uniformly and/or the prior model probabilities are not all equal. This will be relevant later when we consider a direct comparison of one model against another non-nested model, or a situation in which there are different ways of measuring some of the factors.

We noted earlier that the data influence the conditional posterior model probabilities through the marginal likelihoods (MLs). The equations in (3.7) show that these MLs, the $P(F|M, w)$ terms, determine the ordering posterior probabilities as well. For a given model M , we can divide the competing models into two groups. In the first group are those models that either nest or are nested in M . The second group consists of models that are non-nested with respect to M . By (3.6), the MLs of models in the first group influence the posterior probability for M through both channels, the

conditional model posterior as well as the ordering posterior. Interestingly, the MLs of competing *non-nested* models exert their influence on $P(M|F)$ only through the orderings posterior.

Posterior Model Probabilities for the Three Fama-French Factors

Before we examine the posterior probabilities, it will be informative to look over the annualized alpha estimates (with t-statistics in parentheses). Recall that the BF in favor of the zero-null hypothesis is a decreasing function of the conventional t-statistic for the zero-intercept restriction. For the first ordering, $w_1 = (\text{Mkt HML SMB})$, the alpha of HML on Mkt is a large 3.65% (2.85), while for SMB on Mkt and HML it is just 1.18% (1.04). The large HML alpha is evidence against CAPM, but is consistent with both the two-factor model $\{\text{Mkt HML}\}$ and FF3. In this context, the modest SMB alpha is reasonably consistent with the two-factor model. Given these observations, the conditional posterior probabilities make sense: $P(\{\text{Mkt HML}\} | w_1, F) = 72\%$, $P(\text{FF3} | w_1, F) = 25\%$, $P(\text{CAPM} | w_1, F) = 3\%$.

For the second ordering, $w_2 = (\text{Mkt SMB HML})$, the alpha of SMB on Mkt is 1.34% (1.18) and for HML on Mkt and SMB it is 3.57% (2.79). This latter large alpha at the end of the sequence is evidence against both of the nested models in w_2 , CAPM and $\{\text{Mkt SMB}\}$. Accordingly, the conditional posterior probabilities favor FF3: $P(\text{FF3} | w_2, F) = 75\%$, $P(\{\text{Mkt SMB}\} | w_2, F) = 16\%$, $P(\text{CAPM} | w_2, F) = 9\%$. The higher probability for $\{\text{Mkt SMB}\}$ as compared to CAPM, despite the modest SMB alpha, is partly a reflection of the higher conditional *prior* probabilities assigned to the two-factor models (which occur in just one ordering each).

Now let us turn to the posterior probabilities for the orderings, w_1 and w_2 . These depend on the MLs for models in each ordering relative to those for models in the other ordering. Since CAPM and FF3 are common to both orderings, the much stronger evidence for $\{\text{Mkt HML}\}$, as compared to $\{\text{Mkt SMB}\}$, explains the higher probability for the first ordering: $P(w_1 | F) = 70\%$ versus $P(w_2 | F) = 30\%$. By Proposition 3, the overall (conditional only on F) posterior model probabilities can now be obtained as weighted averages of the posterior probabilities within each ordering. For FF3, which occurs in both orderings, this probability is $P(\text{FF3} | F) = 0.70 \times 25\% + 0.30 \times 75\% = 40\%$. The lower conditional posterior probability for FF3 in w_1 , 25% vs. 75% in w_2 , is due to the stiffer model competition in w_1 , but the final probability of 40% reflects the model's competitiveness overall. Since $\{\text{Mkt HML}\}$ occurs only in w_1 , $P(\{\text{Mkt HML}\} | F) = 0.70 \times 72\% + 0.30 \times 0 = 50\%$. The corresponding probabilities for CAPM and $\{\text{Mkt SMB}\}$, obtained similarly, both happen to be about 5%.

In general, distinct orderings will always have some models in common, but other models that differ. Thus, the set of restrictions considered differs with the given ordering. As a result, evidence for the same model can vary substantially across orderings, as we saw for FF3. In this example, the restricted model {Mkt HML} came out on top, but in other situations the (unrestricted) model that includes all of the factors can dominate the competition.

Perspective on the Methodology

There is a literature on employing Bayesian model selection to determine which subset of a given vector X of exogenous explanatory variables to include in the regression equation for a dependent variable, Y . An example in finance is Avramov (2002), who analyzes which variables should be used to track expected return predictability over time. In our cross-sectional asset-pricing context, the challenge is to decide which *factors* to include in the pricing model. This involves a multivariate system, with a time-series regression equation for each left-hand-side return. But what makes our experimental design fundamentally different from the more typical statistical problem is the following: the factors that are not included as right-hand-side *explanatory* variables for a given model necessarily play the role of left-hand-side *dependent* returns whose pricing must be explained by the model's factors.¹⁶ While this is essentially dictated by economic considerations, it also plays an important role in the Bayesian posterior analysis: all models are evaluated conditional on the same data, consisting of “test asset” returns and factor returns, whether the latter are on the left side or the right. Given this feature of the problem, the model-comparison procedure that we develop, while perhaps complicated in some respects, almost seems inevitable.

The approach of assigning probabilities to orderings may not be very intuitive, but an ordering here simply amounts to a grouping of models. The groups are formed in such a way that (conditional) posterior probabilities can readily be obtained for the models within each group. A given model can be included in several groups, however, so an important feature of our methodology is keeping track of model-group combinations throughout. The prior over these combinations is determined by treating all the groups (orderings) symmetrically and requiring that the sum of prior probabilities across groups be the same for each model. Proposition 3 is then the essential tool for backing out posterior model probabilities from the model-group probabilities and marginal likelihoods, leaning heavily on a version of the law of iterated expectations.

¹⁶ The phrase, “either you’re part of the problem or part of the solution” comes to mind.

It might seem, however, that a more straightforward method of extrapolating from the analytically easier nested case could have been used. Consider again the task of comparing two non-nested models like {Mkt HML} and {Mkt SMB}. Let the associated BFs for comparing the “reference model” CAPM to each model (CAPM in numerator) be BF_1 and BF_2 , respectively. It is tempting to take the product $(1/BF_1) \times BF_2$ as the BF for {Mkt HML} vs. {Mkt SMB} in order to back out the ratio of MLs for these two models, but this does not work.¹⁷ As we have seen, the intercept restrictions common to CAPM and each two-factor model cancel out in BF_1 and BF_2 . For example, the SMB intercept on Mkt and HML is restricted under both the CAPM and {Mkt SMB} models and so cancels out in BF_1 . Consequently, the intercept evidence for SMB on {Mkt HML} and, similarly, for HML on {Mkt SMB}, essential for comparing the fit of the *non-nested* two-factor models, is not reflected in BF_1 or BF_2 . Our procedure of placing a prior on the different factor orderings enables us to get around this problem.¹⁸

4. Empirical Results

In this section, we present model-comparison evidence and B-GRS test results. The model probabilities are shown at each point in time to provide an historical perspective on how posterior beliefs would have evolved as the series of available returns has lengthened. Examining different sets of factors provides additional perspective, as the collection of factors considered in the research community has expanded over time. Thus it is interesting to see how this affects posterior beliefs about the models.

A total of ten candidate factors are considered. First, there are the traditional FF3 factors Mkt, HML and SMB plus the momentum factor UMD. To these, we add the investment factor CMA and the profitability factor RMW of Fama and French (2015a). Finally, we also include the size ME, investment IA and profitability ROE factors in Hou, Xue and Zhang (2015a, 2015b), as well as the value factor HML^m from Asness and Frazzini (2013). The size factors SMB and ME, profitability factors RMW and ROE, and investment factors CMA and IA differ based on the type of stock sorts used in their construction. Fama and French create factors in three different ways. We use what they refer to as their “benchmark” factors. Similar to the construction of HML, these are based on independent (2x3) sorts, interacting size with operating profitability for the construction of RMW,

¹⁷ Technically, the problem is that we would be ignoring the fact that we have a different prior on the joint distribution of factor returns when comparing CAPM to the different two-factor models.

¹⁸ A frequentist approach to asset-pricing model comparison might be developed along the lines of Vuong (1989), but we leave that to future work.

and separately with investments to create CMA. RMW is the average of the two high profitability portfolio returns minus the average of the two low profitability portfolio returns. Similarly, CMA is the average of the two low investment portfolio returns minus the average of the two high investment portfolio returns. Finally, SMB is the average of the returns on the nine small-stock portfolios from the three separate 2x3 sorts minus the average of the returns on the nine big-stock portfolios.

Hou, Xue and Zhang (2015a) construct their size, investment and profitability factors from a triple (2 x 3 x 3) sort on size, investment-to-assets, and ROE. More importantly, the HXZ factors use different measures of investment and profitability. Fama and French (2015a) measure operating profitability as NI_{t-1}/BE_{t-1} , where NI_{t-1} is earnings for the fiscal year ending in calendar year $t-1$, and BE_{t-1} is the corresponding book equity. HXZ use a more timely measure of profitability, ROE, which is income before extraordinary items taken from the most recent public quarterly earnings announcement divided by one-quarter-lagged book equity. IA is the annual change in total assets divided by one-year-lagged total assets, whereas investment used by Fama and French is the same change in total assets from the fiscal year ending in year $t-2$ to the fiscal year in $t-1$, divided by total assets from the fiscal year ending in $t-1$, rather than $t-2$. In terms of value factors, HML^m is based on book-to-market rankings that use the most recent monthly stock price in the denominator. This is in contrast to Fama and French (1993), who use annually updated lagged prices in constructing HML. The sample period for our data is January 1972 to December 2013. Some factors are available at an earlier date, but the HXZ factors start in January of 1972 due to the limited coverage of earnings announcement dates and book equity in the Compustat quarterly files.

4.1 Model Comparison Results

Now we present model comparison results. First, we simultaneously compare all the models that can be formed using the FF3 factors Mkt, SMB and HML. This small example extends the results shown in Section 3 and serves as a good illustration of our methodology. We then conduct our main model empirical analysis, which compares models that can be formed from the ten factors mentioned at the beginning of the section. As will be explained in greater detail below, we consider models containing up to six factors, with at most one factor in each of the categories: size, value, investment and profitability.

Regarding priors, our benchmark scenario assumes that $Sh_{max} = 1.5 \times Sh(Mkt)$ in (2.2), i.e., the square root of the prior expected squared Sharpe ratio for the tangency portfolio based on all six factors is 50% higher than the Sharpe ratio for the market. Here, we refer to the prior for the

unrestricted six-factor model. Given the discussion in Section 2, this is sufficient to determine the sequence of implied Sharpe ratios as we expand the set of included factors from one to all six. For a three-factor model, the corresponding multiple of $Sh(Mkt)$ is 1.27, with four factors it is 1.35, it is 1.43 for five factors and 1.5 for six. We think of the 1.5 choice as a prior with a risk-based orientation, assigning little probability to extremely large Sharpe ratios. Later, we examine the sensitivity of posterior beliefs to this assumption, as we also explore multiples corresponding to a more behavioral perspective.

Relative Model Probabilities with the Three Fama-French (1993) Factors

In previous sections, we presented results using these three factors to illustrate our methodology. Recall that there are four models in all: CAPM, FF3 and the two-factor models $\{Mkt\ HML\}$ and $\{Mkt\ SMB\}$. We now report results of the formal model comparison among these four competing models over time. As noted above, our prior incorporates a 1.27 multiple of the Mkt Sharpe ratio, just slightly larger than that used earlier. We employ data from January 1927 to December 2013.

Figure 1 presents the results of this exercise. The top panel shows the model probabilities while the bottom panel gives cumulative factor probabilities, i.e., the probability that each factor is included in the true model. Since we start with equal prior probabilities for each model, it is not surprising that it takes a while to see a substantial spread in the posterior probabilities. The best-performing model since the mid-1980s has been $\{Mkt\ HML\}$, followed closely by FF3. The probabilities for these models are 51.3% and 39.1%, respectively, at the end of the sample. It is also of interest to note that the full model (FF3) need not have the highest probability. The CAPM and $\{Mkt\ SMB\}$ probabilities generally decline after 1980 and are quite low at the end. However, CAPM would have been perceived as the best-performing model in the 1950s and 1960s, which interestingly was a time when the Fama-French model ranked last. In related evidence, Ang and Chen (2007) and Fama and French (2006) find that CAPM works well for B/M-sorted portfolios before 1963.

The cumulative factor probabilities are shown in the bottom panel. For each factor, this is the sum of the posterior probabilities for models that include that factor. The probabilities at the end of our sample are 90.4% for HML, reflecting its inclusion in the two top models, and 43.5% for SMB. Of course, the probability is one for Mkt by assumption.

[Figure 1]

The empirical analysis above was based on the prior assumption $Sh_{\max} = 1.5 \times Sh(\text{Mkt})$ in (2.2) when working with all six factors. Now we explore the sensitivity of our full-sample results to different prior assumptions. Specifically, we let the Sharpe multiple take on the values 1.25, 1.5, 2, and 3 once five additional factors have been added to the market. The 1.25 multiple is consistent with a view that CAPM is a fairly good model and the financial market is informationally efficient. At the other end of the spectrum, a large multiple like 3 is suggestive of informational inefficiency that is more in line with behavioral perspectives.¹⁹

Table 1 presents the full-sample results for the three Fama-French factors. The sample Sharpe ratio for the market is 0.115 over the 1927-2013 period, while the FF3 Sharpe multiple is 1.23, close to the 1.27 under our baseline prior scenario (a 1.5 multiple for adding five more factors, but only 1.27 when adding two additional factors). With the three Fama-French factors, the top model is always {Mkt HML}, the posterior probabilities rising from 44.9% to 65.2% as Sh_{\max} increases. At the same time, the probabilities for FF3 decline from 42.3% to 23.3%. Overall, although we see some variation in the model probabilities for different priors, the rankings of the models are consistent.

[Table 1]

Model Probabilities with Ten Prominent Factors

For our main empirical application we consider ten factors. Rather than mechanically apply our methodology with all nine of the non-market factors treated symmetrically, it seemed to us more natural to structure the prior so as to recognize that several of the factors are just different versions of the same underlying concept. Therefore, we only consider models that contain one version of the factors in each category: size (SMB or ME), profitability (RMW or ROE), value (HML or HML^m) and investment (CMA or IA). We refer to size, profitability, value and investment as *categorical factors*, in this context, in contrast to the actual factors employed in the various models. Similarly, models in which *some* of the factors are categorical and the rest are standard factors are termed *categorical models*.

¹⁹ MacKinlay (1995) analyzes Sharpe ratios under risk-based and non-risk-based alternatives to the CAPM.

To illustrate the basic idea, consider categorical models based on the standard factors Mkt, HML and the categorical factor Size. As with the three-factor example of Section 3, there are two categorical orderings: (Mkt HML Size) and (Mkt Size HML), and four categorical models: {Mkt}, {Mkt HML}, {Mkt Size}, and {Mkt HML Size} based on these factors. The difference now is that we have two versions, SMB and ME, of the Size factor. Thus, we split the prior probability of $\frac{1}{4}$ for the categorical model {Mkt Size} equally between the two versions of that model, $\frac{1}{8}$ each for {Mkt SMB} and {Mkt ME}. Likewise for the two versions of the three-factor categorical model {Mkt HML Size}. In our actual application, there are four categorical factors, as described above. A categorical model like {Mkt Size HML Profitability}, which contains two of the categorical factors, will have four (2^2) versions, while models that include all four have sixteen (2^4). The uniform prior probabilities over categorical models are then split four or sixteen ways, respectively. The details of how we implement the splits are presented in the Appendix.

Our empirical analysis involves ten factors: Mkt, UMD and eight other non-market factors, two in each of the four categories: size, value, profitability and investment. Since each categorical model has up to six factors and Mkt is always included, there are 32 (2^5) possible categorical models. Given all the possible combinations of UMD and the different types of size, profitability, value and investment factors, we have a total of 162 models under consideration.

The top panel in Figure 2 shows posterior probabilities for the individual models, which were obtained under our baseline prior that allows for a multiple of 1.5 times the market Sharpe ratio. We find that quite a few of the individual models receive non-trivial probability, the best (highest probability) model being {Mkt IA ROE SMB HML^m UMD}. The second-best individual model replaces IA with CMA, the fourth-best uses ME instead of SMB and the sixth one uses both CMA and ME, as opposed to IA and SMB. Both the third and fifth best models are five-factor models that do not have a size factor and differ only in their investment factor choice. The top six models all include UMD, while the seventh-best model does not. All of these models fare better than FF5 and the four-factor model of HXZ.

Figure 3 provides another perspective on the evidence, aggregating posterior results over the different versions of each categorical model. Similar to the findings in the previous figure, by the end of the sample, the six-factor categorical model {Mkt Value Size Profitability Investment UMD} comes in first with posterior probability close to 60% and the five-factor model that excludes size is next, but with probability only slightly over 20%. The third best categorical model consists of the same five categories as in FF5, while the fourth best contains the same four categories as the four-

factor model. However, it is essential that the more timely versions of value and profitability are employed in these models. Specifically, in untabulated calculations, the probability share for HML^m in the FF5 categorical model is 76.1%. Similarly, the shares for ROE are 98.9% in the categorical FF5 model and 96.5% in the categorical four-factor model.

In terms of cumulative probabilities aggregated over *all* models, we see from the bottom panel of Figure 3 that the recently proposed categories, profitability and investment, rank highest. Interestingly, value is third with over 90% cumulative probability. Consistent with the findings in Figure 2, the *categorical share* for HML^m , i.e., the proportion of the cumulative probability for value from models that include HML^m , as opposed to HML, is 95%. Similarly, the categorical share of profitability is 97% for ROE. There is less dominance in the size and investment categories, with shares of 68% for SMB and 62% for IA.

While the analysis above simultaneously considered all 162 possible models, we have also conducted direct tests that compare one model to another. In particular, we test the superiority of our six-factor model to the recently proposed models of HXZ and Fama and French. Such a test can easily be computed with our methodology by assigning zero prior probability to the excluded models. Comparing the top individual model found above, {Mkt IA ROE SMB HML^m UMD}, to the four-factor model of HXZ, the direct test assigns 98.6% probability to the six-factor model. The six-factor model probability is greater than 99% when compared to FF5. With the size factor deleted from the six-factor model, the probabilities still exceed 95%.

The model comparison above was based on a prior assumption that $Sh_{max} = 1.5 * Sh(Mkt)$ in (2.5) when working with six factors. We next compute results in which we let the Sharpe multiple take on the values 1.25, 1.5, 2, and 3 once five additional factors have been added to the market. The 1.25 multiple is consistent with a view that the CAPM is a fairly good model and the financial market is informationally efficient. At the other end of the spectrum, a large multiple like 3 is suggestive of an informational inefficiency that is more in line with behavioral perspectives.

Tables 2 and 3 presents the results for the individual and categorical models, respectively. Both tables show the probabilities for the top seven models under the 1.5 multiple specification. The two best models, {Mkt SMB ROE IA HML^m UMD} and {Mkt SMB ROE CMA HML^m UMD}, are also the two best under the more behavioral priors that allow for increases in the Sharpe ratio of 2 and 3 times the market ratio. These two models are among the top four under the lower-multiple specification, though the posterior probabilities are more diffuse in this case. It is also worth noting that the probabilities for the two best models rise from 23% to 42% and from 14.4% to 25.7% as the

multiple increases from 1.5. The top model rankings for the categorical models are also stable across the different priors. The six-factor categorical model {Mkt SIZE PROF INV VAL MOM} is always at the top regardless of the prior and its posterior probability increases substantially as the multiple increases. The categorical model that excludes size comes in second for all but the lowest multiple, where it trails the model that excludes momentum just slightly.

We also conduct prior sensitivity analysis with regard to which versions of the factors should be chosen for a given category - what we have referred to as the categorical shares. We noted above, that the more timely HML^m accounts for 95% of the cumulative probability for the value category. Table 4 shows that timely value remains responsible for the lion's share of the cumulative value probability across the different priors, especially at higher multiples. Varying the prior also yields fairly similar results for timely profitability (ROE), IA and SMB.

4.2 Relative Tests: Are Value and Momentum Redundant?

Barillas and Shanken (2015) show that when comparing two asset-pricing models, all that matters is the extent to which each model prices the factors in the other model. Hou, Xue and Zhang (2015b) and Fama and French (2015a) regress HML on their models that exclude value and cannot reject the hypothesis that HML's alpha is zero, thus concluding that HML is redundant. In addition, HXZ show that their model renders the momentum factor, UMD, redundant. On the other hand, our results above show that the model {Mkt, SMB ROE IA UMD HML^m}, which receives highest posterior probability, contains both a value (HML^m) and a momentum factor (UMD).

To shed further light on this finding, Table 5 shows the annualized intercept estimates for each factor in the top model when it is regressed on the other five factors. We observe that the intercepts for HML^m and UMD are large and statistically significant, rejecting the hypothesis of redundancy. HML^m has an alpha of 6.1% (t-stat 5.26) and UMD has an alpha of 6.7% (t-stat 3.96). When we regress the standard value factor, HML, on the non-value factors {Mkt, SMB ROE IA UMD} in our top model we find, as in the earlier studies, that it is redundant. The intercept is 0.99% with a t-stat of 0.81. The different results for the two value factors is largely driven by the fact that HML^m is strongly negatively correlated (-0.65) with UMD, whereas the correlation is only -0.15 for HML²⁰. The negative loading for HML^m when UMD is included lowers the model expected return and raises the HML^m alpha, so that this timely value factor is not redundant.

²⁰ Asness and Frazzini (2013) argue that the use of less timely price information in HML "reduces the natural negative correlation of value and momentum."

We now evaluate the hypothesis that HML^m is redundant from a Bayesian perspective. Figure 4 shows the results for the Bayesian intercept test on the other factors. As discussed above, the prior under the alternative follows a normal distribution with zero mean and standard deviation σ_α . The larger the value of σ_α , the higher the increase in the Sharpe ratio that one can expect to achieve by adding a position in HML^m to investment in the other factors. The horizontal axis in each panel of the figure shows the prior multiple. This is a multiple of the Sharpe ratio for the factors in the null model that excludes HML^m . The Bayes factor in favor of the null is plotted in the top left panel, while the top right panel gives the posterior probability. Both quickly decrease to zero as the prior Sharpe multiple under the alternative increases, strongly supporting the conclusion that HML^m is not redundant.

Although the inference is not sensitive to the prior here, in other cases it may well be. The lower panels of Figure 4 provide information about the prior that should be helpful in identifying the range of multiples that correspond to one's own belief. The bottom left panel shows σ_α , which gives an idea of the likely magnitude of α 's envisioned under the alternative.²¹ For example, to get an increase in the Sharpe ratio of 25% we would need a very large σ_α of about 7.5% per year. Finally, for additional perspective, the bottom right panel gives the prior Sharpe ratio expressed as a multiple of the *market's* ratio. A multiple of one corresponds to the Sharpe ratio under the null model {Mkt SMB ROE IA UMD}, which in this case is around 4.1 times the market's ratio.

[Figure 4]

The Bayesian analysis for UMD redundancy (not shown) looks much the same as Figure 4. To highlight the role of HML^m in this finding, we exclude that factor and show that the evidence then favors the conclusion that UMD *is* redundant with respect to the remaining factors {Mkt SMB IA ROE}. This essentially confirms the earlier finding of Hou, Xue and Zhang (2015b), but with SMB as the size factor, rather than ME. In Figure 5, we observe that for any value of the prior, the BF is above one, representing support for the null hypothesis. Accordingly, the posterior probability that the null hypothesis (UMD alpha is zero) is true is always above 50%, with values over 80% for Sharpe ratio multiples around 1.15 (see top right panel). The conventional p-value also exceeds 50% here, as indicated by the horizontal line in the figure.

²¹ In general, the plot of σ_α is based on the average residual variance estimate for the left-hand-side assets.

[Figure 5]

4.3 Absolute Test Results

We saw above that over the sample period 1972-2013, the model with the highest posterior probability is the six-factor model {Mkt IA ROE SMB HML^m UMD}. Now we will evaluate this model, as well as the four-factor model of HXZ, from the absolute perspective by challenging them to explain the average returns on test assets. The horizontal axis in the B-GRS absolute test figures show the multiple of the Sharpe ratio for the factors in the given model. This is the multiple under the alternative that the left-hand-side assets are not priced by the model.

Although a wide variety of test-asset portfolios have been examined, we present results for two representative sets that serve to illustrate some interesting findings. The first set of portfolios is based on independent stock sorts on size and momentum, whereas the second set is constructed by sorting stocks on book-to-market and investment. Strictly speaking, the two models considered in this section are not nested because the HXZ model uses ME, whereas our top model uses SMB. However, the results are nearly identical whether one uses ME or SMB in the models.

To test the HXZ model, we initially follow common practice and only employ the test-asset portfolios. Then we add in the excluded factors UMD and HML^m as left-hand-side assets. Using the 25-size/momentum portfolios from January 1972 to December 2013, the GRS statistic for the HXZ model is 2.72 with a p-value of 2e-5, rejecting the model. A descriptive statistic that has also been used to judge model performance is the average of the absolute values of the test-asset alphas, e.g., Fama and French (2015a). The HXZ model produces average absolute alpha estimates of 1.42% per annum. When we add the excluded factors UMD and HML^m as left-hand-side assets, the GRS statistic is 10.5 with p-value virtually zero, but the average absolute alpha estimate increases only slightly to 1.45%.

The B-GRS results with size/momentum portfolios are given in Figure 6. The blue line in each panel shows the results without the excluded factors UMD and HML^m, whereas the red dashed line adds those factors as left-hand-side assets. We see in the top right panel that the probability for the HXZ model is close to zero for Sharpe multiples in the range of 1.1 to 1.6 (blue), but when UMD and HML^m are added, there is even stronger evidence against the model, with the probability close to zero for a much wider range of priors (red).

[Figure 6]

Next, we examine the absolute performance of the six-factor model with the same 25-size/momentum portfolios. The GRS statistic is 3.5, the corresponding p-value of 5e-8 strongly rejecting the model in a classical sense. The B-GRS results also provide strong evidence (not shown) against the null hypothesis. The probability of null curve looks very similar to the red line in Figure 6, quickly declining to an extended zero-probability range. Interestingly, in this case the average absolute alpha is 1.93% per annum, which is much higher than the 1.42%/1.45% under the HXZ model (with/without UMD and HML^m). Yet, we know from our model comparison analysis that the six-factor model is strongly preferred to the HXZ model. This is an example of the sort of conflict discussed in Barillas and Shanken (2015), who show that model comparison must be based on excluded-factor restrictions and that test-asset metrics can be misleading.

According to Proposition 2, however, we should be able to back out the relevant information about excluded-factor constraints by taking the ratio of Bayes factors (BFs) for the absolute tests of the HXZ model and the six-factor model. In this context, the appropriate version of the BF for HXZ includes the restrictions for UMD and HML^m. We have verified that $BF_{HXZ}^{abs} / BF_{6\text{-factor}}^{abs}$ is indeed very small, strongly favoring the six-factor model, except for Sharpe multiples close to one. While this calculation is strictly justified only for nested models, it should provide a good approximation here, given the lack of sensitivity to which size factor is employed in the models.

Now we turn to the results for the 25 portfolios formed on sorts by book-to-market and investment. For the HXZ model, the GRS statistic is 1.53 with a p-value of 0.05. The average absolute alpha is 1.44% per annum. Adding the excluded factors UMD and HML^m increases the GRS statistic to 1.94. The p-value is now much smaller, 0.004, but the average absolute alphas are only slightly higher and equal to 1.47%. The Bayesian results are plotted in Figure 7. For all but the tightest priors, the probability for the null is close to zero with HML^m and UMD considered (red line), whereas the probability based solely on test assets (blue line) is substantial and never below 20%.

[Figure 7]

Next, we use the same 25 book-to-market and investment portfolios to evaluate the six-factor model. The GRS statistic is 2.89 with a p-value of 5e-6. More interestingly, the average absolute

alpha is 2.88% in this case, which is double the value under the HXZ model. As with the size/momentum test portfolios, focusing on this test-asset metric would incorrectly give the impression that the six-factor model is inferior to the HXZ model. Figure 8 plots the Bayesian GRS results. Both the BF and the probability of the null for this model look better (for the null) than the red lines in the top two panels of Figure 7 and worse than the blue lines in that figure. Again, this shows that when the excluded factors are incorporated in the analysis, the absolute tests are consistent with the results of the relative test and favor the six-factor model.

[Figure 8]

We conclude this section with some additional observations about the Bayesian analysis. First, note that the probability for the models in Figures 7 and 8 rebounds from zero in each case and becomes substantial, approaching one (apparent for blue line in Figure 7) as the Sharpe multiple and prior standard deviation for alpha get large. This is an example of Bartlett’s paradox. Understanding what drives this result is a useful exercise in “looking under the hood” of the Bayesian approach.²² When the Sharpe multiple is very large, substantial prior probability under the alternative is assigned to very large alphas. Now suppose the OLS *estimate* of alpha is 4% (annualized) with a standard error of 1.5%, and think about the likelihood for alphas that exceed, say, 50% in magnitude. This will be much lower than the likelihood for $\alpha = 0$, even though the estimate is “significantly” different from 0. The result is a Bayes factor that strongly favors the null hypothesis. Thus, in evaluating pricing hypotheses of this sort, it is essential to form a “reasonable” *a priori* judgment about the magnitude of plausible alphas (reflected in the choice of the parameter k).

The differing classical and Bayesian views about model validity that emerge in Figure 7 also deserve further comment. The p-value of 5% in the evaluation based solely on test assets (blue line) would typically be interpreted as evidence against the null. However, the posterior probability for the null is substantially higher than 5% for all priors, exceeding 50% for some more behavioral priors. This finding is consistent with Lindley’s paradox: in sufficiently large samples, the posterior probability corresponding to a fixed p-value will be close to one, even if the p-value is small.²³ This divergence reflects the fact that posterior probabilities and p-values are fundamentally different

²² The likelihoods are averaged over values of beta and the residual covariance parameters as well as alpha.

²³ Intuitively, an alpha estimate with a t-statistic of two, say, will be close to zero when the sample is very large. The likelihood for $\alpha = 0$ will be quite high in this case, whereas the likelihood for alternative values that still have substantial prior probability but are further from zero, will be much lower. As a result, the Bayes factor (ratio of marginal likelihoods) will strongly favor the null.

measures. Whereas the former reflects likelihoods under both the null and alternatives, the latter is a tail probability under the null that makes no reference to the distribution under the alternative. Nonetheless, p-values are often treated in practice *as if* they are posterior probabilities. The findings in Figure 7 serve as a reminder that this can lead to less than sensible conclusions and highlights what some perceive as an advantage of the Bayesian approach.²⁴

4.4 Results for an Approximate Model

We have learned that for priors moderate to fairly large Sharpe multiples, the evidence favors an unrestricted model over the restricted pricing model {Mkt SMB ROE IA UMD HML^m}, the “winner” in our model comparison contest. But perhaps the model, nonetheless, provides a good *approximation* to the data. After all, one might argue that models, by their nature, always leave out some features of reality and so cannot plausibly be expected to hold *exactly*.²⁵ We address these issues in the Bayesian framework by modifying the prior for a model to accommodate relatively small deviations from the exact specification. The modified BF was given earlier in Proposition 1.

The blue line in Figure 9 shows the results of an analysis in which the prior under the *null* assumes that $\sigma_\alpha = 1\%$ (annualized). This allows for deviations from the exact version of {Mkt SMB ROE IA UMD HML^m} that, on average, have expected value of about 0.8% in magnitude. These deviations give rise to a higher Sharpe ratio under the approximate null hypothesis, about 10% larger than that for the exact null. Thus, whereas the starting point earlier was at a Sharpe ratio multiple of 1, the blue line now starts at a ratio just over 1.1. The test assets are the 25 size/momentum portfolios. Not surprisingly, the posterior probabilities in Figure 9 for the less restrictive model are higher than the probabilities obtained earlier for the exact model. There is no longer a “zero probability range” for the approximate null, but the model probabilities are still less than 0.5 over what we would consider the relevant range of prior Sharpe multiples. Thus, the unrestricted alternative is still favored.

To further explore the fit of the six-factor model, we increase σ_α to 1.5% (expected alpha about 1.2%), which corresponds to a Sharpe multiple just over 1.2. The probability for this level of approximation, shown by the black dashed curve is now greater than 0.5 over most of the prior range.

²⁴ Sample size is automatically incorporated in the BF. While it is sometimes recognized that the significance level in a classical test should be adjusted to reflect sample size, this can be difficult to operationalize and is generally ignored.

²⁵ In the case of exact models, BFs still provide an indication of the “relative success” of the models at predicting the data, e.g., Kass and Raftery (1995), or the “comparative support” the data provide for the models, e.g., Berger and Pericchi (1996).

This sort of sensitivity analysis provides a computationally simple and conceptually appealing Bayesian complement to the descriptive statistics employed by Fama and French (2015a) to evaluate “goodness of fit” for a misspecified model. An advantage of this extension of the B-GRS framework over the conventional F test is that it allows more subtle and informative inferences to be obtained in situations where the sample size is large and models are routinely rejected at conventional levels as above or, e.g., in Fama and French (2015b).

5. Conclusion

We have developed a Bayesian asset-pricing test that requires a prior judgment about the magnitude of plausible model deviations or "alphas" and is easily calculated from the GRS F-statistic. Given a set of candidate traded factors, we show how this test can be adapted to permit an analysis of Bayesian model comparison, i.e., the computation of model probabilities for the collection of all possible pricing models that are based on subsets of the given factors.

Our work is clearly in the tradition of the literature on asset-pricing tests. However, Bayesian analysis has also been used to address other kinds of questions in finance. For example, Pastor and Stambaugh (2000) are interested in comparing models too, but from a different perspective. As they note, the objective of their study “is not to choose one pricing model over another.” Rather, they examine the extent to which investors’ prior beliefs about alternative pricing models (one based on stock characteristics and another on a stock’s factor betas) impact the utility derived from the implied portfolio choices. Utility-based metrics are undoubtedly important, but complementary to our focus on *inference* about models in this paper, and we hope to turn our attention to them in future work.

While we have analyzed the “classic” statistical specification with returns that are independent and identically normally distributed over time (conditional on the market), extensions to accommodate time-variation in parameters and conditional heteroskedasticity of returns would be desirable. The factors examined in Assness and Frazzini (2013), Hou, Xue and Zhang (2015a,b) and Fama and French (2015a,b) have been studied in our preliminary empirical exploration, but other factors related to short and long reversals, the levels of beta and idiosyncratic volatility, and various measures of liquidity could be considered in future work as well.

References

- Ang and Chen, 2007, CAPM over the long run: 1926-2001, *Journal of Empirical Finance*, 14, 1-40.
- Avramov, Doron, 2002, Stock Return Predictability and Model Uncertainty, *Journal of Financial Economics* 64, 423-458.
- Barillas, Francisco and Jay Shanken, 2015, Which Alpha?, Working paper, Emory University.
- Berger J.O. and Pericchi L.R. 1996, The Intrinsic Bayes Factor for Model Selection and Prediction. *Journal of the American Statistical Association*, 91, 433, p. 109-122.
- Black, Fisher, Michael C. Jensen and Myron Scholes, 1972, The Capital Asset Pricing Model: Some Empirical Tests, in *Studies in the Theory of Capital markets*. Michael C. Jensen, ed. New York: Praeger, pp. 79-121
- Breeden, Douglas, 1979, An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities, *Journal of Financial Economics* 7, 265-296.
- Carhart, Mark, 1997, "On Persistence in Mutual Fund Performance". *Journal of Finance* 52, 57-82.
- Dybvig, Phillip H, 1983, An explicit bound on individual assets' deviations from APT pricing in a finite economy, *Journal of Financial Economics* 12, 483-196.
- Fama, Eugene F., 1996, Multifactor Portfolio Efficiency and Multifactor Asset Pricing, *Journal of Financial and Quantitative Analysis* 31, 441-65.
- Fama, Eugene F., 1998, Determining the Number of Priced State Variables in the ICAPM, *Journal of Financial and Quantitative Analysis* 33, 217-31.
- Fama, Eugene F., And Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3-56.
- Fama, Eugene F., And Kenneth R. French, 1996, Multifactor Explanations of Asset Pricing Anomalies, *Journal of Finance* 51, 55-84.
- Fama, Eugene F., And Kenneth R. French, 2006, The Value Premium and the CAPM, *Journal of Finance* 61, 2163-2185.
- Fama, Eugene F., And Kenneth R. French, 2015a, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1-22.

- Fama, Eugene F., And Kenneth R. French, 2015b, Dissecting anomalies with a five-factor model, Manuscript, Booth School of business, University of Chicago.
- Grinblatt, Mark and Sheridan Titman, 1983, Factor pricing in a finite economy, *Journal of Financial Economics* 12, 497-507.
- Gibbons, M., Ross, S., Shanken, J., 1989, A test of the efficiency of a given portfolio. *Econometrica* 57, 1121-1152.
- Harvey, Campbell R. and Guofu Zhou, 1990, Bayesian inference in asset pricing tests. *Journal of Financial Economics* 26, 221-254.
- Hou, Kewei, Xue Chen, and Zhang, Lu, 2015a, Digesting Anomalies: An Investment Approach, *Review of Financial Studies* 28, 650-705.
- Hou, Kewei, Xue Chen, and Zhang, Lu, 2015b, A Comparison of New Factor Models, Working paper.
- Jeffreys, H. 1961, *Theory of Probability*, 3rd ed. Oxford University Press, New York.
- Jegadeesh, Narasimhan and Sheridan Titman, 1993, Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. *Journal of Finance*, 48, 65-91.
- Jensen, Michael C., 1968, The Performance of Mutual Funds in the Period 1945-1964., *Journal of Finance*, 23, 389-416.
- Jobson, D. and R. Korkie, 1982, Potential performance and tests of portfolio efficiency, *Journal of Financial Economics*, 10, 433-466.
- Kan Raymond, Cesare Robotti and Jay Shanken, 2013, Pricing Model Performance and the Two-Pass Cross-Sectional Regression Methodology, *Journal of Finance*, 68, 2617-2649.
- Kandel, Shmuel and Robert F. Stambaugh, 1987, On correlations and inferences about mean-variance efficiency, *Journal of Financial Economics*, 18, 61-90.
- Kass, Robert E. and Adrian E. Raftery, 1995, Bayes Factors, *Journal of the American Statistical Association* 90, 773-795.
- Lintner, John, 1965, The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets, *Review of Economics and Statistics*, 47, 13-27.

- Malatesta, Paul H. and Rex Thompson, 1993. Government Regulation and Structural Change in the Corporate Acquisitions Market: The Impact of the Williams Act, *Journal of Financial and Quantitative Analysis* 28, 363-379.
- McCulloch, Robert and Peter E. Rossi, 1991, A Bayesian approach to testing the arbitrage pricing theory, *Journal of Econometrics*, 49, 141-68.
- MacKinlay, A. Craig, 1995 Multifactor models do not explain deviations form the Capital Asset pricing Models, *Journal of Financial Economics*, 38, 3-28..
- Merton, Robert, 1973, An intertemporal capital asset pricing model, *Econometrica*, 41, 867-887.
- Pastor, Lubos, 2000, Portfolio selection and asset pricing models, *Journal of Finance*, 55, 179-223.
- Pastor, Lubos and Robert F. Stambaugh, 1999, Costs of equity capital and model mispricing, *Journal of Finance* 54, 67-121.
- Pastor, Lubos and Robert F. Stambaugh, 2000, Comparing asset pricing models: an investment perspective, *Journal of Financial Economics*, 56, 335-381.
- Rao, Radhakrishna , 1973, Linear Statistical Inference and its Applications: Second Edition, Wiley Series in Probability and Statistics.
- Sharpe, F. William, 1964, Capital asset prices: a theory of market equilibrium under conditions of risk, *Journal of Finance*, 19, 425-442.
- Shanken, Jay, 1987a, Multivariate proxies and asset pricing relations: Living with the Roll critique, *Journal of Financial Economics*, 18, 91-110.
- Shanken, Jay, 1987b, A Bayesian approach to testing portfolio efficiency, *Journal of Financial Economics*, 19, 195-216.
- Stewart, Kenneth, 1995, The functional equivalence of the W, LR, and LM statistics, *Economics Letters* 49, 109-112.
- Vuong, Quang, 1989, Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses, *Econometrica* 57, pp. 307-333.

Derivation of the formula for Q.

Let $x = (\alpha - \hat{\alpha})'(k\Sigma)^{-1}(\alpha - \hat{\alpha})$. We treat $\hat{\alpha}$ as a fixed vector in this analysis. The inner integral in the definition of Q in (1.10) can be viewed as the expectation of $\exp(tx)$, a function of the random variable α , with $t = -k/(2a)$. Given the prior $P(\alpha | \beta, \Sigma) = \text{MVN}(0, k\Sigma)$, it follows that the conditional distribution of x given Σ is noncentral chi-square with N degrees of freedom and noncentrality parameter $\lambda_\alpha = k^{-1}\hat{\alpha}'\Sigma^{-1}\hat{\alpha}$. Thus, the inner integral is equal to the moment-generating function of this noncentral chi-square evaluated at t .²⁶

$$\frac{\exp\left(\frac{\lambda_\alpha t}{1-2t}\right)}{(1-2t)^{N/2}} = \frac{\exp\left(\frac{-\hat{\alpha}'\Sigma^{-1}\hat{\alpha} / (2a)}{1+k/a}\right)}{(1+k/a)^{N/2}} = \frac{\exp\left(\frac{-\hat{\alpha}'\Sigma^{-1}\hat{\alpha}}{2(a+k)}\right)}{(1+k/a)^{N/2}} \quad (\text{A.1})$$

Next, we need to evaluate the integral of the product of (A.1) and the posterior density for Σ . Given the distributional assumptions and the prior in Section 1, the posterior distribution of Σ is inverted Wishart, $W^{-1}(S, T-K)$, and so the posterior for Σ^{-1} is $W(S^{-1}, T-K)$. Therefore, by the result on p.535 of Rao (1973), $\hat{\alpha}'\Sigma^{-1}\hat{\alpha}$ is distributed as $\hat{\alpha}'S^{-1}\hat{\alpha}$ times a chi-square variable with $N-K$ degrees of freedom. Thus the desired integral is $(1+k/a)^{-N/2}$ times the expectation of $\exp(tx)$, where x now refers to the chi-square variable and $t = -\hat{\alpha}'S^{-1}\hat{\alpha} / [2(a+k)]$. Hence we need to evaluate the moment-generating function of the (central) chi-square.²⁷

$$(1-2t)^{-(T-K)/2} = \left(1 + \hat{\alpha}'S^{-1}\hat{\alpha} / (a+k)\right)^{-(T-K)/2}$$

Since $W = \hat{\alpha}'(S/T)^{-1}\hat{\alpha} / a$ by (1.3), it follows that

$$Q = \left(1 + \frac{a}{(a+k)}(W/T)\right)^{-(T-K)/2} \left(1 + \frac{k}{a}\right)^{-N/2} \quad (\text{A.2})$$

Proof of (1.11) when $k_0 = 0$

²⁶ http://en.wikipedia.org/wiki/Noncentral_chi-squared_distribution#Moment_generating_function

²⁷ http://en.wikipedia.org/wiki/Chi-squared_distribution

It remains to evaluate the ratio of determinants, $|S|/|S_R|$ in (1.9). It is a standard result in the multivariate literature that the likelihood ratio test statistic is $LR = T \ln(|S_R|/|S|)$. Gibbons, Ross and Shanken (1989) further note that $LR = T \ln(1 + W/T)$ so that $\exp(LR/T) = 1 + W/T = |S_R|/|S|$.²⁸ Therefore,

$$(|S|/|S_R|)^{(T-K)/2} = (1 + W/T)^{-(T-K)/2} = Q_0 \quad (\text{A.3})$$

and (1.11) follows from (A.2) and (A.3).

Allowing for $k_0 \neq 0$

Let BF and BF_{k_0} be the Bayes factors for comparing the prior value 0 with the values k and k_0 , respectively. The Bayes factor for the approximate null is the ratio of MLs based on the prior values k_0 and k . This is $(1/BF_0)/(1/BF) = BF/BF_0$, where the ML for the exact null cancels out since it is in the numerators of both BF and BF_0 . By (1.9), this is Q_{k_0}/Q . as given in equation (1.11).

²⁸ This also follows from Stewart (1995) by formulating the zero-alpha restriction in terms of his equation (7) with $q=1$. Let C be a $1 \times K$ vector with 1 in the (1,1) position and zeroes elsewhere, M an $N \times N$ identity matrix and D a $1 \times N$ zero vector.

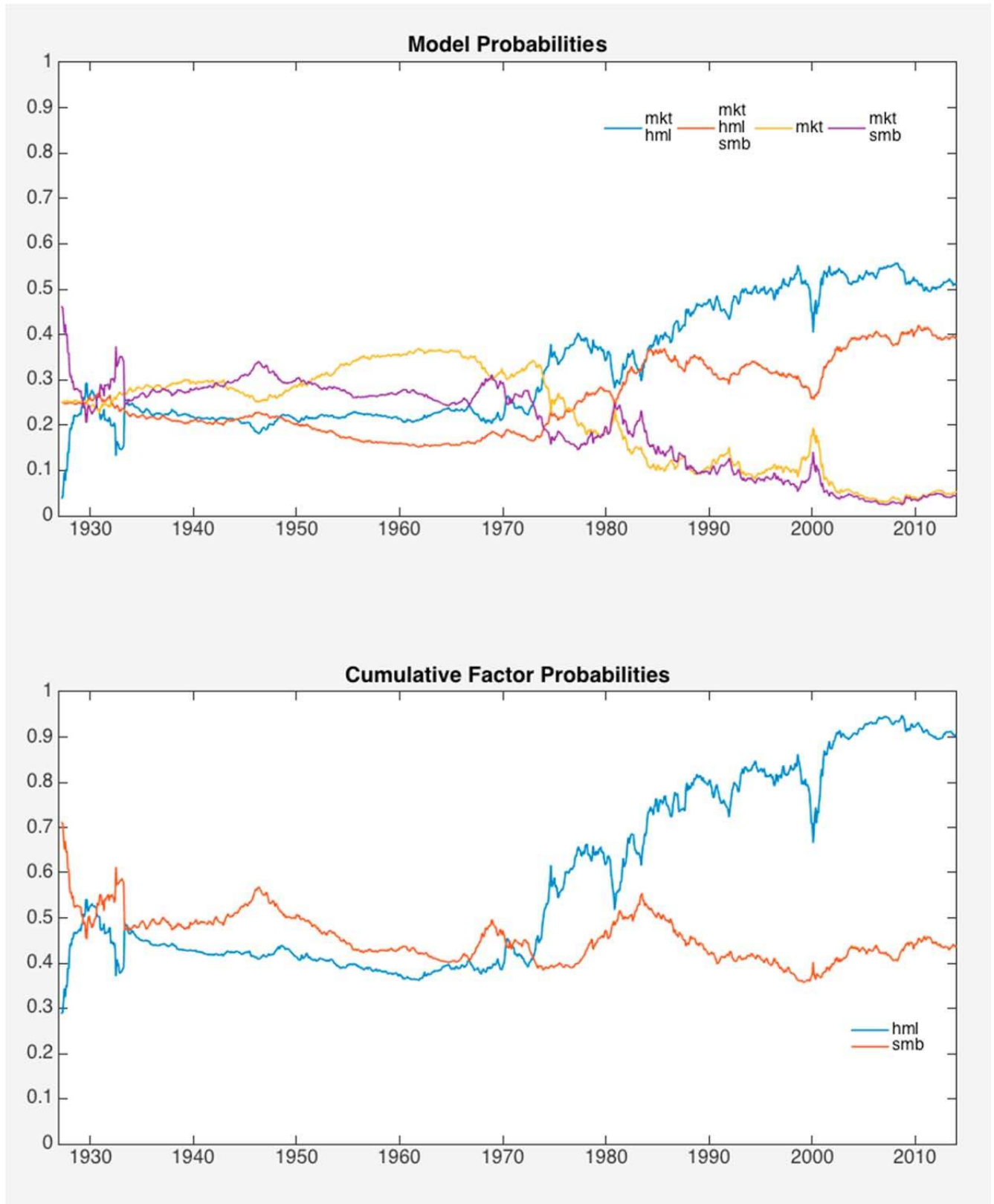


Figure 1: FF3 factors 1927-2013, $Sh_{\max} = 1.27 \times Sh(\text{Mkt})$.

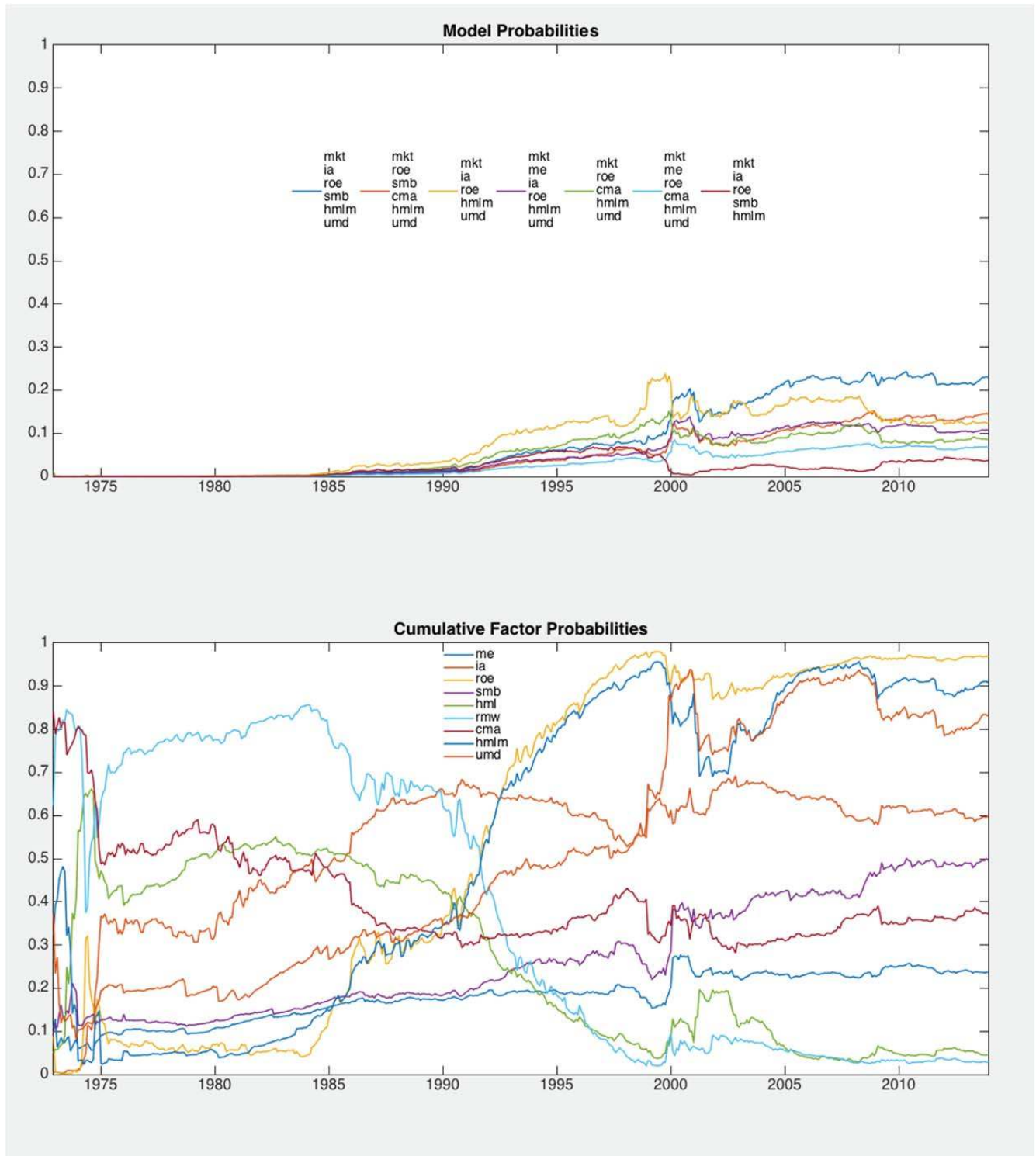


Figure 2: Sample 1972-2013, $Sh_{\max} = 1.5 \times Sh(\text{Mkt})$.

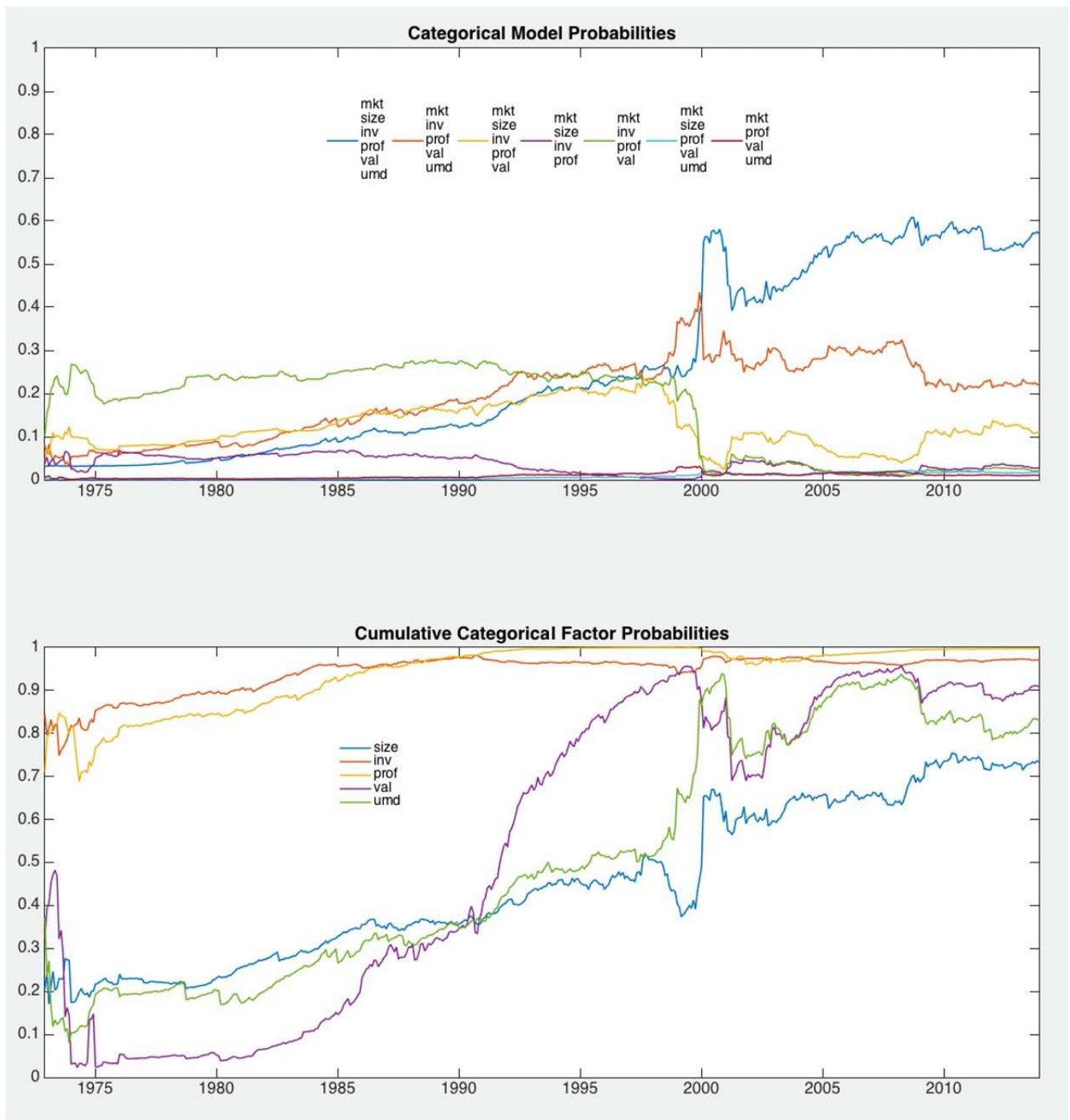


Figure 3: Sample 1972-2013, $Sh_{\max} = 1.5 \times Sh(\text{Mkt})$.

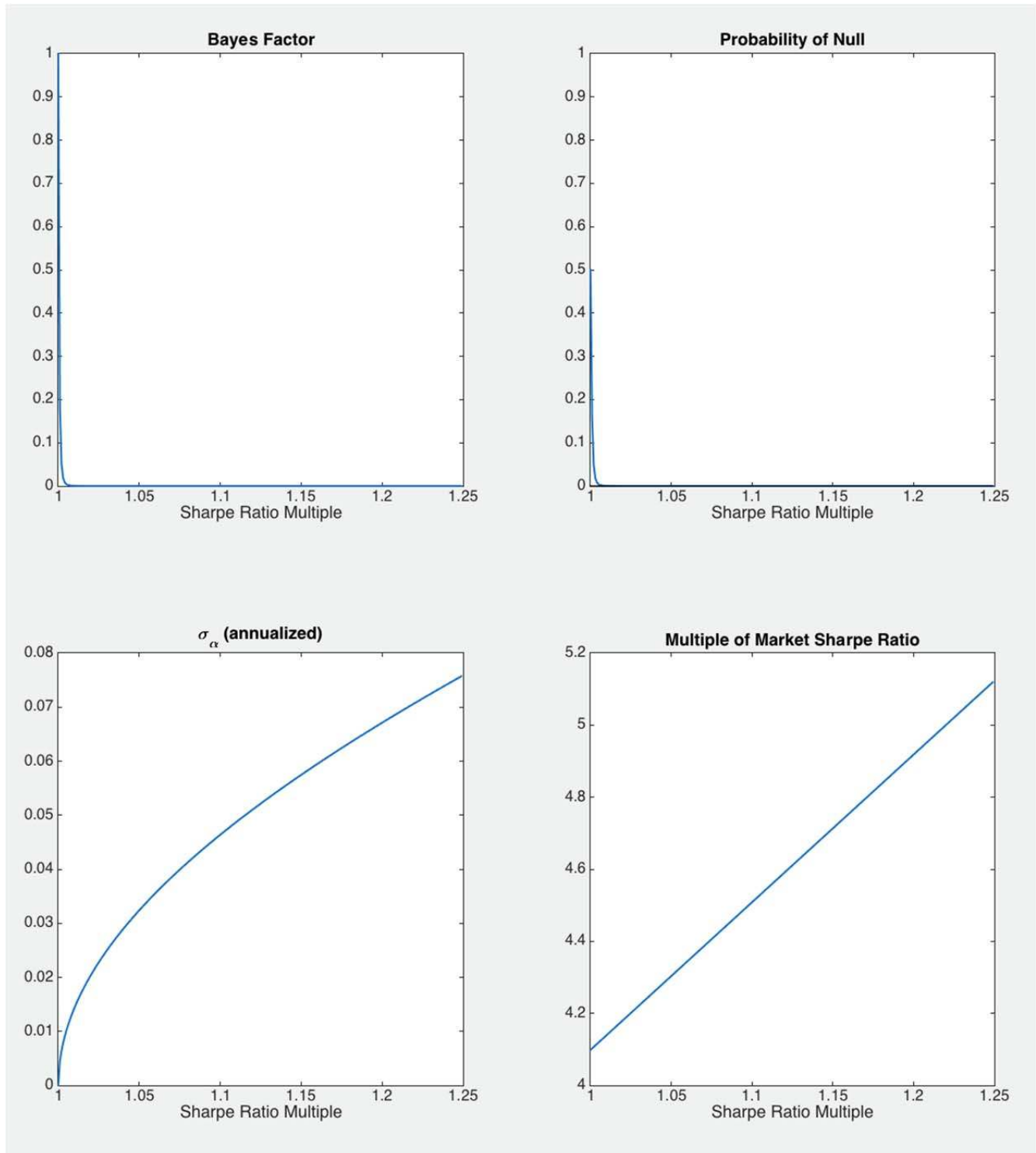


Figure 4: HML^m is not redundant in relation to the other factors {Mkt SMB ROE IA UMD} in the top model. Bayesian GRS intercept test for HML^m . Sharpe ratio for alternative as multiple of ratio under null hypothesis.

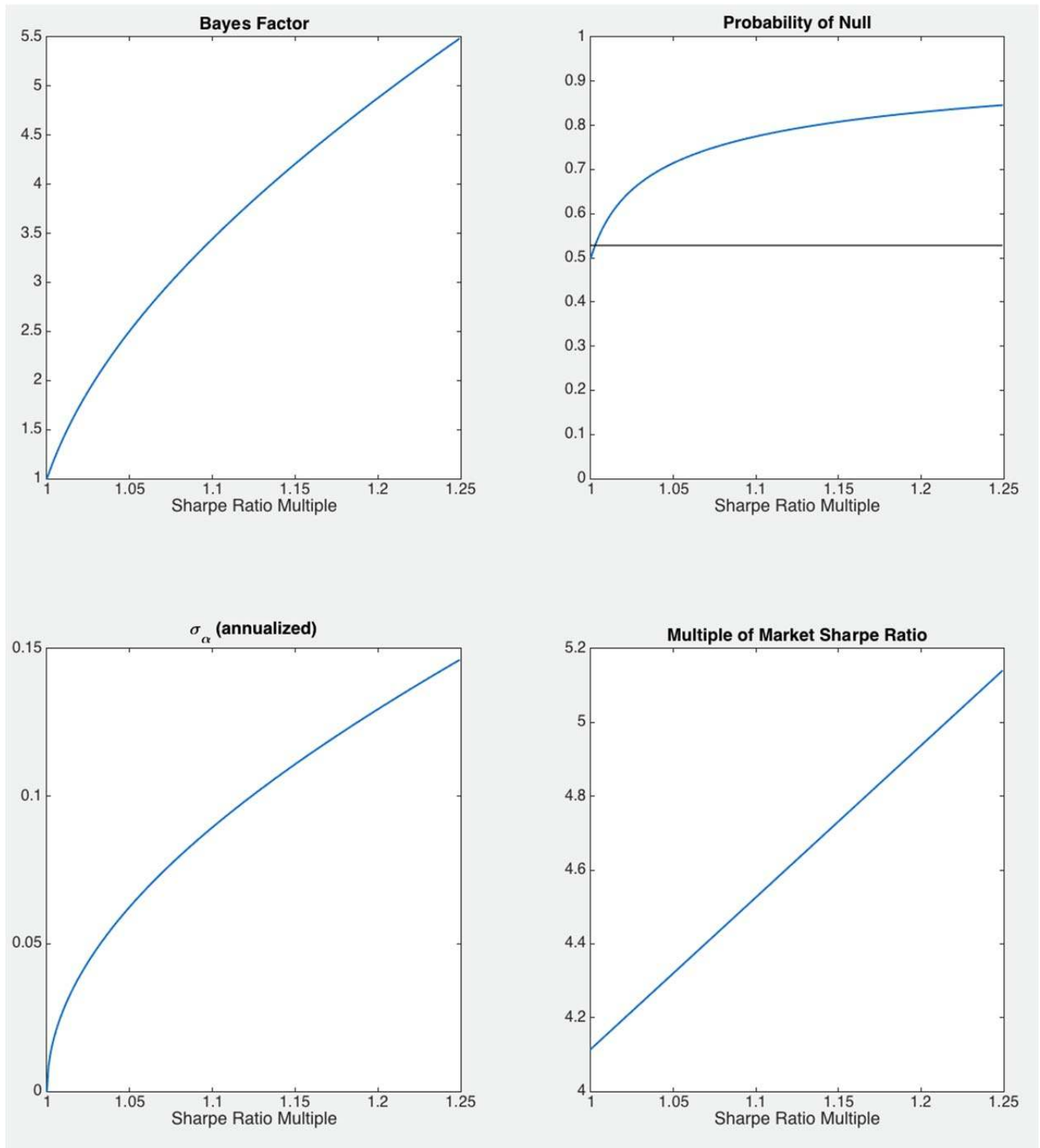


Figure 5: UMD is redundant in relation to the model {Mkt SMB ROE IA}. Bayesian GRS intercept test for UMD. Sharpe ratio for alternative as multiple of ratio under null hypothesis.

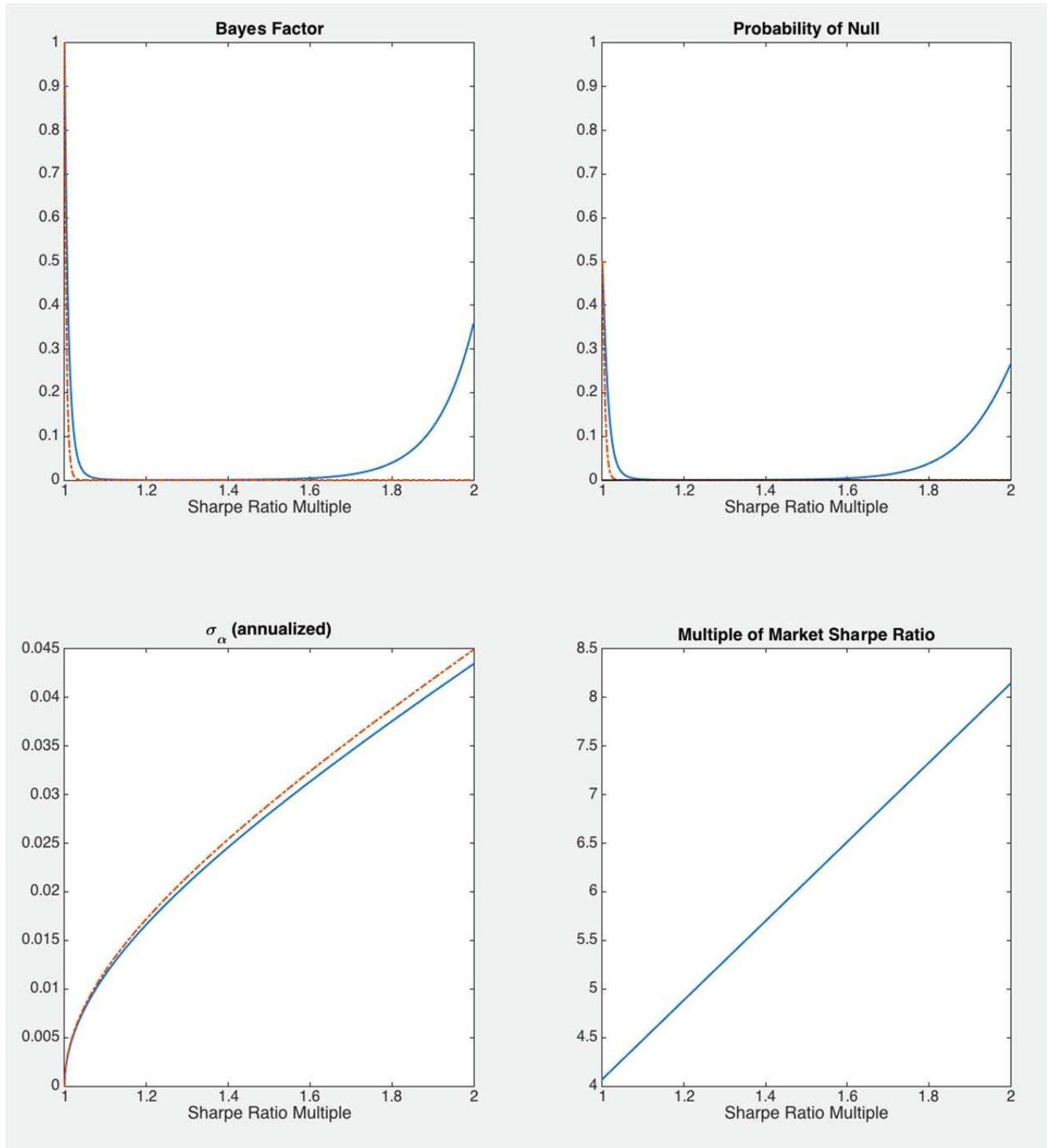


Figure 6: Sample 1972-2013, Model = {Mkt ME IA ROE}. Test assets = 25 size-momentum portfolios (blue line) plus UMD, HML^m (red line). Sharpe ratio for alternative as multiple of ratio under null hypothesis.

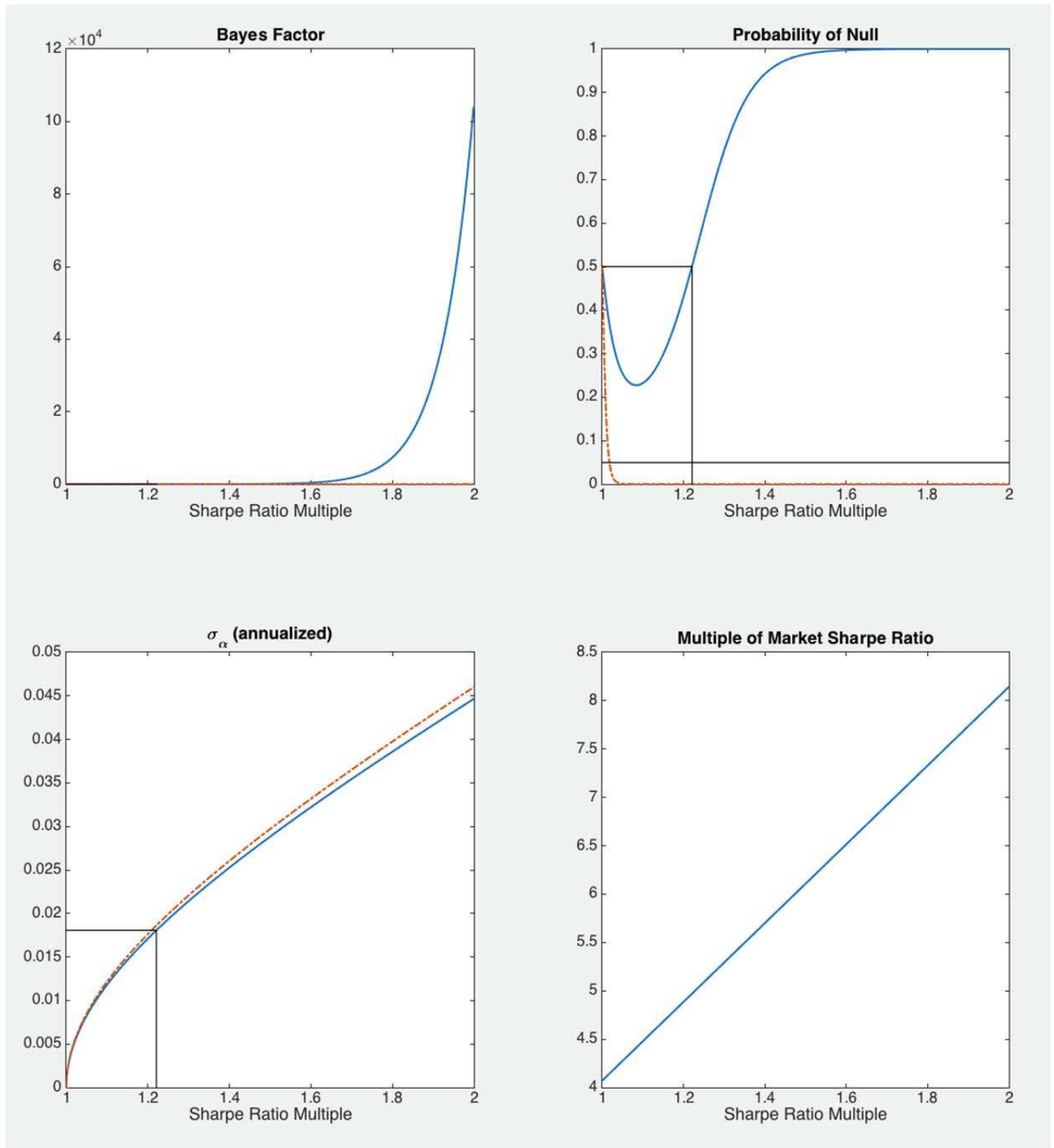


Figure 7: Sample 1972-2013, Model = {Mkt ME IA ROE}. Test assets = 25 Book-to-market/investment portfolios (blue line) plus UMD, HML^m (red line). Sharpe ratio for alternative as multiple of ratio under null hypothesis.

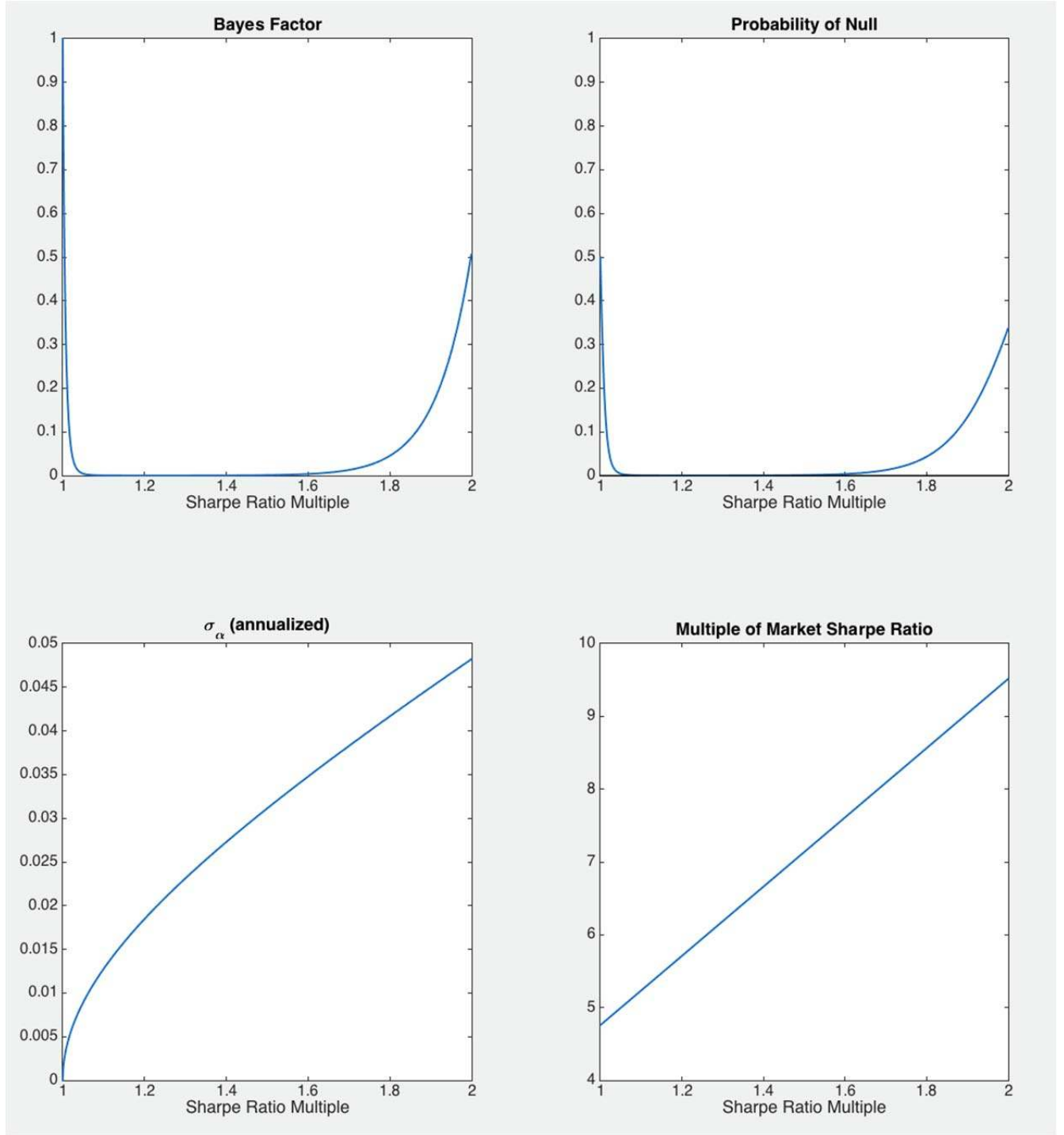


Figure 8: Sample 1972-2013, Model = {Mkt SMB IA ROE UMD HML^m}. Test assets: 25 Book-to-market/investment portfolios. Sharpe ratio for alternative as multiple of ratio under null hypothesis.

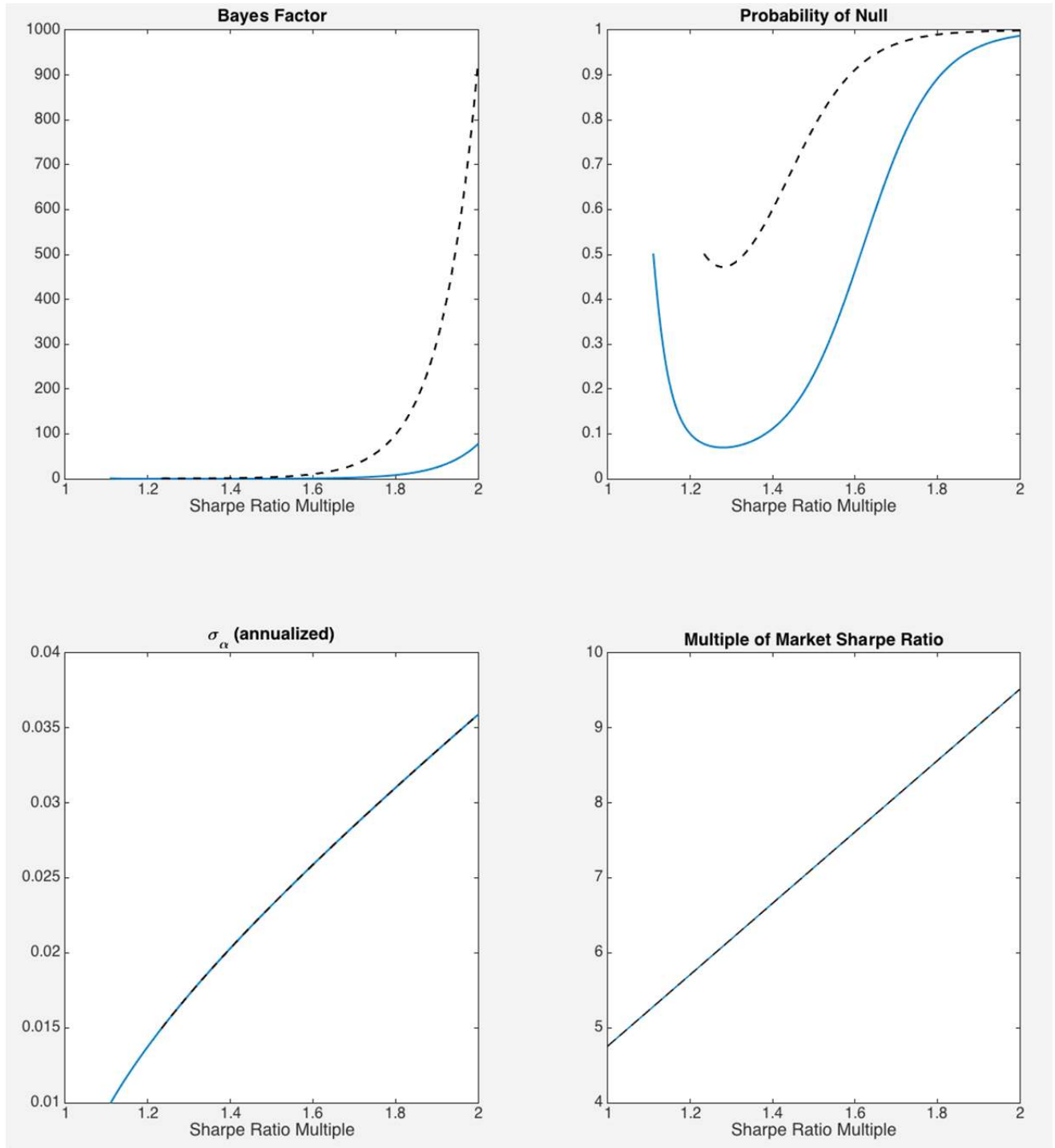


Figure 9: Sample 1972-2013, Model = {Mkt SMB IA ROE UMD HML^m}, Test assets = 25 size-momentum portfolios. $\sigma_\alpha = 1\%$ (blue line) or 1.5% (black dashed) under the approximate null hypothesis. Sharpe ratio for alternative as multiple of ratio under null hypothesis.

Table 1**Fama-French factors - Posterior Model Probabilities for different Prior Sharpe Multiples**

Data Sample: Jan 1927 to Dec 2013

Market Sharpe Ratio = 0.115

3-factor Sharpe Ratio = 0.142 or 1.23*Market Sharpe Ratio

Model/Prior Multiple*	1.13	1.27	1.58	2.24
Mkt HML	44.9	51.3	58.5	65.2
Mkt HML SMB	42.3	39.1	32.2	23.3
Mkt	6.5	5.2	5.7	8.2
Mkt SMB	6.3	4.4	3.6	3.4

*Multiple of Mkt Sharpe ratio under 3-factor alternative.

Table 2**10 factors - Posterior Model Probabilities for different Prior Sharpe Multiples**

Data Sample: Jan 1972 to Dec 2013

Market Sharpe Ratio = 0.113

6-factor (best model) Sharpe Ratio = 0.51 or 4.5*Market Sharpe Ratio

Model/Prior Multiple*	1.25	1.5	2	3
Mkt SMB ROE IA HML ^m UMD	7.8	23.0	35.7	42.0
Mkt SMB ROE CMA HML ^m UMD	5.0	14.4	22.1	25.7
Mkt ROE IA HML ^m UMD	8.3	12.4	8.4	5.2
Mkt ME ROE IA HML ^m UMD	4.5	10.8	12.6	11.4
Mkt ROE CMA HML ^m UMD	5.5	8.6	6.2	4.1
Mkt ME ROE CMA HML ^m UMD	2.9	6.8	7.9	7.0
Mkt SMB ROE IA HML ^m	3.9	3.7	1.3	0.5

*Multiple of Mkt Sharpe ratio under 6-factor alternative.

Table 3: Categorical Models - Posterior Probabilities for different Prior Sharpe Multiples

Data Sample: Jan 1972 to Dec 2013

Market Sharpe Ratio = 0.113

6-factor (best model) Sharpe Ratio = 0.51 or 4.5 *Market Sharpe Ratio

Model/Prior Multiple*	1.25	1.5	2	3
Mkt SIZE PROF INV VAL MOM	26.8	57.0	78.6	86.2
Mkt PROF INV VAL MOM	18.4	22.2	14.8	9.3
Mkt SIZE PROF INV VAL	19.9	11.0	3.0	0.9
Mkt SIZE PROF INV	9.4	2.8	0.3	0.0
Mkt PROF INV VAL	10.1	2.1	0.2	0.0
Mkt SIZE PROF VAL MOM	0.0	1.7	2.4	3.0
Mkt PROF VAL MOM	0.8	1.0	0.7	0.5

Table 4: Relative Probabilities for each Categorical Factor in the 10-factor Analysis

Factor/Prior Multiple*	1.25	1.5	2	3
ROE	83.8	97.1	99.7	100
IA	62.3	61.7	61.2	61.4
SMB	61.8	67.8	73.9	78.6
HML ^m	70.9	95.2	99.7	100

* Multiple of Mkt Sharpe ratio under 6-factor alternative.

Remaining PROF, INV, SIZE and VAL probability goes to RMW, CMA, ME and HML, respectively.

Table 5

Intercepts for each Factor in the Highest-Probability Model on the other Five-factors

This table presents annualized alphas from regressions of each factor on the other factors in the model {Mkt SMB ROE IA UMD HML^m}. Sample period Jan 1972 to Dec 2013

Factor	SMB	ROE	IA	UMD	HML ^m
Alpha (t-statistic)	5.09 (3.14)	6.97 (6.08)	1.20 (1.50)	6.60 (3.96)	6.07 (5.26)