

## **BAYESIAN TESTS OF GLOBAL FACTOR MODELS**

Jonathan Fletcher

University of Strathclyde

Key words: Model Comparison, Bayesian Analysis, Factor Models

JEL classification: G11, G12

The author is from the University of Strathclyde.

I am grateful to the data being provided by Ken French's data library, and AQR. I am grateful to F. Barillas for providing Matlab code to conduct the Bayesian tests. I am grateful for the comments from an anonymous reviewer.

This draft: June 2018

Address correspondence to Professor J. Fletcher, Department of Accounting and Finance, University of Strathclyde, Stenhouse Building, Cathedral Street, Glasgow, G4 0LN, United Kingdom, phone: +44 (0) 141 548 4963, fax: +44 (0) 552 3547, email: [j.fletcher@strath.ac.uk](mailto:j.fletcher@strath.ac.uk)

## **BAYESIAN TESTS OF GLOBAL FACTOR MODELS**

### **ABSTRACT**

I use the Bayesian approach of Barillas and Shanken(2018) to examine the mean-variance efficiency of nine global factor models in global stock returns and to conduct multiple model comparison tests. The mean-variance efficiency of each factor model is strongly rejected. In the multiple model comparison tests, the three-factor model of Asness, Moskowitz and Pedersen(2013) has the best performance at higher prior maximum Sharpe(1966) ratio multiples and significantly outperforms all the other factor models. However, in out-of-sample tests, the AMP model significantly underperforms the best performing models that can be formed among the set of all factors used by the global factor models.

## **I Introduction**

In a world of integrated capital markets, the relevant risk factors are global factors rather than domestic factors. Global versions of domestic factor models have been developed for the capital asset pricing model (CAPM) (Grauer, Litzenberger and Stehle(1976)), arbitrage pricing theory (APT) (Solnik(1983)), and empirical factor models such as Fama and French(1998) among others. Models that include exchange rate risk factors are Adler and Dumas(1983), Dumas and Solnik(1995), DeSantis and Gerard(1998), and Zhang(2006) among others<sup>1</sup>.

A number of recent studies advocate different factor models. Hou, Karolyi and Kho(2011) find that the best performing model is a three-factor model including the market, value (based on cash flow to price), and momentum factors. An alternative three-factor model is proposed by Asness, Moskowitz and Pedersen(2013) where the value and momentum factors are formed differently to Hou et al. Fama and French(2012,2017) examine the performance of global and regional factor models based on the models of Fama and French(1993), Carhart(1997), and Fama and French(2015). Fama and French(2012,2017) find that regional versions of their models outperform global versions of the models (see also Cakici(2015)). Brusa, Ramadorai and Verdelhan(2014) propose a three-factor model, which includes the world market index in local currency, and two currency risk factors (Dollar and Carry)<sup>2</sup>. Cooper, Mitache and Priestley(2017) propose a global version of the macroeconomic factor model of Chen, Roll and Ross(1986).

---

<sup>1</sup> See Karolyi and Stulz(2003) and Lewis(2011) for excellent reviews of international asset pricing.

<sup>2</sup> These factors are common factors in currency markets as in Lustig, Roussanov, and Verdelhan(2011).

When comparing models, empirical studies often focus on specification tests such as Gibbons, Ross and Shanken(1989) test of mean-variance efficiency, and metrics based on the pricing errors of the test assets (Fama and French(2012,2015,2016,2017)). A recent study by Barillas and Shanken(2017) points out that such an approach is problematic as it ignores the pricing of the excluded factors from the model. Barillas and Shanken show that when comparing models using a number of metrics, the choice of test assets is irrelevant. The only relevant comparison is how well models price the factors not included in the model. Barillas and Shanken(2018) develop a Bayesian approach to examine the mean-variance efficiency of linear factor models and to conduct model comparison tests.

I use the Bayesian approach of Barillas and Shanken(2018) to test the mean-variance efficiency of nine global factor models in global stock returns and to conduct multiple model comparison tests. My sample period covers November 1990 and December 2016. To test the mean-variance efficiency of the global factor models, I use two sets of global test assets, which are the 25 size/book-to-market (BM) portfolios, and 25 size/momentum portfolios. I also examine performance tests of the best performing model among the nine factor models, to the top three models among the set of all models that can be formed from the union of all the factors included in the nine global factor models.

My initial choice of models is motivated by the recent studies of Fama and French(2012,2015,2018), Asness et al(2013), and Asness, Frazzini, Israel and Moskowitz(2015). The first group of models include the CAPM, the empirical factor models of Fama and French(1993), Carhart(1997), Fama and French(2015)<sup>3</sup> models, and the Fama and French(2015) model with a momentum factor (Fama and French(2018)). Fama and

---

<sup>3</sup> Fama and French(2015) augment their Fama and French(1993) model with profitability and investment factors.

French(2018) also extend their models to include only the small spread factors for the value, profitability, investment, and momentum factors. Asness et al(2013) propose a three-factor model which includes value and momentum factors in addition to the global market index. My final model is based on the Asness et al(2015) six-factor model<sup>4</sup>

There are four main findings in my study. First, the mean-variance efficiency of each global factor model is rejected using both the Gibbons et al(1989) and Barillas and Shanken(2018) tests. Second, in the multiple model comparison tests among the nine global factor models, the three-factor model of Asness et al(2103) has the best performance at the higher prior maximum Sharpe ratio multiples. Third, none of the factors in the Asness et al model are redundant and all make a reasonable contribution to the maximum squared Sharpe(1966) performance of the other factors in the model. Fourth, in out-of-sample performance tests, the AMP model provides a significant lower Sharpe performance than the best performing models that can be formed from the union of all the factors used in the nine global factor models.

My study makes two contributions to the literature. First, I extend the evidence in Barillas and Shanken(2018) by applying their Bayesian approach to evaluate global factor models. Recent studies by Hou, Xue and Zhang(2017) and Harvey(2017) highlight the importance of replication studies in Finance, which is common in other fields of science. Second, I extend the prior empirical evidence of global factor models such as Fama and French(2012,2017), Cakici(2015), and Asness et al(2013) among others by using a Bayesian

---

<sup>4</sup> The Asness et al(2015) model replaces the value factor in the six-factor model with a more timely version of the factor as in Asness and Frazzini(2013). Barillas, Kan, Robotti and Shanken(2017) find that this six-factor model performs well in relative model comparison tests in U.S. stock returns.

approach to test the mean-variance efficiency of the global factor models and conducting Bayesian model comparison tests

My study is organized as follows. Section II describes the research method used in the study. Section III discusses the data. Section IV reports the empirical results and the final section concludes.

## II Research Method

Linear factor models such as the CAPM and APT predict:

$$E(r_i) = \sum_{k=1}^K \beta_{ik} f_k \quad (1)$$

where  $r_i$  is the excess return on asset  $i$ ,  $f_k$  is the factor risk premium on the  $k$ th factor,  $\beta_{ik}$  is the beta on asset  $i$  relative to factor  $k$  for  $k=1, \dots, K$ , and  $K$  is the number of factors in the model. The prediction in equation (1) can be evaluated from the time-series regression:

$$r_{it} = \alpha_i + \sum_{k=1}^K \beta_{ik} f_{kt} + u_{it} \text{ for } i=1, \dots, N \quad (2)$$

where  $r_{it}$  is the excess return on asset  $i$  at time  $t$ ,  $f_{kt}$  is the excess return on factor  $k$  at time  $t$ ,  $\alpha_i$  is the alpha of asset  $i$ ,  $u_{it}$  is a random error term on asset  $i$  at time  $t$  with  $E(u_{it}) = 0$  and  $E(u_{it} f_{kt}) = 0$  for each  $k$ th factor, and  $N$  is the number of risky assets. The beta model in equation (1) imposes testable restrictions in equation (2) as:

$$H_0: \alpha_i = 0, \text{ for } i=1, \dots, N \quad (3)$$

Gibbons et al(1989) derive a multivariate F test of the null hypothesis in equation (3). Assuming the residuals from equation (2) have a multivariate normal distribution with zero mean and constant covariance matrix, the test statistic is given by:

$$GRS = [(T-N-K)/N] * (\alpha' \Sigma^{-1} \alpha) / (1 + Sh(F)^2) \quad (4)$$

where  $\alpha$  is a  $(N,1)$  vector of  $\alpha_i$ 's from equation (2),  $\Sigma$  is the Maximum Likelihood (ML) estimate of  $(N,N)$  residual covariance matrix,  $Sh(F)^2$  is the maximum squared Sharpe(1966) performance of the  $K$  factors in the linear factor model, and  $T$  is the number of observations.

Under the null hypothesis of the ex-ante mean-variance efficiency of the K-factor model, the GRS test has a central F distribution with N and T-N-K degrees of freedom.

Gibbons et al(1989) show that the quadratic form  $\alpha'\Sigma^{-1}\alpha$  can be written as:

$$\alpha'\Sigma^{-1}\alpha = \text{Sh}(F,R)^2 - \text{Sh}(F)^2 \quad (5)$$

where  $\text{Sh}(F,R)^2$  is the maximum squared Sharpe performance of the N+K assets. If some combination of the K factor portfolios lies on the mean-variance frontier of the N+K assets, then  $\text{Sh}(F,R)^2 = \text{Sh}(F)^2$ . Under the alternative hypothesis  $\alpha_i \neq 0$  and the GRS test has a non-central F distribution and is left unspecified<sup>5</sup>.

Bayesian tests of equation (3) have been developed by Shanken(1987), Harvey and Zhou(1990), and McCulloch and Rossi(1990, 1991) among others. Barillas and Shanken(2018) build on these earlier studies and derive a Bayesian test of mean-variance efficiency of equation (3) that can be solved analytically. The Bayesian test is a function of the GRS test. Barillas and Shanken assume a diffuse prior for  $\beta$  and  $\Sigma$ , where  $\beta$  is a (N,K) matrix of factor betas, and an informative prior for  $\alpha$ . The diffuse prior is given by:

$$P(\beta,\Sigma) \propto |\Sigma|^{-(N+1)/2} \quad (6)$$

The informative prior for  $\alpha$  under the alternative hypothesis is given by:

$$P(\alpha|\beta,\Sigma) = \text{MVN}(0,k\Sigma) \quad (7)$$

where MVN is the multivariate normal distribution and k captures our view of the likely size of the expected excess return deviations for the model. Barillas and Shanken(2018) point out that using a prior of  $\alpha$  linked to the residual covariance matrix is followed in a number of studies such as MacKinlay(1995) and Pastor and Stambaugh(2000). The link between  $\alpha$  and  $\Sigma$  makes extremely high Sharpe ratios unlikely. Barillas and Shanken show that k is given by:

---

<sup>5</sup> Fama(2015) provides a review of the GRS test and the alternative two-pass cross-sectional regression approach of Fama and MacBeth(1973) to evaluate linear factor models.

$$k = (\text{Sh}_{\max}^2 - \text{Sh}(F)^2)/N \quad (8)$$

where  $\text{Sh}_{\max}^2$  is the maximum expected squared Sharpe performance under the alternative hypothesis. The researcher specifies the size of  $\text{Sh}_{\max}^2$ .

Imposing an upper bound on  $\text{Sh}_{\max}^2$  has been used in different applications. Cochrane and Saa-Requejo(2000) rule out Sharpe ratios greater than twice the market Sharpe ratio as “good-deals” to place tighter bounds on option prices beyond no arbitrage bounds. Chretien and Kammoun(2017) impose an upper bound  $\text{Sh}_{\max}^2$  to estimate good-deal performance measures for mutual funds. Huang and Zhou(2017) derive an upper bound on the  $R^2$  from the predictive regression implied by a given asset pricing model by using an upper bound on  $\text{Sh}_{\max}^2$ .

Proposition 1 in Barillas and Shanken(2018) shows that the Bayes factor (BF) of  $H_0: \alpha = 0$  against  $H_1: \alpha \neq 0$  is given by:

$$\text{BF} = (1/Q) * (|S|/|S_R|)^{(T-K)/2} \quad (9)$$

where  $S$  is the  $(N,N)$  cross-product matrix of the residuals from equation (2), and  $S_R$  is the  $(N,N)$  cross-product matrix of the residuals from equation (2) when the  $N$   $\alpha_i$ 's are constrained to be zero. The  $Q$  term is given by:

$$Q = (1+(a/(a+k)) * (W/T))^{-(T-K)/2} * (1+(k/a))^{-N/2} \quad (10)$$

where  $a=(1+\text{Sh}(F)^2)/T$ , and  $W=\text{GRS} * ((N*T)/(T-N-K))^6$ . The posterior probability of the null hypothesis is given by  $\text{BF}/(1+\text{BF})$ . I refer to the Bayesian test in equations (9) and (10) as the B-GRS test. The null hypothesis of the B-GRS test can be written as  $k=0$  against  $k$ . Barillas and Shanken also show that the B-GRS test can be used to test the approximate fit of a model, where a degree of mispricing is allowed under the null hypothesis. The approximate fit can be viewed as a test of  $k=k_0$  against  $k$ . Barillas and Shanken show that the test of the approximate

---

<sup>6</sup> The  $Q$  statistic was initially derived by Harvey and Zhou(1990). Barillas and Shanken(2018) derive the closed-form solution to it.



fit of a model is given by  $BF = Q_{k0}/Q$ , where  $Q_{k0}$  is the value of  $Q$  when  $k$  is set equal to  $k_0$ . In this study, I only focus on testing the absolute fit of the models i.e.  $k=0$ . I set  $Sh^2_{max}$  in the tests of mean-variance efficiency to be equal to multiples of 1.2, 1.4, 1.6, 1.8, and 2 of  $Sh(F)^2$  of the factors in the model taken with respect to the prior under the alternative that the alphas of the test assets are non-zero.

Fama and French(2012,2015,2016) compare the performance of different linear factor models based on metrics of the pricing errors of the  $N$  test assets, such as the mean absolute alpha. Barillas and Shanken(2017) point out that when comparing linear factor models based on the pricing errors of the  $N$  test assets ignores the issue of whether the factor models can price the factors not included in the model. Barillas and Shanken show for a number of metrics, that the choice of test assets is irrelevant for relative model comparison tests. The only relevant issue is how well models price factors not included in the model. Barillas and Shanken provide examples where a factor model can outperform another model in terms of pricing errors of the  $N$  test assets but underperform when comparing models by how the model prices the excluded factors from the model.

Barillas and Shanken(2018) develop a Bayesian approach for model comparison tests<sup>7</sup>. The approach can be used to test nested or non-nested models and can be applied to multiple model comparison tests. The Bayesian approach of Barillas and Shanken assumes that the

---

<sup>7</sup> Gospodinov, Kan and Robotti(2013) develop model comparison tests based on the Hansen and Jagannathan(1997) distance measure. Kan, Robotti and Shanken(2013) propose model comparison tests based on the cross-sectional  $R^2$ . Barillas et al(2017) develop classical asymptotic tests of model comparison based on the Sharpe ratio.

market index is included in each factor model<sup>8</sup>. Barillas and Shanken point out that the motivation for doing this is that the market portfolio plays a central role in the CAPM and ICAPM asset pricing models, and is the aggregate supply of securities (see also Fama and French(2016)). Define  $L$  as the number of non-market factors in a model denoted by  $f$ , which is a  $(T,L)$  matrix of excess returns, and  $K$  is the union of separate factors across all the models. The excluded  $K-L$  factors from a model is denoted by  $f^*$ , which is a  $(T,K-L)$  matrix. Consider the following regressions:

$$f = \alpha + \beta \text{Mkt} + e \quad (11)$$

$$f^* = \alpha^* + \beta^* [\text{Mkt}, f] + e^* \quad (12)$$

$$r = \alpha_r + \beta_r [\text{Mkt}, f, f^*] + e_r \quad (13)$$

where  $\text{Mkt}$  is  $(T,1)$  vector of excess market returns. Proposition 3 in Barillas and Shanken show that the marginal likelihood (ML) of each model is given by:

$$\text{ML} = \text{ML}_U(f|\text{Mkt}) * \text{ML}_R(f^*|\text{Mkt}, f) * \text{ML}_R(r|\text{Mkt}, f, f^*) \quad (14)$$

where  $\text{ML}_U(f|\text{Mkt})$  is the unrestricted ML from equation (11),  $\text{ML}_R(f^*|\text{Mkt}, f)$  is the restricted ML from equation (12) by setting  $\alpha^* = 0$ , and  $\text{ML}_R(r|\text{Mkt}, f, f^*)$  is the restricted ML from equation (13) by setting  $\alpha_r = 0$ . Since  $\text{ML}_R(r|\text{Mkt}, f, f^*)$  is common across all models, this term drops out and so the choice of test assets is irrelevant for model comparison.

Given the ML for each model, the posterior probability for each model  $i$  can be calculated as in Barillas and Shanken(2018) assuming the prior probability of each model is equal given by:

$$P(M_i|D) = \text{ML}_i / \sum_{i=1}^I \text{ML}_i \quad (15)$$

---

<sup>8</sup> Barillas and Shanken(2018) generalize their approach where the market index is not automatically included in each model.

where  $D$  is the data used by the test assets and factors, and  $I$  is the number of factor models. When calculating the  $ML_U(f|Mkt)$ , we need to specify  $Sh_{max}^2$  to calculate  $k$ . I follow Barillas and Shanken and consider  $Sh_{max} = \text{prior multiple} * Sh(mkt)$ , where  $Sh(mkt)$  is the sample Sharpe ratio of the market index, and  $Sh_{max}$  is the square root of the expected squared Sharpe performance with  $K$  factors taken with respect to the prior under the alternative that the non-market factors have non-zero alphas. I use prior multiples of 1.25, 1.5, 2, and 3 and  $k$  is fixed across models. The lower multiples are more consistent with risk-based explanations of the non-zero alphas (MacKinlay(1995)).

### **III Data**

#### **A) Test Assets**

I test the mean-variance efficiency and conduct multiple model comparison tests of nine global factor models between November 1990 and December 2016. For the mean-variance efficiency tests, I use two sets of test assets, which are the 25 size/BM portfolios, and 25 size/momentum portfolios. All of the return data is in U.S. dollars. All of the data is collected from Ken French's web site unless otherwise specified. I use the one-month U.S. Treasury Bill as the risk-free asset. The size/BM portfolios are formed along size (Small to Big) and the BM ratio (Growth to Value). The size/momentum portfolios are formed along size and momentum (Losers to Winners). Table 1 reports the monthly average excess returns (%) for the size/BM portfolios (panel A) and the size/momentum portfolios (panel B).

Table 1 here

Panel A of Table 1 shows that there is a wide spread in the mean excess returns in the global size/BM portfolios. The mean excess returns range between 0.165% (Small/Growth) and 1.028% (Small/Value). There is a value effect across all size groups. The value effect is

strongest among the smallest companies and less noticeable among the largest companies. The size effect varies across the BM groups. There is a strong size effect in the Value portfolios and a reverse size effect in the Growth portfolios.

The spread in mean excess returns of the size/momentum portfolios in panel B of Table 1 is wider than the size/BM portfolios. The mean excess returns range between 0.104% (Small/Losers) and 1.449% (Small/Winners). There is a strong momentum effect across all size groups. The momentum effect is strongest among the smallest companies. Excluding the Losers portfolios, there is a strong size effect across the momentum groups. The size effect is strongest among the Winners portfolios.

## B) Factor Models

My main empirical tests focus on nine global factor models. There are 13 ( $K=13$ ) separate factors across all the models and so each model can be considered a subset of the 13-factor model. The models include:

### 1. CAPM

This model is a single-factor model that uses the excess returns of the world stock market index (Mkt) as the proxy for aggregate wealth.

### 2. Fama and French(1993) (FF)

The FF model is a three-factor model. The factors are the excess return on the market index and two zero-cost portfolios that capture the global size (SMB) and value/growth (HML) effects in stock returns.

### 3. Carhart(1997)

The Carhart model is a four-factor model. The factors are the three factors in the FF model and a zero-cost portfolio that captures the momentum effect (WML) in global stock returns.

### 4. Fama and French(2015) FF5

This model is a five-factor model. The factors include the factors in the FF model and two zero-cost portfolios that capture the profitability (RMW) and investment growth (CMA) effects in global stock returns. I use the SMB factor in the FF5 model as the size factor across all models.

5. Fama and French(2018) (FF5s)

This model is a five-factor model as in FF5, but includes the small ends of the HML, RMW, and CMA spread factors denoted as HML<sub>S</sub>, RMW<sub>S</sub>, and CMA<sub>S</sub>.

6. Fama and French(2018) (FF6)

This model is a six-factor model, which augments the FF5 model with the WML factor.

7. Fama and French(2018) (FF6s)

This model is a six-factor model, which augments the FF5s model with the small end of the WML factor (WML<sub>S</sub>).

8. Asness, et al(2015) (AFIM)

This model is a six-factor model which replaces the HML factor in the FF6 model with a more timely version of the HML factor (HML<sub>T</sub>) as in Asness and Frazzini(2013). The HML<sub>T</sub> factor is collected from the AQR data library.

9. Asness, et al(2013) (AMP)

This model is a three-factor model. The factors include the market index, and global value (VAL) and momentum (MOM) factors across all asset classes. The VAL and MOM factors come from the AQR data library.

Table 2 reports summary statistics of the factor excess returns. The summary statistics include the mean and standard deviation of monthly factor excess returns (%). The final column reports the *t*-statistic of the null hypothesis that the average excess returns on the factors equal zero.

Table 2 here

Table 2 shows that a number of factors have significant positive average excess returns. The  $WML_S$  factor has the highest mean excess return. The large significant positive average excess returns of the  $WML$ ,  $WML_S$ , and  $MOM$  factors confirms the strong momentum effects in global stock returns. The higher mean excess return of the  $WML_S$  factor confirms the stronger momentum effect in smaller companies. It is interesting to note that the Asness et al(2013)  $MOM$  factor has a considerable lower mean excess return than the  $WML$  factor. The  $SMB$  size factor has an insignificant average excess return. The  $HML$ ,  $HML_S$ ,  $HML_T$ , and  $VAL$  factors all have significant positive average excess returns confirming the strong value effect in global stock returns. The mean excess returns between the  $HML$  and  $HML_T$  factors are similar to one another. The higher mean excess return of the  $HML_S$  factor relative to the  $HML$  factor confirms the result that the value effect is stronger among smaller companies in global stock returns.

There are significant profitability and investment effects in global stock returns as reflected in the significant mean excess returns of the  $RMW$ ,  $RMW_S$ ,  $CMA$ , and  $CMA_S$  factors. The investment effect is stronger among smaller companies in global stock returns but this pattern is less noticeable in the profitability effect. This pattern differs from Fama and French(2015) who find that both the investment and profitability effects are both stronger in smaller companies in U.S. stock returns.

#### **IV Empirical Results**

I begin my empirical tests by examining the mean-variance efficiency of the nine global factor models using the GRS and B-GRS tests. Table 3 reports the empirical results for the size/BM portfolios in panel A, and size/momentum portfolios in panel B. The table includes the GRS F test and the corresponding p value ( $p(\text{GRS})$ ), and the posterior probabilities of the

null hypothesis using the B-GRS test ( $k=0$ ) under the prior  $Sh_{\max} = \text{prior multiple} * Sh(F)$ . The final column reports the mean absolute alpha ( $A|\alpha_i|$ ) of the test assets.

Table 3 here

Table 3 shows that the mean-variance efficiency of each factor model is rejected using the GRS test. This result holds across both sets of test assets. A similar finding exists with the B-GRS test. The vast majority of the posterior probabilities equal zero. It is only for the CAPM, when the prior multiple of 1.2 is used, that the posterior probability is larger than 0.05. Using the  $A|\alpha_i|$  measure, the global factor models have smaller average mispricing in the size/BM portfolios compared to the size/momentum portfolios. The  $A|\alpha_i|$  measures range between 0.094% (Carhart) and 0.166% (CAPM) for the size/BM portfolios and between 0.120% (Carhart) and 0.365% (FF5s) for the size/momentum portfolios. The factor models with the momentum factor (Carhart, FF6, FF6s, AFIM, and AMP) have a lower average mispricing in the size/momentum portfolios compared to the alternative models, which is not surprising as the same characteristic is used to form the test assets.

The rejection of mean-variance efficiency in Table 3 for the CAPM, FF, and Carhart models is consistent with Fama and French(2012) and the FF5 model is similar to Cakici(2015). The tests in Table 3 are useful to examine the absolute fit of each model. However for comparing factor models, Barillas and Shanken(2017) show that the relevant consideration is how well the models price the factors not included in the model. I next examine the Bayesian multiple model comparison tests among the nine global factor models. Table 4 reports the posterior probabilities of each model. I set  $Sh_{\max} = \text{prior multiple} * Sh(\text{Mkt})$ , where  $Sh_{\max}$  is the square root of the expected maximum squared Sharpe performance with 13 factors included,

taken with respect to the prior under the alternative that the alphas of the non-market factors are non-zero. I use prior multiples of 1.25, 1.5, 2, and 3.

Table 4 here

Table 4 shows that, with the exception of the CAPM and AMP models, the models have posterior probabilities of zero. The models with the momentum factor (Carhart, FF6, FFs, and AFIM), which perform better in pricing the test assets in Table 3, perform poorly in the model comparison tests. This result is consistent with Barillas and Shanken(2017) and highlights the importance of taking account of how well the models price the excluded factors when comparing models. At the prior maximum Sharpe ratio multiples of 1.25 and 1.5, the CAPM is the surprising winner. This result might be driven by the fact that although the CAPM alphas are far away from zero, they are closer to the alphas under the alternative hypothesis compared to the other models (Barillas and Shanken(2018)). At the higher prior maximum Sharpe ratio multiples of 2 and 3, the AMP model is the clear winner. The AMP model has a posterior probability of 0.918 and 0.998 respectively at the prior maximum Sharpe ratio multiples of 2 and 3.

The results in Table 4 show the posterior probability of each factor model for the whole sample period. To provide a different perspective, Figure 1 reports the model probabilities over time as in Barillas and Shanken(2018)<sup>9</sup> for the prior maximum Sharpe ratio multiple of 2. Using all the data from November 1990, with recursive windows, the posterior probabilities of each model is calculated. Figure 1 plots the model probabilities of the two best performing

---

<sup>9</sup> I am grateful to the reviewer for suggesting this topic.



models at the end of the sample period (AMP and CAPM). All of the other models have posterior probabilities close to zero through time.

Figure 1 here

Figure 1 shows that there is considerable variation through time of the model probabilities of the AMP and CAPM models. Prior to November 2006, the CAPM model has the highest posterior probability. The CAPM again has a higher posterior probability than the AMP model between September 2011 and August 2012. The AMP model has a spike in model probabilities in the late 1990s. It is only really from September 2012, that the AMP model has begun to dominate the alternative factor models, which has increased as the sample period has gone on. This finding suggests that it takes a considerable amount of time before the AMP model comes to dominate the alternative models.

I next examine factor redundancy regressions for the AMP model, which has the best performance at the higher prior multiples, and estimate the marginal contribution of each factor to the maximum squared Sharpe performance of the other factors in the model. Fama and French(2015) show that factor redundancy can be tested by running a time-series regression of the excess factor returns on a constant and the other factors in the model. A factor is redundant when the alpha ( $\alpha$ ) from the regression equals zero. Fama and French(2018) show that the marginal contribution of each factor to the maximum squared Sharpe performance of the other factors in the model is given by  $\alpha^2/\sigma(u)^2$ , where  $\sigma(u)$  is the residual volatility from the factor redundancy regressions.

Panel A of Table 5 reports the alphas, betas, and  $t$ -statistics in parentheses from the factor redundancy regressions using the factors in the AMP model. Panel B reports the posterior probabilities of the Bayesian test of the zero alpha null hypothesis for each factor in

the factor redundancy regressions. Panel C reports the residual volatility ( $\sigma(u_i)$ ) from the factor redundancy regressions and the marginal contribution of each factor to the maximum squared Sharpe performance of the other factors in the model ( $\alpha^2/\sigma(u)^2$ ).

Table 5 here

Panel A of Table 5 shows that none of the factors in the AMP model are redundant. All three factors have large significant positive alphas relative to the other factors in the model. Likewise the Bayesian test in panel B strongly rejects the null hypothesis of zero alphas for each factor. The posterior probabilities for a zero alpha null hypothesis are tiny. Each factor in the factor redundancy regressions has significant negative betas on the other two factors in the model. The negative factor betas coupled with the positive average excess returns on the factors lowers the expected excess returns on each factor, which leads to the positive alphas for each factor.

All three factors have a reasonable marginal contribution to the maximum squared Sharpe performance of the other factors in the model. The momentum factor has the largest marginal contribution. This result is driven by a large positive alpha and a low residual volatility. A similar pattern exists for the value factor. In contrast, the market factor has the largest positive alpha but has the smallest marginal contribution due to the higher residual volatility. Table 5 suggests that all three factors play an important role in the performance of the AMP model.

The analysis so far has compared the performance of a small number of global factor models and has not considered the performance of other factor models that could be formed

from the set of factors included in all the models<sup>10</sup>. I next examine performance tests of the AMP model compared to the best performing models that can be formed from the 13 factors used in the nine global models following a similar approach to Barillas and Shanken(2018). I split the sample period into two equal subperiods with an estimation window between November 1990 and November 2003 and an evaluation window between December 2003 and December 2016. I run the Bayesian multiple model comparison tests among all possible models that can be formed from the 13 factors in the estimation window and identify the top three models with the highest posterior probability. I use all four prior Sharpe multiples used in Table 4. I estimate the optimal tangency portfolio for the factors of the three best performing models and the AMP model in the estimation window. Using the optimal weights, I calculate the portfolio excess returns of each model in the evaluation window and then estimate the Sharpe performance for each model. I test for equal Sharpe performance measures between the AMP model and for each of the three models using the Ledoit and Wolf(2008)  $z$ -test.

Table 6 reports the performance tests of the AMP model relative to the best performing models in the estimation window. The table includes the posterior probabilities of the models in the estimation window, the maximum Sharpe performance in the estimation window (Shest), the Sharpe performance of the each model in the evaluation window (Shout), and the corresponding  $z$ -tests of equal Shout performance measures between the AMP model and the best performing models.

Table 6 here

---

<sup>10</sup> A recent study by Kozak, Nagel and Santosh(2018) argue against linear factor models with a small number of factors and estimate a stochastic discount factor model with a large number of characteristic factors.

Table 6 shows that the AMP model does not perform well in the Bayesian model comparison tests compared to the best performing models in the estimation window. The posterior probability of the AMP model is zero for every prior multiple Sharpe ratio used. The posterior probabilities of the best performing models are in excess of 0.134. The AMP model performs better using the Shest measure. At the prior multiple of 1.25, the AMP model has a higher Shest measure than Model 1. Comparing the posterior probabilities of the best performing models shows that selecting models by the posterior probability is not the same thing as selecting models with the highest maximum Sharpe performance (Barillas and Shanken(2018)). At the lower prior multiple Sharpe ratios, the best performing models tend to include fewer factors. However the model with the smallest number of factors in the best performing models (Model 1 at the prior multiple of 1.25) contains more factors than the AMP model at 7 factors. The most common factors excluded from the best performing models are the SMB factor and the two momentum factors WML and WML<sub>S</sub>. This result suggests that there is redundancy among the momentum factors.

Focusing on the out-of-sample Sharpe performance of the models, Table 6 shows that there is a substantial drop in the out-of-sample Sharpe performance relative to the Shest measures across all models. This pattern is consistent with the poor out-of-sample Sharpe performance of the sample tangency portfolio (DeMiguel, Garlappi and Uppal(2009)). The drop in Sharpe performance for the best performing models tends to be larger when the model contains more factors due to the higher estimation error due to the larger number of inputs required to estimate the tangency portfolio. Although there is a substantial drop in the Sharpe performance for the best performing models, all of the best models provide a significant higher Sharpe performance than the AMP model in the evaluation window. This result holds across all prior multiple Sharpe ratios. This finding suggests that although the AMP model performs

the best among the nine models considered, better performing models can be constructed from among the 13 factors. This result would caution the use of the AMP model in practical applications.

## **V Conclusions**

I use the Bayesian approach of Barillas and Shanken(2018) to examine the mean-variance efficiency of nine global factor models in global stock returns and to conduct model comparison tests. There are four main findings in my study. First, the mean-variance efficiency of each global factor model is rejected using the both the GRS test and B-GRS tests in for both sets of test assets. The factor models do a better job in pricing the size/BM portfolios compared to the size/momentum portfolios. Models that include a momentum factor have lower average mispricing in the size/momentum portfolios, which stems from the fact that the same characteristic is used to form the factor and test assets. The rejection of the mean-variance efficiency of global factor models is consistent with Fama and French(2012) and Cakici(2015) among others.

Second, in the Bayesian multiple model comparison tests among the nine global factor models, the best performing model is the AMP model at the higher prior multiple Sharpe ratios. The posterior probability of the AMP model is 0.918 at the prior multiple = 2 and 0.998 at the prior multiple = 3. At lower prior multiples the CAPM has the highest posterior probability. It takes a considerable amount of time for the AMP model to begin to dominate the alternative models. The Bayesian model comparison tests of relative model performance provides a different picture to comparing models by the average mispricing of the test assets. A number of models, e.g. Carhart, FF6, and AFIM have lower average mispricing in the test assets than the AMP and CAPM models but have a posterior probability of zero in the Bayesian tests across all prior Sharpe ratio multiples. This result is consistent with Barillas and

Shanken(2017) and highlights the importance of taking account of how well factor models price excluded factors in relative model comparison tests.

Third, in the factor redundancy tests of the AMP model, none of the factors are redundant. All three factors have large significant positive alphas. This result is driven by the negative betas each factor has on the other factors in the model. All three factors in the AMP models make a reasonable marginal contribution to the maximum squared Sharpe performance of the other factors in the models. The MOM and VAL factors make the largest marginal contribution due to the lower residual volatility.

Fourth, the AMP model significantly underperforms the best performing models that can be formed among the 13 factors used in the nine global factor models. The best performing models not only have a much higher posterior probability in the estimation window but also provides a significant higher out-of-sample Sharpe performance than the AMP model in the evaluation window. This finding suggests that although the AMP model performs the best in the model comparison tests among the nine global models at the higher prior multiple Sharpe ratios, there are other models that can be formed among the global factors that deliver superior performance. This finding would caution the use of the AMP model in practical applications.

My study has primarily focused on nine global factor models. An interesting extension would be to look at additional global factor models such as Cooper et al(2017) or a global version of Hou, Xue and Zhang(2015) or Stambaugh and Yuan(2017) models. The Bayesian model comparison tests could be adapted to compare global models versus regional versions of the models along the lines of Fama and French(2012, 2017). I have not considered the categorical factor model approach of Barillas and Shanken(2018). An examination of this approach among a large number of global factors is another interesting extension. I leave these issues to future research.

Table 1 Summary Statistics of the Test Assets

Panel A:					
<b>Size/BM</b>	Growth	2	3	4	Value
Small	0.165	0.444	0.693	0.742	1.028
2	0.201	0.451	0.572	0.666	0.747
3	0.267	0.432	0.545	0.584	0.736
4	0.422	0.476	0.514	0.622	0.643
Big	0.394	0.438	0.496	0.522	0.520
Panel B:					
<b>Size/Momentum</b>	Losers	2	3	4	Winners
Small	0.104	0.644	0.818	1.091	1.449
2	0.183	0.519	0.640	0.804	1.078
3	0.279	0.503	0.585	0.605	0.840
4	0.255	0.478	0.596	0.582	0.834
Big	0.154	0.379	0.466	0.583	0.628

The table reports the monthly average excess returns (%) of 25 global size/BM portfolios (Panel A) and 25 global size/momentum portfolios between November 1990 and June December 2016. The portfolios are sorted by size (Small to Big) in the rows and the BM ratio (Growth to Value) in the columns of panel A and momentum (Losers to Winners) in the columns of panel B.

Table 2 Summary Statistics of Factors

<b>Factor</b>	Mean	Standard Deviation	<i>t</i> -statistic
Market	0.479	4.232	2.00 <sup>1</sup>
SMB	0.147	1.942	1.34
HML	0.358	2.298	2.76 <sup>1</sup>
RMW	0.332	1.466	4.01 <sup>1</sup>
CMA	0.248	1.904	2.30 <sup>1</sup>
WML	0.584	3.944	2.62 <sup>1</sup>
HML <sub>S</sub>	0.570	2.627	3.84 <sup>1</sup>
RMW <sub>S</sub>	0.368	1.457	4.48 <sup>1</sup>
CMA <sub>S</sub>	0.339	1.777	3.38 <sup>1</sup>
WML <sub>S</sub>	0.789	3.814	3.66 <sup>1</sup>
HML <sub>T</sub>	0.376	2.941	2.26 <sup>1</sup>
VAL	0.224	1.724	2.31 <sup>1</sup>
MOM	0.322	2.169	2.63 <sup>1</sup>

<sup>1</sup> Significant at 5%

The table reports summary statistics of the factors in the global factor models between November 1990 and December 2016. The summary statistics include the mean and standard deviation of monthly excess returns (%). The final column reports the *t*-statistic of the null hypothesis that the average excess return of the factor equals zero.



Table 3 Tests of Mean-Variance Efficiency

Panel A:								
Size/BM	GRS	p(GRS)	1.2	1.4	1.6	1.8	2	$A \alpha_i $
CAPM	4.124	0	0.131	0.021	0.003	0.000	0.000	0.166
FF	3.644	0	0.006	0.000	0	0	0	0.115
Carhart	3.196	0	0.001	0	0	0	0	0.094
FF5	3.314	0	0	0	0	0	0.000	0.106
FF5s	3.773	0	0	0	0	0	0	0.157
FF6	3.116	0	0.000	0	0	0.000	0.001	0.095
FF6s	3.024	0	0.000	0	0.000	0.001	0.009	0.123
AFIM	3.585	0	0	0	0	0	0	0.121
AMP	3.196	0	0.000	0	0	0	0.000	0.129
Panel B:								
Size/Mom	GRS	p(GRS)	1.2	1.4	1.6	1.8	2	$A \alpha_i $
CAPM	5.241	0	0.083	0.007	0.000	0	0	0.302
FF	5.049	0	0.000	0	0	0	0	0.303
Carhart	4.246	0	0	0	0	0	0	0.120
FF5	5.220	0	0	0	0	0	0	0.245
FF5s	5.609	0	0	0	0	0	0	0.365
FF6	4.865	0	0	0	0	0	0	0.137
FF6s	4.230	0	0	0	0	0	0	0.158
AFIM	5.067	0	0	0	0	0	0	0.139
AMP	4.316	0	0	0	0	0	0	0.153

The table reports the GRS and B-GRS tests of mean-variance efficiency of nine global factor models in two different sets of test assets between November 1990 and December 2016. Panel A refers to the results using the 25 size/BM portfolios, and panel B refers to the results using 25 size/momentum (Mom) portfolios. The table includes the GRS test and the corresponding p value (p(GRS)). The next five columns include the posterior probabilities of the B-GRS test of the mean-variance efficiency of each factor model. The  $Sh_{max}$  is set equal to the prior multiple of  $Sh(F)$ , where  $Sh_{max}$  is the square root of the expected squared Sharpe performance with respect to the prior under the alternative that the alphas of the test assets are non-zero. The prior multiples are set equal to 1.2, 1.4, 1.6, 1.8, and 2. The final column reports the average absolute alpha ( $A|\alpha_i|$ ) of the test assets.

Table 4 Multiple Model Comparison Tests

<b>Model/Prior</b>	<b>1.25</b>	<b>1.5</b>	<b>2</b>	<b>3</b>
<b>Multiple</b>				
CAPM	0.991	0.867	0.080	0.001
FF	0	0	0	0
Carhart	0	0	0	0
FF5	0	0	0.001	0.0001
FF5s	0	0	0.000	0
FF6	0	0	0	0
FF6s	0	0	0	0
AFIM	0	0	0	0
AMP	0.009	0.133	0.918	0.998

The table reports the posterior probabilities from the Bayesian multiple model comparison tests between the nine global factor models between November 1990 and December 2016. The prior for  $Sh_{max}$  is set equal to the prior multiple\* $Sh(Mkt)$ .  $Sh_{max}$  is the square root of the expected maximum squared Sharpe performance with 13 factors included, taken with respect to the prior under the alternative that the alphas of the non-market factors are non-zero. Prior multiples of 1.25, 1.5, 2, and 3 are used.

Table 5 Factor Redundancy Tests of the AMP Model

Panel A	Market	VAL	MOM
$\alpha$	0.882 (3.63) <sup>1</sup>	0.439 (6.24) <sup>1</sup>	0.568 (6.59) <sup>1</sup>
$\beta$			
Market		-0.062 (-3.75) <sup>1</sup>	-0.105 (-5.25) <sup>1</sup>
VAL	-0.690 (-3.75) <sup>1</sup>		0.870 (-17.63) <sup>1</sup>
MOM	-0.767 (-5.25) <sup>1</sup>	-0.572 (-17.63) <sup>1</sup>	
Panel B			
Prior Multiple	Market	VAL	MOM
1.2	0.009	0	0
1.4	0.011	0	0
1.6	0.013	0	0
1.8	0.015	0	0
2	0.017	0	0
Panel C	Market	VAL	MOM
$\sigma(u)$	0.040	0.012	0.015
$\alpha^2/\sigma(u)^2$	0.047	0.130	0.143

This table reports factor redundancy tests for each factor in the AMP model between November 1990 and December 2016. Panel A reports the alpha ( $\alpha$ ), betas ( $\beta$ ), and  $t$ -statistics in parentheses from the time-series regressions of the factors in the columns on a constant and the excess returns of the other two factors of the model. Panel B reports the posterior probabilities of the Bayesian test of the null hypothesis that the alpha of the factor equals zero. The prior for  $Sh_{\max}$  is set equal to the prior multiple of the maximum Sharpe performance of the other factors in the model, where  $Sh_{\max}$  is the square root of the expected squared Sharpe performance with respect to the prior under the alternative that the alpha of the factor is non-zero. The prior multiples are set equal to 1.2, 1.4, 1.6, 1.8, and 2. Panel C includes the residual volatility ( $\sigma(u)$ ) from each regression, and the marginal contribution of each factor to the maximum squared Sharpe performance of the other factors in the model ( $\alpha^2/\sigma(u)^2$ ).

Table 6 Performance Tests of the AMP Model Against the Best Performing Models

Panel A:				
<b>Prior Multiple=1.25</b>	Prob	Shest	Shout	z-test
Model 1	0.244	0.498	0.402	1.92 <sup>2</sup>
Model 2	0.193	0.581	0.420	2.43 <sup>1</sup>
Model 3	0.181	0.706	0.442	3.61 <sup>1</sup>
AMP	0	0.558	0.274	
Panel B:				
<b>Prior Multiple=1.5</b>	Prob	Shest	Shout	z-test
Model 1	0.254	0.706	0.442	3.61 <sup>1</sup>
Model 2	0.190	0.618	0.366	2.41 <sup>1</sup>
Model 3	0.179	0.498	0.402	1.92 <sup>2</sup>
AMP	0	0.558	0.274	
Panel C:				
<b>Prior Multiple=2</b>	Prob	Shest	Shout	z-test
Model 1	0.359	0.706	0.442	3.61 <sup>1</sup>
Model 2	0.205	0.618	0.366	2.41 <sup>1</sup>
Model 3	0.135	0.820	0.499	4.48 <sup>1</sup>
AMP	0	0.558	0.274	
Panel D:				
<b>Prior Multiple=3</b>	Prob	Shest	Shout	z-test
Model 1	0.417	0.820	0.499	4.48 <sup>1</sup>
Model 2	0.243	0.706	0.442	3.61 <sup>1</sup>
Model 3	0.105	0.740	0.443	3.65 <sup>1</sup>
AMP	0	0.558	0.274	

<sup>1</sup> Significant at 5%

<sup>2</sup> Significant at 10%

The table conducts performance tests between the AMP model and the three best performing models that can be formed from the 13 global factors in the Bayesian model comparison tests in the November 1990 and November 2003 estimation window. Models 1 to 3 are ranked by the highest posterior probability in the Bayesian tests. The table reports the posterior probabilities (Prob) and maximum Sharpe performance for each model in the estimation window (Shest). The Shout column is the Sharpe performance of each model in the December 2003 and December 2016 evaluation window using the optimal tangency portfolio weights from the estimation window. The final column is the z-test of Ledoit and Wolf(2008) of equal Sharpe performance measures between the AMP model and each of the best performing models. The prior for  $Sh_{\max}$  is set equal to the prior multiple\* $Sh(\text{Mkt})$ .  $Sh_{\max}$  is the square root of the expected maximum squared Sharpe performance with 13 factors included, taken with respect to the prior under the alternative that the alphas of the non-market factors are non-zero. Prior multiples of 1.25, 1.5, 2, and 3 are used.

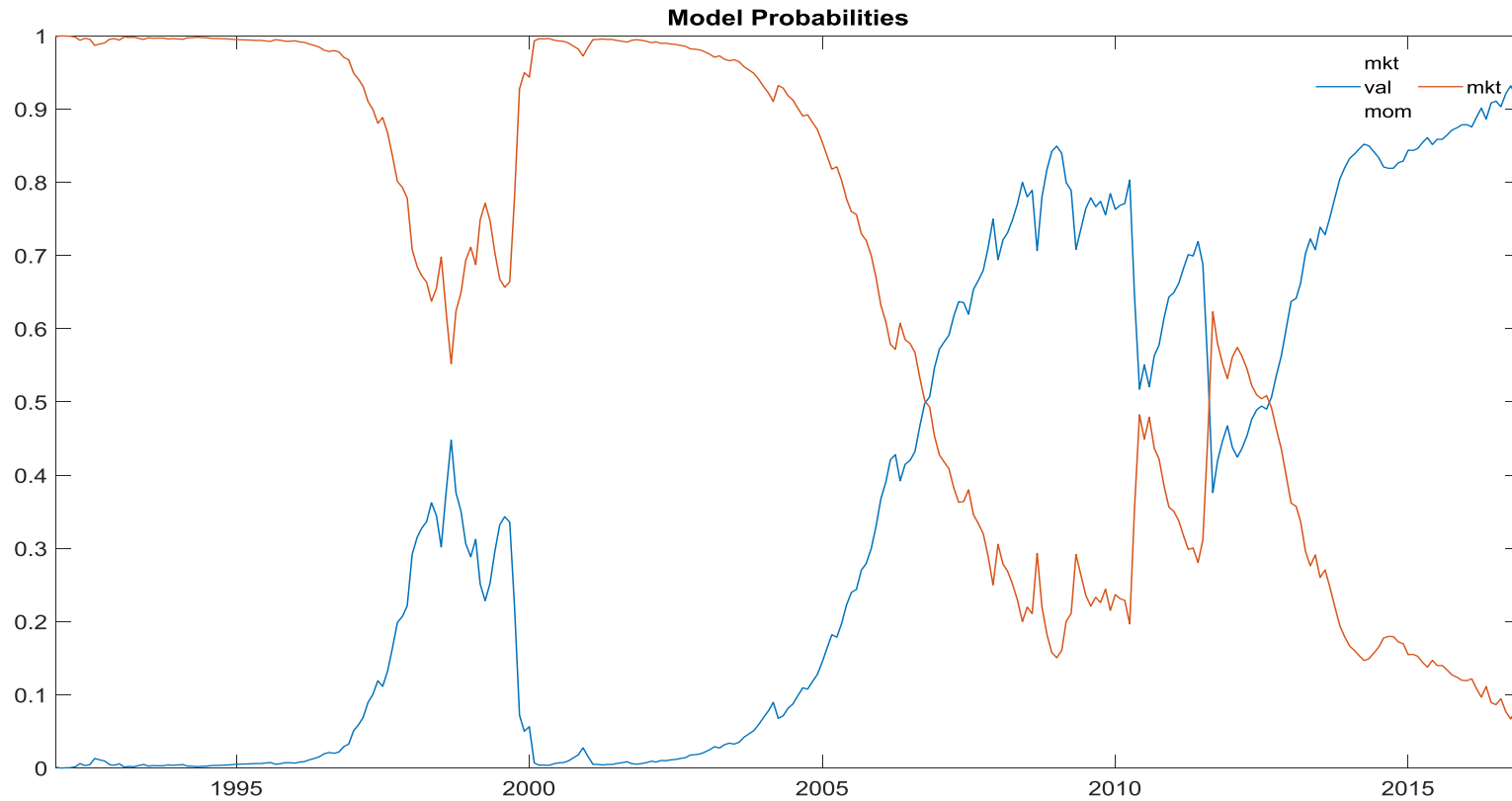


Figure 1. Model Probabilities. The figure plots the time-series of posterior probabilities for the two models with the highest probability at the end of the sample period. Recursive sample periods are used beginning in December 1991 and December 2016. The prior for  $Sh_{\max}$  is set equal to  $2 \cdot Sh(\text{Mkt})$ .  $Sh_{\max}$  is the square root of the expected maximum squared Sharpe performance with 13 factors included, taken with respect to the prior under the alternative that the alphas of the nonmarket factors are non-zero.

## References

- Adler, M. and B. Dumas, 1983, International portfolio selection and corporation finance: A synthesis, *Journal of Finance*, 38, 925-984.
- Asness, C. and A. Frazzini, 2013, The devil in HML's details, *Journal of Portfolio Management*, 39, 49-68.
- Asness, C.S., Frazzini, A., Israel, R. and T. Moskowitz, 2015, Fact, fiction, and value investing, *Journal of Portfolio Management*, 42, 34-52.
- Asness, C.S., Moskowitz, T. and L.H. Pedersen, 2013, Value and momentum everywhere, *Journal of Finance*, 68, 929-985.
- Barillas, F., Kan, R., Robotti, C. and J. Shanken, 2017, Model comparison with Sharpe ratios, *Working Paper*, University of Toronto.
- Barillas, F. and J. Shanken, 2017, Which alpha?, *Review of Financial Studies*, 30, 1316-1338.
- Barillas, F. and J. Shanken, 2018, Comparing asset pricing models, *Journal of Finance*, forthcoming.
- Brusa, F., Ramadorai, T. and A. Verdelhan, 2014, The international CAPM redux, *Working Paper*, University of Oxford.
- Cakici, N., 2015, The five-factor Fama-French model: International evidence, *Working Paper*, Fordham University.
- Carhart, M. M., 1997, Persistence in mutual fund performance. *Journal of Finance*, 52, 57-82.
- Chen, N.F., Roll, R. and S.A. Ross, 1986, Economic forces and the stock market, *Journal of Business*, 59, 383-403.
- Chretien, S. and M. Kammoun, 2017, Mutual fund performance evaluation and best clienteles, *Journal of Financial and Quantitative Analysis*, 52, 1577-1604.
- Cochrane, J.H. and J.Saa-Requejo, 2000, Beyond arbitrage: Good-deal asset price bounds in incomplete markets, *Journal of Political Economy*, 108, 79-119.

Cooper, I., Mittrache, A. and R. Priestley, 2017, A global macroeconomic risk model for value, momentum, and other asset classes, *Working Paper*, BI Norwegian Business School.

DeMiguel, V., Garlappi, L. and R. Uppal, 2009, Optimal versus naïve diversification: How inefficient is the 1/N portfolio strategy?. *Review of Financial Studies*, 22, 1915,1953.

De Santis, G. and B. Gerard, 1998, How big is the premium for currency risk?, *Journal of Financial Economics*, 49, 375-412.

Dumas, B. and B. Solnik, 1995, The world price of foreign exchange rate risk, *Journal of Finance*, 50, 445-479.

Fama, E.F., 2015, Cross-section versus time-series tests of asset pricing models, *Working Paper*, University of Chicago.

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

Fama, E.F. and K.R. French, 1998, Value versus growth: The international evidence, *Journal of Finance*, 53, 1975-1999.

Fama, E.F. and K.R. French, 2012, Size, value, and momentum in international stock returns, *Journal of Financial Economics*, 105, 457-472.

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

Fama, E.F. and K.R. French, 2016, Dissecting anomalies with a five-factor model, *Review of Financial Studies*, 29, 69-103.

Fama, E.F. and K.R. French, 2017, International tests of a five-factor asset pricing model, *Journal of Financial Economics*, 123, 441-463.

Fama, E.F. and K.R. French, 2018, Choosing factors, *Journal of Financial Economics*, 128, 234-252.

Fama, E.F., and J.D. MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of Political Economy*, 71, 607-636.

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

Gospodinov, N., Kan, R. and C. Robotti, 2013, Chi-squared tests for evaluation and comparison of asset pricing models, *Journal of Econometrics*, 173, 108-125.

Grauer, F.L.A., Litzenberger, R.H. and R.E. Stehle, 1976, Sharing rules and equilibrium in an international capital market under uncertainty, *Journal of Financial Economics*, 3, 233-256.

Hansen, L.P. and R. Jagannathan, 1997, Assessing specification errors in stochastic discount factor models, *Journal of Finance* 52, 591-607.

Harvey, C.R., 2017, The scientific outlook in financial economics, *Journal of Finance*, 72, 1399-1440.

Harvey, C.R. and G. Zhou, 1990, Bayesian inference in asset pricing tests, *Journal of Financial Economics*, 26, 221-254.

Hou, K., Karolyi, G.A., and B.C. Kho, 2011, What factors drive global stock returns?, *Review of Financial Studies*, 24, 2527-2574.

Hou, K., Xue, C. and L. Zhang, 2015, Digesting anomalies: An investment approach, *Review of Financial Studies*, 28, 650-705.

Hou, K., Xue, C. and L. Zhang, 2017, Replicating anomalies, *Working Paper*, Ohio State University.

Huang, D. and G. Zhou, 2017, Upper bounds on return predictability, *Journal of Financial and Quantitative Analysis*, 52, 401-425.

Kan, R., Robotti, C. and J. Shanken, 2013, Pricing model performance and the two-pass cross-sectional regression methodology, *Journal of Finance*, 68, 2617-2649.



Karolyi, G.A. and R. Stulz, 2003, Are financial assets priced locally or globally?, In the Handbook of the Economics of Finance, Eds Constantinides, G., Harris, M. and R.M. Stulz, 975-1020, Elsevier.

Kozak, S., Nagel, S. and S. Santosh, 2018, Shrinking the cross-section, *Working Paper*, University of Chicago.

Ledoit, O. and M. Wolf, 2008, Robust performance hypothesis testing with the Sharpe ratio, *Journal of Empirical Finance*, 15, 850-859.

Lewis, K.K., 2011, Global asset pricing, *Annual Review of Financial Economics*, 3, 435-466.

Lustig, H., Roussanov, N. and A. Verdelhan, 2011, Common risk factors in currency markets, *Review of Financial Studies*, 24, 3731-3777.

MacKinlay, A.C. 1995, Multifactor models do not explain deviations from the capital asset pricing model, *Journal of Financial Economics*, 38, 3-28.

McCulloch, R. and P.E. Rossi, 1990, Posterior, predictive, and utility-based approaches to testing the arbitrage pricing theory, *Journal of Financial Economics*, 28, 7-38.

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

Pastor, L. and R.F. Stambaugh, 2000, Comparing asset pricing models: An investment perspective, *Journal of Financial Economics*, 56, 335-381.

Shanken, J., 1987, A Bayesian approach to testing portfolio efficiency, *Journal of Financial Economics*, 19, 195-216.

Sharpe, W.F., 1966, Mutual fund performance, *Journal of Business* 39, 119-138.

Solnik, B., 1983, International arbitrage pricing theory, *Journal of Finance*, 38, 449-457.

Stambaugh, R.F. and Y. Yuan, 2017, Mispricing factors, *Review of Financial Studies*, 30, 1270-1315.

Zhang, X., 2006, Specification tests of international asset pricing models, *Journal of International Money and Finance*, 25, 275-307.