School of Finance

University of St.Gallen

TESTING COMPETING FACTOR PRICING MODELS

PAUL SÖDERLIND

WORKING PAPERS ON FINANCE NO. 2015/24

SWISS INSTITUTE OF BANKING AND FINANCE (S/BF – HSG)

NOVEMBER 2015
THIS VERSION: MAY 2016

Testing Competing Factor Pricing Models

Paul Söderlind*

First draft: 15 June 2015
Current draft: 23 May 2016

Abstract

A GMM-based system for two different linear factor pricing models is used to test if the pricing errors are the same. Simulations demonstrate the small sample properties. As an illustration, the test is applied to the Fama-French (1996, 2015) models.

Keywords: GMM, pricing errors, Bonferroni test, bootstrap

JEL: G12, C33, G10

1 Introduction

Empirical asset pricing studies often compare competing factor pricing models by reporting their respective average pricing errors and Gibbons, Ross, and Shanken (1989) statistics. However, the relative performance is tested only rarely.

Several recent papers have discussed how statistical tests can be used to compare models. For instance, De Moor, Dhaene, and Sercu (2015) address an important problem in comparing across model: a worse fit may drive down the significance of the pricing error test—and thus make the model look successful. They suggest putting the two models on equal footing by adding noise to the more precise model, which leads to an adjusted p-value. Barillas and Shanken (2015a) use a Bayesian approach to compute probabilities for all different models based on a set of pricing factors. This gives a rich testing framework, but requires formulating priors on the pricing errors. Li, Xu, and Zhang (2010)

*University of St. Gallen. *Address:* s/bf-HSG, Rosenbergstrasse 52, CH-9000 St. Gallen, Switzerland. *E-mail:* Paul.Soderlind@unisg.ch. I thank Nina Karnaukh and Igor Pozdeev for comments. Julia and Matlab routines for implementing these tests are available via https://sites.google.com/site/paulsoderlindecon/home.

develop a model selection approach that tests whether two models have the same (second) Hansen-Jagannathan distances. Harvey, Liu, and Zhu (2016) emphasize the importance of controlling the rate of false rejections in multiple testing by using a Bonferroni correction. Harvey and Liu (2014) use a bootstrap approach to analyze the properties of asset pricing tests, including the difference of (scaled) absolute pricing errors of two different models.

The contribution of the current paper is to suggest a simple and direct comparison of the pricing errors of two models. It is a combination two GMM systems: one for each factor model. This systems approach is straightforward and overcomes many of the issues with model comparisons. It shares several features with the studies discussed above: putting the models of equal footing, using (also) a Bonferroni correction and bootstrap simulations to gauge the small sample properties. However, it does not address the data mining issue, nor does it allow for using the information in prior beliefs.

The structure of the paper is as follows. Section 2 sets up the tests for the case when the factors are excess returns; Section 3 treats general (non-excess return) factors; Section 4 studies the properties of the tests by a series of Monte Carlo simulations; Section 5 is an empirical application to the Fama and French (1996, 2015) models; and Section 6 concludes. An appendix contains some technical details.

2 Testing with Excess Return Factors

This section sets up the tests for the case when the factors are excess returns.

2.1 Competing Factor Models

Consider a factor model for the excess returns of an $n \times 1$ vector of test assets (R_t^e)

$$R_t^e = \alpha + \beta f_t + \varepsilon_t, \tag{1}$$

where f_t is a $K \times 1$ vector of excess return factors and β is an $n \times K$ matrix. The equation for test asset i is found in row i of this system. We will henceforth call this the "f-model." As usual, the $n \times 1$ vector α measures the pricing errors. See Campbell, Lo, and MacKinlay (1997) 6 and Cochrane (2005) 12 for details on such models.

The competing model (the "h-model") is

$$R_t^e = \delta + \gamma h_t + u_t, \tag{2}$$

where h_t is an $L \times 1$ vector of excess return factors, δ is $n \times 1$ and γ is $n \times L$. In some cases h_t is a superset of f_t, but that is not necessary. We are interested in comparing the fit of the two models, the null hypothesis being that they are equal ($\alpha = \delta$).

2.2 Moment Conditions and Test

The idea of the test is to combine the f- and h-models into one system and then test if the intercepts are the same in the two models.

In spite of the simplicity of the test, it is useful to be explicit about the implementation. To estimate the two models as a joint system with GMM, we use the moment conditions

$$\frac{1}{T}\sum\nolimits_{t=1}^{T}\begin{bmatrix}\tilde{f}_t \otimes (R^e_t - \alpha - \beta f_t)\\ \tilde{h}_t \otimes (R^e_t - \delta - \gamma h_t)\end{bmatrix} = \mathbf{0}_{n(1+K+1+L)\times 1}, \tag{3}$$

where $\tilde{f}_t$ stacks 1 and f_t and $\tilde{h}_t$ stacks 1 and h_t, $\mathbf{0}_d$ denotes a vector or matrix of zeros with the dimensions d, and $\otimes$ denotes a Kronecker product.

There are $n(1+K)$ moment conditions in the first set and as many parameters. Similarly, there are $n(1+L)$ moment conditions in the second set and again as many parameters. The system is thus exactly identified and will give the OLS estimates of the parameters.[1]

Let $\hat{\theta} = [\hat{\alpha}, \operatorname{vec}\hat{\beta}, \hat{\delta}, \operatorname{vec}\hat{\gamma}]$ denote the GMM estimate of the true parameter vector θ_0. From the general properties of GMM, we know that $\sqrt{T}(\hat{\theta} - \theta_0)$ typically converges (as T increases) to a normal distribution

$$\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, V), \text{ with } V = D^{-1}S(D^{-1})'. \tag{4}$$

In this expression D is the probability limit of the Jacobian matrix of the moment conditions, whose inverse has a very simple structure.[2] Also, S is the covariance matrix of ($\sqrt{T}\times$) the sample moment conditions, which can be estimated by a standard approach (giving a White (1980) covariance matrix of $\hat{\theta}$) or, for instance, by the Newey and West (1987) method.

We consider two different versions of the test of the joint hypothesis that $\alpha = \delta$, that is, $\alpha_i = \delta_i$ for $i = 1...n$. The *first version* (called the "joint test" below) writes the restrictions

[1] An alternative approach would be to impose $\delta = \alpha$ and get a system with n overidentifying restrictions.

[2] It is straightforward to show that $D^{-1} = \operatorname{diag}([\mathrm{E}(\tilde{f}_t\tilde{f}_t')]^{-1} \otimes I_n, [\mathrm{E}(\tilde{h}_t\tilde{h}_t')]^{-1} \otimes I_n)$, where $\operatorname{diag}(A, B)$ creates a block-diagonal matrix with the matrices A and B along the diagonal and zeros elsewhere. In practice, we replace the expectation with a sample average.

as

$$P\theta = \mathbf{0}_{n\times 1}, \text{ where } P = [\ I_n \quad \mathbf{0}_{n\times nK} \quad -I_n \quad \mathbf{0}_{n\times nL}\ ], \tag{5}$$

where I_n is an $n \times n$ identity matrix—and then calculates the test statistic

$$\Xi = T\hat{\theta}' P'(PVP')^{-1} P\hat{\theta}. \tag{6}$$

Under the null hypothesis, Ξ typically converges in distribution to a χ^2_n variable, since there is one restriction for each test asset. For instance, with five test assets and a 5% significance level, we would compare Ξ with 11.07. However, the simulation results below show that this is not always correct.

To develop intuition for the test, consider the case of only one test asset. The variance of $\hat{\alpha} - \hat{\delta}$ (the PVP' term in (6)) is then $\text{Var}(\hat{\alpha}) + \text{Var}(\hat{\delta}) - 2\,\text{Cov}(\hat{\alpha}, \hat{\delta})$. The first two terms capture the uncertainty about both $\hat{\alpha}$ and $\hat{\delta}$ (not just one of them), which is similar in spirit to the approach in De Moor, Dhaene, and Sercu (2015) who stress the importance of not relying on just the least (or the most) uncertain model. The covariance term captures the joint uncertainty. For instance, positively correlated noise in $\hat{\alpha}$ and $\hat{\delta}$ tend to cancel out in $\hat{\alpha} - \hat{\delta}$.

The *second version* (called the “Bonferroni test” below) instead studies each of the n assets separately. Clearly, if we can safely reject the null hypothesis for at least one asset, then the joint hypothesis is also rejected. However, this cannot be implemented with traditional critical values since the chance of at least one false rejection increases with the number of test assets (see Harvey, Liu, and Zhu, 2016, for a careful analysis of this in the context of asset pricing).

To control this “family-wise error rate,” a Bonferroni correction is applied. To do this, let P_i be row i of P as defined in (5) and calculate the test statistic

$$\Xi_i = T\hat{\theta}' P_i'(P_i V P_i')^{-1} P_i\hat{\theta}, \tag{7}$$

which is just a squared t-stat of $\alpha_i - \delta_i$. In most cases, Ξ_i converges in distribution to a χ^2_1 variable. Redo this for each asset—and reject the joint hypothesis on the family-wise significance level of p if at least one of the individual test statistics exceeds the p/n critical value. For instance, with five test assets and a 5% significance level, we compare the highest individual test statistic (across $i = 1...n$) with the 1% critical value from a χ^2_1 distribution which is 6.63 (cf. the 5% critical value which is 3.84).[3]

[3] Since we focus on the highest individual test statistic, the Bonferroni and the Holm-Bonferroni (Holm, 1979) methods give the same result. This would be different if we wanted to see how many of the alphas

3 Testing with General Factors

This section sets up the tests for the case when the factors are not excess returns.

With general (not excess return) factors, the two models imply that

$$\mathrm{E}\, R_t^e = \alpha + \beta\lambda \text{ and } \mathrm{E}\, R_t^e = \delta + \gamma\psi, \tag{8}$$

where β and γ are the coefficients from the time series regressions (1) and (2) and where λ are the risk premia for the f_t factors and ψ for the h_t factors.

To estimate the factor risk premia (λ and ψ) from the cross-section of test assets, we set up a joint GMM system as to mimic the traditional two-step approach (see Cochrane (2005) 12 for the case of one model). That is, we combine the moment conditions for the time series regressions (3) with the sample analogues of the cross-sectional equations (8). The latter add $2n$ moment conditions, but only $K + L$ parameters (in λ and ψ), so there are overidentifying restrictions.

To estimate the parameters, we follow the standard approach of forming linear combinations of the cross-sectional moment conditions to get an exactly identified system. In practice, this means that the time series parameters $(\alpha, \beta, \delta, \gamma)$ are still estimated by OLS and that the factor risk premia are estimated by

$$\lambda = (B\beta)^{-1} B\bar{R}_t^e \text{ and } \psi = (G\gamma)^{-1} G\bar{R}_t^e, \tag{9}$$

where $\bar{R}_t^e$ is the $n \times 1$ vector of average excess returns, while B and G are matrices that define the linear combinations discussed above. For instance, $B = \beta'$ and $G = \gamma'$ give the same estimates as a traditional cross-sectional regression (see Cochrane (2005)).

With these point estimates we can test if the pricing errors are the same ($\bar{R}_t^e - \beta\lambda = \bar{R}_t^e - \gamma\psi$) by testing the cross-sectional moment conditions directly (see the appendix). The sampling uncertainty in this test is driven by the uncertainty about the estimated parameters $(\beta, \lambda, \gamma, \psi)$, since the average excess returns of the test assets cancel out.[4]

4 Simulation Results

This section studies the properties of the tests by a series of Monte Carlo simulations.

that differ from the deltas. In that case the Holm-Bonferroni method is more powerful.

[4] A way to double check the logic of this approach is apply it on excess return factors: include the excess return factors also as test assets and define the B and G matrices in (9) so that we effectively do cross-sectional GLS. In this setting, the test of $\beta\lambda = \gamma\psi$ should coincide with the joint test in (6), which indeed is the case.

4.1 Nested Models

An important special case is where the models are nested and the aim is to ascertain whether the extra factors affect the average returns of the test assets. In this case, the h-model (2) can be written

$$R_t^e = \delta + \gamma_f f_t + \gamma_z z_t + u_t, \tag{10}$$

where f_t are the common factors and z_t are the extra factors in the h-model. γ_f and γ_z are matrices of coefficients.

It is clear that some of the slopes on the extra factors (γ_z) must be non-zero for those extra factors to matter for the average return of the test assets. However, it is well known that this is not sufficient (see, for instance, Barillas and Shanken, 2015b). To see this, regress the extra factors on the common factors

$$z_t = \mu_z + \rho f_t + v_t, \tag{11}$$

where μ_z is a vector of pricing errors (intercepts), ρ is a matrix of coefficients and v_t is a vector of residuals.

Using (11) to substitute for z_t in (10) gives

$$R_t^e = \delta + \gamma_z(\mu_z + v_t) + (\gamma_f + \gamma_z \rho) f_t + u_t. \tag{12}$$

Estimating this equation gives the same slope coefficients on f_t as β in the f-model (1), since f_t is uncorrelated to both v_t and u_t.[5] For this reason (and because all residuals have zero means), the expected values of the f-model (1) and the rewritten h-model (12) imply that

$$\delta + \gamma_z \mu_z = \alpha. \tag{13}$$

This shows that at least some elements in the $n \times 1$ vector $\gamma_z \mu_z$ must be non-zero for the extra factors in the h-model to matter for the pricing of the test assets: testing $\alpha = \delta$ is (in this setting) the same as testing $\gamma_z \mu_z = 0$.

To see the practical importance of this, consider a testing framework that only tests whether the f-factors can price the z-factors ("factor redundancy"), that is, whether $\mu_z = 0$. Clearly, if this null hypothesis cannot be rejected, then we cannot reject the hypothesis

[5] It can also be noticed that (12) gives the same estimate (and t-stats) of γ_z as from estimating (10). This is explained by the Frisch-Waugh theorem on the properties of linear regressions, see, for instance, Davidson and MacKinnon (2004) p 67.

that f- and h-models have the same pricing errors for the test assets. However, things are different when we reject the null hypothesis: even if $\mu_z \neq 0$, the pricing errors for the test assets can be the same (if the non-redundant factors have zero loadings). This points to the importance of a direct test of $\alpha = \delta$ (or equivalently $\gamma_z \mu_z = 0$).

In the first Monte Carlo simulations presented below the f-model is the CAPM, while the h-model is the 3-factor model (market, SMB, HML) from Fama and French (1996), henceforth called "FF3." The simulations are based on estimates (regressions coefficients and covariance matrices) on monthly data July 1963 to December 2015, using data from Kenneth French's data library.

The *first step* of the simulations is to generate artificial data of f_t by drawing iid normally distributed random numbers with the same mean and standard deviation as observed for the U.S. equity market. The *second step* is to generate data for SMB and HML factors from (11), using the same means, regression coefficients and residual covariance matrix as in the sample. However, in some simulations, the pricing error of the extra factors (μ_z) are are tweaked to simulate either the size ($\mu_z = 0$) or the power ($\mu_z \neq 0$). The *third step* is to generate returns for the test assets according to (10) where the regression coefficients and the covariance matrix are from the sample. The number of test assets (n) is initially 5: the "diagonal" of the 25 FF test assets (small growth to large value). Results for all 25 FF portfolio are discussed later.

The upper panel of *Figure 1* shows the rejection frequencies of the tests, using asymptotic 5% critical values. The upper left subfigure shows simulated rejection frequencies for a medium-sized sample ($T = 500$) at different values of the pricing error of the extra factors (μ_z values from 0 to 3/12 per month on the horizontal axis). The power of the joint test is around 40% at $\mu = 3/12$, while the Bonferroni test has a power of 65%. The right subfigure shows the simulation results for $T = 1000$: the power is clearly stronger.

The overall impression from these figures is that the Bonferroni test performs well and that the joint test has low power at small pricing errors. However, the power increases rapidly as the pricing error and sample size increase.[6] Results for all 25 FF test assets suggest that the power is generally lower (similar to the findings in Campbell, Lo, and MacKinlay (1997) 6), but that the Bonferroni test suffers less than the joint test.

[6] It can also be noticed that both the joint test and the Bonferroni test have too low rejection rates at $\mu_z = 0$. The reason is that when the loadings on the extra factors are clearly non-zero ($\gamma_z \neq 0$), then the pricing errors depend only on whether the extra factors are priced by the original factors or not. For that reason, the joint test converges to a χ^2_2 (not χ^2_5) variable when μ_z is exactly zero. In contrast, with $\mu_z \neq 0$ but $\gamma_z = 0$ the joint test converges to a χ^2_5. See also Li, Xu, and Zhang (2010) for an analysis of the distribution of the HJ distance under the null and alternative hypotheses.

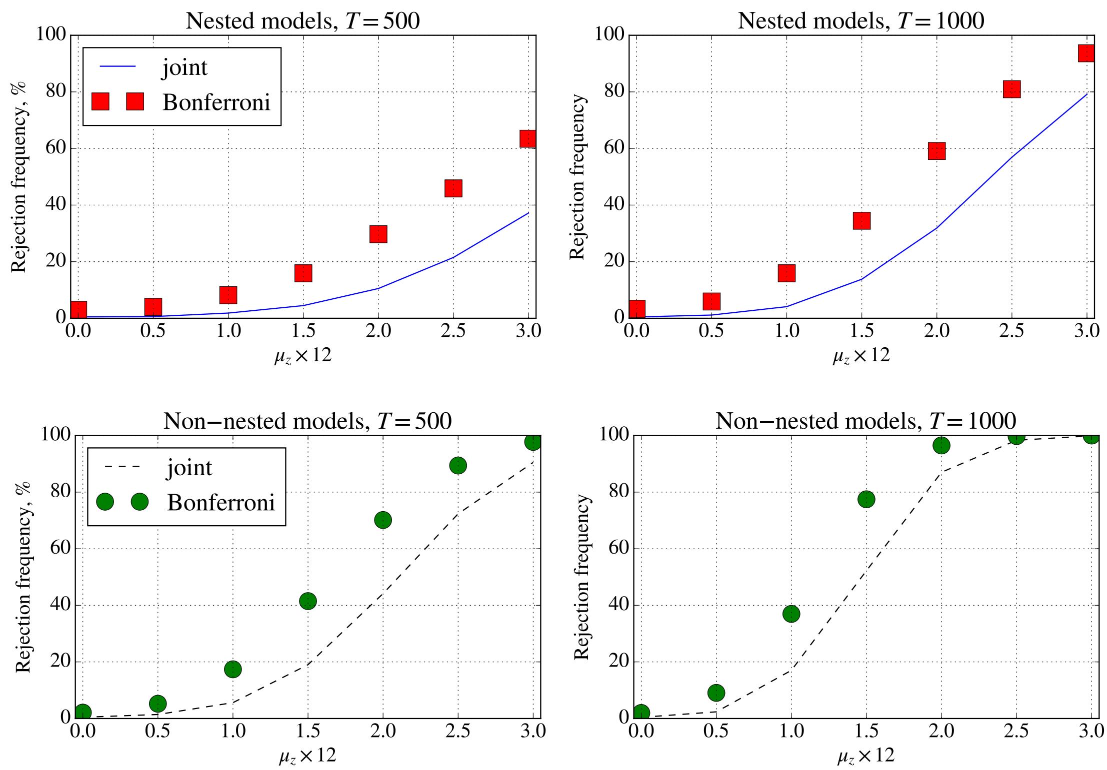

Figure 1: **Rejection frequencies of the tests, five test assets** The figures shows the rejection frequencies at different values of the pricing errors of the extra factors μ_z, using the asymptotic 5% critical values. The test statistics are calculated as in equations (6) and (7). Except for the μ_z values all simulation parameters are from estimating the CAPM and the (market, SMB, HML) models in the upper panel and the (market, SMB, HML) and (market, SMB, RMW, CMA) models in the lower panel on monthly data 1963:07–2015:12. The number of simulations is 5000.

Simulations (not shown) where we instead vary γ_z (while μ_z is from the sample and thus non-zero) are similar, but the power increases somewhat quicker with the sample size and pricing errors.

4.2 Non-Nested Models

We now consider non-nested models. In these simulations, the f-model is the FF3 model, while the h-model is a 4-factor model (market, SMB, RMW, CMA) from Fama and French (2015), henceforth called "FF4." The RMW factor captures profitability and CMA captures investments. These models are not nested since SMB enters only the f-model. As before, the simulations are based on estimates on monthly data July 1963 to December

2015 and the data is from Kenneth French's data library.

To understand the importance of the factors and to calibrate the simulations, we once again apply a projection approach. To do that, let x_t denote the factors that enter both models (that is, the common factors in f_t and h_t), y_t the factors that are unique to the f-model (1) and z_t those that are unique to the h-model (2). Then, regress y_t on a constant and the common factors x_t and let μ_y denote the estimated intercept. Do the same for z_t and let μ_z denote the intercept. It is then straightforward to show that

$$\alpha + \beta_y \mu_y = \delta + \gamma_z \mu_z, \tag{14}$$

where β_y are the loadings on y_t in the f-model and γ_z are the loadings on z_t in the h-model.[7]

Equation (14) shows that testing $\alpha = \delta$ is the same as testing $\beta_y \mu_y = \gamma_z \mu_z$, which is more involved than in the nested case (which is a special case of y_t being an empty vector so the test simplifies to $\gamma_z \mu_z = 0$). In the simulations, we set $\mu_y = 0$ and vary μ_z in the same way as for the nested case.

The lower panel of *Figure 1* shows the simulation results. The overall pattern is similar to the nested case, except that the power is generally higher. The reason is that the extra factors improve the fit of the models (lower variance of the residuals).

To sum up, the simulations show that both tests have reasonable properties, but that the Bonferroni test may have better power in realistic settings.

5 An Empirical Illustration

This section presents an example of how to implement the tests in empirical work. We use the same models and FF data as in the Monte Carlo simulation. While the 25 FF portfolios may have a particular structure (see, for instance, Lewellen, Nagel, and Shanken, 2010), they are still commonly used—and therefore provide a good illustration of the performance of the tests discussed here.

To allow for non-normally distributed factors and heteroskedasticity, the inference is based on a bootstrap (but the results are similar to the Monte Carlo simulations discussed above). In this bootstrap, we first draw (with replacement) $T = 630$ values of (R^e_t, f_t, h_t)

[7] Write the f-model (1) as $R^e_t = \alpha + \beta_x x_t + \beta_y y_t + \varepsilon_t$ and the h_t-model (2) as $R^e_t = \delta + \gamma_x x_t + \gamma_z z_t + u_t$. Then regress $y_t = \mu_y + \pi x_t + w_t$ and $z_t = \mu_z + \rho x_t + v_t$ and use these to substitute for y_t and z_t. In both equations, the estimated slope on x_t will the same as from estimating $R^e_t = a + b x_t + \xi_t$. Taking expectations therefore gives (14).

to create a new artificial sample of the same length as the empirical sample. Notice that we draw time periods (with replacement), so the covariance structure is preserved and any heteroskedasticity is accounted for (similar to Harvey and Liu, 2014).

Each such artificial sample gives a point estimate θ^* from the GMM system (3). To construct appropriate critical values, we test (in each simulated sample) the hypothesis

$$P(\theta - \overline{\theta^*}) = \mathbf{0}_{n\times 1}, \tag{15}$$

where P is defined in (5) and where $\overline{\theta^*}$ denotes the average value of θ^* across the simulations. Subtracting $\overline{\theta^*}$ accounts for the possibility that the null hypothesis may not be true in the observed data.

Table 1 shows results for the five diagonal FF portfolios (small growth to large value) as well as all the 25 FF portfolio for the sample July 1963 to December 2015. The left panel shows tests of the CAPM against the FF3 (market, SMB, HML), the middle panel shows tests of the FF3 model against the FF4 (market, SMB, RMW, CMA) and the right panel shows tests of the CAPM against the FF4.

For instance, with 5 test assets, the test statistics for CAPM vs. FF3 (left panel) are 16.4 for the joint test in equation (6) and 15.1 for the Bonferroni test in equation (7), with bootstrapped p-values of 3.9% and 0.05%. The null hypothesis that the two models give the same pricing errors is thus be rejected by both tests.

With 25 test assets, only the Bonferroni test rejects the null that CAPM and FF3 give the same pricing errors. The previous Monte Carlo simulations suggest that this may be due to the low power of the joint test in large systems and realistic sample sizes. The other panels show similar results, although the rejection of the null hypothesis is a bit stronger.

6 Concluding Remarks

This paper suggests a simple GMM system of two competing linear factor models as a vehicle for testing whether the pricing errors are the same in two models. Two versions of the test are used: a joint Wald test and a Bonferroni-corrected test that focuses on the highest test statistic across the assets. Simulation evidence shows that both versions have reasonable properties, but that the Bonferroni-corrected test may have larger power when the pricing errors are moderate and the sample small.

An empirical application to the Fama and French (1996, 2015) models illustrates how the tests can be used.

	CAPM vs. FF3		FF3 vs. FF4		CAMP vs. FF4	
$n \downarrow$	Joint	Bonferroni	Joint	Bonferroni	Joint	Bonferroni
5	16.4	15.1	16.7	10.5	41.4	21.4
	(3.9)	(0.0)	(1.9)	(0.7)	(0.0)	(0.0)
25	21.9	16.6	35.9	19.5	57.9	25.6
	(79.4)	(0.0)	(43.5)	(0.1)	(18.4)	(0.0)

Table 1: **Empirical test statistics with bootstrapped p-values in %.** The table presents results for CAPM against the FF3 (market, SMB, HML) models in the left panel, the FF3 against "FF4" (market, SMB, RMW, CMA) models in the middle panel and CAPM against FF4 in the right panel. The sample is monthly data for 1963:07–2015:12. The test statistics are calculated as in equations (6) and (7). The p-values (in parentheses) are from a bootstrap where the returns and factors are drawn randomly (pair-wise) from the sample. The number of simulations is 5000.

A Appendix: Technical Details of Section 3

The moment conditions for the case of general factors are

$$\bar{m} = \frac{1}{T}\sum_{t=1}^{T} \begin{bmatrix} \tilde{f}_t \otimes (R_t^e - \alpha - \beta f_t) \\ R_t^e - \beta\lambda \\ \tilde{h}_t \otimes (R_t^e - \delta - \gamma h_t) \\ R_t^e - \gamma\psi \end{bmatrix} = \mathbf{0}_{n(\tilde{K}+1+\tilde{L}+1)\times 1}. \tag{16}$$

To estimate the parameters, premultiply (16) by $A = \text{diag}(I_{n\tilde{K}}, B, I_{n\tilde{L}}, G)$, where B is $K \times n$ and G is $L \times n$. $A\bar{m}$ has as many elements as there are parameters, so calculating the point estimates is straightforward.

The Jacobian of the moment conditions (16) with respect to the parameters $\theta = [\alpha, \text{vec}\,\beta, \lambda, \delta, \text{vec}\,\gamma, \psi]$ is $D = \text{diag}(D_f, D_h)$ where

$$D_f = -\begin{bmatrix} \mathrm{E}(\tilde{f}_t\tilde{f}_t') \otimes I_n & \mathbf{0}_{n\tilde{K}\times K} \\ \begin{bmatrix} 0 & \lambda' \end{bmatrix} \otimes I_n & \beta_{n\times K} \end{bmatrix} \text{ and } D_h = -\begin{bmatrix} \mathrm{E}(\tilde{h}_t\tilde{h}_t') \otimes I_n & \mathbf{0}_{n\tilde{L}\times L} \\ \begin{bmatrix} 0 & \psi' \end{bmatrix} \otimes I_n & \gamma_{n\times L} \end{bmatrix},$$

and where diag(A, B) creates a block-diagonal matrix with the matrices A and B along the diagonal and zeros elsewhere

Let $\bar{m}$ denote the moment conditions (16) evaluated at the point estimates. It is well known (see, for instance, Cochrane (2005)) that $\sqrt{T}\bar{m}$ has an asymptotic normal distribution with zero means (under the null hypothesis) and a (reduced rank) covariance matrix equal to

$$\Psi = \tilde{\Psi} S \tilde{\Psi}', \text{ where } \tilde{\Psi} = I - D\,(AD)^{-1} A.$$

To test if the pricing errors are the same, we formulate the linear restrictions

$$P\bar{m} = \mathbf{0}_{n\times 1}, \text{ where } P = \begin{bmatrix} \mathbf{0}_{n\times n\tilde{K}} & I_n & \mathbf{0}_{n\times n\tilde{L}} & -I_n \end{bmatrix}$$

and calculate the test statistics

$$T\bar{m}' P' (P\Psi P')^{-1} P\bar{m} \xrightarrow{d} \chi^2_q,$$

where we expect the degrees of freedom to equal $\max(n-K, n-L)$, since a linear combination does not reduce the rank of the covariance matrix.

To test coefficients, we use the fact that $\sqrt{T}(\theta - \theta_0)$ also has an asymptotic normal distribution with the covariance matrix

$$V = \tilde{V} S \tilde{V}', \text{ where } \tilde{V} = (AD)^{-1} A.$$

References

Barillas, F., and J. Shanken, 2015a, "Comparing asset pricing models," Working paper, Goizueta Business School, Emory University.

Barillas, F., and J. Shanken, 2015b, "Which alpha?," Working paper, Goizueta Business School, Emory University.

Campbell, J. Y., A. W. Lo, and A. C. MacKinlay, 1997, *The econometrics of financial markets*, Princeton University Press, Princeton, New Jersey.

Cochrane, J. H., 2005, *Asset pricing*, Princeton University Press, Princeton, New Jersey, revised edn.

Davidson, R., and J. G. MacKinnon, 2004, *Econometric theory and methods*, Oxford University Press, Oxford.

De Moor, L., G. Dhaene, and P. Sercu, 2015, "On comparing zero-alpha tests across multifactor asset pricing models," *Journal of Banking and Finance*, 62, S235–240.

Fama, E. F., and K. R. French, 1996, "Multifactor explanations of asset pricing anomalies," *Journal of Finance*, 51, 55–84.

Fama, E. F., and K. R. French, 2015, “A five-factor asset pricing model,” *Journal of Financial Economics*, 116, 1–22.

Gibbons, M., S. Ross, and J. Shanken, 1989, “A test of the efficiency of a given portfolio,” *Econometrica*, 57, 1121–1152.

Harvey, C. R., and Y. Liu, 2014, “Lucky factors,” Working paper, Duke University.

Harvey, C. R., Y. Liu, and H. Zhu, 2016, “...and the cross-section of expected returns,” *The Review of Financial Studies, forthcoming*.

Holm, S., 1979, “A simple sequentially rejective multiple test procedure,” *Scandinavian Journal of Statistics*, 6, 65–70.

Lewellen, J., S. Nagel, and J. Shanken, 2010, “A skeptical appraisal of asset pricing tests,” *Journal of Financial Economics*, 96, 175–194.

Li, H., Y. Xu, and X. Zhang, 2010, “Evaluating asset pricing models using the second Hansen-Jagannathan distance,” *Journal of Financial Economics*, 97, 279–301.

Newey, W. K., and K. D. West, 1987, “A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix,” *Econometrica*, 55, 703–708.

White, H. L., 1980, “A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity,” *Econometrica*, 48, 817–838.