HOW MANY FACTORS ARE IMPORTANT IN U.K. STOCK RETURNS?

Jonathan Fletcher

The author is from the University of Strathclyde.

Key Words: Linear factor models, Multiple Testing, Bootstrap

JEL Classification: G11, G12

Current draft: February 2019

I am grateful for the use of the Matlab code on Campbell Harvey's web site. I am grateful for the comments of two anonymous reviewers and an Associate Editor.

Address correspondence to Professor J. Fletcher, Department of Accounting and Finance, University of Strathclyde, Stenhouse Building, 199 Cathedral Street, Glasgow, G4 0QU, United Kingdom, phone: +44 (0) 141 548 4963, fax: +44 (0) 552 3547, email: j.fletcher@strath.ac.uk

HOW MANY FACTORS ARE IMPORTANT IN U.K. STOCK RETURNS?

ABSTRACT

I use the sequential approach of Harvey and Liu(2018) to build linear factor models in U.K. stock returns among a set of 13 candidate factors using individual stocks and three groups of test portfolios between July 1983 and December 2017. My study finds that the Market factor is the dominant factor in reducing mispricing in individual stocks and test portfolios regardless of the pricing error metric used. The Market factor has a bigger impact when using a value weighting pricing error metric. Whether a second factor is used or not depends upon which metric is used for mispricing and the time period examined. My study finds support for a two-factor model for the whole sample period of the Market factor and the Conservative Minus Aggressive (CMA) factor of Fama and French(2015) when giving greater weight to the mispricing of larger companies.

I Introduction

Linear factor models such as the capital asset pricing model (CAPM) and arbitrage pricing theory (APT) play an important role in a number of practical applications in Finance. These applications include evaluating fund performance (Fama and French(2010)), estimating expected returns (Sarisoy, Goeij and Werker(2017)), optimal portfolio choice (Raponi, Uppal and Zaffaroni(2018)), factor investing (Ang(2014)), and cost of equity capital estimation. A central prediction of the CAPM and APT is that either the market portfolio (Roll(1977)) or a combination of the K factor portfolios lie on the ex ante mean-variance frontier.

The Gibbons, Ross and Shanken(1989) test is used most often to examine the mean-variance efficiency of different factor models (Fama and French(2015,2016,2018)). Empirical studies by Fama and French and Hou, Xue and Zhang(2015,2019) among others also compare different linear factor models using summary metrics of the pricing errors of the test assets. A recent study by Barillas and Shanken(2017) provide a critique of using summary metrics of pricing errors of test assets for relative model comparison. Barillas and Shanken show that when conducting relative model comparison tests the choice of test assets is irrelevant for a number of metrics[1]. The only relevant issue in relative model comparison tests is how well a factor model prices factors not included in the model[2].

---

[1] The metrics include evaluating models by the increase in Sharpe(1966) performance, the modified Hansen and Jagannathan(1997) distance measure, and statistical likelihoods.

[2] Building on Barillas and Shanken(2017), Barillas and Shanken(2018) develop a Bayesian model comparison testing framework. Barillas, Kan, Robotti and Shanken(2019) derive the asymptotic distribution of comparing linear factor models using the Sharpe ratio metric. See also Fama and French(2018) who use a bootstrap approach for comparing different six-factor models.

A recent study by Harvey and Liu(2018) argues that the choice of test assets is important when the number of assets (N) is large contrary to Barillas and Shanken(2017). When N is large relative to the number of time-series observations (T), in-sample arbitrage opportunities can arise and the ex post maximum squared Sharpe performance of the test assets and factors can become infinitely large and so the comparison of factor models using the Sharpe ratio metric becomes less meaningful. Harvey and Liu propose a sequential approach to build a factor model that does not assume correct model specification and controls for multiple testing[3]. A factor is only included in a model if it leads to a significant reduction in the mispricing of the test assets.

I use the Harvey and Liu(2018) sequential approach to build linear factor models in U.K. stock returns between July 1983 and December 2017 using individual stocks and three groups of different test portfolios. I build the factor model from a set of 13 candidate factors. My choice of factors follows from recent studies of Fama and French(2015, 2018), Frazzini and Pedersen(2014), and Daniel, Hirshleifer and Sun(2018). To calculate the reduction in pricing errors, I use the same metrics as in Harvey and Liu, which are based on the percentage difference between the scaled alphas of an augmented model and a baseline model. I use the mean scaled absolute alpha (SI_{ew}), the median scaled absolute alpha (SI_{med}), and the mean value weighted scaled absolute alpha (SI_{vw}) measures[4].

[3] Controlling for multiple testing is important given the large number of factors which have been found to be important in cross-sectional stock returns (Harvey, Liu and Zhu(2016), Green, Hand and Zhang(2017)).

[4] Harvey and Liu(2018) divide the alphas from the augmented and baseline models by the standard error of the alpha from the baseline model. This scaling has a number of advantages, which will be discussed later in the paper.

There are three main findings in my study. First, using either the SI_{ew} or SI_{med} measures, the Market factor is the dominant factor in reducing the scaled pricing errors of the individual stocks and test portfolios. For the individual stocks and two out of the three groups of test portfolios, none of the additional candidate factors make a significant reduction in scaled pricing errors beyond the baseline model of the Market factor after controlling for multiple testing. Second, using the SI_{vw} measure, the Market factor is again the dominant factor in reducing the scaled pricing errors of the individual stocks. The magnitude of the SI_{vw} measure is more than double the SI_{ew} or SI_{med} measures using individual stocks. This result suggests that the Market factor does a better job in reducing mispricing in larger stocks (Harvey and Liu(2018)). Third, using the SI_{vw} measure, there is some support for the inclusion of a second factor into the factor model but this result depends upon the sample period. My results suggest that there are at most two factors, of which one is the Market factor, that are important in reducing the mispricing in U.K. stock returns after controlling for multiple testing.

My study makes two main contributions to the literature. First, I complement the recent study of Harvey and Liu(2018) by using their sequential approach to build linear factor models in U.K. stock returns. Recent studies by Harvey(2017), Hou et al(2019) highlight the importance of replication studies in Finance, which is common in other fields of science. Second, I extend the prior literature of asset pricing studies in U.K. stock returns such as Fletcher(1994,2001,2018), Davies, Fletcher and Marshall(2015), Gregory, Tharyan and Christidis(2013), and Michou and Zhou(2016) among others. I extend this literature by using the sequential approach of Harvey and Liu to build a factor model and also controlling for multiple testing.

The paper is organized as follows. Section II describes the research method. Section III discusses the data. Section IV presents the empirical results and the final section concludes.

II Research Method

Linear factor models such as the CAPM and APT imply the following relation between expected excess returns and risk as:

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

where r_i is the excess return on asset i, β_{ik} is the beta of asset i with respect to factor k, λ_k is the risk premium of factor k, and K is the number of factors in the model. When the K factors are portfolio returns, equation (1) can be tested by the following time-series regression[5]:

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

where r_{it} is the excess return of asset i at time t, r_{kt} is the excess return of factor k at time t, α_i is the pricing error of asset i[6], N is the number of test assets, and u_{it} is a random error term at time t with $E(u_{it}) = 0$, and $E(u_{it}r_{kt}) = 0$ for all K factors. In the time-series regression approach, $\lambda_k = E(r_{kt})$, i.e. the average excess return of the factor (Shanken(1992)). If the factor model is well specified, equation (1) implies the following null hypothesis for the N α_i's in equation (2) given by:

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

The most popular test of equation (3) is the Gibbons et al(1989) test. The Gibbons et al test is given by:

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

where α is a (N,1) vector of the N α_i's, Σ is the (N,N) residual covariance matrix from equation (2) (Maximum Likelihood (ML)) estimate, and $Sh^2(f)$ is the maximum squared Sharpe(1966) performance of the K factors. Gibbons et al(1989) show that the $\alpha'\Sigma^{-1}\alpha$ term in equation (4) can be written as $Sh^2_{max} - Sh^2(f)$, where Sh^2_{max} is the maximum squared Sharpe performance

[5] Fama(2015) reviews the time-series regression approach and the alternative two-pass cross-sectional regression approach (Fama and MacBeth(1973)) to evaluate linear factor models.

[6] In fund performance, α_i is known as the Jensen(1968) performance measure.

of the N+K risky assets. The GRS test compares the maximum squared Sharpe performance attainable by the K factors relative to the maximum squared Sharpe performance attainable by the N+K assets. Under the assumption that the residuals from equation (2) have a multivariate normal distribution with zero mean and constant covariance matrix, the GRS test has a central F distribution with N and T-N-K degrees of freedom.

A number of recent empirical studies such as Fama and French(2015,2016,2018), and Hou et al(2015,2018) use the GRS test and summary metrics based on the α_i's of the N test assets to compare the performance of different factor models. Barillas and Shanken(2017) provide a critique of using the N test asset α_i's for relative model comparison. Barillas and Shanken show that the choice of test assets is irrelevant in relative model comparison tests when using the time-series regression approach to evaluate factor models. To illustrate, with the $Sh^2_{max} – Sh^2(f)$ metric, one model performs better than another model when their $Sh^2_{max} – Sh^2(f)$ measure is lower. If we include all factors from the different models as part of the investment universe, then Sh^2_{max} is fixed across models. As a result, models will only differ by their $Sh^2(f)$ measures and so the choice of test assets is irrelevant.

Harvey and Liu(2018) provide a critique of the GRS test and a counter-argument to the test asset irrelevance of Barillas and Shanken(2017). The first issue of the GRS test is that when N > T, the inverse of the residual covariance matrix does not exist. A second problem is that the GRS test cannot be used to compare models as Σ^{-1} varies across models. A third problem is that when N > T, in-sample arbitrage opportunities can exist and so Sh^2_{max} can approach ∞. Harvey and Liu also point out that possible in-sample arbitrage opportunities makes the choice of test assets important in relative model comparison. Consider two models A and B. Model A has K1 factors and model B has K2 factors, and K is the union of all the factors in each model. When N > T, Sh^2_{max} can be infinite and so the difference $Sh^2_{max} – Sh^2(f)$

for both models can be infinite and so we can no longer discern whether one model is better than another.

Rather than trying to conduct model specification tests as in equation (2), Harvey and Liu(2018) develop a sequential approach to build a factor model, by only including factors which lead to a significant reduction in pricing errors of the test assets. Their approach builds a factor model which only includes useful factors but does not require correct model specification[7]. The key issue is being able to control for multiple testing (White(2000)) when selecting factors given the large number of factors found to be important in explaining cross-sectional stock returns. The test assets can either be individual stocks[8] or portfolios.

Define K as the number of candidate factors. Harvey and Liu(2018) build their factor model by starting with a baseline model of a Constant. Harvey and Liu then create K augmented factor models by adding each factor separately to the baseline model. The alphas from equation (2) of the N test assets are estimated for the baseline and augmented models. The factor with the largest reduction in alphas relative to the baseline model is identified based on some metric of the pricing errors of the test assets. The statistical significance of this factor is then evaluated using a bootstrap approach. If the factor provides a significant reduction in pricing errors, it is added to the baseline model. We repeat the steps above with a new baseline model and only add a factor to the baseline model if it provides a significant incremental

[7] Empirical methods that allow for model misspecification are advocated by Kan and Robotti(2009), Ludvigson(2013), and Gospodinov, Kan and Robotti(2013) among others.

[8] There have been a number of empirical studies using individual stocks under the two-pass cross-sectional regression method including Jegadeesh, Noh, Pukthuanthong, Roll and Wang(2018), Bai and Zhou(2015), Kim and Skoulakis(2018) and Raponi, Robotti and Zaffaroni(2018) among others.

reduction in pricing errors beyond the baseline model after controlling for multiple testing. This process continues until there is no longer any factor that provides a significant incremental reduction in pricing errors beyond the baseline model after controlling for multiple testing.

Harvey and Liu(2018) propose three different pricing error metrics to evaluate adding candidate factors to a baseline model as:

$$SI_{ew} = [(1/N)\Sigma_{i=1}^{N}(|\alpha_i^a|-|\alpha_i^b|)/s_i^b]/[(1/N)\ \Sigma_{i=1}^{N}|\alpha_i^b|/s_i^b] \qquad (5)$$

where SI is the scaled intercept, α_i^b and α_i^a are the alphas from equation (2) under the baseline model and the augmented model, and s_i^b is the standard error of α_i^b. The SI_{ew} measure is the percentage difference of the average scaled absolute alphas between the augmented and baseline models. If a candidate factor contributes to reducing the scaled pricing errors in the test assets beyond a baseline model then SI_{ew} will be negative.

The second metric used is:

$$SI_{med} = [(median(|\alpha_i^a|/s_i^b)-median(|\alpha_i^b|/s_i^b\ )]/[median(|\alpha_i^b|/s_i^b)] \qquad (6)$$

where med is the median. The SI_{med} measure is the percentage difference in the median absolute scaled alphas between the augmented and baseline models. If a factor makes an incremental contribution to reducing the scaled pricing errors beyond the baseline model, then SI_{med} will be negative. The SI_{med} measure is more robust to outliers than the SI_{ew} measure, which is particularly important when the test assets are individual stocks (Harvey and Liu(2018)).

The third metric is given by:

$$SI_{vw}=[((1/T)\Sigma_{t=1}^{T}\Sigma_{i=1}^{N}w_{it}*sa_i^a)-((1/T)\Sigma_{t=1}^{T}\Sigma_{i=1}^{N}(w_{it}*sa_i^b)]/[((1/T)\ \Sigma_{t=1}^{T}\Sigma_{i=1}^{N}(w_{it}*sa_i^b)] \quad (7)$$

where vw is value weighting, $sa_i^a = |\alpha_i^a|/s_i^b$, $sa_i^b = |\alpha_i^b|/s_i^b$, and $w_{it} = x_{it}/M_t$, where x_{it} is the Market value of asset i at time t, and M_t is the Market value of all risky assets. The SI_{vw} measure is the percentage difference in average value weighted scaled absolute alphas between the augmented and baseline models. If a factor makes an incremental contribution to pricing beyond the baseline model, SI_{vw} measure should be negative. The SI_{vw} measure places a larger weight on the mispricing of larger companies, compared to either the SI_{ew} or the SI_{med} metrics.

Harvey and Liu(2018) point out that using scaled alphas has a number of advantages compared to using metrics based on actual alphas, such as the mean absolute alpha. Using scaled alphas takes account of heterogeneity in the residual volatilities across the test assets and allows the test assets to have a different number of time-series observations[9]. By scaling alphas by the standard errors from the baseline model ensures that all three metrics equal zero under the null hypothesis that the candidate factors have a zero incremental contribution in reducing the mispricing of the test assets beyond the baseline model. Harvey and Liu also find using scaled alphas is more powerful in finding useful factors than metrics based on alphas.

To evaluate the statistical significance of the SI_{ew}, SI_{med}, and SI_{vw} measures, Harvey and Liu(2018) use a bootstrap approach[10]. The bootstrap approach derives the empirical distributions of the SI_{ew}, SI_{med}, and SI_{vw} measures for each factor. This test gives us a single-test p value for the null hypothesis that the factor has a zero incremental contribution of reducing the pricing errors of the test assets relative to the baseline model. The bootstrap approach also derives the empirical distribution of the minimum SI_{ew}, SI_{med}, and SI_{vw} measures across factors. This test gives us a multiple-test p value of the null hypothesis that all of the factors have a zero incremental reduction in pricing errors beyond the baseline model.

III Data

---

[9] Studies by Kosowski, Timmermann, Wermers and White(2006), Fama and French(2010), and Ferson and Chen(2015) among others use the *t*-statistic of alpha when evaluating cross-sectional U.S. mutual fund performance.

[10] See Harvey and Liu(2018) for more details.

In most of my empirical analysis, I use individual stocks as the test assets. My sample period covers January 1983 and December 2017 due to the availability of some of the factors. All of my data is collected from the London Share Price Database (LSPD) provided by the London Business School, unless otherwise specified. The LSPD provides monthly returns on all companies listed on the U.K. markets. I collect the one-month U.K. Treasury Bill return as the risk-free asset from LSPD and Thompson Financial Datastream. I exclude investment trusts, foreign companies, and secondary shares. I run the tests using all stocks with at least 36 return observations either in the actual data and the simulated data as in Harvey and Liu(2018). I also use three groups of test portfolios as the test assets. I use 16 size/book-to-Market (BM) portfolios, 10 industry portfolios, and 16 size/momentum portfolios. Details on how the test portfolios are formed is available on request.

My set of factors is motivated by the recent studies of Fama and French(2015,2018), Frazzini and Pedersen(2014), and Daniel et al(2018). I use 13 candidate factors in my study. Details on how the factors are formed is available on request. The first group of factors stem from the Fama and French(1993,2015) and the Carhart(1997) models and include the excess returns on the Market factor and zero-cost portfolios for the size (SMB), value (HML), profitability (RMW), investment (CMA), and momentum (WML) effects in stock returns. The second group of factors stem from Fama and French(2018) and include the small spread factors for value (HML_S), profitability (RMW_S), investment (CMA_S), and momentum (WML_S) factors. The third group of factors include the betting against beta (BAB) factor of Frazzini and Pedersen(2014), and the financial (FIN) and post earnings announcement drift (PEAD) factors similar to Daniel et al(2018).

My study only captures a subset of the factors that have been proposed in the literature. Pastor and Stambaugh(2003) propose an aggregate liquidity factor. An alternative model to the Fama and French(2015) five-factor model is the four-factor model proposed by Hou et al(2015) based on the q-theory of investment. Hou, Mo, Xue and Zhang(2019) examine the empirical performance of their q-factor model relative to a number of alternative factor models. Hou, Mo, Xue and Zhang(2018) extend the q-factor model to include an expected growth factor. Stambaugh and Yuan(2017) propose a four-factor model, which includes the Market and size factors, along with two mispricing factors constructed from 11 Market anomalies. All of these factors are worth exploring in a future study.

Table 1 reports the mean and standard deviation of the factor excess returns. The final column reports the *t*-statistic of the null hypothesis that the average excess factor returns equal zero.

Table 1 here

Table 1 shows that the vast majority of factors have significant positive mean excess returns at the 5% significance level, except the SMB and RMW factors. The RMW factor has a significant positive mean excess return at the 10% level. The WML_S and WML factors have the highest positive mean excess returns at 1.209% and 0.909% respectively. The difference in mean excess returns between the WML_S and WML factors highlights the stronger momentum effect in smaller companies. There are significant value, profitability, and investment factors. The mean excess returns are higher for the small spread factors. Among the BAB, FIN, and PEAD factors, the BAB factor has the highest mean excess return but also the highest volatility across all factors. Although most of the factors have significant positive average excess returns, it is only the CMA, CMA_S, WML, WML_S, and FIN factors which have a *t*-statistic higher than 3, which is recommended by Harvey et al(2016) to take account of multiple testing.

IV Empirical Results

I begin my empirical analysis by using the sequential approach of Harvey and Liu(2018) to build a factor model using the size/BM, industry, and size/momentum portfolios as the test assets. Tables 2 and 3 report the results using the SI_{ew} and SI_{med} measures. To conserve space, I only report the results here for where there is a factor which leads to a significant reduction in the scaled pricing errors relative to the baseline model. Table 2 includes the results for the size/BM portfolios (panel A), and industry portfolios (panel B), and Table 3 includes the results of the size/momentum portfolios. In Table 2 and panel A of Table 3, the baseline model is a Constant. Each panel reports the SI_{ew} and SI_{med} measures (%) for each factor, the fifth (5%) percentile from the bootstrap distribution of SI_{ew} and SI_{med} measures of each factor, and the corresponding single-test p values. The bottom of the panel reports the 5% percentile from the bootstrap distribution of the minimum SI^m_{ew} and SI^m_{med} measures across all the candidate factors and the multiple-test p value.

Table 2 here

Table 3 here

Table 2 and panel A of Table 3 shows that the Market factor has the largest reduction in scaled pricing errors among the candidate factors for all three groups of test portfolios. This result holds using both the SI_{ew} and SI_{med} measures. The reduction in scaled pricing errors is massive for the Market factor and equal to 67.4% (size/BM), 81.3% (industry), and 32.1% (size/momentum) for the SI_{ew} measure. The large reduction in scaled pricing errors for the Market factor stands in sharp contrast to the poor performance of the CAPM in the size/BM portfolios as in Fama and French(1993).

The SI_{ew} and SI_{med} measures for the Market factor in Table 2 and panel A of Table 3 are all highly statistically significant using the single-test p values. There are very few other factors, which have significant SI_{ew} and SI_{med} measures across the three groups of test portfolios. The HML factor has a significant SI_{ew} measure for the industry portfolios and a significant SI_{med} measure for the size/momentum portfolios. The BAB factor has significant SI_{ew} and SI_{med} measures for all three groups of test portfolios. The multiple-test p values in Table 2 and panel A of Table 3 show that the Market factor has a significant reduction in scaled pricing errors beyond the baseline model of a Constant after controlling for multiple testing.

I repeat the tests using the Market factor as the baseline model for the three groups of test portfolios. For the size/BM and industry portfolios, no other factor makes a significant incremental reduction in scaled pricing errors beyond the baseline model of the Market factor. As a result, for these test portfolios our final model is a single-factor model. For the size/momentum portfolios in panel B of Table 3, there are a number of factors with significant SI_{ew} and SI_{med} measures relative to the baseline model of the Market factor. The RMW, RMW_S, WML, WML_S, and PEAD factors all have significant SI_{ew} and SI_{med} measures. The WML_S factor has the largest reduction in scaled pricing errors among the factors and is highly statistically significant even after controlling for multiple testing. As a result, the WML_S factor is added to the baseline model.

I repeat the tests in Table 3 using the baseline model of the Market and WML_S factors. It is only the HML factor that has a significant SI_{ew} and SI_{med} measures using the single-test p values. However the statistical significance of the HML factor does not hold after controlling for multiple testing. The final model for the size/momentum portfolios is a two-factor model.

Tables 2 and 3 suggest that the Market factor is the dominant factor in reducing the scaled pricing errors of the test portfolios. This finding is consistent with Harvey and Liu(2018). The addition of the WML_S factor when the test assets are the size/momentum portfolios is not a surprise given that the same characteristic is used to form the test assets and the factor[11]. I next proceed to build a factor model using individual stocks as the test assets. Table 4 reports the results using the SI_{ew} and SI_{med} measures. As before in Tables 2 and 3, in panel A of Table 4 the baseline model is a Constant.

Table 4 here

Panel A of Table 4 shows that the Market factor has the largest reduction in pricing errors across the candidate factors using both the SI_{ew} and SI_{med} measures. The reduction in scaled absolute alphas by the Market factor is 17.8% for the SI_{ew} measure and 22.1% for the SI_{med} measure. Both reductions are significant with the single-test p values. The only other factor which has a significant reduction in scaled pricing errors relative to the baseline model is the BAB factor. Controlling for multiple testing, the Market factor continues to provide a significant reduction in scaled pricing errors with a p value of 0.003 for the SI_{ew} measure and 0.001 for the SI_{med} measure. The dominant role of the Market factor using the SI_{ew} and SI_{med} measures is consistent with Harvey and Liu(2018).

In panel B of Table 4, the tests are repeated but this time the baseline model includes the Market factor. Panel B shows that only two factors, SMB and BAB, have negative SI_{ew} and SI_{med} measures and all are tiny. None of the factors lead to a significant reduction in scaled

[11] Using test portfolios has a number of problems in asset pricing tests. The mispricing of individual stocks can cancel out in test portfolios, results can be sensitive to the portfolio formation method used, portfolios can be subject to the data snooping bias of Lo and MacKinlay(1990), and test portfolios can have a strong covariance structure due to a small number of factors which makes it hard to discriminate between models (Lewellen, Nagel and Shanken(2010)).

pricing errors relative to the baseline model of the Market factor using either the single-test p values or the multiple test p value. This result differs from Harvey and Liu(2018) who find support for the SMB and HML factors to be included in the factor model using the SI_{ew} and SI_{med} measures in U.S. stock returns.

One important feature of the Harvey and Liu(2018) approach is that we select factors based on the largest reduction of scaled alphas and not on the most significant single-test p value. It is possible that the factor with the largest reduction in scaled alphas does not have the lowest single-test p value. Harvey and Liu argue that it is important to select factors using the largest reduction in scaled pricing errors rather than the p value as it focuses on economic significance rather than just statistical significance.

Table 4 suggests that the Market factor is the dominant factor in reducing mispricing in U.K. stock returns. This finding is consistent with Harvey and Liu(2018)[12] and contrasts with studies suggesting that the Market factor does not play an important role in explaining cross-sectional stock returns such as Fama and French(1992), Kan, Robotti and Shanken(2013), and Chordia, Goyal and Shanken(2015) among others. Using the SI_{ew} and SI_{med} measures to capture the incremental contribution of each factor beyond the baseline model places a larger weight on smaller stocks. Linear factor models tend to find their greatest challenges among smaller stocks (e.g. Fama and French(2015,2016)). Harvey and Liu(2018) argue that if there is a large stock and small stock with equal alphas, we should place a greater weight on the large stock due to its' greater economic significance for investors.

I repeat the tests of Table 4 but this time use the SI_{vw} measure. Table 5 reports the results. In panel A of Table 5, the baseline model is a Constant.

---

[12] See also He, Huang and Zhou(2018).

Table 5 here

Panel A of Table 5 shows that the Market, HML, and BAB factors lead to a significant reduction in scaled alphas relative to the baseline model of a Constant using the single-test p values. The Market factor leads to the largest reduction in scaled alphas at 47.6%. The reduction using the SI_{vw} measure by the Market factor is a lot larger than observed for the SI_{ew} and SI_{med} measures at 17.8% and 22.1%. This result is consistent with Harvey and Liu(2018) in U.S. stock returns who point out that this result implies that the candidate factors play a bigger role in larger stocks[13]. The reduction in scaled alphas by the Market factor remains highly significant even after controlling for multiple testing.

In panel B of Table 5, the tests are repeated using a baseline model of the Market factor. There are five factors which have a negative SI_{vw} measure, the HML, HML_S, CMA, CMA_S, and FIN factors. Using the single-test p values, the HML_S, CMA, and CMA_S factors have significant negative SI_{vw} measures at the 5% significance level and the HML and FIN factors have significant negative SI_{vw} measures at the 10% significance level. The magnitude of the SI_{vw} measures compared to the SI_{vw} measure for the Market factor in panel A are a lot lower, suggesting that once we include the Market factor in the model, the incremental contribution of additional factors becomes more marginal. Among the five factors with negative SI_{vw} measures, the CMA factor has the largest reduction in the scaled pricing errors. The SI_{vw} measure of the CMA factor remains highly significant even after controlling for multiple testing with a multiple-test p value of 0.008. The importance of the CMA factor differs from Harvey and Liu(2018) in U.S. stock returns, who finds that the second factor is the Quality

[13] Plyakha, Uppal and Vilkov(2016) find that the alphas from linear factor models are higher under EW compared to VW.

minus Junk (QMJ) factor of Asness, Frazzini and Pedersen(2018) but also point out that the alternative profitability factors do well in U.S. stock returns.

In panel C of Table 5, I repeat the tests but this time the baseline model includes the Market and CMA factors. Panel C shows that only the HML_S, CMA_S, and PEAD factors have negative SI_{vw} measures but all are tiny and none are statistically significant. As a result, we end with a two-factor model including the Market and CMA factors. An interesting result in Tables 4 and 5 is that neither of the momentum factors (WML and WML_S) play a significant role in reducing the scaled pricing errors of the individual stocks, even although they have the highest mean excess returns among the factors.

Tables 4 and 5 suggest that after controlling for multiple testing only a small number of factors are significant at reducing the pricing errors of individual stocks in U.K. stock returns. The findings here and in Harvey and Liu(2018)[14] stand in sharp contrast to the model comparison tests of Barillas and Shanken(2018), and Barillas et al(2019) who find support for different six-factor models as does Fama and French(2018). The results are even more striking when compared to recent studies which advocate high dimensional factor models such as Kozak, Nagel and Santosh(2018) and Feng, Giglio and Xiu(2017). The differences between the studies can be due to a number of reasons such as the choice of the empirical method, the choice of test assets, and what we are looking for in a linear factor model. The attraction of the Harvey and Liu approach is that we can use individual stocks and we only include useful factors in the model that actually lead to a significant reduction in pricing errors rather than searching for a correct model specification.

[14] He et al(2018) also find that the CAPM performs as well as alternative factor models in U.S. stock returns.

I next examine whether including a global Market factor[15] in addition to the domestic Market factor has any impact on the results[16]. I repeat the tests of Tables 4 and 5 but this time include the world Market excess returns as an additional factor. Using the SI_{ew} and SI_{med} measures, the World factor does have a significant SI_{ew} and SI_{med} measures using the single-test p value, but the domestic Market factor still has the largest reduction in scaled pricing errors across factors. When comparing the results to a baseline model of the Market factor, the World factor does not lead to a significant incremental reduction in scaled pricing errors. A similar pattern emerges when using the SI_{vw} measure. My results in Tables 4 and 5 are robust to including the World Market factor as an additional candidate factor.

I conduct two robustness tests for the results in Table 5. First, I examine the impact of only using the 500 largest companies by Market value each month to calculate the SI_{vw} measure for each factor. I find similar results to those of Table 5. Second, I examine the impact of using a larger minimum number of returns of observations for stocks to be included in the tests. I repeat the tests of Table 5 but this time set the minimum number of return observations to 60. I find similar results to Table 5 that the relevant factor model only includes the Market and CMA factors.

The analysis in Tables 2 to 5 evaluate the performance of the factors over the whole sample period. The main downside of this approach is the assumption of constant factor betas. I next examine how well the factors perform over two different subperiods. I split the overall

---

[15] I use the excess returns of the Datastream world Market factor as the global Market factor.

[16] I thank one of the reviewers for suggesting this issue. Results are available on request.

sample period into two subperiods[17] using the 2007/2008 financial crisis to split the sample into July 1983 and December 2006, and January 2007 and December 2017. Tables 6 and 7 report the results over the two subperiods.

Table 6 here

Table 7 here

Panel A of Table 6 shows that the Market factor again leads to the largest reduction in the scaled pricing errors across all factors using the SI_{vw} measure in the July 1983 and December 2006 period. The magnitude of the reduction using the SI_{vw} measure for the Market factor is 46.4%. This reduction is just marginally below that for the overall sample period in panel A of Table 5. The Market, HML, and BAB factors all provide a significant reduction in the scaled pricing errors when the baseline model is a Constant using the single-test p values. This pattern is consistent with the overall sample period. The Market factor remains highly significant even after controlling for multiple testing with a multiple-test p value of 0.

In panel B of Table 6, I repeat the tests but this time use a baseline model including the Market factor. There are a number of factors that lead to a significant reduction in the scaled pricing errors relative to the baseline model of the Market factor during the July 1983 and December 2006 subperiod. The HML, HML_S, CMA, CMA_S, FIN, and PEAD factors all have significant negative SI_{vw} measures using the single-test p values. The magnitude of the SI_{vw} measures are larger than the corresponding SI_{vw} measures across the whole sample period,

[17] I also examine the performance of the factors over non-overlapping 5 year subperiods. There is much less evidence in finding any significant factors over short time periods. This might reflect a lack of power due to the high sampling error in individual stocks over shorter periods.

suggesting that additional factors play a greater role in reducing the mispricing of individual stocks in U.K. stock returns in the earlier subperiod. The HML_S factor has the largest negative SI_{vw} measure across factors and is highly significant even after controlling for multiple testing with a multiple-test p value of 0.

I next add the HML_S factor to the baseline model and repeat the tests in panel C of Table 6. Panel C shows that none of candidate factors make a significant reduction in the scaled pricing errors of the individual stocks relative to the baseline two-factor model. There are only two factors with negative SI_{vw} measures and both are tiny. None of the single-test p values are significant. The results in Table 6 suggest a two-factor model in the first subperiod.

Panel A of Table 7 shows that the Market factor is again the dominant factor in reducing the scaled pricing errors in the January 2007 and December 2017 subperiod relative to the baseline model of the Constant. None of the alternative factors have a negative SI_{vw} measure. The magnitude of the SI_{vw} measure of the Market factor is large in economic terms of 35% and is statistically significant even after controlling for multiple testing. The magnitude of the SI_{vw} measure for the Market factor is considerably lower than for the first subperiod, suggesting that the Market factor has less of an impact in the second subperiod compared to the first subperiod.

In panel B of Table 7, I repeat the tests for the second subperiod using the baseline model of a Market factor. There are only four candidate factors with negative SI_{vw} measures, including the HML, HML_S, WML, and WML_S factors. The HML factor has the largest negative SI_{vw} measure and is statistically significant at the 10% level using the single-test p value. However the HML factor is no longer significant after controlling for multiple testing and so our final model in the second subperiod is a single factor model including the Market factor. Tables 6 and 7 show that the selected factors in the factor models do a better job in reducing the scaled pricing errors of the individual stocks in the earlier part of the sample period.

V Conclusions

I use the sequential approach of Harvey and Liu(2018) to build linear factor models in U.K. stock returns among a set of 13 candidate factors. I use both individual stocks and test portfolios. There are three main findings in my study.

First, with the SI_{ew} or SI_{med} measures, the Market factor is the dominant factor in reducing the scaled pricing errors of the individual stocks and test portfolios among the 13 candidate factors. For the size/BM portfolios, industry portfolios, and individual stocks, none of the additional factors make a significant incremental contribution to the reduction of the scaled pricing errors beyond the baseline model of the Market factor. For the size/momentum portfolios, the WML_S factor is also included in the factor model. The importance of the Market factor is consistent with Harvey and Liu(2018) and He et al(2018) and contrasts with studies suggesting that the Market factor does not play an important role in explaining cross-sectional stock returns such as Fama and French(1992), Kan, Robotti and Shanken(2013), and Chordia, Goyal and Shanken(2015) among others. The significance of the WML_S factor in the size/momentum portfolios is not surprising as the same characteristic is used to form both the test assets and factor (Harvey and Liu).

Second, with the SI_{vw} measure, the Market factor is again the dominant factor in reducing the scaled pricing errors of the individual stocks. The magnitude of the SI_{vw} measure is more than double the SI_{ew} or SI_{med} measures, suggesting that the Market factor does a better job in reducing mispricing in larger stocks (Harvey and Liu(2018)). It is also consistent with studies that show the problems of asset pricing models are often concentrated in smaller stocks (Fama and French(2015,2016)). The Market factor has a larger impact in reducing the mispricing in the July 1983 and December 2006 compared to the January 2007 and December 2017 period.

Third, with the SI_{vw} measure, there is some support for the inclusion of a second factor into the factor model but this result depends upon the sample period. For the whole sample period, the second factor is the CMA factor of Fama and French(2015). In the July 1983 and December 2006 subperiod, the second factor is the HML_S factor. This finding of a two-factor model is consistent with Harvey and Liu(2018) but differs in that they find support for a profitability factor. The difference in findings between the CMA factor and a profitability factor might stem from the use of a shorter sample period here or by focusing on a different market. The relation between profitability and average returns in U.K. stock returns is sensitive as to how profitability is measured (Fletcher(2017)).

The findings here and in Harvey and Liu(2018) stand in sharp contrast to other empirical studies that provide support for larger factor models such as Barillas and Shanken(2018), Fama and French(2018), Kozak, Nagel and Santosh(2018), and Feng, Giglio and Xiu(2019) among others. The approach of Harvey and Liu does not necessarily look for a well specified linear factor model but rather a model which only includes useful factors that actually play a significant role in reducing the scaled pricing errors after controlling for multiple testing.

My study suggests that the Market factor is the most important factor in reducing mispricing among 13 candidate factors. The importance of this result is that the Market factor should be included in every linear factor model in practical applications. It also provides a measure of support of the ways practitioners use the CAPM in practical applications such as cost of equity capital estimation (Graham and Harvey(2001)). My study has only focused on a small number of candidate factors. A wider range of factors could be included such as the factors in Hou et al(2015) or the mispricing factors in Stambaugh and Yuan(2017), or the Quality Minus Junk (QMJ) factor of Asness et al(2018). My study has only considered unconditional factor models and could be extended to consider conditional factor models. I leave an examination of these issues to future research.

Table 1 Summary Statistics of Factors

Factors	Mean	Standard Deviation	*t*-statistic
Market	0.451	4.134	2.22[1]
SMB	0.032	2.913	0.22
HML	0.281	2.540	2.25[1]
HML_S	0.339	3.013	2.28[1]
RMW	0.168	1.993	1.72[2]
RMW_S	0.256	2.227	2.34[1]
CMA	0.375	1.890	4.03[1]
CMA_S	0.429	2.009	4.34[1]
WML	0.909	3.745	4.94[1]
WML_S	1.209	3.614	6.80[1]
BAB	0.585	6.077	1.96[1]
FIN	0.433	1.760	5.01[1]
PEAD	0.245	2.244	2.22[1]

[1] Significant at 5%
[2] Significant at 10%

The table reports summary statistics of the excess factor returns between July 1983 and December 2017. The summary statistics include the mean and standard deviation of excess returns (%), and the *t*-statistic of the null hypothesis that the average excess factor return equals zero.

Table 2 Performance of Factors using Size/BM and Industry Portfolios as Test Assets using the SI_{ew} and SI_{med} Measures

Panel A:						
Size/BM	SI^m	5%	p-value	SI_{med}	5%	p value
Market	-0.674	-0.384	0	-0.803	-0.411	0
SMB	-0.032	-0.223	0.429	-0.023	-0.225	0.4
HML	-0.044	-0.073	0.098	-0.016	-0.086	0.259
HML_S	0.163	-0.098	0.994	0.168	-0.133	0.969
RMW	0.143	-0.121	0.932	0.145	-0.144	0.915
RMW_S	0.058	-0.051	0.902	0.056	-0.064	0.878
CMA	0.559	-0.173	0.999	0.576	-0.200	0.994
CMA_S	0.553	-0.163	0.999	0.523	-0.187	0.996
WML	0.459	-0.127	0.999	0.480	-0.146	0.992
WML_S	0.647	-0.133	1	0.682	-0.159	0.998
BAB	-0.357	-0.259	0.015	-0.326	-0.262	0.029
FIN	0.924	-0.218	0.999	0.828	-0.245	0.992
PEAD	0.296	-0.173	0.981	0.250	-0.190	0.951
	5%	p value			5%	p value
Multiple Test	-0.411	0		Multiple Test	-0.441	0

Panel B:						
Industry	SI^m	5%	p-value	SI_{med}	5%	p value
Market	-0.813	-0.386	0	-0.870	-0.399	0
SMB	-0.004	-0.054	0.351	0.001	-0.124	0.508
HML	-0.074	-0.064	0.034	-0.024	-0.112	0.251
HML_S	0.081	-0.070	0.947	0.129	-0.128	0.949
RMW	0.138	-0.127	0.926	0.141	-0.163	0.891
RMW_S	0.030	-0.042	0.874	0.019	-0.074	0.723
CMA	0.410	-0.152	0.992	0.487	-0.187	0.983
CMA_S	0.342	-0.129	0.993	0.479	-0.170	0.989
WML	0.483	-0.161	0.993	0.584	-0.193	0.985
WML_S	0.677	-0.161	1	0.875	-0.199	0.998
BAB	-0.313	-0.207	0.006	-0.336	-0.247	0.012
FIN	0.610	-0.160	0.991	0.795	-0.208	0.985
PEAD	0.226	-0.152	0.955	0.299	-0.183	0.949
	5%	p value			5%	p value
Multiple Test	-0.390	0		Multiple Test	-0.431	0

The table reports the reduction in scaled absolute pricing errors of individual factors relative to a baseline model between July 1983 and December 2017. The test assets are 16 size/BM portfolios (panel A), and 10 industry portfolios (panel B). The baseline model is a Constant. The SI_{ew} measure is the percentage difference (%) in the mean scaled absolute alphas between the augmented and baseline models. The SI_{med} measure is the percentage difference (%) in the median scaled absolute alphas between the augmented and baseline models. The 5% and p value columns are the fifth percentile and the corresponding singe-test p value of each factor under the null hypothesis that the factor has a zero incremental contribution of reducing scaled absolute alphas relative to the baseline model. The final row of each panel is the fifth percentile (5%) and p value from the multiple test under the null hypothesis that none of the factors lead to a significant reduction in scaled absolute alphas relative to the baseline model.

Table 3 Performance of Factors using Size/Momentum Portfolios as Test Assets using the SI_{ew} and SI_{med} Measures

Panel A:						
Constant	SI^m	5%	p-value	SI_{med}	5%	p value
Market	-0.321	-0.264	0.015	-0.399	-0.327	0.014
SMB	-0.012	-0.103	0.424	-0.015	-0.119	0.358
HML	-0.018	-0.029	0.117	-0.080	-0.058	0.019
HML_S	0.042	-0.032	0.954	0.040	-0.034	0.924
RMW	0.072	-0.060	0.952	0.118	-0.097	0.958
RMW_S	0.028	-0.029	0.888	0.054	-0.049	0.915
CMA	0.258	-0.085	1	0.326	-0.114	0.998
CMA_S	0.259	-0.087	0.999	0.355	-0.114	1
WML	0.254	-0.062	1	0.245	-0.107	0.995
WML_S	0.405	-0.060	1	0.326	-0.104	1
BAB	-0.231	-0.168	0.01	-0.302	-0.197	0.005
FIN	0.475	-0.100	1	0.481	-0.127	1
PEAD	0.127	-0.085	0.991	0.155	-0.112	0.975
	5%	p value			5%	p value
Multiple Test	-0.270	0.015		Multiple Test	-0.330	0.015
Panel B:						
Market	SI^m	5%	p-value	SI_{med}	5%	p value
SMB	-0.018	-0.119	0.384	-0.023	-0.119	0.269
HML	0.042	-0.036	0.963	0.021	-0.058	0.797
HML_S	0.031	-0.030	0.955	-0.020	-0.048	0.135
RMW	-0.026	-0.022	0.034	-0.081	-0.050	0.013
RMW_S	-0.018	-0.017	0.048	-0.055	-0.039	0.021
CMA	0.010	-0.014	0.892	-0.068	-0.042	0.009
CMA_S	0.106	-0.031	1	-0.013	-0.047	0.211
WML	-0.309	-0.122	0	-0.352	-0.176	0.001
WML_S	-0.350	-0.117	0	-0.367	-0.164	0
BAB	-0.085	-0.112	0.098	-0.113	-0.136	0.077
FIN	0.073	-0.023	1	-0.153	-0.064	0
PEAD	-0.036	-0.028	0.03	-0.150	-0.070	0.004
	5%	p value			5%	p value
Multiple Test	-0.154	0		Multiple Test	-0.211	0.002

The table reports the reduction in scaled absolute pricing errors of individual factors relative to a baseline model between July 1983 and December 2017. The test assets are 16 size/momentum portfolios. In panel A, the baseline model is a Constant, in panel B, the baseline model is the Market factor. The SI_{ew} measure is the percentage difference (%) in the mean scaled absolute alphas between the augmented and baseline models. The SI_{med} measure is the percentage difference (%) in the median scaled absolute alphas between the augmented and baseline models. The 5% and p value columns are the fifth percentile and the corresponding singe-test p value of each factor under the null hypothesis that the factor has a zero incremental contribution of reducing scaled absolute alphas relative to the baseline model. The final row of each panel is the fifth percentile (5%) and p value from the multiple test under the null hypothesis that none of the factors lead to a significant reduction in scaled absolute alphas relative to the baseline model.

Table 4 Performance of Factors using Individual Stocks as Test Assets using the SI_{ew} and SI_{med} Measures

Panel A:						
Constant	SI_{ew}	5%	p-value	SI_{med}	5%	p value
Market	-0.178	-0.108	0.003	-0.221	-0.118	0.001
SMB	-0.003	-0.046	0.48	-0.002	-0.060	0.5
HML	0.021	0.005	0.293	0.031	0	0.425
HML_S	0.051	0.006	0.911	0.072	0.005	0.943
RMW	0.072	-0.005	0.998	0.083	-0.009	0.996
RMW_S	0.108	0.023	0.989	0.129	0.023	0.99
CMA	0.125	-0.024	1	0.138	-0.028	1
CMA_S	0.115	-0.019	1	0.145	-0.025	1
WML	0.146	-0.030	1	0.152	-0.038	1
WML_S	0.249	-0.031	1	0.244	-0.037	1
BAB	-0.114	-0.074	0.006	-0.136	-0.090	0.008
FIN	0.249	-0.032	1	0.275	-0.042	1
PEAD	0.081	-0.036	0.999	0.084	-0.050	0.999
	5%	p value			5%	p value
Multiple Test	-0.110	0.003		Multiple Test	-0.122	0.001

Panel B:						
Market	SI^m	5%	p-value	SI_{med}	5%	p value
SMB	-0.013	-0.040	0.307	-0.012	-0.051	0.384
HML	0.023	0.003	0.687	0.053	-0.002	0.972
HML_S	0.034	0.004	0.913	0.058	-0.001	0.988
RMW	0.030	0.000	0.959	0.049	-0.005	0.99
RMW_S	0.056	0.016	0.955	0.073	0.010	0.976
CMA	0.045	-0.005	1	0.055	-0.013	1
CMA_S	0.041	-0.004	1	0.054	-0.010	1
WML	0.090	-0.007	1	0.144	-0.013	1
WML_S	0.173	-0.008	1	0.233	-0.018	1
BAB	-0.009	-0.016	0.136	-0.006	-0.026	0.28
FIN	0.097	-0.009	1	0.134	-0.019	1
PEAD	0.032	-0.007	0.997	0.070	-0.016	1
	5%	p value			5%	p value
Multiple Test	-0.040	0.343		Multiple Test	-0.052	0.582

The table reports the reduction in scaled absolute pricing errors of individual factors relative to a baseline model between July 1983 and December 2017. The test assets are individual stocks, with at least 36 return observations. In panel A, the baseline model is a Constant, in panel B, the baseline model is the Market factor. The SI_{ew} measure is the percentage difference (%) in the mean scaled absolute alphas between the augmented and baseline models. The SI_{med} measure is the percentage difference (%) in the median scaled absolute alphas between the augmented and baseline models. The 5% and p value columns are the fifth percentile and the corresponding singe-test p value of each factor under the null hypothesis that the factor has a zero incremental contribution of reducing scaled absolute alphas relative to the baseline model. The final row of each panel is the fifth percentile (5%) and p value from the multiple test under the null hypothesis that none of the factors lead to a significant reduction in scaled absolute alphas relative to the baseline model.

Table 5 Performance of Factors using Individual Stocks as Test Assets using the SI_{vw} Measure

Panel A: **Constant**	SI_{vw}	5%	p value
Market	-0.476	-0.258	0
SMB	0.003	-0.057	0.614
HML	-0.078	-0.045	0.011
HML_S	-0.002	-0.013	0.183
RMW	0.100	-0.071	0.962
RMW_S	0.083	-0.016	0.974
CMA	0.175	-0.069	1
CMA_S	0.144	-0.049	1
WML	0.298	-0.090	1
WML_S	0.437	-0.088	1
BAB	-0.137	-0.078	0.009
FIN	0.234	-0.060	1
PEAD	0.128	-0.074	0.987
	5%	p value	
Multiple Test	-0.258	0	

Panel B: **Market**	SI_{vw}	5%	p value
SMB	0.011	-0.062	0.739
HML	-0.053	-0.059	0.062
HML_S	-0.087	-0.058	0.013
RMW	0.061	-0.032	0.999
RMW_S	0.074	-0.026	1
CMA	-0.116	-0.042	0
CMA_S	-0.086	-0.032	0
WML	0.142	-0.027	1
WML_S	0.247	-0.030	1
BAB	0.032	-0.022	0.959
FIN	-0.052	-0.062	0.073
PEAD	0.012	-0.032	0.933
	5%	p value	
Multiple Test	-0.087	0.008	

Panel C: **Market/CMA**	SI_{vw}	5%	p value
SMB	0.019	-0.046	0.933
HML	0.004	-0.030	0.552
HML_S	-0.009	-0.021	0.155
RMW	0.036	-0.032	0.981
RMW_S	0.049	-0.020	0.995
CMA_S	-0.017	-0.028	0.122
WML	0.139	-0.030	1
WML_S	0.261	-0.029	1
BAB	0.036	-0.019	0.991
FIN	0.076	-0.037	0.998
PEAD	-0.002	-0.036	0.552
	5%	p value	
Multiple Test	-0.065	0.657	

The table reports the reduction in scaled absolute pricing errors of individual factors relative to a baseline model between July 1983 and December 2017. The test assets are individual stocks, with at least 36 return observations. In panel A, the baseline model is a Constant, in panel B, the baseline model is the Market factor, and in panel C the baseline model is the Market and CMA factors. The SI_{VW} measure is the percentage difference (%) in the average value weighted (VW) scaled absolute alphas between the augmented and baseline models. The 5% and p value columns are the fifth percentile and the corresponding singe-test p value of each factor under the null hypothesis that the factor has a zero incremental contribution of reducing scaled absolute alphas relative to the baseline model. The final row of each panel is the fifth percentile (5%) and p value from the multiple test, under the null hypothesis that none of the factors lead to a significant reduction in scaled absolute alphas relative to the baseline model.

Table 6 Performance of Factors using Individual Stocks as Test Assets using the SI_{vw} Measure: July 1983 and December 2006

Panel A: **Constant**	SI_{vw}	5%	p value
Market	-0.464	-0.306	0
SMB	0.004	-0.075	0.594
HML	-0.066	-0.037	0.01
HML_S	0.047	-0.023	0.987
RMW	-0.012	-0.069	0.326
RMW_S	0.001	-0.020	0.422
CMA	0.129	-0.065	1
CMA_S	0.165	-0.071	1
WML	0.260	-0.083	1
WML_S	0.456	-0.088	1
BAB	-0.196	-0.103	0.001
FIN	0.285	-0.071	0.999
PEAD	0.105	-0.059	0.986
	5%	p value	
Multiple Test	-0.306	0	

Panel B: **Market**	SI_{vw}	5%	p value
SMB	0.016	-0.069	0.736
HML	-0.173	-0.080	0
HML_S	-0.192	-0.082	0
RMW	-0.013	-0.054	0.385
RMW_S	-0.039	-0.061	0.1
CMA	-0.121	-0.059	0.002
CMA_S	-0.107	-0.046	0
WML	0.139	-0.038	1
WML_S	0.275	-0.043	1
BAB	0.050	-0.037	0.97
FIN	-0.096	-0.065	0.015
PEAD	-0.047	-0.044	0.041
	5%	p value	
Multiple Test	-0.108	0	

Panel C: **Market/HML_S**	SI_{vw}	5%	p value
SMB	0.054	-0.066	0.99
HML	0	-0.048	0.506
RMW	-0.003	-0.043	0.586
RMW_S	0.026	-0.022	0.949
CMA	0.022	-0.028	0.945
CMA_S	0.053	-0.020	0.999
WML	0.126	-0.020	1
WML_S	0.329	-0.024	1
BAB	0.064	-0.029	0.999
FIN	0.060	-0.046	0.987
PEAD	-0.035	-0.047	0.105
	5%	p value	
Multiple Test	-0.084	0.432	

The table reports the reduction in scaled absolute pricing errors of individual factors relative to a baseline model between July 1983 and December 2006. The test assets are individual stocks, with at least 36 return observations. In panel A, the baseline model is a Constant, in panel B, the baseline model is the Market factor, and in panel C the baseline model is the Market and HML_S factors. The SI_{vw} measure is the percentage difference (%) in the average value weighted (VW) scaled absolute alphas between the augmented and baseline models. The 5% and p value columns are the fifth percentile and the corresponding singe-test p value of each factor under the null hypothesis that the factor has a zero incremental contribution of reducing scaled absolute alphas relative to the baseline model. The final row of each panel is the fifth percentile (5%) and p value of the multiple test, under the null hypothesis that none of the factors lead to a significant reduction in scaled absolute alphas relative to the baseline model.

Table 7 Performance of Factors using Individual Stocks as Test Assets using SI_{vw} Measure: January 2007 and December 2017

Panel A:			
Constant	SI_{vw}	5%	p value
Market	-0.350	-0.309	0.029
SMB	0.006	-0.089	0.604
HML	0.263	-0.142	0.999
HML_S	0.213	-0.110	0.998
RMW	0.395	-0.117	1
RMW_S	0.493	-0.106	1
CMA	0.246	-0.089	0.994
CMA_S	0.098	-0.042	0.97
WML	0.281	-0.122	1
WML_S	0.304	-0.107	1
BAB	0.005	-0.062	0.506
FIN	0.122	-0.061	0.988
PEAD	0.124	-0.122	0.971
	5%	p value	
Multiple Test	-0.309	0.029	

Panel B:			
Market	SI_{vw}	5%	p value
SMB	0.000	-0.144	0.526
HML	-0.109	-0.119	0.062
HML_S	-0.015	-0.044	0.211
RMW	0.164	-0.084	0.999
RMW_S	0.052	-0.072	0.939
CMA	0.000	-0.064	0.532
CMA_S	0.027	-0.040	0.923
WML	-0.045	-0.079	0.16
WML_S	-0.025	-0.050	0.154
BAB	0.012	-0.033	0.651
FIN	0.021	-0.105	0.794
PEAD	0.020	-0.085	0.876
	5%	p value	
Multiple Test	-0.173	0.208	

The table reports the reduction in scaled absolute pricing errors of individual factors relative to a baseline model between January 2007 and December 2017. The test assets are individual stocks, with at least 36 return observations. In panel A, the baseline model is a Constant, and in panel B, the baseline model is the Market factor. The SI_{vw} measure is the percentage difference in the average value weighted (VW) scaled absolute alphas between the augmented and baseline models. The 5% and p value columns are the fifth percentile and the corresponding singe-test p value of each factor under the null hypothesis that the factor has a zero incremental contribution of reducing scaled absolute alphas relative to the baseline model. The final row of each panel is the fifth percentile and p value from the multiple test, under the null hypothesis that none of the factors lead to a significant reduction in scaled absolute alphas relative to the baseline model.

References

Ang, A., 2014, Asset management a systematic approach to factor investing, *Oxford University Press*.

Asness, C., Frazzini, A. and L.H. Pedersen, 2018, Quality minus junk, *Review of Accounting Studies*, forthcoming.

Bai, J. and G. Zhou, 2015, Fama-MacBeth two-pass regressions: Improving risk premia estimates, *Finance Research Letters*, 15, 31-40.

Barillas, F., Kan, R., Robotti, C. and J. Shanken, 2019, Model comparison with Sharpe ratios, *Working Paper*, University of Toronto.

Barillas, F. and J. Shanken, 2017 Which alpha?, *Review of Financial Studies*, 30,1316-1388.

Barillas, F. and J. Shanken, 2018, Comparing asset pricing models, *Journal of Finance*, 73, 715-754.

Carhart, M. M., 1997, Persistence in mutual fund performance. *Journal of Finance*, 52, 57-82.

Chordia, T., Goyal, A. and J. Shanken, 2015, Cross-sectional asset pricing with individual stocks: Betas versus characteristics, *Working Paper*, Emory University.

Daniel, K., Hirshleifer, D. and L. Sun, 2018, Short- and long-horizon behavioural factors, *Working Paper*, Columbia University.

Davies, J.R., Fletcher, J. and A. Marshall, 2015, Testing factor-based models in U.K. stock returns, *Review of Quantitative Finance and Accounting*, 45, 337-362.

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, 1992, The cross-section of expected stock returns, *Journal of Finance*, 47, 427-465.

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, 2010, Luck versus skill in the cross-section of mutual fund returns, *Journal of Finance*, 65, 1915-1947.

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, 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.

Feng, G., Giglio, S. and D. Xiu, 2019, Taming the factor zoo, *Working Paper*, University of Chicago.

Ferson, W.E. and Y. Chen, 2015, How many good and bad fund managers are there really?, *Working Paper*, University of Southern California.

Fletcher, J., 1994, The mean-variance efficiency of benchmark portfolios: UK evidence, *Journal of Banking and Finance*, 18, 673-685.

Fletcher, J., 2001, An examination of alternative factor models in UK stock returns, *Review of Quantitative Finance and Accounting*, 16, 117-130.

Fletcher, J. 2017, An empirical examination of the incremental contribution of stock characteristics in U.K. stock returns, *International Journal of Financial Studies*, forthcoming.

Fletcher, J., 2018, Model comparison tests of linear factor models in U.K. stock returns, *Finance Research Letters*, forthcoming.

Frazzini, A. and L.H. Pedersen, 2014, Betting against beta, *Journal of Financial Economics*, 111, 1-25.

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.

Graham, J. and C.R. Harvey, 2001, The theory and practice of corporate finance: Evidence from the field, *Journal of Financial Economics*, 60, 187-243.

Green, J., Hand, J.R. and X.F. Zhang, 2017, The characteristics that provide independent information about average U.S. monthly stock returns, *Review of Financial Studies*, 47, 1785-1809.

Gregory, A., Tharyan, R. and A. Christidis, 2013, Constructing and testing alternative versions of the Fama-French and Carhart models in the UK, *Journal of Business Finance and Accounting*, 40, 172-214.

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 Y. Liu, 2018, Lucky factors, *Working Paper*, Duke University.

Harvey, C.R., Liu, Y. and H. Zhu, 2016, … and the cross-section of expected returns, *Review of Financial Studies*, 29, 5-68.

He, A., Huang, D. and G. Zhou, 2018, Failure of existing factor models in explaining individual stock returns: Evidence from a predictability test, *Working Paper*, Singapore Management University.

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, 2019, Replicating anomalies, *Review of Financial Studies*, forthcoming.

Hou, K., Mo, H., Xue, C. and L. Zhang, 2019, Which factors, *Review of Finance*, 23, 1-35.

Hou, K., Mo, H., Xue, C. and L. Zhang, 2018, q^5, *Working Paper*, Ohio State University.

Jegadeesh, N., Noh, J., Pukthuanthong, K., Roll, R. and J.L. Wang, 2018, Empirical tests of asset pricing models with individual assets: Resolving the errors-in-variables bias in risk premium estimation, *Journal of Financial Economics*, forthcoming.

Jensen, M.C., 1968, The performance of mutual funds in the period 1945-1964. *Journal of Finance* 23, 389-416.

Kan, R. and C. Robotti, 2009, Model comparison using the Hansen-Jagannathan distance, *Review of Financial Studies*, 22, 3449-3490.

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.

Kim, S. and G. Skoulakis, 2018, Ex-post risk premia estimation and asset pricing using large cross sections: The regression calibration approach, *Journal of Econometrics*, 204, 159-188.

Kozak, S., Nagel, S. and S. Santosh, 2018, Shrinking the cross section, *Journal of Financial Economics*, forthcoming.

Kowowski, R., Timmermann, A., Wermers, R. and H. White, 2006, Can mutual fund "stars" really pick stocks? New evidence from a bootstrap analysis, *Journal of Finance*, 61, 2551-2595.

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

Lo, A. and A.C. MacKinlay, 1990, Data-snooping biases in tests of financial asset pricing models, *Review of Financial Studies*, 3, 431-468.

Ludvigson, S., 2013, Advances in consumption-based asset pricing: Empirical tests, in Constantinides, G., Harris, M. and R. Stulz (Eds), Handbook of the Economics of Finance, 2, 799-906.

Michou, M. and H. Zhou, 2016, On the information content of new asset pricing factors in the UK, *Working Paper*, University of Edinburgh.

Pastor, L. and R.F. Stambaugh, 2003, Liquidity risk and expected stock returns, *Journal of Political Economy*, 111, 642-685.

Plyakha, Y., Uppal, R. and G. Vilkov, 2016, Equal or value weighting? Implications for asset-pricing tests, *Working Paper*, Edhec.

Raponi, V., Robotti, C. and P. Zaffaroni, 2018, Testing beta-pricing models using large cross-sections, *Working Paper*, Imperial College London.

Raponi, V., Uppal, R. and P. Zaffaroni, 2018, Portfolio choice with model misspecification: A foundation for alpha and beta portfolios, *Working Paper*, Edhec.

Roll, R., 1977, A critique of the asset pricing theory's test; Part I: On past and potential testability of the theory, *Journal of Financial Economics*, 4, 129-176.

Sarisoy, C., de Goeij, and B.J.M. Werker, 2017, Linear factor models and the estimation of expected returns, *Academic Paper*, Netspar.

Shanken, J., 1992, On the estimation of beta-pricing models, *Review of Financial Studies*, 5, 1-33.

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

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

White, H., 2000, A reality check for data snooping, *Econometrica*, 68, 1097-1126.