Model Comparison with Sharpe Ratios

Francisco Barillas, Raymond Kan, Cesare Robotti, and Jay Shanken*

*Barillas, e-mail: f.barillas@unsw.edu.au, School of Banking and Finance, University of New South Wales; Kan, e-mail: kan@chass.utoronto.ca, Department of Finance, Rotman School of Management, University of Toronto; Robotti, e-mail: cesare.robotti@wbs.ac.uk, Finance Group, Warwick Business School, University of Warwick; Shanken (corresponding author), e-mail: jay.shanken@emory.edu, Department of Finance, Goizueta Business School, Emory University, and NBER. We thank Hendrik Bessembinder (the editor), Wayne Ferson, Chen Xue, an anonymous referee, and participants at the 2018 SoFiE Conference, the 2018 Western Finance Association Meetings, the 2018 China International Conference in Finance, the 2018 IMS Asia Pacific Rim Meeting, and the 2017 workshop on "New Methods for the Empirical Analysis of Financial Markets" in Comillas, Spain, for helpful comments and suggestions. Finally, we thank Kenneth French for providing us with the data on several factor portfolios used in this paper.

Abstract

We show how to conduct asymptotically valid tests of model comparison when the extent of model mispricing is gauged by the squared Sharpe ratio improvement measure. This is equivalent to ranking models on their maximum Sharpe ratios, effectively extending the GRS test to accommodate comparison of non-nested models. Mimicking portfolios can be substituted for any nontraded model factors and estimation error in the portfolio weights is taken into account in the statistical inference. A variant of the Fama and French (2018) six-factor model, with a monthly-updated version of the usual value spread, emerges as the dominant model.

I. Introduction

Financial economists have long sought to explain differences in asset expected returns. The resulting pricing models can be viewed statistically as constrained multivariate linear regressions of asset returns on systematic factors. The constraint requires that asset expected returns be a linear function of the betas (the slope coefficients). When returns in excess of a risk-free rate are employed and the factors are themselves excess portfolio returns or return spreads, the regression intercepts – the investment alphas – must be zero. The capital asset pricing model (CAPM) of Sharpe (1964) and Lintner (1965) was the first such model, with the value-weighted market portfolio of all financial assets serving as the equilibrium-based factor. Equilibrium theory has also given rise to the intertemporal CAPM of Merton (1973) and Long (1974) and the consumption CAPM of Breeden (1979) and Rubinstein (1976). These theories motivate the use of state variable innovations and consumption growth as nontraded asset-pricing factors. However, as Breeden (1979) notes, maximally-correlated portfolios can also serve as the factors in such models and the usual asset-pricing restrictions continue to hold.

The empirically motivated three-factor model (FF3) of Fama and French (1993), with traded size (SMB) and value (HML) factors along with the market excess return (MKT) was, for many years, the premier factor model in the literature, sometimes supplemented by a momentum factor, as suggested by Carhart (1997). In recent years, however, the floodgates have opened and many alternative factor pricing models (to be discussed below) have been explored. In practice, it is unlikely that a model's constraints will hold exactly and so it is of interest to quantify the extent of mispricing for each model. Barillas and Shanken (2017) address the issue of how to compare models under the classic Sharpe ratio improvement metric for evaluating the fit of a model. This is the quadratic form in the alphas that is equivalent to the improvement in the squared Sharpe ratio (expected excess return over standard deviation) obtained when investment in other asset returns is permitted in addition to the given model's factors. This metric is central to the Gibbons, Ross, and Shanken's (GRS) (1989) test of whether a given portfolio attains the maximum possible Sharpe ratio.[1]

[1]This measure of reward to risk was introduced by Sharpe (1966) in the context of mutual-fund performance

A key premise in the analysis of Barillas and Shanken (2017) is that a model should ideally price the traded factors in the various models, as well as the returns designated as "test assets." In this context, they show that *model comparison* under the Sharpe ratio improvement metric is driven by the extent to which each model is able to price the factors in the other models, as reflected in the "excluded-factor" alphas. Surprisingly, the test assets drop out of the analysis and are, therefore, irrelevant for model comparison. It follows that the model whose factors permit the highest squared Sharpe ratio to be achieved is ultimately preferred. The argument is straightforward: for simplicity, consider two models with traded factors, f_1 and f_2, respectively. The extent to which f_1 fails to price f_2 and the test-asset returns, R, is measured by the squared Sharpe increase, $Sh^2(f_1, f_2, R) - Sh^2(f_1)$, that results from exploiting the corresponding alphas of f_2 and R on f_1. Similarly, $Sh^2(f_2, f_1, R) - Sh^2(f_2)$ indicates the degree of misspecification of the model with factors f_2. Taking the difference gives $Sh^2(f_2) - Sh^2(f_1)$ and thus the model with "less mispricing" also has the higher squared Sharpe ratio.

Barillas and Shanken (2017) show that test assets also drop out if models are compared on the basis of their statistical likelihoods. Barillas and Shanken (2018) build on this observation and develop a Bayesian procedure that permits the simultaneous calculation of probabilities for all models derived from a given set of factors. In essence, their procedure seeks to identify a parsimonious model that spans the maximum Sharpe ratio portfolio for the traded factors, but without retaining redundant factors. Direct evidence about the relative magnitudes of the squared Sharpe ratios for different models is not provided, however. In this paper, we focus directly on a comparison of models' maximum squared Sharpe ratios in an asymptotic analysis under very general distributional assumptions. Complementary insights about model comparison can thus be obtained by viewing the evidence from each of these perspectives.

We focus on squared Sharpe ratios for analytical convenience, but it is the Sharpe ratio itself that we want to maximize from an economic perspective. A model's tangency portfolio, which maximizes the squared Sharpe ratio, could lie on the lower portion of the risky-asset frontier and

evaluation and was dubbed the Sharpe ratio in the classic analysis of active-portfolio investment of Treynor and Black (1973).

thus have a negative Sharpe ratio. However, the maximum Sharpe ratio is still the (positive) square root of the maximum squared ratio, attained by shorting the tangency portfolio and investing in the risk-free asset.[2] It follows that the same model rankings are produced by maximum squared Sharpe ratios and maximum Sharpe ratios.

Another criterion for comparison due to Hansen and Jagannathan (HJ) (1997) has frequently been used in the literature. This "HJ-distance" is a measure of model misspecification that indicates how closely a proposed stochastic discount factor (SDF) based on a set of factors comes to being a valid SDF; it can also be regarded as the maximum pricing error of the model over portfolios with unit second moment. When a risk-free asset is available, Kan and Robotti (2008) suggest a modification to the HJ-distance which requires that all competing SDFs assign the same price to the risk-free asset. In this case, the distance compares performance based on pricing errors for *excess* returns. With traded factors, they further note that imposing the restriction that the factors are priced without error yields a distance measure equal to the increase in the squared Sharpe ratio. Thus, our analysis can also be interpreted as a procedure for comparing models in terms of this modified HJ-distance.

When the factors in one model are all contained in the other – the case of nested models – the squared Sharpe ratio of the larger model must be at least as high as that for the nested model. The question then is whether equality holds or the larger model is strictly superior. The statistical analysis for this scenario is a simple application of the GRS test, with the factors that are excluded from the nested model serving as left-hand-side returns. The challenge now is to develop a test for comparing *non-nested* models, the case in which each model contains factors not included in the other model. Although the asymptotic distribution of the Sharpe ratio difference has been derived for a pair of simple trading strategies, the generalization required for model comparison must accommodate the difference for two maximum Sharpe ratio portfolios obtained from different (possibly overlapping) sets of factors.[3] We provide such an analysis, while also adjusting for the well-known small-sample bias in the squared Sharpe ratio estimator, as documented by Jobson and

[2] Multiplying the tangency excess return by a constant c multiplies the mean by c and the standard deviation by $|c|$. The ratio is -1 when $c < 0$.

[3] See, for example, Jobson and Korkie (1981), Memmel (2003), Christie (2005), and Opdyke (2007).

Korkie (1980). Our simulations indicate that the resulting procedure performs well in samples of the sort employed in practice.

For models that include nontraded factors, pricing is typically explored using cross-sectional regression (CSR) analysis. Building on earlier work by Balduzzi and Robotti (2008) and Lewellen, Nagel, and Shanken (2010), Barillas and Shanken (2017) note that comparison in terms of a quadratic form in the generalized least squares (GLS) pricing errors again reduces to examining the difference of squared Sharpe ratios, but with mimicking portfolios now substituted for the nontraded factors. In this context, test assets along with any traded factors serve to identify the mimicking portfolios and the statistical analysis must account for the additional estimation error in the portfolio weights. We provide asymptotic results for this setting as well, while also adjusting for the small-sample bias in the mimicking portfolio squared Sharpe ratio estimator. Thus, analyzing models with nontraded factors again amounts to a comparison of the models' squared Sharpe ratios – an intuitively appealing economic criterion. This complements the more statistically-oriented CSR model R-squared that is often reported and whose asymptotic properties are analyzed by Kan, Robotti, and Shanken (2013).

Our statistical methodology is applied in the comparison of several fairly recent models that have been explored in the literature. We find that the two-factor intermediary asset pricing model of He, Kelly, and Manela (2017)[4] and the "betting-against-beta" CAPM extension of Frazzini and Pedersen (2014) are dominated by the q-theory model of Hou, Xue, and Zhang (2015), the Stambaugh and Yuan (2017) mispricing model, and the Fama and French (2018) five-factor model with cash profitability. A variant of the original Hou, Xue, and Zhang (2015) model that uses the cash profitability factor instead of its original profitability factor (ROE) is superior to the six-factor Fama and French (2018) model that also includes momentum. The best overall performer, however, is a variant of the six-factor Fama and French (2018) model which uses a "timely" value factor due to Asness and Frazzini (2013) instead of the traditional HML factor.

[4]This is true with their traded financial intermediary capital risk factor or a mimicking portfolio constructed from their nontraded factor.

II. Comparing Sharpe Ratios for Models with Traded Factors

We begin Section II with a brief review of the GRS test. First, some definitions and notation. A factor model M is a multivariate linear regression with N excess returns, R, and K traded factors, f. With T observations on f_t and R_t:

$$R_t = \alpha_R + \beta f_t + \epsilon_t, \quad t = 1, \ldots, T, \tag{1}$$

where R_t, ϵ_t, and α_R are N-vectors, β is an $N \times K$ matrix, and f_t is a K-vector. The improvement in the squared Sharpe ratio from adding test assets R to the investment universe is a quadratic form in the test-asset alphas:

$$\alpha_R' \Sigma^{-1} \alpha_R = Sh^2(f, R) - Sh^2(f), \tag{2}$$

where Σ is the invertible population covariance matrix of the zero-mean disturbance ϵ_t.[5] The associated F-statistic is then proportional to the statistic obtained by substituting the sample quantities in equation (2) and dividing by one plus the sample estimate of $Sh^2(f)$.[6] Thus a test of $\alpha_R = 0_N$, where 0_N is an N-vector of zeros, is a test of whether f yields the squared maximum Sharpe ratio.

Next, we consider pricing restrictions for nested models and show how to implement the GRS test in this context, with the factors excluded from the nested model serving as left-hand-side returns.

A. Model Comparison and Alpha-Based Tests

Let A be a pricing model with factors $[f_{1t}', \ f_{2t}']'$ that nests model B with factors f_{1t}, where f_{1t} and f_{2t} are K_1 and K_2-vectors, respectively. In addition, let α_{21} denote the alphas for the factors f_{2t} when they are regressed on f_{1t}. Proposition 1 in Barillas and Shanken (2017) shows that to compare nested models, we need only focus on testing the excluded-factor restriction, $\alpha_{21} = 0_{K_2}$

[5] See GRS and Jobson and Korkie (1982).

[6] With the usual maximum likelihood estimates, the proportionality constant is $(T - N - K)/N$ and the degrees of freedom of the F distribution are N and $T - N - K$. The divisor adjusts for the covariance matrix of the alpha estimates conditional on the factors f.

(test assets are irrelevant). This restriction can be formally evaluated using the basic alpha test.[7] For example, testing the CAPM versus FF3 involves testing whether the CAPM alphas of *HML* and *SMB* are zero. If this joint hypothesis is rejected, we have evidence that FF3 dominates the CAPM and that the (squared) Sharpe ratio achievable with the factors in FF3 is higher than that for the market factor. In this case, the maximum Sharpe ratio portfolio has nonzero weight on *HML* and/or *SMB*.[8]

Comparing non-nested models is less straightforward, however. For example, let model A consist of *MKT* and *SMB* and model B consist of *MKT* and *HML*. Suppose the GRS test indicates that adding *HML* increases the squared Sharpe ratio of model A, while the alpha of *SMB* on model B is not statistically significant. As Barillas and Shanken (2017) note, such findings would be *consistent* with model B having the higher squared Sharpe ratio. But in general, failure to reject either model or finding that both can be rejected does not tell us which model has the higher squared Sharpe ratio.[9] Therefore, in this paper, we develop a direct asymptotic test of this hypothesis.

B. Asymptotic Distribution of the Difference in Squared Sharpe Ratios for Non-Nested Models

Now consider two non-nested models (A and B) with factor returns f_{At} and f_{Bt}, respectively, $t = 1, 2, \ldots, T$. We assume throughout that all time series are jointly stationary and ergodic with finite fourth moments. This includes the traded-factor returns and later, nontraded factors and other basis-asset returns. Denote the squared maximum Sharpe ratios that are attainable from the two sets of factors by $\theta_A^2 = \mu_A' V_A^{-1} \mu_A$ and $\theta_B^2 = \mu_B' V_B^{-1} \mu_B$, where μ_A, μ_B, V_A, and V_B are the nonzero means and invertible covariance matrices of the two sets of factors. Similarly, let the corresponding sample quantities be $\hat{\theta}_A^2 = \hat{\mu}_A' \hat{V}_A^{-1} \hat{\mu}_A$ and $\hat{\theta}_B^2 = \hat{\mu}_B' \hat{V}_B^{-1} \hat{\mu}_B$.[10]

[7] In the empirical section, we employ a version of the test that takes into account residual heteroscedasticity conditional on the factors. We refer to this as the "basic alpha-based test." This is the special case of Shanken (1990) with no conditioning variables.

[8] Confidence intervals for the difference of squared Sharpe ratios with nested models can also be obtained as in Lewellen, Nagel, and Shanken (2010).

[9] Of course, failure to reject a null hypothesis does not imply it is true and so power considerations further complicate the interpretation of results.

[10] In our analysis, $\hat{V}$ is the maximum likelihood estimator of V, the population covariance matrix.

PROPOSITION 1: *The asymptotic distribution of the difference in sample squared Sharpe ratios is given by*

$$\sqrt{T}([\hat{\theta}_A^2 - \hat{\theta}_B^2] - [\theta_A^2 - \theta_B^2]) \overset{A}{\sim} N(0, E[d_t^2]), \tag{3}$$

provided that $E[d_t^2] > 0$, *where*

$$d_t = 2(u_{At} - u_{Bt}) - (u_{At}^2 - u_{Bt}^2) + (\theta_A^2 - \theta_B^2), \tag{4}$$

with $u_{At} = \mu_A' V_A^{-1}(f_{At} - \mu_A)$ and $u_{Bt} = \mu_B' V_B^{-1}(f_{Bt} - \mu_B)$.

Proof: See Appendix.

We prove this result in the Appendix by casting the estimation of the first and second moments of the returns in the generalized method of moments (GMM) framework and using the delta method for functions of these parameters.[11] The validity of our asymptotic approximations requires that at least one of the Sharpe ratios of the models to be compared is different from zero. The analysis in the Appendix (apart from the proofs of the various lemmas below) accommodates serial correlation. However, for simplicity, the statements of this and other results in the body of the paper assume serially uncorrelated time series (factors and returns), a reasonable approximation for many empirical applications. To conduct statistical tests, we need a consistent estimator of $E[d_t^2]$. This can be obtained by replacing each term in d_t with the corresponding sample estimate. We denote the result $\hat{d}_t$ and calculate the sample second moment, $\sum_{t=1}^{T} \hat{d}_t^2 / T$.

To better understand the determinants of the asymptotic variance of the difference in sample squared Sharpe ratios, in the next lemma we assume that the traded-factor returns are multivariate elliptically distributed.

LEMMA 1: *When the traded-factor returns are i.i.d. multivariate elliptically distributed with kurtosis parameter* κ,[12] *the asymptotic variance of the difference in sample squared Sharpe ratios is*

[11]The proof of Proposition 1 relies on the proof of Proposition 2 below. However, for expositional purposes, Proposition 1 was placed first.

[12]The kurtosis parameter for an elliptical distribution is defined as $\kappa = \mu_4/(3\sigma^4) - 1$, where σ^2 and μ_4 are its second and fourth central moments, respectively.

given by

$$(5)\ \ E[d_t^2] = \theta_A^2\left[4 + (2+3\kappa)\theta_A^2\right] + \theta_B^2\left[4 + (2+3\kappa)\theta_B^2\right] - 2\left\{2\rho\theta_A\theta_B[2 + (1+\kappa)\rho\theta_A\theta_B] + \kappa\theta_A^2\theta_B^2\right\},$$

where $\rho = \text{Corr}[u_{At}, u_{Bt}] = E[u_{At}u_{Bt}]/(\theta_A\theta_B)$ *is the correlation between the returns on the maximum Sharpe ratio portfolios of* f_{At} and f_{Bt}.

Proof: See Appendix.

The first term is the asymptotic variance of $\hat{\theta}_A^2$, the second term is the asymptotic variance of $\hat{\theta}_B^2$, and the last term is -2 times the asymptotic covariance between $\hat{\theta}_A^2$ and $\hat{\theta}_B^2$. The variance of d_t depends on ρ, the correlation between the returns on the maximum Sharpe ratio portfolios of the factors of models A and B, and on the kurtosis parameter κ. When $\rho = 1$, that is, the two maximum Sharpe ratio portfolios are identical, $E[d_t^2] = 0$ and the asymptotic normality result in Proposition 1 breaks down. When $\rho = 0$ and the factors are multivariate normally distributed, that is, $\kappa = 0$, the asymptotic variance simplifies to $E[d_t^2] = 2\left[\theta_A^2(2+\theta_A^2) + \theta_B^2(2+\theta_B^2)\right]$. Finally, it can be shown that $E[d_t^2]$ is an increasing function of the kurtosis parameter κ.

The asymptotic variance in Proposition 1 forms the basis for testing non-nested models. When the two models have overlapping factors, however, it is important from both an *economic* and a *statistical* perspective to distinguish between two ways the null hypothesis can hold. One possibility is that the common factors span the (true) maximum Sharpe ratio portfolio based on the factors from both models. If so, the squared Sharpe ratio of each model equals that of the common-factors model and the other factors are redundant. This spanning condition can be evaluated by an alpha-based test, with the factors that are excluded from each model together serving as the left-hand-side returns. If spanning is rejected, some or all of the additional factors contribute to an increase in the squared Sharpe ratio and equality may or may not hold for the two models. In the absence of spanning, $E[d_t^2] > 0$ in equation (4) and one can perform a direct test of $\theta_A^2 = \theta_B^2$ using Proposition 1. Alternatively, given an a priori judgment that exact spanning is implausible and can be ruled out, one can simply use the direct test. In our empirical work, the alpha-based test easily rejects the spanning condition in all cases considered and so we focus on the direct test in applications.

III. Comparing Models with Mimicking Portfolios

Section II dealt with the case in which the factors are excess returns or return spreads. However, some models, e.g., the consumption CAPM and the intertemporal CAPM, include one or more risk factors that are not themselves asset returns. Breeden (1979) points out that such factors can be replaced with portfolios whose weights are proportional to their betas from the projection of the factors on returns and a constant. In this section, we first present the asymptotic distribution of the so-called "mimicking portfolio" squared sample Sharpe ratio and then the distribution of the difference in the sample squared Sharpe ratios for two models that could have mimicking portfolios as factors.

A. Overview of the Mimicking Portfolio Methodology

Suppose that the K-vector f_t consists of some traded and some nontraded factors. Let R_t be a vector of returns that includes the traded-factor returns as well as any basis-asset returns that will be used to specify mimicking portfolios for the nontraded factors. In a typical cross-sectional regression analysis, the basis assets would be the "test assets." For a traded factor, the mimicking portfolio is, of course, simply the factor itself. As noted by Barillas and Shanken (2017), in contrast to the test-asset irrelevance result for traded-factor models, model comparison can depend on the basis assets used to construct the mimicking portfolios for nontraded factors.[13]

We define $Y_t = [f_t',\ R_t']'$ and its population mean and covariance matrix as

$$\mu = E[Y_t] \equiv \begin{bmatrix} \mu_f \\ \mu_R \end{bmatrix}, \tag{6}$$

$$V = \mathrm{Var}[Y_t] \equiv \begin{bmatrix} V_f & V_{fR} \\ V_{Rf} & V_R \end{bmatrix}. \tag{7}$$

In the following analysis, we assume that V_f and V_R are invertible and that V_{Rf} is of full column rank.[14] Consider the projection of f_t on R_t and a constant and denote the resulting mimicking-portfolio returns by $f_t^* = V_{fR}V_R^{-1}R_t \equiv AR_t$ with $\mu^* = E[f_t^*] = A\mu_R$ and $V^* = \mathrm{Var}[f_t^*] = AV_RA' =$

[13] It should also be noted that increasing the number of basis assets used to construct the mimicking portfolio does not lead, in general, to an increase in the squared Sharpe ratio of the mimicking portfolio returns. A proof of this result is available from the authors upon request.

[14] This condition can be evaluated using rank restrictions tests such as the ones proposed by Cragg and Donald (1997), Robin and Smith (2000), and Kleibergen and Paap (2006).

$V_{fR}V_R^{-1}V_{Rf}$. For the mimicking portfolios to exist, the beta sums must not all be zero, i.e., we assume that $A1_N \neq 0_K$, where 1_N is an N-vector of ones and 0_K is a K-vector of zeros.[15] The population squared Sharpe ratio of a set of mimicking portfolios is given by

$$\theta^2 = \mu^{*\prime}V^{*-1}\mu^* \equiv \mu_R' V_R^{-1}V_{Rf}(V_{fR}V_R^{-1}V_{Rf})^{-1}V_{fR}V_R^{-1}\mu_R. \tag{8}$$

Suppose that we have T observations on Y_t and let $\hat{\mu}$ and $\hat{V}$ denote the sample moments of Y_t corresponding to the population moments in equations (6) and (7). The mimicking portfolio methodology estimates the weights of the mimicking portfolios, the matrix A, by running the multivariate regression

$$f_t = a + AR_t + \eta_t, \qquad t = 1, \ldots, T. \tag{9}$$

Let $\hat{\mu}^* = \hat{A}\hat{\mu}_R$ and $\hat{V}^* = \hat{A}\hat{V}_R\hat{A}'$, where $\hat{A} = \hat{V}_{fR}\hat{V}_R^{-1}$. Then, the sample squared Sharpe ratio of a set of mimicking portfolios can be obtained as

$$\hat{\theta}^2 = \hat{\mu}^{*\prime}\hat{V}^{*-1}\hat{\mu}^* \equiv \hat{\mu}_R'\hat{A}'(\hat{A}\hat{V}_R\hat{A}')^{-1}\hat{A}\hat{\mu}_R. \tag{10}$$

B. Asymptotic Distribution of the Sample Squared Sharpe Ratio of a Set of Mimicking Portfolios

Let $v_t = \mu_R' V_R^{-1}(R_t - \mu_R)$, $u_t = \mu^{*\prime}V^{*-1}(f_t^* - \mu^*)$, and $y_t = \mu^{*\prime}V^{*-1}\eta_t$. The following proposition presents a general expression for the asymptotic distribution of $\hat{\theta}^2$.

PROPOSITION 2: *The asymptotic distribution of $\hat{\theta}^2$ is given by*

$$\sqrt{T}(\hat{\theta}^2 - \theta^2) \overset{A}{\sim} N(0, E[h_t^2]), \tag{11}$$

provided that $E[h_t^2] > 0$, where

$$h_t = 2u_t(1 - y_t) - u_t^2 + 2y_t v_t + \theta^2. \tag{12}$$

Proof: See Appendix.

[15] Huberman, Kandel, and Stambaugh (1987) show that this condition is equivalent to assuming that the global minimum-variance portfolio has positive systematic risk.

When the factors are perfectly tracked by the returns, $y_t = 0$ and the h_t expression in the proposition reduces to

$$h_t = 2u_t - u_t^2 + \theta^2, \tag{13}$$

where $u_t = \mu_f' V_f^{-1}(f_t - \mu_f)$ and $\theta^2 = \mu_f' V_f^{-1} \mu_f$.[16]

To conduct statistical tests, we need a consistent estimator of $E[h_t^2]$. This can be obtained by replacing each term in h_t with the corresponding sample estimate. We denote the result $\hat{h}_t$ and calculate the sample second moment, $\sum_{t=1}^{T} \hat{h}_t^2/T$.

Additional insight into the determinants of the asymptotic variance of the mimicking portfolio sample squared Sharpe ratio in Proposition 2 can be obtained by specializing the analysis. The next result examines the case of factors and returns that are multivariate elliptically distributed.

LEMMA 2: *When the factors and returns are i.i.d. multivariate elliptically distributed with kurtosis parameter κ, the asymptotic variance of $\hat{\theta}^2$ is given by*

$$E[h_t^2] = \theta^2 \left[4 + (2 + 3\kappa)\theta^2\right] + 4(1 + \kappa)E[y_t^2]\left(\theta_R^2 - \theta^2\right), \tag{14}$$

where $\theta_R^2 = \mu_R' V_R^{-1} \mu_R$ represents the squared maximum Sharpe ratio of R, $E[y_t^2] = \mu^{*\prime} V^{*-1} V_{f\cdot R} V^{*-1} \mu^*$, and $V_{f\cdot R} = V_f - V_{fR} V_R^{-1} V_{Rf}$ is the covariance matrix of the residuals from projecting the factors on the returns.

Proof: See Appendix.

Note that the first term in equation (14) is all that would be needed to compute the asymptotic variance of $\hat{\theta}^2$ if the mimicking-portfolio weights were known. The second term in equation (14) represents the errors-in-variables (EIV) adjustment required when the weights are estimated. The EIV adjustment term is nonnegative since $1 + \kappa > 0$ and $\theta_R^2 \geq \theta^2$.[17] The latter inequality holds since θ_R^2 is the squared maximum Sharpe ratio over *all* portfolios of R, whereas θ^2 is the squared

[16] In this case, the asymptotic approximation provided by Maller, Durand, and Jafarpour (2010) and Maller, Roberts, and Tourky (2016) could be used to derive the asymptotic variance of the sample squared Sharpe ratio. However, from their expression, it is not clear how to accommodate serial correlation, while it is straightforward from inspection of equation (13).

[17] Bentler and Berkane (1986) show that $1 + \kappa > 0$.

maximum Sharpe ratio over combinations of the mimicking portfolios based on R. The impact of the EIV adjustment term on the asymptotic variance of $\hat{\theta}^2$ can be large when the factors are not well mimicked by the returns, since in this case $E[y_t^2]$ could be very different from zero.

For example, when $K = 1$, we have

$$E[y_t^2] = \frac{(1 - \mathcal{R}^2)\theta^2}{\mathcal{R}^2}, \tag{15}$$

where $\mathcal{R}^2 = V_f^{-\frac{1}{2}} V_{fR} V_R^{-1} V_{Rf} V_f^{-\frac{1}{2}}$ is the coefficient of determination from regressing f_t on R_t. From this expression, it is clear that there is a negative relationship between $E[y_t^2]$ and $\mathcal{R}^2$, which indicates that $E[y_t^2]$ can be large when the factors are poorly mimicked by the underlying basis-asset returns. In contrast, when the factors are perfectly tracked by the basis-asset returns, we have $E[y_t^2] = 0$ and the EIV adjustment term vanishes.[18] The EIV term can also be large when $\theta_R^2 - \theta^2$ is positive, that is, when the K-factor pricing model does not hold. Conversely, when the K-factor pricing holds, i.e., there exists a K-vector λ such that $\mu_R = V_{Rf}\lambda$, then we have $\theta_R^2 = \theta^2$, and the EIV adjustment term will vanish. Finally, $E[h_t^2]$ is increasing in the kurtosis parameter κ.

C. Pairwise Model Comparison with Mimicking Portfolios

Nested models. Without loss of generality, assume that model A has $f_{At}^* = [f_{1t}^{*\prime},\ f_{2t}^{*\prime}]'$, whereas model B has $f_{Bt}^* = f_{1t}^*$. Let $\mu_1^* = E[f_{1t}^*]$ and $\mu_2^* = E[f_{2t}^*]$. Similarly, let $V_{11}^* = \text{Var}(f_{1t}^*)$, $V_{12}^* = \text{Cov}(f_{1t}^*,\ f_{2t}^{*\prime})$, $V_{22}^* = \text{Var}(f_{2t}^*)$, and $V_{21}^* = V_{12}^{*\prime}$. Suppose f_{1t}^* is a K_1-vector and f_{2t}^* is K_2-vector, with $K = K_1 + K_2$.

As with traded-factor models, testing the equality of squared Sharpe ratios of mimicking portfolios when the two models are nested amounts to evaluating the hypothesis that the alphas of the mimicking portfolios excluded from the smaller model (f_{2t}^*) are zero when regressed on the mimicking portfolios common to both models (f_{1t}^*). Paralleling the notation in Section II.A, the hypothesis is $\alpha_{21}^* = 0_{K_2}$. In this case, we can no longer use a basic alpha-based test since we have generated regressors (the portfolio weights).

[18] See Jobson and Korkie (1980) for a derivation of the asymptotic distribution of the sample squared Sharpe ratio under the assumption that the traded factors (returns) are multivariate normally distributed.

PROPOSITION 3: *Under the null hypothesis* $H_0 : \alpha^*_{21} = 0_{K_2}$,

$$T\hat{\alpha}^{*\prime}_{21}\hat{V}\left(\hat{\alpha}^*_{21}\right)^{-1}\hat{\alpha}^*_{21} \overset{A}{\sim} \chi^2_{K_2}, \tag{16}$$

where $\hat{V}(\hat{\alpha}^*_{21})$ *is a consistent estimator of*

$$V(\hat{\alpha}^*_{21}) = E[q_t q_t'], \tag{17}$$

with

$$q_t = \xi_t(1 - y_{1t}) + w_t(v_t - u_{1t}), \tag{18}$$

$\xi_t = (f^*_{2t} - \mu^*_2) - V^*_{21}V^{*-1}_{11}(f^*_{1t} - \mu^*_1)$, $y_{1t} = \mu^{*\prime}_1 V^{*-1}_{11}(f_{1t} - \mu_1)$, $\eta_{1t} = (f_{1t} - \mu_1) - (f^*_{1t} - \mu^*_1)$, $\eta_{2t} = (f_{2t} - \mu_2) - (f^*_{2t} - \mu^*_2)$, $u_{1t} = \mu^{*\prime}_1 V^{*-1}_{11}(f^*_{1t} - \mu^*_1)$, *and* $w_t = \eta_{2t} - V^*_{21}V^{*-1}_{11}\eta_{1t}$.

Proof: See Appendix.

If $K_2 = 1$, we can simply rely on the t-ratio associated with $\hat{\alpha}^*_{21}$ to perform the test. In the traded-factor case, we can employ the basic alpha-based test for the purpose of testing $\alpha_{21} = 0_{K_2}$, since in this case we have no generated regressors. We also show in the Appendix that the zero-intercept restriction is equivalent to a restriction in the GLS cross-sectional regression framework, but with excess returns (the vector R) projected on covariances with the factors, instead of betas.

Non-nested models. Now consider two non-nested models, A and B, with mimicking portfolios f^*_{At} and f^*_{Bt}, respectively. Let $\mu^*_A = E[f^*_{At}]$ and $\mu^*_B = E[f^*_{Bt}]$. Similarly, let $V^*_A = \text{Var}(f^*_{At})$ and $V^*_B = \text{Var}(f^*_{Bt})$. Finally, denote the nonzero population squared Sharpe ratios that are attainable from the two sets of mimicking portfolios by θ^2_A and θ^2_B, with sample counterparts $\hat{\theta}^2_A$ and $\hat{\theta}^2_B$.

PROPOSITION 4: *The asymptotic distribution of the difference in sample squared Sharpe ratios is given by*

$$\sqrt{T}\left([\hat{\theta}^2_A - \hat{\theta}^2_B] - [\theta^2_A - \theta^2_B]\right) \overset{A}{\sim} N\left(0, E[d^2_t]\right), \tag{19}$$

provided that $E[d^2_t] > 0$, *where*

$$d_t = h_{At} - h_{Bt}, \tag{20}$$

with $u_{At} = \mu_A^{*\prime} V_A^{*-1}(f_{At}^* - \mu_A^*)$, $y_{At} = \mu_A^{*\prime} V_A^{*-1}\eta_{At}$, $h_{At} = 2u_{At}(1 - y_{At}) - u_{At}^2 + 2y_{At}v_t + \theta_A^2$, *and similarly for model B. As defined earlier,* $\eta_{jt} = (f_{jt} - \mu_j) - (f_{jt}^* - \mu_j^*)$ *for* $j = A, B$.

Proof: See Appendix.

Proposition 4 reveals that when the factors of models A and B are perfectly spanned by the basis-asset returns, that is, $y_{At} = y_{Bt} = 0$, then $E[d_t^2]$ collapses to the asymptotic variance provided in Proposition 1 for the traded-factor case. Typically, y_{At} and y_{Bt} are different from zero, and the EIV adjustment term can be a main driver of the asymptotic variance of the difference in sample squared Sharpe ratios of two sets of mimicking-portfolio returns. As earlier, when the factors and returns are i.i.d. multivariate elliptically distributed, additional insights can be obtained.[19] For example, if the returns on the maximum Sharpe ratio portfolios of f_{At}^* and f_{Bt}^* are perfectly correlated, then $E[d_t^2]$ is zero and the asymptotic normality result in Proposition 4 breaks down. Perfect correlation occurs, in particular, when both models A and B price the basis-asset returns correctly so that the maximum Sharpe ratio portfolios for A and B both equal the maximum Sharpe ratio portfolio for the basis-asset returns. This is unlikely to be true in practice, however.

Similar to the traded-factors scenario, it is important when evaluating two non-nested models to test whether the common mimicking portfolios (if any) span the maximum Sharpe ratio portfolio based on the mimicking portfolios for both models. If so, the mimicking portfolios specific to each model are redundant and the models deliver the same squared Sharpe ratio. Equivalently, the alphas of those redundant portfolios must be zero. Testing this hypothesis again boils down to an extension of the basic alpha-based test to accommodate estimation error in the mimicking portfolio weights – in this case, with model-specific mimicking portfolios as the left-hand-side returns (see Proposition 5 in the Appendix).

IV. Multiple Model Comparison

Suppose a researcher is considering more than two models and wants to test whether one of the models – the "benchmark" – is at least as good (it has at least as high squared Sharpe ratio) as

[19] Lemma 3 in the Appendix provides an explicit expression for $E[d_t^2]$ under a multivariate elliptical assumption on the factors and the returns.

the others. In such a case, the relevant *significance level* for a series of pairwise comparisons will not be clear and so a joint test is needed. The analysis with traded factors is outlined here.[20] We begin with the simple case of nested models. Then we turn to the more challenging examination of non-nested models.

Nested models. Consider a benchmark model that is nested in a series of alternative models. We form a single alternative model that includes all of the factors contained in the models that nests the benchmark. It is then easily demonstrated that the expanded model dominates the benchmark model if and only if one or more of the "larger" models dominates it. Thus, the null hypothesis that the benchmark model has the same (it cannot be higher) squared Sharpe ratio as these alternatives can be tested using the methodology developed for pairwise nested-model comparison. Specifically, we examine the alphas from projecting all the factors excluded from the benchmark model onto the benchmark factors and test whether these alphas are jointly zero. If we reject the null of zero alphas, then we conclude that the benchmark model is dominated by one or more of the larger models. Otherwise, we fail to reject the hypothesis that the benchmark model performs as well as the other models.

Non-nested models. Our multiple model comparison test for non-nested models is based on the multivariate inequality test of Wolak (1987), (1989). Suppose we have p models. Let $\delta = (\delta_2, \ldots, \delta_p)$ and $\hat{\delta} = (\hat{\delta}_2, \ldots, \hat{\delta}_p)$, where $\delta_i = \theta_1^2 - \theta_i^2$ and $\hat{\delta}_i = \hat{\theta}_1^2 - \hat{\theta}_i^2$ for $i = 2, \ldots, p$. We are interested in testing

$$H_0 : \delta \geq 0_r \quad \text{vs.} \quad H_1 : \delta \in \Re^r, \tag{21}$$

where $r = p - 1$ is the number of non-negativity restrictions. Thus, under the null hypothesis, model 1 (the benchmark) performs at least as well as models 2 to p (the competing models).

The test is based on the sample counterpart of δ, $\hat{\delta} = (\hat{\delta}_2, \ldots, \hat{\delta}_p)$, which has an asymptotic normal distribution with mean δ and covariance matrix $\Sigma_{\hat{\delta}}$ (conditions for this are provided in the Online Appendix to Kan, Robotti, and Shanken (2013)). The test statistic is constructed by first

[20] Details are available from the authors upon request along with the extension to accommodate mimicking portfolios.

solving the quadratic programming problem

$$\min_{\delta}(\hat{\delta}-\delta)'\hat{\Sigma}_{\hat{\delta}}^{-1}(\hat{\delta}-\delta) \quad \text{s.t.} \quad \delta \geq 0_r, \tag{22}$$

where $\hat{\Sigma}_{\hat{\delta}}$ is a consistent estimator of $\Sigma_{\hat{\delta}}$. Let $\tilde{\delta}$ be the optimal solution of the problem in equation (22). The likelihood ratio test of the null hypothesis is given by

$$LR = T(\hat{\delta}-\tilde{\delta})'\hat{\Sigma}_{\hat{\delta}}^{-1}(\hat{\delta}-\tilde{\delta}). \tag{23}$$

A large value of LR suggests that the non-negativity restrictions do not all hold. To conduct statistical inference, we need the asymptotic distribution of LR. We refer the readers to Kan, Robotti, and Shanken (2013) for its derivation and a discussion of numerical methods for calculating the p-value.

In comparing a benchmark model with a set of alternative models, we first remove those alternative models i that are nested by the benchmark model since by construction the null hypothesis, $\delta_i \geq 0$, holds in this case. If any of the remaining alternatives is nested by another alternative model, we remove the "smaller" model since the squared Sharpe ratio of the "larger" model will be at least as big. Finally, we also remove from consideration any alternative models that nest the benchmark, since for nested models the asymptotic normality assumption on $\hat{\delta}_i$ does not hold under the null hypothesis that $\delta_i = 0$.

V. Empirical Results

We start by describing the factors and the various empirical asset-pricing specifications. Next, we summarize the empirical findings for the tests of equality of squared Sharpe ratios for competing traded-factor models. Finally, we explore model comparison for the mimicking-portfolio case.

A. Factors and Pricing Models

We analyze eight asset-pricing models starting with a two-factor model which, in addition to the value-weighted market excess return (MKT), includes a traded financial intermediary capital risk factor ($FIRFT$) developed by He, Kelly, and Manela (2017) (MKT+FIRFT). $FIRFT$ is the monthly value-weighted equity excess return for the New York Fed's primary dealer sector and it does not include new equity issuance. The authors find significant pricing results for this factor in cross-sectional regression analysis. Therefore, it is of interest to further examine the performance of their model with the Sharpe metric employed. They obtain stronger cross-sectional results for a nontraded version of the factor that we will consider later in Section V.C. Second is the Frazzini and Pedersen (2014) model, which extends the CAPM with the betting-against-beta factor (BAB) – long low-beta assets and short high-beta assets (MKT+BAB).

The third model is the Fama and French (2018) five-factor model (FF5CP), which adds an investment factor (CMA) and a cash profitability factor ($RMWCP$) to the FF3 model. 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 (2×3) sorts, interacting size with cash profitability for the construction of $RMWCP$, and separately with investments to create CMA. $RMWCP$ 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 2×3 sorts minus the average of the returns on the nine big-stock portfolios.

Note that FF5CP differs from the original Fama and French (2015) five-factor model which constructs the profitability factor using an *accruals-based* operating profitability measure suggested by Novy-Marx (2013). Ball, Gerakos, Linnainmaa, and Nikolaev (2016) argue that a cash-based measure of profitability yields a factor that better accounts for average return differences in sorts on accruals. Following Fama and French (2018), our fourth model adds the up-minus-down (UMD) momentum factor motivated by the work of Jegadeesh and Titman (1993) to the FF5CP model (FF5CP+UMD).

The fifth model is the Hou, Xue, and Zhang (2015) four-factor model (HXZ), which includes size (ME), investment (IA), and profitability (ROE) factors in addition to the market. In contrast to Fama and French (2018), HXZ construct their factors from a triple ($2 \times 3 \times 3$) sort on these characteristics. Moreover, their profitability measure is based on income before extraordinary items taken from the most recent public quarterly earnings announcement. Our sixth model is the four-factor model of Stambaugh and Yuan (SY) (2015), which extends the CAPM by adding a size factor (*SMBSY*) and two mispricing factors, "management" and "performance" (*MGMT* and *PERF*), that aggregate information across 11 prominent anomalies by averaging rankings within two clusters exhibiting the greatest return co-movement.

Given that the choice of profitability factor is a key to the performance of the five-factor model of Fama and French, our seventh model substitutes *RMWCP* for *ROE* in the HXZ model (HXZCP). Our final model (FF5CP*+UMD) includes the more timely value factor HML^m from Asness and Frazzini (2013) instead of the standard *HML*. HML^m is based on book-to-market rankings that use the most recent monthly stock price in the denominator, whereas *HML* uses annually updated lagged prices. The sample period for our data is January 1972 to December 2015. 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.

Panel A of Table 1 presents summary statistics for each monthly factor return – mean, standard deviation, and t-statistic. The latter is, of course, proportional to the factor Sharpe ratio.[21]

Table 1 about here

All factors have positive and sizable average returns. The factor with the highest return premium is *BAB*, followed by *UMD*, *PERF*, and *MGMT*. The size factors, *SMB* and *ME*, have the smallest return premiums. The financial intermediary capital risk factor has the highest volatility of all the factors. All premiums, with the exception of *SMB*, have t-statistics larger than 2. The cash profitability factor, *RMWCP*, has the lowest standard deviation, which partly explains why it has the highest t-statistic (6.67).

Panel B of Table 1 provides the factor correlations. Naturally, different versions of the same factor tend to be highly correlated. We make a few additional observations about the factors that are newer to the factor-pricing literature. As noted by Asness and Frazzini (2013), *UMD* is much

[21] Return spreads, such as *HML*, can be viewed as long \$1 in the risk-free asset and \$1 on each side of the spread. Then, *HML* is the excess return on that unit investment factor.

more negatively correlated with timely value, HML^m (-0.654), than with HML (-0.168). On the other hand, correlations between the value, investment, and $MGMT$ factors are strong and positive, but weaker for HML^m than HML. The correlations between profitability, momentum, and $PERF$ are also high. These mispricing factor correlations make sense insofar as the $MGMT$ cluster includes the investment/assets anomaly, while the $PERF$ cluster includes momentum and gross profitability. Finally, the intermediary factor of He, Kelly, and Manela (2017) is highly correlated with the market factor (0.819).

B. Tests of Equality of Squared Sharpe Ratios for Competing Traded-Factor Models

In Table 2, we report pairwise tests of equality of the squared Sharpe ratios for different models, some nested and others non-nested.[22] The models are presented from left to right and top to bottom in order of increasing squared Sharpe ratios. Panel A shows the differences between the (bias-adjusted) sample squared Sharpe ratios (column model − row model) for various pairs of models. In Panel B, we report p-values for the tests of equality of the squared Sharpe ratios. The estimated squared Sharpe ratio for each model is modified so as to be unbiased in small samples under joint normality. This entails multiplying $\hat{\theta}^2$ by $(T-K-2)/T$ and subtracting K/T, eliminating the upward bias, while leaving the asymptotic distribution unchanged. We use * to highlight those cases that are significant at the 5% level and ** for the 1% level.

Table 2 about here

The diagonal elements of Panel A are the sample squared Sharpe ratio differences between

[22]The required condition mentioned earlier, that a model's Sharpe ratio is nonzero, can be evaluated using a chi-squared test. Specifically, under $H_0: \theta^2 = 0$, $T\hat{\theta}^2 \overset{A}{\sim} \chi^2_K$. In our empirical application, we reject this null for all of our models at the 1% level, except for MKT+FIRFT where the null is rejected at the 5% level.

Although we pointed out that maximizing the squared Sharpe ratio is equivalent to maximizing the Sharpe ratio if we allow shorting of the tangency portfolio, sometimes (e.g., CAPM) we may want to interpret the upper tangency as part of the restriction of the model. A model's tangency portfolio is on the upper half of the minimum-variance frontier when $b = e'V_f^{-1}\mu_f \geq 0$, where e is a K-vector of initial costs of the factor portfolios. Specifically, $e_i = 1$ for factor returns in excess of the risk-free rate (e.g., $FIRFT$ and MKT) and $e_i = 0$ for differences in returns on two factors (e.g., SMB and HML). The condition $b = e'V_f^{-1}\mu_f \geq 0$ can be tested by considering $\hat{b} = e'\hat{V}_f^{-1}\hat{\mu}_f$ and its associated t-statistic. A derivation of the asymptotic distribution of $\hat{b}$ is available from the authors upon request. In the data, the b estimates are positive for all models and the associated t-ratios range from 2.41 (for MKT+FIRFT) to 5.36 (for FF5CP*+UMD), thus suggesting that the b's for the various models are reliably positive.

the model in that column and the next best model.[23] As previously discussed, p-values must be computed differently depending on whether the models to be compared are nested or non-nested. In the case of nested models, we test whether the factors in the larger model that are excluded from the smaller model have zero alphas when regressed on the smaller model. For example, since FF5CP is nested in FF5CP+UMD, the corresponding p-value reported in Panel B is for the intercept in the regression of *UMD* on FF5CP.

When the models are non-nested, which is the case for the rest of our comparisons, we use our sequential test. We first check whether the difference in squared Sharpe ratios between the model composed of the common factors and the one that includes all the factors from both models is different from zero. This is a test of whether the alphas of the non-common factors on the common ones are zero. If this test fails to reject, then the evidence is consistent with the common-factors model being as good as the model that adds the non-overlapping factors. Thus, the two non-nested models are equivalent as well under this null. However, if the preliminary test rejects, then we proceed to directly test whether the squared Sharpe ratios of the non-nested models are different by computing the p-value based on the results in Proposition 1.

For example, in comparing the two non-nested models, HXZ and HXZCP, we first run the alpha-based test for the different profitability factors, *ROE* and *RMWCP*, regressed on the three-factor model (with *MKT*, *ME*, and *IA*) that is nested in these two models. This test easily rejects the joint hypothesis that both alphas are zero with p-value virtually zero. In fact, this is the case for the preliminary test in all our non-nested pairwise model comparisons. Had the preliminary test not rejected in this example, the evidence would be consistent with the three-factor model being as good as either of the two four-factor models. However, since it did reject, the next step is to divide the (bias-adjusted) squared Sharpe ratio difference, $0.273 - 0.166 = 0.107$, by its standard error, 0.038, which is the square root of the asymptotic variance given in Proposition 1 divided by the number of monthly observations ($\sqrt{0.777/528}$). This yields a t-statistic of 2.78, with p-value 0.005, as reported in Panel B.

[23] The bias-adjusted sample squared Sharpe ratio for MKT+FIRFT, not shown, is 0.0095.

The main empirical findings can be summarized as follows. First, the results show that the MKT+FIRFT and MKT+BAB models are outperformed by the other models, with significance at the 1% level except for HXZ which outperforms MKT+BAB with a 3% level of significance.[24] Next, FF5CP has a higher sample squared Sharpe ratio than both SY and HXZ, but the difference between them is not statistically significant. When we add the momentum factor to the FF5CP model, it outperforms HXZ at the 5% level, but it still does not dominate the SY model, which includes the related factor, *PERF*. Moreover, adding momentum to FF5CP does not result in a statistically significant increase in the squared Sharpe ratio. Replacing the original profitability factor (ROE) in the HXZ model with the cash-based profitability factor ($RMWCP$) results in a substantial increase in the squared Sharpe ratio, that is statistically significant at the 1% level. This version of HXZ, HXZCP, now outperforms the SY model as well as FF5CP and FF5CP+UMD, but the differences are not reliably different from zero. Finally the choice of value factor in the six-factor Fama and French (2018) model is important. In fact, with the more timely value factor (HML^m), the model FF5CP*+UMD outperforms all of the other models at the 5% level.[25]

Thus far, we have considered comparisons of two competing models. Statistical significance may be overstated, however, by the inevitable process of "searching" for comparisons that lead to rejection. Therefore, given a set of models of interest, one may want to test whether a single model, the "benchmark," has the highest squared Sharpe ratio of all the models. To explore this issue, we use the test for non-nested models based on the multivariate inequality analysis of Wolak (1989), outlined in Section IV. The null hypothesis in this joint test is that none of the other models is superior to the benchmark. The alternative is that some other model has a higher (population) θ^2 than the benchmark. The empirical results are presented in Table 3.

[24] Given the poor performance of the He, Kelly, and Manela (2017) two-factor model when using monthly data, we also compared MKT+FIRFT with FF3 using quarterly data from 1970:Q1 to 2012:Q4 (from Asaf Manela's website). Our findings indicate that the difference between the squared sample Sharpe ratios of FF3 and MKT+FIRFT is 0.0743 with a p-value of 0.0945 for the normal test (the spanning condition is rejected at the 1% significance level). Therefore, moving from monthly to quarterly data seems to partially help, and MKT+FIRFT performs about the same as the popular Fama and French (1993) three-factor model at the 5% significance level based on our Sharpe ratio improvement metric.

[25] If we exclude CMA and SMB from the six-factor model FF5CP*+UMD, the sample squared Sharpe ratio of the resulting four-factor model is still higher than that of HXZCP by 0.05. However, the difference is no longer statistically significant (p-value of 0.224).

Table 3 about here

Naturally, since FF5CP*+UMD has the highest sample squared Sharpe ratio, the p-value for this model in the joint test is very large, consistent with the conclusion that FF5CP*+UMD performs at least as well in population as the other models. More interesting is the case in which HXZCP is the benchmark. Whereas FF5CP*+UMD was superior (p-value of 0.043) to this model in the pairwise comparisons, the p-value for the joint test with benchmark HXZCP is 0.120. Thus, we miss rejecting the hypothesis that HXZCP has a squared Sharpe ratio at least as big as those for the alternative models. However, we do continue to reject the remaining models with p-values close to zero in the joint test except for SY, which we can only reject at the 5% level.

C. Model Comparison with a Nontraded Financial Intermediary Capital Risk Factor

Section III develops a test for comparing competing models when one or both models contain mimicking portfolios. As an application of that methodology, we explore the nontraded financial intermediary capital risk factor of He, Kelly, and Manela (2017). The theory behind the intermediary factors assumes that the intermediaries chosen are marginal investors for all asset classes. Although the authors note that this is least likely for equities (the intermediaries are more active in other assets), cross-sectional regression analysis with the zero-beta rate constrained to equal the risk-free rate produces a positive risk premium for the nontraded factor that is economically large, with a t-statistic of 3.81. The actual series of nontraded factor values, η_t^{Δ}, is defined in terms of AR(1) innovations in the capital ratio, scaled by the lagged capital ratio. The capital ratio is the ratio of total market equity (measured monthly) to total market assets (book debt plus market equity) of New York Fed's primary dealer holding companies, where book debt is the latest quarterly observation.

We first construct a mimicking portfolio ($FIRFM$) by regressing η_t^{Δ} on a constant and all of the traded-factor returns considered above. Thus, $R = (MKT, SMB, HML, CMA, RMWCP, ME, IA, ROE, UMD, HML^m, BAB, SMBSY, MGMT, PERF, FIRFT)$ includes all the factors in the models that we wish to compare. Additional basis assets could be considered, but are not required.

Although some of these returns are highly correlated, we are interested in the fitted value (the overall mimicking return), not the individual weights. The sample period is again January 1972 to December 2015.

There is no requirement for asset pricing or the asymptotic analysis that the mimicking portfolio be highly correlated with the underlying factors. However, the correlation should be significantly different from zero so as to avoid complications akin to the "useless factor" problems in cross-sectional regressions (see Kan and Zhang (1999)). The mimicking portfolio regression for η^{Δ} has an adjusted R^2 of 0.85. Furthermore, the F test of joint significance of the slope coefficients in the mimicking portfolio projection yields a p-value which is essentially zero. Thus, the evidence indicates that these asset returns are able to mimic the nontraded factor to a very large degree. Not surprisingly, the contribution of the traded intermediary factor, *FIRFT*, to the mimicking portfolio is reliably different from zero.[26]

Insofar as marginal utility is high when intermediaries experience a negative shock to their equity capital, asset-pricing theory suggests a positive premium for intermediary risk. The intermediary mimicking portfolio, *FIRFM*, has an average risk premium of 0.58% per month over our sample period. The associated t-statistic is 2.15, so the estimate is reliably positive.[27] Similarly, *FIRFT* has an average premium of 0.61% or 7.32% annualized, with a t-statistic of 2.06. The correlation between *FIRFT* and *FIRFM* is almost 100%, whereas the correlation of *FIRFM* with market excess returns is 0.817. The (jackknife bias-adjusted) squared sample Sharpe ratio for MKT+FIRFM is 0.0085.[28] The jackknife estimate entails multiplying $\hat{\theta}^2$ by T and subtracting $(T-1)\hat{\theta}^2_{(\cdot)}$, where

[26]Panel B of Table 1 indicates that in absolute value, the correlation between *FIRFT* and several other traded factors is quite high as well.

[27]The t-statistic is computed based on the asymptotic distribution of $\hat{\mu}^*$, which is given by

$$\sqrt{T}(\hat{\mu}^* - \mu^*) \overset{A}{\sim} N(0_K, E[q_t q_t']), \tag{24}$$

where

$$q_t = f_t^* - \mu^* + \eta_t v_t. \tag{25}$$

A proof of this result is available from the authors upon request. To conduct statistical tests, we need a consistent estimator of $E[q_t q_t']$. This can be obtained, as earlier, by replacing all quantities in q_t by their sample counterparts and taking the time-series sample second moment.

[28]Based on a chi-squared test with 2 degrees of freedom, we reject the null of a zero squared Sharpe ratio for MKT+FIRFM at the 5% level.

$\hat{\theta}^2_{(\cdot)} = \frac{1}{T}\sum_{t=1}^{T}\hat{\theta}^2_{(t)}$ and $\hat{\theta}^2_{(t)}$ is the squared Sharpe ratio estimated by removing the t-th observation from the data.[29]

Next, we compare the performance of this nontraded financial intermediary model to that of the traded-factor models considered earlier, again taking into account estimation error in the mimicking portfolio weights. Accordingly, Panel A of Table 4 reports the differences in (bias-adjusted) squared sample Sharpe ratios.

Table 4 about here

As earlier, models are presented in order of increasing squared Sharpe ratio from left to right. Finally, we assess the statistical significance of these differences using the result in Proposition 4, which provides the asymptotic variance of the difference in sample squared Sharpe ratios for two models with mimicking portfolios. In this application, some terms drop out, since MKT+FIRFM is being compared to models with all traded factors. Panel B of Table 4 reports the p-values. MKT+FIRFM is dominated by all models except for MKT+FIRFT.[30] Thus, recalling the evidence in Table 2, neither the traded nor the nontraded intermediary models fare well in our tests.[31]

VI. Simulation Evidence

In this section, we explore the small-sample properties of our various test statistics via Monte Carlo simulations. The time-series sample size is taken to be $T = 540$, close to the actual sample size of 528 in our empirical work. The factor and basis-asset returns are drawn from a multivariate normal distribution. We compare actual rejection rates over 100,000 iterations to the nominal 5%

A generalization of the test discussed in footnote 22 takes into account the estimation error in the weights of the mimicking portfolio and is available upon request. As for the models with traded factors only, we find no evidence of a negative b. In the data, the b estimate for MKT+FIRFM is positive (2.55) and the associated t-ratio is 2.34.

[29]The difference between the jackknife-adjusted estimator and the sample estimator is of order $O(1/T)$, so it does not have an impact on the asymptotic variance.

[30]Using quarterly data from Asaf Manela's website over the period 1970:Q1 – 2012:Q4, we find slight improvements in performance for MKT+FIRFM. The squared sample Sharpe ratio of the model is now 0.0920, and the difference in squared sample Sharpe ratios of the two-factor model and FF3 is negligible, 0.0024, with a p-value of 0.7092 for the normal test (the spanning condition is rejected at the 1% level).

[31]We also have explored the traded and nontraded versions of the Pástor and Stambaugh (2003) liquidity factors with similar results. The mimicking regression $\mathcal{R}^2$ for the nontraded factor is just 17% in this case and surprisingly, the contribution of the traded liquidity factor to the mimicking portfolio is not reliably different from zero.

level of our tests. A more detailed description of the various simulation designs can be found in the Appendix.

We start by considering models with traded factors only. As emphasized in Section II.A, the null hypothesis of equal squared Sharpe ratios for nested models can be tested using the alpha-based test. Here, the size of the alpha-based test, with FF3 nested in FF5CP, is inferred from simulations in which *RMWCP* and *CMA* are exactly priced by the three common factors, *MKT*, *SMB*, and *HML*. The alpha-based test performs very well, with a rejection rate of 5%. Power for the nested-models test is evaluated by simulating data for which the true squared Sharpe ratios equal the sample values and thus FF3 is dominated by FF5CP. The rejection rate for this scenario is 100%.

Next, we turn to non-nested models and consider MKT+FIRFT vs. HXZCP. This is an example of non-nested models with a common factor, *MKT*. In this case, as emphasized in Section II.B, the null of equal squared Sharpe ratios can hold when the common factor, *MKT*, spans the maximum Sharpe ratio portfolio based on the factors from both models (*FIRFT* for MKT+FIRFT, and *ME*, *RMWCP*, and *IA* for HXZCP). Again, this condition can be tested using the alpha-based test. This test is right on the money with rejection rates of 5.0% and 100% under the null and alternative hypotheses, respectively. If we reject this spanning condition, then we can still have equality of squared Sharpe ratios and this equality can be tested using the normal test in Proposition 1. In this experiment, the factor means are specified in such a way that the squared Sharpe ratio is the same for MKT+FIRFT and HXZCP, that is, 0.284. The size property of the normal test is excellent (5.0%). The power of the normal test is explored using the sample squared Sharpe ratios of MKT+FIRFT and HXZCP as the population squared Sharpe-ratio values. These are 0.013 and 0.284, so the null hypothesis of equivalent model performance is false in these simulations. The rejection rate of 100% reflects the large differences in sample squared Sharpe ratios across models and the high precision of these estimates.

We also examine the small-sample properties of the multiple-comparison inequality test for non-nested models. Recall that the composite null hypothesis for this test maintains that θ^2 for the benchmark model is at least as high as that for all other models under consideration. Therefore, to evaluate size, we consider the case in which all models have the same θ^2 value, so as to maximize the likelihood of rejection under the null. We simulate six different single-factor models corresponding to the factors *MKT*, HML^m, *RMWCP*, *UMD*, *IA*, and *FIRFT*, and implement the likelihood ratio test with $r = 5$. Since we calibrate the parameters to the market factor, *MKT*, the implied common θ^2 for the various models is 0.013. The rejection rates range from 3.3% to 6.7%. Thus, the test is fairly well specified under the null of equivalent model performance. To examine power, we simulate four of our original models, MKT+FIRFT, HXZ, FF5CP, and FF5CP*+UMD, with the sample squared Sharpe ratios serving as the population θ^2s. Since FF5CP*+UMD has the highest θ^2, we let each of the remaining models serve as the null model in a multiple comparison test against three alternative models. Thus, we evaluate power for three different scenarios. The rejection rates for the test are very high: 100% for MKT+FIRFT, 99.8% for FF5CP, and 95.5% for HXZ.

Turning to the analysis with mimicking portfolios, we set $R = (MKT, SMB, HML, CMA, RMWCP, ME, IA, ROE, UMD, HML^m, BAB, SMBSY, MGMT, PERF, FIRFT)$, that is, R contains all the traded-factor returns considered in the empirical section of the paper. We start from the nested-model case. As emphasized in Section III.C, this is a situation in which we can no longer employ the basic alpha-based test to implement nested-model comparison since the mimicking portfolio weights need to be estimated. Instead, we rely on the chi-squared test in Proposition 3. The size of this test, with the CAPM nested in FF3+FIRFM, is inferred from simulations in which the intermediary mimicking portfolio, *SMB*, and *HML* are exactly priced by the common factor, *MKT*, and the mean returns, μ_R, also incorporate the constraint $\alpha_{21}^* = 0_{K_2}$. Our new test performs very well, with a rejection rate of 5.1%. The power properties of our chi-squared test are analyzed by simulating data for which the true squared Sharpe ratios equal the sample values and thus the CAPM is dominated by FF3+FIRFM. Since the difference in true squared Sharpe ratios is substantial (0.032), the rejection rate of the test is 94.5%. Naturally, "good" power requires that the differences in model performance are fairly large.

As for non-nested models, we consider MKT+FIRFM vs. HXZCP, and test the spanning condition using our result in Proposition 5 in the Appendix. The chi-squared test enjoys excellent size and power properties with a rejection rate of 5.2% under the null of spanning and a rejection rate of 100% under the alternative of no spanning. Equality of squared Sharpe ratios can occur also when the spanning condition is rejected. In this scenario, the normal test in Proposition 4 should be used. To investigate the size properties of the normal test, the factor means are specified in such a way that the squared Sharpe ratio is the same for MKT+FIRFM and HXZCP, that is, 0.078. The normal test is found to perform well under the null, with a rejection rate of 4.1%. The power of the normal test is explored using the sample squared Sharpe ratios of MKT+FIRFM and HXZCP as the population squared Sharpe-ratio values. These are 0.013 and 0.284, respectively. The rejection rate for the normal test is 100%. However, in general, power can be affected by the limited precision of the sample squared Sharpe ratios of the models, given the residual in the projection of the nontraded factors on the basis-asset returns.

Finally, in order to analyze the size properties of the multiple-model comparison test, we again simulate six different single-factor models corresponding to the factors *MKT*, HML^m, *RMWCP*, *UMD*, *IA*, and the intermediary mimicking portfolio *FIRFM*. Similar to the traded-factor case, we calibrate the parameters to the market factor, *MKT*. The implied common θ^2 for the various models is therefore 0.013. The rejection rates range from 3.2% to 6.1%. Thus, the test is fairly well specified under the null of identical model performance. To examine power, we simulate four of our original models, MKT+FIRFM, FF5CP, HXZ, and FF5CP*+UMD, with the sample squared Sharpe ratios serving as the population θ^2s. Since FF5CP*+UMD has the highest θ^2, we let each of the remaining models serve as the null model in a multiple comparison test against three alternative models. The rejection rates for the test are 100% for MKT+FIRFM, 99.7% for FF5CP, and 94.5% for HXZ.

In summary, our Monte Carlo simulations suggest that the proposed tests should be fairly reliable for the sample size encountered in our empirical work.

VII. Conclusion

Barillas and Shanken (2017) analyze model comparison with the extent of model mispricing measured by the improvement in the squared Sharpe ratio. This is the increase obtained when investment in other returns (traded factors and test assets) is considered in addition to a model's factors. In this framework, model comparison is equivalent to identifying the model whose factors yield the highest squared Sharpe ratio. Moreover, this result extends to models that include nontraded factors, with mimicking portfolios substituted for those factors.

We have shown how to conduct asymptotically valid tests for such model comparisons and apply these methods in an analysis of a variety of factor-pricing models. A variant of the six-factor model of Fama and French (2018), with a monthly-updated version of the usual value spread, emerges as the dominant model over the period 1972–2015.

Clearly, some factor models in the literature, like the CAPM or the financial intermediary model, have stronger theoretical motivation than others. As Harvey, Liu, and Zhu (2016) have emphasized, the statistical hurdle for testing such specifications should be lower than that used for testing more empirically-driven models. In the context of our model comparison tests, the factor alphas and resulting Sharpe ratios of the more empirically-based models can be biased upward due to data mining, thus enhancing model performance. In principle, it would be desirable to develop some sort of bias adjustment for data mining, but this remains a challenging topic for future research.

Appendix

Proof of Proposition 2:

The proof relies on the fact that $\hat{\theta}^2$ is a smooth function of $\hat{\mu}$ and $\hat{V}$. Therefore, once we have the asymptotic distribution of $\hat{\mu}$ and $\hat{V}$, we can use the delta method to obtain the asymptotic distribution of $\hat{\theta}^2$. Let

$$\text{(A-1)} \qquad \varphi = \begin{bmatrix} \mu \\ \text{vec}(V) \end{bmatrix}, \qquad \hat{\varphi} = \begin{bmatrix} \hat{\mu} \\ \text{vec}(\hat{V}) \end{bmatrix}.$$

We first note that $\hat{\mu}$ and $\hat{V}$ can be written as the GMM estimator that uses the moment conditions $E[r_t(\varphi)] = 0_{(N+K)(N+K+1)}$, where

$$\text{(A-2)} \qquad r_t(\varphi) = \begin{bmatrix} Y_t - \mu \\ \text{vec}((Y_t - \mu)(Y_t - \mu)' - V) \end{bmatrix}.$$

Since this is an exactly identified system of moment conditions, it is straightforward to verify that under the assumption that Y_t is stationary and ergodic with finite fourth moment, we have

$$\text{(A-3)} \qquad \sqrt{T}(\hat{\varphi} - \varphi) \overset{A}{\sim} N(0_{(N+K)(N+K+1)}, S_0),$$

where

$$\text{(A-4)} \qquad S_0 = \sum_{j=-\infty}^{\infty} E[r_t(\varphi) r_{t+j}(\varphi)'].$$

Note that S_0 is a singular matrix as $\hat{V}$ is symmetric, so there are redundant elements in $\hat{\varphi}$. We could have written $\hat{\varphi}$ as $[\hat{\mu}', \ \text{vech}(\hat{V})']'$, but the results are the same under both specifications.

Using the delta method, the asymptotic distribution of $\hat{\theta}^2$ is given by

$$\text{(A-5)} \qquad \sqrt{T}(\hat{\theta}^2 - \theta^2) \ \overset{A}{\sim} \ N\left(0, \left[\frac{\partial \theta^2}{\partial \varphi'}\right] S_0 \left[\frac{\partial \theta^2}{\partial \varphi'}\right]'\right).$$

It is straightforward to obtain

$$\text{(A-6)} \qquad \frac{\partial \theta^2}{\partial \mu_f'} = 0_K', \qquad \frac{\partial \theta^2}{\partial \mu_R'} = 2\mu^{*\prime} V^{*-1} A.$$

The derivative of θ^2 with respect to $\text{vec}(V)$ is more involved and is given by

$$\text{(A-7)} \qquad \frac{\partial \theta^2}{\partial \text{vec}(V)'} = \left[0_K', \ \mu^{*\prime} V^{*-1} A\right] \otimes \left[0_K', \ -\mu^{*\prime} V^{*-1} A\right] + \left[0_K', \ \mu_R' \left(V_R^{-1} - A' V^{*-1} A\right)\right] \otimes \left[2\mu^{*\prime} V^{*-1}, \ -2\mu^{*\prime} V^{*-1} A\right].$$

Using the expression for $\partial\theta^2/\partial\varphi'$, we can simplify the asymptotic variance of $\hat{\theta}^2$ to

$$\text{(A-8)} \qquad V(\hat{\theta}^2) = \sum_{j=-\infty}^{\infty} E[h_t(\varphi)h_{t+j}(\varphi)],$$

where

$$\text{(A-9)} \qquad \begin{aligned} h_t(\varphi) &= \frac{\partial\theta^2}{\partial\varphi'} r_t(\varphi) \\ &= 2\mu^{*\prime}V^{*-1}(f_t^* - \mu^*) - \mu^{*\prime}V^{*-1}(f_t^* - \mu^*)(f_t^* - \mu^*)'V^{*-1}\mu^* \\ &\quad + 2\mu^{*\prime}V^{*-1}(f_t - \mu_f)(R_t - \mu_R)'V_R^{-1}\mu_R \\ &\quad - 2\mu^{*\prime}V^{*-1}(f_t^* - \mu^*)(R_t - \mu_R)'V_R^{-1}\mu_R \\ &\quad - 2\mu^{*\prime}V^{*-1}(f_t - \mu_f)(f_t^* - \mu^*)'V^{*-1}\mu^* \\ &\quad + 2\mu^{*\prime}V^{*-1}(f_t^* - \mu^*)(f_t^* - \mu^*)'V^{*-1}\mu^* + \theta^2 \\ &= 2u_t - u_t^2 + 2\mu^{*\prime}V^{*-1}\eta_t v_t - 2\mu^{*\prime}V^{*-1}\eta_t u_t + \theta^2 \\ &= 2u_t(1 - \mu^{*\prime}V^{*-1}\eta_t) - u_t^2 + 2\mu^{*\prime}V^{*-1}\eta_t v_t + \theta^2 \\ &= 2u_t(1 - y_t) - u_t^2 + 2y_t v_t + \theta^2. \end{aligned}$$

In particular, if h_t is uncorrelated over time, then we have $V(\hat{\theta}^2) = E[h_t^2]$, and its consistent estimator is given by

$$\text{(A-10)} \qquad \hat{V}(\hat{\theta}^2) = \frac{1}{T}\sum_{t=1}^{T}\hat{h}_t^2.$$

When h_t is autocorrelated, one can use Newey and West's (1987) method to obtain a consistent estimator of $V(\hat{\theta}^2)$.

This completes the proof of Proposition 2.

Proof of Lemma 2

In our proof, we rely on the mixed moments of multivariate elliptical distributions. Lemma 2 of Maruyama and Seo (2003) shows that if (X_i, X_j, X_k, X_l) are jointly multivariate elliptically distributed and with mean zero, we have

$$\text{(A-11)} \qquad E[X_iX_jX_k] = 0,$$

$$\text{(A-12)} \qquad E[X_iX_jX_kX_l] = (1+\kappa)(\sigma_{ij}\sigma_{kl} + \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}),$$

where $\sigma_{ij} = \text{Cov}[X_i, X_j]$. Consider

$$h_t = 2u_t(1 - y_t) - u_t^2 + 2y_t v_t + \theta^2$$

from Proposition 2. It is straightforward to show that

$$E[u_t] = 0, \tag{A-13}$$

$$E[v_t] = 0, \tag{A-14}$$

$$E[y_t] = 0, \tag{A-15}$$

$$E[u_t^2] = \theta^2, \tag{A-16}$$

$$E[v_t^2] = \theta_R^2, \tag{A-17}$$

$$E[y_t^2] = \mu^{*\prime} V^{*-1} V_{f \cdot R} V^{*-1} \mu^*, \tag{A-18}$$

$$E[u_t v_t] = \theta^2, \tag{A-19}$$

$$E[u_t y_t] = 0, \tag{A-20}$$

$$E[v_t y_t] = 0. \tag{A-21}$$

With these results and under the multivariate elliptical assumption on Y_t, we can show that

$$\begin{aligned} E[h_t^2] &= 4E[u_t^2(1-y_t)^2] + E[u_t^4] + 4E[y_t^2 v_t^2] - 4E[u_t^3(1-y_t)] \\ &\quad + 8E[u_t v_t y_t (1-y_t)] - 4E[u_t^2 v_t y_t] - 2\theta^4 + \theta^4 \\ &= 4\theta^2 + 4(1+\kappa)\theta^2 E[y_t^2] + 3(1+\kappa)\theta^4 \\ &\quad + 4(1+\kappa)\theta_R^2 E[y_t^2] - 0 - 8(1+\kappa)\theta^2 E[y_t^2] - 0 - \theta^4 \\ &= \theta^2[4 + (2+3\kappa)\theta^2] + 4(1+\kappa)E[y_t^2](\theta_R^2 - \theta^2). \end{aligned} \tag{A-22}$$

This completes the proof of Lemma 2.

Proofs of Propositions 1 and 4:

Using Proposition 2, we obtain the following expressions for models A and B:

$$h_{At} = \left[\frac{\partial \theta_A^2}{\partial \varphi}\right]' r_t = 2u_{At}(1 - y_{At}) - u_{At}^2 + 2y_{At} v_t + \theta_A^2, \tag{A-23}$$

$$h_{Bt} = \left[\frac{\partial \theta_B^2}{\partial \varphi}\right]' r_t = 2u_{Bt}(1 - y_{Bt}) - u_{Bt}^2 + 2y_{Bt} v_t + \theta_B^2. \tag{A-24}$$

By the delta method and equations (A-1)–(A-4), the asymptotic distribution of $\hat{\theta}_A^2 - \hat{\theta}_B^2$ is given by

$$\text{(A-25)} \qquad \sqrt{T}([\hat{\theta}_A^2 - \hat{\theta}_B^2] - [\theta_A^2 - \theta_B^2]) \overset{A}{\sim} N\left(0, \left[\frac{\partial(\theta_A^2 - \theta_B^2)}{\partial\varphi}\right]' S_0 \left[\frac{\partial(\theta_A^2 - \theta_B^2)}{\partial\varphi}\right]\right).$$

With the analytical expressions of h_{At} and h_{Bt}, the asymptotic variance of $\sqrt{T}(\hat{\theta}_A^2 - \hat{\theta}_B^2)$ can be written as

$$\text{(A-26)} \qquad \sum_{j=-\infty}^{\infty} E[d_t d_{t+j}],$$

where

$$\text{(A-27)} \qquad d_t = \left(\frac{\partial\theta_A^2}{\partial\varphi} - \frac{\partial\theta_B^2}{\partial\varphi}\right)' r_t = h_{At} - h_{Bt}.$$

This completes the proof of Proposition 4.

Note that when the factors are perfectly tracked by the returns, we have that η_{jt} is a zero vector and $y_{jt} = 0$ for $j = A, B$. Hence, the asymptotic variance in Proposition 4 reduces to that in Proposition 1 for models with traded factors.

This completes the proof of Proposition 1.

Lemma 3 and Proof of Lemma 1

LEMMA 3: *When the factors and returns are i.i.d. multivariate elliptically distributed with kurtosis parameter κ, the asymptotic variance of the difference in sample squared Sharpe ratios of two sets of mimicking portfolios, f_{At}^* and f_{Bt}^*, is given by*

$$\text{(A-28)} \qquad E[d_t^2] = E[h_{At}^2] + E[h_{Bt}^2] - 2E[h_{At}h_{Bt}],$$

with

$$\begin{aligned}
\text{(A-29)} \qquad E[h_{At}^2] &= \theta_A^2\left[4 + (2+3\kappa)\theta_A^2\right] + 4(1+\kappa)E[y_{At}^2]\left(\theta_R^2 - \theta_A^2\right), \\
\text{(A-30)} \qquad E[h_{Bt}^2] &= \theta_B^2\left[4 + (2+3\kappa)\theta_B^2\right] + 4(1+\kappa)E[y_{Bt}^2]\left(\theta_R^2 - \theta_B^2\right), \\
\text{(A-31)} \qquad E[h_{At}h_{Bt}] &= 2\rho\theta_A\theta_B\left[2 + (1+\kappa)\rho\theta_A\theta_B\right] + \kappa\theta_A^2\theta_B^2 \\
&\quad + 4(1+\kappa)E[y_{At}y_{Bt}](\theta_R^2 + \rho\theta_A\theta_B - \theta_A^2 - \theta_B^2),
\end{aligned}$$

where $\rho = \text{Corr}[u_{At}, u_{Bt}] = E[u_{At}u_{Bt}]/(\theta_A\theta_B)$ *is the correlation between the returns on the maximum Sharpe ratio portfolios of* f^*_{At} and f^*_{Bt}, $E[y^2_{At}] = \mu^{*\prime}_A V^{*-1}_A V_{f_A\cdot R} V^{*-1}_A \mu^*_A$, $E[y^2_{Bt}] = \mu^{*\prime}_B V^{*-1}_B V_{f_B\cdot R} V^{*-1}_B \mu^*_B$, $V_{f_A\cdot R} = V_{f_A} - V_{f_A R} V^{-1}_R V_{R f_A}$, $V_{f_B\cdot R} = V_{f_B} - V_{f_B R} V^{-1}_R V_{R f_B}$, and $E[y_{At}y_{Bt}] = \mu^{*\prime}_A V^{*-1}_A \text{Cov}[\eta_{At}, \eta'_{Bt}] V^{*-1}_B \mu^*_B$.

Proof of Lemma 3:

Since the $E[h^2_t]$ expressions for models A and B have already been derived in Lemma 2, we only need to compute $E[h_{At}h_{Bt}]$. It can be shown that

$$\text{(A-32)}\quad E[h_{At}h_{Bt}] = 4E[u_{At}u_{Bt}(1-y_{At})(1-y_{Bt})] - 2E[u_{At}u^2_{Bt}(1-y_{At})] + 4E[u_{At}y_{Bt}(1-y_{At})v_t] + 2\theta^2_B E[u_{At}(1-y_{At})] - 2E[u^2_{At}u_{Bt}(1-y_{Bt})] + E[u^2_{At}u^2_{Bt}] - 2E[u^2_{At}y_{Bt}v_t] - \theta^2_B E[u^2_{At}] + 4E[y_{At}u_{Bt}(1-y_{Bt})v_t] - 2E[y_{At}u^2_{Bt}v_t] + 4E[y_{At}y_{Bt}v^2_t] + 2\theta^2_B E[y_{At}v_t] + 2\theta^2_A E[u_{Bt}(1-y_{Bt})] - \theta^2_A E[u^2_{Bt}] + 2\theta^2_A E[y_{Bt}v_t] + \theta^2_A\theta^2_B.$$

Under the multivariate elliptical assumption on Y_t, we obtain

$$\text{(A-33)}\quad E[h_{At}h_{Bt}] = 4\rho\theta_A\theta_B + 4(1+\kappa)\rho\theta_A\theta_B E[y_{At}y_{Bt}] + 0 - 4(1+\kappa)E[y_{At}y_{Bt}]\theta^2_A + 0 + 0 + (1+\kappa)(\theta^2_A\theta^2_B + 2\rho^2\theta^2_A\theta^2_B) + 0 - \theta^2_A\theta^2_B - 4(1+\kappa)E[y_{At}y_{Bt}]\theta^2_B + 0 + 4(1+\kappa)E[y_{At}y_{Bt}]\theta^2_R + 0 + 0 - \theta^2_A\theta^2_B + 0 + \theta^2_A\theta^2_B.$$

After simplification, we have

$$\text{(A-34)}\quad E[h_{At}h_{Bt}] = 2\rho\theta_A\theta_B[2 + (1+\kappa)\rho\theta_A\theta_B] + \kappa\theta^2_A\theta^2_B + 4(1+\kappa)E[y_{At}y_{Bt}](\theta^2_R + \rho\theta_A\theta_B - \theta^2_A - \theta^2_B).$$

This completes the proof of Lemma 3.

When $y_{At} = y_{Bt} = 0$, we have

$$\text{(A-35)}\quad E[h^2_{At}] = \theta^2_A\left[4 + (2+3\kappa)\theta^2_A\right],$$

$$\text{(A-36)}\quad E[h^2_{Bt}] = \theta^2_B\left[4 + (2+3\kappa)\theta^2_B\right],$$

$$\text{(A-37)}\quad E[h_{At}h_{Bt}] = 2\rho\theta_A\theta_B[2 + (1+\kappa)\rho\theta_A\theta_B] + \kappa\theta^2_A\theta^2_B.$$

This completes the proof of Lemma 1.

Remarks and proof of Proposition 3:

There are cases in which $u_{At} = u_{Bt}$ and the normal approximations in Propositions 1 and 4 break down. This occurs when the models are nested. Let

$$\text{(A-38)} \qquad \mu_A^* = \begin{bmatrix} \mu_1^* \\ \mu_2^* \end{bmatrix}, \qquad \mu_B^* = \mu_1^*,$$

and

$$\text{(A-39)} \qquad V_A^* = \begin{bmatrix} V_{11}^* & V_{12}^* \\ V_{21}^* & V_{22}^* \end{bmatrix}, \qquad V_B^* = V_{11}^*.$$

We have

$$\begin{aligned} \text{(A-40)} \qquad u_{At} &= \mu_A^{*\prime} V_A^{*-1}(f_{At}^* - \mu_A^*) \\ &= \begin{bmatrix} \mu_1^* \\ \mu_2^* \end{bmatrix}' \begin{bmatrix} V_{11}^{*-1} + V_{11}^{*-1}V_{12}^*V_{22\cdot1}^{*-1}V_{21}^*V_{11}^{*-1} & -V_{11}^{*-1}V_{12}^*V_{22\cdot1}^{*-1} \\ -V_{22\cdot1}^{*-1}V_{21}^*V_{11}^{*-1} & V_{22\cdot1}^{*-1} \end{bmatrix} \begin{bmatrix} f_{1t}^* - \mu_1^* \\ f_{2t}^* - \mu_2^* \end{bmatrix} \\ &= [\mu_1^{*\prime}V_{11}^{*-1} - \alpha_{21}^{*\prime}V_{22\cdot1}^{*-1}V_{21}^*V_{11}^{*-1},\ \alpha_{21}^{*\prime}V_{22\cdot1}^{*-1}] \begin{bmatrix} f_{1t}^* - \mu_1^* \\ f_{2t}^* - \mu_2^* \end{bmatrix}, \end{aligned}$$

where $V_{22\cdot1}^* = V_{22}^* - V_{21}^*V_{11}^{*-1}V_{12}^*$ and $\alpha_{21}^* = \mu_2^* - V_{21}^*V_{11}^{*-1}\mu_1^*$. Note that $\alpha_{21}^* = 0_{K_2}$ implies

$$\text{(A-41)} \qquad u_{At} = \mu_1^{*\prime}V_{11}^{*-1}(f_{1t}^* - \mu_1^*) \equiv \mu_B^{*\prime}V_B^{*-1}(f_{Bt}^* - \mu_B^*) = u_{Bt},$$

and $y_{At} = y_{Bt}$. Conversely $u_{At} = u_{Bt}$ implies that $\alpha_{21}^* = 0_{K_2}$. Similarly, $\alpha_{21}^* = 0_{K_2}$ implies $\theta_A^2 = \theta_B^2 = \mu_1^{*\prime}V_{11}^{*-1}\mu_1^*$, and conversely $\theta_A^2 = \theta_B^2$ implies $\alpha_{21}^* = 0_{K_2}$. This suggests that for the nested-model case, we only need to test $H_0 : \alpha_{21}^* = 0_{K_2}$.[32]

Proof of Proposition 3: We first show that

$$\text{(A-42)} \qquad \sqrt{T}(\hat{\alpha}_{21}^* - \alpha_{21}^*) \overset{A}{\sim} N(0_{K_2}, V(\hat{\alpha}_{21}^*)).$$

Using the delta method, the asymptotic distribution of $\hat{\alpha}_{21}^*$ is given by

$$\text{(A-43)} \qquad \sqrt{T}(\hat{\alpha}_{21}^* - \alpha_{21}^*) \ \overset{A}{\sim}\ N\left(0_{K_2}, \left[\frac{\partial \alpha_{21}^*}{\partial \varphi'}\right] S_0 \left[\frac{\partial \alpha_{21}^*}{\partial \varphi'}\right]'\right).$$

[32]Note that for nested models we do not need to perform the normal test because $\alpha_{21}^* \neq 0_{K_2}$ implies that the squared Sharpe ratio of model A is larger than the squared Sharpe ratio of model B.

It is straightforward to obtain

$$\text{(A-44)} \qquad \frac{\partial \alpha_{21}^*}{\partial \mu_f'} = 0_{K_2\times K}, \qquad \frac{\partial \alpha_{21}^*}{\partial \mu_R'} = (V_{f_2R} - V_{21}^* V_{11}^{*-1} V_{f_1R}) V_R^{-1}.$$

The derivative of α_{21}^* with respect to vec(V) is given by

$$\begin{aligned}\text{(A-45)} \ \frac{\partial \alpha_{21}^*}{\partial \mathrm{vec}(V)'} &= \left[0_K',\ (\mu_R - V_{Rf_1} V_{11}^{*-1} \mu_1^*)' V_R^{-1}\right] \\ &\quad \otimes \left[-V_{21}^* V_{11}^{*-1},\ I_{K_2},\ (V_{21}^* V_{11}^{*-1} V_{f_1R} - V_{f_2R}) V_R^{-1}\right] \\ &\quad + \left[\mu_1^{*\prime} V_{11}^{*-1},\ 0_{K_2}',\ 0_N'\right] \otimes \left[0_{K_2\times K},\ (V_{21}^* V_{11}^{*-1} V_{f_1R} - V_{f_2R}) V_R^{-1}\right] \mathcal{K}_{N+K},\end{aligned}$$

where $\mathcal{K}_m$ is an $m^2 \times m^2$ commutation matrix defined as $\mathcal{K}_m \mathrm{vec}(X) = \mathrm{vec}(X')$ for an $m \times m$ matrix X. Using the expression for $\partial \alpha_{21}^*/\partial \varphi'$, we can simplify the asymptotic variance of $\hat{\alpha}_{21}^*$ to

$$\text{(A-46)} \qquad V(\hat{\alpha}_{21}^*) = \sum_{j=-\infty}^{\infty} E[q_t(\varphi) q_{t+j}(\varphi)'],$$

where

$$\begin{aligned}\text{(A-47)} \qquad q_t(\varphi) &= \frac{\partial \alpha_{21}^*}{\partial \varphi'} r_t(\varphi) \\ &= (f_{2t}^* - \mu_2^*) - V_{21}^* V_{11}^{*-1} (f_{1t}^* - \mu_1^*) \\ &\quad + V_{21}^* V_{11}^{*-1} (f_{1t}^* - \mu_1^*)(f_{1t} - \mu_1)' V_{11}^{*-1} \mu_1^* \\ &\quad - (f_{2t}^* - \mu_2^*)(f_{1t} - \mu_1)' V_{11}^{*-1} \mu_1^* \\ &\quad + [-V_{21}^* V_{11}^{*-1} [(f_{1t} - \mu_1) - (f_{1t}^* - \mu_1^*)] \\ &\quad + [(f_{2t} - \mu_2) - (f_{2t}^* - \mu_2^*)]](v_t - u_{1t}) \\ &= \xi_t (1 - y_{1t}) + w_t (v_t - u_{1t}).\end{aligned}$$

Let $\hat{V}(\hat{\alpha}_{21}^*)$ be a consistent estimator of $V(\hat{\alpha}_{21}^*)$. Then, under the null hypothesis,

$$\text{(A-48)} \qquad T \hat{\alpha}_{21}^{*\prime} \hat{V}(\hat{\alpha}_{21}^*)^{-1} \hat{\alpha}_{21}^* \overset{A}{\sim} \chi_{K_2}^2,$$

and this statistic can be used to test $H_0 : \theta_A^2 = \theta_B^2$. This completes the proof of Proposition 3.

An alternative test of $\alpha_{21}^* = 0_{K_2}$ can be obtained by establishing a connection between the mimicking portfolio framework and the following GLS two-pass cross-sectional regression framework. Consider the second-pass projection with covariances instead of betas and assume that the zero-beta rate is zero. Then, the "price of covariance risk" parameters are given by

$$\lambda = (V_{fR}V_R^{-1}V_{Rf})^{-1}V_{fR}V_R^{-1}\mu_R. \tag{A-49}$$

It is immediately evident that the λ vector for model A is given by

$$\lambda_A = \begin{bmatrix} \lambda_{A,1} \\ \lambda_{A,2} \end{bmatrix} = V_A^{*-1}\mu_A^* = \begin{bmatrix} V_{11}^{*-1}\mu_1^* - V_{11}^{*-1}V_{12}^*V_{22\cdot 1}^{*-1}\alpha_{21}^* \\ V_{22\cdot 1}^{*-1}\alpha_{21}^* \end{bmatrix}. \tag{A-50}$$

It follows that $\alpha_{21}^* = 0_{K_2}$ if and only if $\lambda_{A,2} = 0_{K_2}$. Therefore, nested model comparison can also be conducted by testing whether $\lambda_{A,2}$ is a zero vector. If we choose this approach, then we can use the results in Proposition 21 and Lemma 9 of the Online Appendix of Kan, Robotti, and Shanken (2013) to implement the test.

Remarks and Proposition 5:

The normal approximations in Propositions 1 and 4 can break down also in the non-nested model case. Without loss of generality, assume model A has mimicking portfolios $f_{At}^* = [f_{1t}^{*\prime},\ f_{2t}^{*\prime}]'$ and model B has mimicking portfolios $f_{Bt}^* = [f_{1t}^{*\prime},\ f_{3t}^{*\prime}]'$, where f_{3t}^* is a K_3-vector. Consider a model C which has only the factors $f_{Ct}^* = f_{1t}^*$ (the common mimicking portfolios). Let $\mu_1^* = E[f_{1t}^*]$, $\mu_2^* = E[f_{2t}^*]$, $\mu_3^* = E[f_{3t}^*]$, $V_{11}^* = \text{Var}(f_{1t}^*)$, $V_{12}^* = \text{Cov}(f_{1t}^*, f_{2t}^{*\prime})$, $V_{21}^* = V_{12}^{*\prime}$, $V_{22}^* = \text{Var}(f_{2t}^*)$, $V_{13}^* = \text{Cov}(f_{1t}^*, f_{3t}^{*\prime})$, $V_{31}^* = V_{13}^{*\prime}$, $V_{33}^* = \text{Var}(f_{3t}^*)$, and define

$$\mu_A^* = \begin{bmatrix} \mu_1^* \\ \mu_2^* \end{bmatrix}, \qquad \mu_B^* = \begin{bmatrix} \mu_1^* \\ \mu_3^* \end{bmatrix}, \qquad \mu_C^* = \mu_1^*. \tag{A-51}$$

Similarly, let

$$V_A^* = \begin{bmatrix} V_{11}^* & V_{12}^* \\ V_{21}^* & V_{22}^* \end{bmatrix}, \qquad V_B^* = \begin{bmatrix} V_{11}^* & V_{13}^* \\ V_{31}^* & V_{33}^* \end{bmatrix}, \qquad V_C^* = V_{11}^*. \tag{A-52}$$

Define $u_{At} = \mu_A^{*\prime}V_A^{*-1}(f_{At}^* - \mu_A^*)$, $u_{Bt} = \mu_B^{*\prime}V_B^{*-1}(f_{Bt}^* - \mu_B^*)$, and $u_{Ct} = \mu_C^{*\prime}V_C^{*-1}(f_{Ct}^* - \mu_C^*) \equiv \mu_1^{*\prime}V_{11}^{*-1}(f_{1t}^* - \mu_1^*)$. Using the same proof as for Proposition 3, we have

$$u_{At} = u_{Ct} = u_{Bt} \tag{A-53}$$

if and only if $\alpha_{21}^* = \mu_2^* - V_{21}^*V_{11}^{*-1}\mu_1^* = 0_{K_2}$ and $\alpha_{31}^* = \mu_3^* - V_{31}^*V_{11}^{*-1}\mu_1^* = 0_{K_3}$. Note that when $u_{At} = u_{Bt}$, we also have $y_{At} = y_{Bt}$, $\theta_A^2 = \theta_B^2$, and the normality result in Proposition 4 breaks down (similarly, in the traded-factor case, when $\alpha_{21} = 0_{K_2}$ and $\alpha_{31} = 0_{K_3}$, the normality result in Proposition 1 breaks down). In the following proposition, we show how to jointly test $\alpha^*_{21} = 0_{K_2}$ and $\alpha^*_{31} = 0_{K_3}$. Let $\psi = [\alpha^{*\prime}_{21},\ \alpha^{*\prime}_{31}]'$ and $\hat{\psi} = [\hat{\alpha}^{*\prime}_{21},\ \hat{\alpha}^{*\prime}_{31}]'$.

PROPOSITION 5: *Under the null hypothesis $H_0 : \psi = 0_{K_2+K_3}$,*

$$T\hat{\psi}'\hat{V}(\hat{\psi})^{-1}\hat{\psi} \overset{A}{\sim} \chi^2_{K_2+K_3}, \tag{A-54}$$

where $\hat{V}(\hat{\psi})$ is a consistent estimator of

$$V(\hat{\psi}) = \sum_{j=-\infty}^{\infty} E[\tilde{q}_t\tilde{q}'_{t+j}], \tag{A-55}$$

and $\tilde{q}_t$ is a $(K_2 + K_3)$-vector obtained by stacking up the q_t's for models A and B, respectively (the q_t for model A is given in Proposition 3 and the q_t for model B is similarly defined).

Proof of Proposition 5:

The proof of this result relies on the proof of Proposition 3 for the determination of the q_t's for models A and B. Let $\hat{V}(\hat{\psi})$ be a consistent estimator of $V(\hat{\psi})$. Then, under the null hypothesis $H_0 : \psi = 0_{K_2+K_3}$,

$$T\hat{\psi}'\hat{V}(\hat{\psi})^{-1}\hat{\psi} \overset{A}{\sim} \chi^2_{K_2+K_3}, \tag{A-56}$$

and this statistic can be used to test $H_0 : \theta^2_A = \theta^2_B$.

This completes the proof of Proposition 5.

In the traded-factor case, we can simply use the basic alpha-based test for the purpose of testing $\alpha_{21} = 0_{K_2}$ and $\alpha_{31} = 0_{K_3}$, since in this case we have no generated regressors. An alternative test of $\alpha^*_{21} = 0_{K_2}$ and $\alpha^*_{31} = 0_{K_3}$ can be obtained by focusing on the GLS two-pass cross-sectional regression framework. The λ vector for model A is given by

$$\lambda_A = \begin{bmatrix} \lambda_{A,1} \\ \lambda_{A,2} \end{bmatrix} = V_A^{*-1}\mu^*_A = \begin{bmatrix} V_{11}^{*-1}\mu^*_1 - V_{11}^{*-1}V^*_{12}V_{22\cdot 1}^{*-1}\alpha^*_{21} \\ V_{22\cdot 1}^{*-1}\alpha^*_{21} \end{bmatrix}. \tag{A-57}$$

It follows that $\alpha^*_{21} = 0_{K_2}$ if and only if $\lambda_{A,2} = 0_{K_2}$. Similarly, the λ vector for model B is given by

$$\lambda_B = \begin{bmatrix} \lambda_{B,1} \\ \lambda_{B,3} \end{bmatrix} = V_B^{*-1}\mu^*_B = \begin{bmatrix} V_{11}^{*-1}\mu^*_1 - V_{11}^{*-1}V^*_{13}V_{33\cdot 1}^{*-1}\alpha^*_{31} \\ V_{33\cdot 1}^{*-1}\alpha^*_{31} \end{bmatrix}, \tag{A-58}$$

where $V^*_{33\cdot 1} = V^*_{33} - V^*_{31} V^{*-1}_{11} V^*_{13}$. It follows that $\alpha^*_{31} = 0_{K_3}$ if and only if $\lambda_{B,3} = 0_{K_3}$. Therefore, non-nested model comparison can also be conducted by testing $\lambda_{A,2} = 0_{K_2}$ and $\lambda_{B,3} = 0_{K_3}$. If we choose this approach, then we can use the results in Proposition 21 and Lemma 10 of the Online Appendix of Kan, Robotti, and Shanken (2013) to implement the test.

In summary, for the non-nested model case with overlapping mimicking portfolios, we first need to jointly test $\alpha^*_{21} = 0_{K_2}$ and $\alpha^*_{31} = 0_{K_3}$. If we reject the null, we need to perform the normal test. Therefore, for non-nested models with overlapping mimicking portfolios, the test of $H_0 : \theta^2_A = \theta^2_B$ is a sequential test. For the non-nested model case with non-overlapping mimicking portfolios, we can simply perform the normal test in order to test $H_0 : \theta^2_A = \theta^2_B$.

Simulation designs for models with traded factors only

In all simulations, we set the true variance-covariance matrix of the factor returns equal to its sample estimate from the data. In order to impose the various null hypotheses and investigate the size properties of the tests, we constrain the means of the factor returns as described below.

Nested models

Define $\mu_1 = E[f_{1t}]$, $\mu_2 = E[f_{2t}]$, $V_{11} = \mathrm{Var}(f_{1t})$, and $V_{21} = \mathrm{Cov}(f_{2t}, f'_{1t})$. To investigate the size properties of the alpha-based test for pairwise nested-model comparison, we impose the null hypothesis $H_0 : \alpha_{21} = 0_{K_2}$ which can be rewritten as

(A-59) $$\mu_2 = V_{21} V^{-1}_{11} \mu_1.$$

Therefore, in the simulations, we set $\mu_1 = \hat{\mu}_1$ and $\mu_2 = \hat{V}_{21} \hat{V}^{-1}_{11} \hat{\mu}_1$, where $\hat{\mu}_1$, $\hat{V}_{21}$, and $\hat{V}_{11}$ are the sample counterparts of μ_1, V_{21}, and V_{11}, respectively. To investigate the power properties of the test, we simply set $\mu_1 = \hat{\mu}_1$ and $\mu_2 = \hat{\mu}_2$, where $\hat{\mu}_2$ is the sample counterpart of μ_2.

Non-nested models

For pairwise non-nested model comparison with overlapping factors, we first need to test whether $\alpha_{21} = 0_{K_2}$ and $\alpha_{31} = 0_{K_3}$. Define $\mu_3 = E[f_{3t}]$ and $V_{31} = \mathrm{Cov}[f_{3t}, f'_{1t}]$. In order to impose the null hypothesis and examine the size properties of the alpha-based test, we let $\mu_1 = \hat{\mu}_1$, $\mu_2 = \hat{V}_{21} \hat{V}^{-1}_{11} \hat{\mu}_1$, and $\mu_3 = \hat{V}_{31} \hat{V}^{-1}_{11} \hat{\mu}_1$, where $\hat{V}_{31}$ is the sample counterpart of V_{31}. To examine power, we set $\mu_1 = \hat{\mu}_1$,

$\mu_2 = \hat{\mu}_2$, and $\mu_3 = \hat{\mu}_3$, where $\hat{\mu}_3$ is the sample counterpart of μ_3.

If we reject $\alpha_{21} = 0_{K_2}$ and $\alpha_{31} = 0_{K_3}$, then we need to implement the normal test described in Section II.B. To impose $\theta_A^2 = \theta_B^2$ when $u_{At} \neq u_{Bt}$ is more complicated. Note that

$$\theta_A^2 = \mu_1' V_{11}^{-1} \mu_1 + \alpha_{21}' V_{22\cdot 1}^{-1} \alpha_{21}, \tag{A-60}$$

where $V_{22\cdot 1} = V_{22} - V_{21} V_{11}^{-1} V_{12}$ and $\alpha_{21} = \mu_2 - V_{21} V_{11}^{-1} \mu_1$. Similarly, we have

$$\theta_B^2 = \mu_1' V_{11}^{-1} \mu_1 + \alpha_{31}' V_{33\cdot 1}^{-1} \alpha_{31}, \tag{A-61}$$

where $V_{33\cdot 1} = V_{33} - V_{31} V_{11}^{-1} V_{13}$ and $\alpha_{31} = \mu_3 - V_{31} V_{11}^{-1} \mu_1$. Therefore, $\theta_A^2 = \theta_B^2$ if and only if

$$\alpha_{21}' V_{22\cdot 1}^{-1} \alpha_{21} = \alpha_{31}' V_{33\cdot 1}^{-1} \alpha_{31}. \tag{A-62}$$

Set $\mu_1 = \hat{\mu}_1$, $\mu_2 = \hat{\mu}_2$ and $\hat{\alpha}_{21} = \hat{\mu}_2 - \hat{V}_{21} \hat{V}_{11}^{-1} \hat{\mu}_1$. then we need to choose α_{31} such that $\alpha_{31}' V_{33\cdot 1}^{-1} \alpha_{31} = c$, where $c \equiv \hat{\alpha}_{21}' \hat{V}_{22\cdot 1} \hat{\alpha}_{21}$. There are many solutions to this equation, but we can set up the following minimization problem:

$$\min_{\alpha_{31}} (\alpha_{31} - \hat{\alpha}_{31})'(\alpha_{31} - \hat{\alpha}_{31}) \quad \text{s.t.} \quad \alpha_{31}' V_{33\cdot 1}^{-1} \alpha_{31} = c, \tag{A-63}$$

where $\hat{\alpha}_{31} = \hat{\mu}_3 - \hat{V}_{31} \hat{V}_{11}^{-1} \hat{\mu}_1$. This way we set α_{31} as close as possible to $\hat{\alpha}_{31}$. Once the minimizer α_{31}^* is obtained, we can recover μ_3 as

$$\mu_3 = \alpha_{31}^* + \hat{V}_{31} \hat{V}_{11}^{-1} \hat{\mu}_1. \tag{A-64}$$

So, in summary, to analyze the size properties of the normal test, we can set $\mu_1 = \hat{\mu}_1$, $\mu_2 = \hat{\mu}_2$, and $\mu_3 = \alpha_{31}^* + \hat{V}_{31} \hat{V}_{11}^{-1} \hat{\mu}_1$. To analyze the power properties of the normal test, we set $\mu_1 = \hat{\mu}_1$, $\mu_2 = \hat{\mu}_2$, and $\mu_3 = \hat{\mu}_3$. A similar simulation design can be implemented to investigate the size and power properties of the normal test when the two models do not have common factors.

To evaluate the size properties of the multiple model comparison test described in Section IV, we consider the case in which all models have the same θ^2 value, so as to maximize the likelihood of rejection under the null. We now explain how we can set the means of the factors such that the squared Sharpe ratio for each single-factor model is the same. Suppose that model 1 is the benchmark model and that the number of models is equal to p. In the single-factor setting, equality of squared Sharpe ratios requires that

$$\theta_1^2 \equiv c = \frac{\mu_i^2}{\sigma_i^2} \tag{A-65}$$

for $i = 2, \ldots, p$, where μ_i and σ_i^2 are the mean and variance of factor i, respectively. Now set $\mu_1 = \hat{\mu}_1$, $\sigma_1^2 = \hat{\sigma}_1^2$, $c = \frac{\hat{\mu}_1^2}{\hat{\sigma}_1^2}$, and $\sigma_i^2 = \hat{\sigma}_i^2$ for $i = 2, \ldots, p$, where $\hat{\sigma}_i^2$ is the sample counterpart of σ_i^2. In order to make the squared Sharpe ratios of the various models identical, we can set

$$\mu_i \quad = \quad \sqrt{c}\hat{\sigma}_i, \tag{A-66}$$

for $i = 2, \ldots, p$. This guarantees that we maximize the likelihood of rejection under the null. To examine the power properties of the multiple model comparison test, we can simply set the means of the factors equal to their sample estimates from the data.

Simulation designs for models with mimicking portfolios

In all simulations, we set the true variance-covariance matrix of the factors and basis-asset returns equal to its sample estimate from the data. In order to impose the various null hypotheses and investigate the size properties of the tests, we constrain the means of the factor and basis-asset returns as described below.

Nested models

To impose the null $\alpha_{21}^* = 0_{K_2}$ and study the size properties of the chi-squared test in Proposition 3, we set $\mu_1 = \hat{a} + \hat{\mu}_1^*$ and $\mu_2 = \hat{b} + \hat{\mu}_2^* = \hat{b} + \hat{V}_{21}^* \hat{V}_{11}^{*-1} \hat{\mu}_1^*$, where $\hat{a}$ and $\hat{b}$ are the estimated intercepts from regressing f_{1t} and f_{2t} on the augmented span of R. The constraint $\alpha_{21}^* = 0_{K_2}$ also imposes some restrictions on μ_R. Given $\hat{\mu}^* = [\hat{\mu}_1^{*\prime},\ \hat{\mu}_1^{*\prime} \hat{V}_{11}^{*-1} \hat{V}_{12}^*]'$, we can solve the following constrained minimization problem to set μ_R:

$$\min_{\mu_R} (\mu_R - \hat{\mu}_R)' \hat{V}_R^{-1} (\mu_R - \hat{\mu}_R) \quad \text{s.t.} \quad \hat{\mu}^* = \hat{A} \mu_R, \tag{A-67}$$

where $\hat{A} = \hat{V}_{fR} \hat{V}_R^{-1}$. This way we set μ_R as close as possible to $\hat{\mu}_R$ in a GLS sense. Denote by μ_R° the minimizer of equation (A-67). Then we can set $\mu_R = \mu_R^\circ$ and generate factors and returns under the constrained mean vector $[\mu_1',\ \mu_2',\ \mu_R^{\circ\prime}]'$. To analyze the power properties of the test, we can simply leave the mean vector unrestricted, that is, set $\mu_1 = \hat{\mu}_1$, $\mu_2 = \hat{\mu}_2$, and $\mu_R = \hat{\mu}_R$.

Non-nested models

In the presence of overlapping mimicking portfolios, we first need to test whether $\alpha_{21}^* = 0_{K_2}$ and $\alpha_{31}^* = 0_{K_3}$ using Proposition 5 in this Appendix. In order to impose this null and examine the size properties of our chi-squared test, we set $\mu_1 = \hat{a} + \hat{\mu}_1^*$, $\mu_2 = \hat{b} + \hat{V}_{21}^* \hat{V}_{11}^{*-1} \hat{\mu}_1^*$, and $\mu_3 = \hat{c} + \hat{V}_{31}^* \hat{V}_{11}^{*-1} \hat{\mu}_1^*$, where $\hat{a}$, $\hat{b}$, and $\hat{c}$ are the estimated intercepts from regressing f_{1t}, f_{2t}, and f_{3t} on the augmented span of the basis-asset returns. Given $\hat{\mu}^* = [\hat{\mu}_1^{*\prime},\ \hat{\mu}_1^{*\prime} \hat{V}_{11}^{*-1} \hat{V}_{12}^*,\ \hat{\mu}_1^{*\prime} \hat{V}_{11}^{*-1} \hat{V}_{13}^*]'$, we can solve the following constrained minimization problem to constrain the μ_R vector:

$$\min_{\mu_R} (\mu_R - \hat{\mu}_R)' \hat{V}_R^{-1} (\mu_R - \hat{\mu}_R) \quad \text{s.t.} \quad \hat{\mu}^* = \hat{A} \mu_R, \tag{A-68}$$

where $\hat{A} = \hat{V}_{fR} \hat{V}_R^{-1}$. Denote by μ_R° the minimizer of equation (A-68). Then we can set $\mu_R = \mu_R^{\circ}$ and generate factor and basis-asset returns using the mean vector $[\mu_1',\ \mu_2',\ \mu_3',\ \mu_R^{\circ\prime}]'$. To examine power, we set the means of the factors and the returns equal to their sample estimates from the data.

If we reject $\alpha_{21}^* = 0_{K_2}$ and $\alpha_{31}^* = 0_{K_3}$, then we need to implement the normal test in Proposition 4. To study the size properties of the normal test, we need to impose $\theta_A^2 = \theta_B^2$ when $u_{At} \neq u_{Bt}$. Note that

$$\theta_A^2 = \mu_1^{*\prime} V_{11}^{*-1} \mu_1^* + \alpha_{21}^{*\prime} V_{22\cdot 1}^{*-1} \alpha_{21}^*, \tag{A-69}$$

where $V_{22\cdot 1}^* = V_{22}^* - V_{21}^* V_{11}^{*-1} V_{12}^*$ and $\alpha_{21}^* = \mu_2^* - V_{21}^* V_{11}^{*-1} \mu_1^*$. Similarly, we have

$$\theta_B^2 = \mu_1^{*\prime} V_{11}^{*-1} \mu_1^* + \alpha_{31}^{*\prime} V_{33\cdot 1}^{*-1} \alpha_{31}^*, \tag{A-70}$$

where $V_{33\cdot 1}^* = V_{33}^* - V_{31}^* V_{11}^{*-1} V_{13}^*$ and $\alpha_{31}^* = \mu_3^* - V_{31}^* V_{11}^{*-1} \mu_1^*$. Therefore, $\theta_A^2 = \theta_B^2$ if and only if

$$\alpha_{21}^{*\prime} V_{22\cdot 1}^{*-1} \alpha_{21}^* = \alpha_{31}^{*\prime} V_{33\cdot 1}^{*-1} \alpha_{31}^*. \tag{A-71}$$

Then we can write equation (A-71) as a function of μ_R:

$$\mu_R' \hat{E} \mu_R = 0, \tag{A-72}$$

where $\hat{E} = \hat{C}'\hat{V}_{22\cdot 1}^{*-1}\hat{C} - \hat{D}'\hat{V}_{33\cdot 1}^{*-1}\hat{D}$, $\hat{C} = \hat{V}_{f_2R}\hat{V}_R^{-1} - \hat{V}_{21}^*\hat{V}_{11}^{*-1}\hat{V}_{f_1R}\hat{V}_R^{-1}$, and $\hat{D} = \hat{V}_{f_3R}\hat{V}_R^{-1} - \hat{V}_{31}^*\hat{V}_{11}^{*-1}\hat{V}_{f_1R}\hat{V}_R^{-1}$. There are many solutions to equation (A-72), but we can set up the following minimization problem:

(A-73)
$$\min_{\mu_R}(\mu_R - \hat{\mu}_R)'\hat{V}_R^{-1}(\mu_R - \hat{\mu}_R) \quad \text{s.t.} \quad \mu_R'\hat{E}\mu_R = 0.$$

This way we set μ_R as close as possible to $\hat{\mu}_R$ in a GLS sense. Denote by μ_R° the minimizer of this constrained optimization problem. Then, we set $\mu_R = \mu_R^\circ$. When R contains the set of traded factors (as is the case in our empirical work and simulation experiments), we can set the means of the nontraded factors equal to their sample estimates from the data. Since the results are independent of the means of the nontraded factors, we set the means equal to their sample estimates when R does not contain the set of traded factors. To analyze power, we set the means of the factors and the returns equal to their sample estimates from the data.

Similar to the traded-factor case, to evaluate the size properties of the multiple model comparison test with mimicking portfolios, we consider the situation in which all models have the same θ^2 value. The squared Sharpe ratio of the single-factor model with mimicking portfolio i is given by

(A-74)
$$\theta_i^2 = \frac{(V_{f_iR}V_R^{-1}\mu_R)^2}{(V_{f_iR}V_R^{-1}V_{Rf_i})}.$$

Let

(A-75)
$$V_{f_iR}^n = \frac{V_{f_iR}}{(V_{f_iR}V_R^{-1}V_{Rf_i})^{\frac{1}{2}}}.$$

Then we can write

(A-76)
$$\theta_i^2 = (V_{f_iR}^nV_R^{-1}\mu_R)^2.$$

To ensure that all models have the same θ^2, a sufficient condition is

(A-77)
$$V_{f_iR}^nV_R^{-1}\mu_R = c,$$

where c is a constant. Let $V_{fR}^n = [V_{f_1R}^n, \ldots, V_{f_KR}^n]$. We have

(A-78)
$$V_{fR}^nV_R^{-1}\mu_R = c1_K.$$

In order to constrain μ_R, we consider the following minimization problem:

$$\min_{\mu_R}(\mu_R - \hat{\mu}_R)'\hat{V}_R^{-1}(\mu_R - \hat{\mu}_R) \quad \text{s.t.} \quad \hat{V}_{fR}^{n}\hat{V}_R^{-1}\mu_R = \hat{c}1_K, \tag{A-79}$$

where $\hat{c} = \hat{V}_{f_iR}^{n}\hat{V}_R^{-1}\hat{\mu}_R$, with f_i being the single factor of model i. We choose the market factor as factor i. Denote by μ_R° the minimizer of this constrained optimization problem. Then we set $\mu_R = \mu_R^{\circ}$.[33] To examine the power properties of the test, we set $\mu_R = \hat{\mu}_R$ and $\mu_f = \hat{\mu}_f$, so that the population squared Sharpe ratio of each model is set equal to its sample θ^2.

[33]Since the results are independent of the choice of the mean of the factors, we set the means equal to their sample estimates.

References

Asness, C., and A. Frazzini. "The Devil in HML's Details." *Journal of Portfolio Management*, 39 (2013), 49–68.

Ball, R.; J. Gerakos; J. T. Linnainmaa; and V. Nikolaev. "Accruals, Cash Flows, and Operating Profitability in the Cross Section of Stock Returns." *Journal of Financial Economics*, 121 (2016), 28–45.

Balduzzi, P., and C. Robotti. "Mimicking Portfolios, Economic Risk Premia, and Tests of Multi-Beta Models." *Journal of Business and Economic Statistics*, 26 (2008), 354–368.

Barillas, F., and J. Shanken. "Which Alpha?" *Review of Financial Studies*, 30 (2017), 1316–1338.

Barillas, F., and J. Shanken. "Comparing Asset Pricing Models." *Journal of Finance*, 73 (2018), 715–754.

Bentler, P. M., and M. Berkane. "Greatest Lower Bound to the Elliptical Theory Kurtosis Parameter." *Biometrika*, 73 (1986), 240–241.

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

Carhart, M. M. "On Persistence in Mutual Fund Performance." *Journal of Finance*, 52 (1997), 57–82.

Christie, S. "Is the Sharpe Ratio Useful in Asset Allocation?" Working Paper, Macquarie University (2005).

Cragg, J. G., and S. G. Donald. "Inferring the Rank of a Matrix." *Journal of Econometrics*, 76 (1997), 223–250.

Fama, E. F., and K. R. French. "Common Risk Factors in the Returns on Stocks and Bonds." *Journal of Financial Economics*, 33 (1993), 3–56.

Fama, E. F., and K. R. French. "A Five-Factor Asset Pricing Model." *Journal of Financial Economics*, 116 (2015), 1–22.

Fama, E. F., and K. R. French. "Choosing Factors." *Journal of Financial Economics*, 128 (2018), 234–252.

Frazzini, A., and L. Pedersen. "Betting against Beta." *Journal of Financial Economics*, 111 (2014), 1–25.

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

Hansen, L. P., and R. Jagannathan. "Assessing Specification Errors in Stochastic Discount Factor Models." *Journal of Finance*, 52 (1997), 557–590.

Harvey, C.; Y. Liu; and H. Zhu. "... and the Cross-Section of Expected Returns." *Review of Financial Studies*, 29 (2016), 5–68.

He, Z.; B. Kelly; and A. Manela. "Intermediary Asset Pricing: New Evidence from Many Asset Classes." *Journal of Financial Economics*, 126 (2017), 1–35.

Hou, K.; C. Xue; and L. Zhang. "Digesting Anomalies: An Investment Approach." *Review of Financial Studies*, 28 (2015), 650–705.

Huberman, G.; S. Kandel; and R. F. Stambaugh. "Mimicking Portfolios and Exact Arbitrage Pricing." *Journal of Finance*, 42 (1987), 1–9.

Jegadeesh, N., and S. Titman. "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency." *Journal of Finance*, 48 (1993), 65–91.

Jobson, J. D., and B. M. Korkie. "Estimation of Markowitz Efficient Portfolios." *Journal of the American Statistical Association*, 75 (1980), 544–554.

Jobson, J. D., and B. M. Korkie. "Performance Hypothesis Testing with the Sharpe and Treynor Measures." *Journal of Finance*, 36 (1981), 889–908.

Jobson, J. D., and B. M. Korkie. "Potential Performance and Tests of Portfolio Efficiency." *Journal of Financial Economics*, 10 (1982), 433–466.

Kan, R., and C. Robotti. "Specification Tests of Asset Pricing Models Using Excess Returns." *Journal of Empirical Finance*, 15 (2008), 816–838.

Kan, R.; C. Robotti; and J. Shanken. "Pricing Model Performance and the Two-Pass Cross-Sectional Regression Methodology." *Journal of Finance*, 68 (2013), 2617–2649.

Kan, R., and C. Zhang. "Two-Pass Tests of Asset Pricing Models with Useless Factors. *Journal of Finance*, 54 (1999), 203–235.

Kleibergen, F., and R. Paap. "Generalized Reduced Rank Tests Using the Singular Value Decomposition." *Journal of Econometrics*, 133 (2006), 97–126.

Lewellen, J. W.; S. Nagel; and J. Shanken. "A Skeptical Appraisal of Asset Pricing Tests." *Journal of Financial Economics*, 96 (2010), 175–194.

Lintner, J. "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets." *Review of Economics and Statistics*, 47 (1965), 13–37.

Long, J. B. "Stock Prices, Inflation, and the Term Structure of Interest Rates." *Journal of Financial Economics*, 1 (1974), 131–170.

Maller, R. A.; R. B. Durand; and H. Jafarpour. "Optimal Portfolio Choice Using the Maximum Sharpe Ratio." *Journal of Risk*, 12 (2010), 49–73.

Maller, R. A.; S. Roberts; and R. Tourky. "The Large-Sample Distribution of the Maximum Sharpe Ratio with and without Short Sales." *Journal of Econometrics*, 194 (2016), 138–152.

Maruyama, Y., and T. Seo. "Estimation of Moment Parameter in Elliptical Distributions." *Journal of the Japan Statistical Society*, 33 (2003), 215–229.

Memmel, C. "Performance Hypothesis Testing with the Sharpe Ratio." *Finance Letters*, 1 (2003), 21–23.

Merton, R. C. "An Intertemporal Capital Asset Pricing Model." *Econometrica*, 41 (1973), 867–887.

Newey, W. K., and K. D. West. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica*, 55 (1987), 703–708.

Novy-Marx, R. "The Other Side of Value: The Gross Profitability Premium." *Journal of Financial Economics*, 108 (2013), 1–28.

Opdyke, J. D. "Comparing Sharpe Ratios: So where Are the p-values?" *Journal of Asset Management*, 8 (2007), 308–336.

Pástor, L., and R. F. Stambaugh. "Liquidity Risk and Expected Stock Returns." *Journal of Political Economy*, 111 (2003), 642–685.

Robin, J. M., and R. J. Smith. "Tests of Rank." *Econometric Theory*, 16 (2000), 151–175.

Rubinstein, M. "The Valuation of Uncertain Income Streams and the Pricing of Options." *Bell Journal of Economics*, 7 (1976), 407–425.

Shanken, J. "Intertemporal Asset Pricing: An Empirical Investigation." *Journal of Econometrics*, 45 (1990), 99–120.

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

Sharpe, W. F. "Mutual Fund Performance." *Journal of Business*, 39 (1966), 119–138.

Stambaugh, R. F., and Y. Yuan. "Mispricing Factors." *Review of Financial Studies*, 30 (2017), 1270–1315.

Treynor, J. L., and F. Black. "How to Use Security Analysis to Improve Portfolio Selection." *Journal of Business*, 46 (1973), 66–86.

Wolak, F. A. "An Exact Test for Multiple Inequality and Equality Constraints in the Linear Regression Model." *Journal of the American Statistical Association*, 82 (1987), 782–793.

Wolak, F. A. "Testing Inequality Constraints in Linear Econometric Models." *Journal of Econometrics*, 41 (1989), 205–235.

TABLE 1

Summary Statistics for Monthly Factor Returns

In Table 1, we report the sample summary statistics for the traded factors. The sample period for our data is January 1972 to December 2015 (528 observations). *MKT* is the difference between the value-weighted market return and the one-month U.S. Treasury bill rate. *SMB* and *HML* are the small minus big size factor and high minus low book-to-market value factor of Fama and French (1993). *CMA* is the conservative minus aggressive investment factor of Fama and French (2015). *RMWCP* is the robust minus weak cash profitability factor of Fama and French (2018). *ME*, *IA*, and *ROE* are the size, investment, and profitability factors in Hou, Xue, and Zhang (2015). *UMD* is the up-minus-down momentum factor. HML^m is the more timely value factor from Asness and Frazzini (2013). *BAB* is the betting-against-beta factor in Frazzini and Pedersen (2014). *SMBSY*, *MGMT*, and *PERF* are the size and the two anomaly factors in Stambaugh and Yuan (2017). *FIRFT* is the financial intermediary capital risk factor in He, Kelly, and Manela (2017).

Panel A: Means, Standard Deviations, and t-statistics

	Mean	Standard Deviation	t-statistic
MKT	0.53%	4.55%	2.66
SMB	0.21%	3.11%	1.51
HML	0.35%	2.98%	2.70
CMA	0.33%	1.97%	3.89
RMWCP	0.40%	1.39%	6.67
ME	0.27%	3.11%	2.01
IA	0.41%	1.85%	5.09
ROE	0.56%	2.58%	4.99
UMD	0.72%	4.42%	3.73
HML^m	0.34%	3.57%	2.21
BAB	0.90%	3.39%	6.08
SMBSY	0.38%	2.81%	3.15
MGMT	0.64%	2.81%	5.23
PERF	0.68%	3.84%	4.06
FIRFT	0.61%	6.77%	2.06

Panel B: Correlations

	SMB	*HML*	*CMA*	*RMWCP*	*ME*	*IA*	*ROE*	*UMD*	HML^m	*BAB*	*SMBSY*	*MGMT*	*PERF*	*FIRFT*
MKT	0.241	−0.316	−0.389	−0.277	0.236	−0.362	−0.191	−0.142	−0.115	−0.099	0.213	−0.524	−0.260	0.819
SMB		−0.129	−0.050	−0.312	0.973	−0.165	−0.405	−0.017	−0.018	−0.036	0.937	−0.320	−0.134	0.120
HML			0.700	−0.201	−0.067	0.688	−0.082	−0.168	0.767	0.388	−0.056	0.716	−0.267	−0.052
CMA				−0.062	−0.011	0.904	−0.063	0.019	0.483	0.307	0.008	0.766	−0.047	−0.222
RMWCP					−0.337	−0.052	0.492	0.297	−0.372	−0.001	−0.246	0.077	0.622	−0.417
ME						−0.117	−0.316	0.006	0.004	0.012	0.931	−0.279	−0.125	0.164
IA							0.059	0.033	0.479	0.333	−0.088	0.758	−0.054	−0.164
ROE								0.495	−0.437	0.274	−0.292	0.093	0.631	−0.198
UMD									−0.654	0.191	0.031	0.048	0.716	−0.270
HML^m										0.111	−0.010	0.482	−0.635	0.140
BAB											0.040	0.318	0.136	−0.007
SMBSY												−0.242	−0.065	0.119
MGMT													0.013	−0.289
PERF														−0.467

TABLE 2

Tests of Equality of Squared Sharpe Ratios

In Table 2, we present pairwise tests of equality of the squared Sharpe ratios of the eight asset-pricing models. The models include the He, Kelly, and Manela (2017) two-factor model (MKT+FIRFT), the betting-against-beta extension of the CAPM (MKT+BAB) of Frazzini and Pedersen (2014), the Hou, Xue, and Zhang (2015) four-factor model (HXZ), the Stambaugh and Yuan (2017) mispricing model (SY), the Fama and French (2018) five-factor model with cash profitability (FF5CP) as well as its extension with the momentum factor (FF5CP+UMD), the Hou, Xue, and Zhang (2015) four-factor model with *RMWCP* instead of *ROE* (HXZCP), and a six-factor model of Fama and French (2018) that replaces *HML* with HML^m (FF5CP*+UMD). The models are presented from left to right and top to bottom in order of increasing squared Sharpe ratios. The sample period for our data is January 1972 to December 2015 (528 observations). We report in Panel A the difference between the (bias-adjusted) sample squared Sharpe ratios of the models in column i and row j, $\hat{\theta}_i^2 - \hat{\theta}_j^2$, and in Panel B the associated p-value (in parentheses) for the test of $H_0 : \theta_i^2 = \theta_j^2$. * indicates significance at the 5% level and ** indicates significance at the 1% level.

Panel A: Differences in Sample Squared Sharpe Ratios

	MKT+BAB	HXZ	SY	FF5CP	FF5CP+UMD	HXZCP	FF5CP*+UMD
MKT+FIRFT	0.076**	0.157**	0.212**	0.233**	0.242**	0.263**	0.333**
MKT+BAB		0.080*	0.136**	0.157**	0.166**	0.187**	0.257**
HXZ			0.056	0.077	0.086*	0.107**	0.176**
SY				0.021	0.030	0.051	0.121*
FF5CP					0.009	0.030	0.100**
FF5CP+UMD						0.021	0.090**
HXZCP							0.070*

Panel B: p-values

	MKT+BAB	HXZ	SY	FF5CP	FF5CP+UMD	HXZCP	FF5CP*+UMD
MKT+FIRFT	0.007	0.000	0.000	0.000	0.000	0.000	0.000
MKT+BAB		0.027	0.002	0.001	0.000	0.000	0.000
HXZ			0.122	0.075	0.042	0.005	0.001
SY				0.616	0.430	0.238	0.015
FF5CP					0.054	0.136	0.000
FF5CP+UMD						0.346	0.000
HXZCP							0.043

TABLE 3

Multiple Model Comparison Tests

In Table 3, we present multiple model comparison tests of the squared Sharpe ratios of eight asset-pricing models. The models include the He, Kelly, and Manela (2017) two-factor model (MKT+FIRFT), the betting-against-beta extension of the CAPM (MKT+BAB) of Frazzini and Pedersen (2014), the Hou, Xue, and Zhang (2015) four-factor model (HXZ), the Stambaugh and Yuan (2017) mispricing model (SY), the Fama and French (2018) five-factor model with cash profitability (FF5CP) as well as its extension with the momentum factor (FF5CP+UMD), the Hou, Xue, and Zhang (2015) four-factor model with $RMWCP$ instead of ROE (HXZCP), and a six-factor model of Fama and French (2018) that replaces HML with HML^m (FF5CP*+UMD). The models are estimated using monthly returns from January 1972 to December 2015 (528 observations). We report the benchmark models in column 1 and their (bias-adjusted) sample squared Sharpe ratio ($\hat{\theta}^2$) in column 2. r in column 3 denotes the number of alternative models in each multiple non-nested model comparison. LR in column 4 is the value of the likelihood ratio statistic with p-value given in column 5. Finally $\hat{\theta}^2_M - \hat{\theta}^2$ in column 6 denotes the difference between the (bias-adjusted) sample squared Sharpe ratios of the expanded model (M) and the benchmark model, with p-values given in column 7.

Benchmark	$\hat{\theta}^2$	r	LR	p-value	$\hat{\theta}^2_M - \hat{\theta}^2$	p-value
MKT+FIRFT	0.010	6	44.026	0.000		
MKT+BAB	0.086	6	21.195	0.000		
HXZ	0.166	6	10.580	0.005		
SY	0.222	6	5.893	0.042		
FF5CP	0.243	6	15.025	0.001	0.009	0.054
FF5CP+UMD	0.252	6	13.783	0.002		
HXZCP	0.273	6	4.092	0.120		
FF5CP*+UMD	0.342	6	0.000	0.787		

TABLE 4

Model Comparisons with a Nontraded Financial Intermediary Capital Risk Factor Model

In Table 4, we present pairwise tests of equality of the squared Sharpe ratios between the CAPM model augmented with the financial intermediary mimicking portfolio (MKT+FIRFM) vs. the eight asset-pricing models with traded factors only. The eight models include the He, Kelly, and Manela (2017) two-factor model (MKT+FIRFT), the betting-against-beta extension of the CAPM (MKT+BAB) of Frazzini and Pedersen (2014), the Hou, Xue, and Zhang (2015) four-factor model (HXZ), the Stambaugh and Yuan (2017) mispricing model (SY), the Fama and French (2018) five-factor model with cash profitability (FF5CP) as well as its extension with the momentum factor (FF5CP+UMD), the Hou, Xue, and Zhang (2015) four-factor model with *RMWCP* instead of *ROE* (HXZCP), and a six-factor model of Fama and French (2018) that replaces *HML* with HML^m (FF5CP*+UMD). The models are presented from left to right in order of increasing squared Sharpe ratios. The sample period for our data is January 1972 to December 2015 (528 observations). In Panel A, we report the difference between the (bias-adjusted) sample squared Sharpe ratios of the given models and that of MKT+FIRFM, and in Panel B the associated *p*-value (in parentheses) for the test of equality (zero difference). ** indicates significance at the 1% level.

Panel A: Differences in Sample Squared Sharpe Ratios

	MKT+FIRFT	MKT+BAB	HXZ	SY	FF5CP	FF5CP+UMD	HXZCP	FF5CP*+UMD
MKT+FIRFM	0.001	0.077**	0.158**	0.213**	0.234**	0.243**	0.264**	0.334**

Panel B: p-values

	MKT+FIRFT	MKT+BAB	HXZ	SY	FF5CP	FF5CP+UMD	HXZCP	FF5CP*+UMD
MKT+FIRFM	0.120	0.007	0.000	0.000	0.000	0.000	0.000	0.000