Choosing factors: the international evidence

Author(s): Grobys, Klaus; Kolari, James W.

Title: Choosing factors: the international evidence

Year: 2021

Version: Accepted version

Please cite the original version:

Grobys, K. & Kolari, J. W. (2021). Choosing factors: the international evidence. *Applied Economics*. https://doi.org/10.1080/00036846.2021.1967865

Choosing Factors: The International Evidence

Klaus Grobys[1a] and James Kolari[2a]

University of Vaasa

This draft: 4. August 2021

Abstract

Extending Fama and French's (2018) U.S. study on choosing factors to international equity markets, we test nested and non-nested asset pricing models for North America, Europe, Asia excluding Japan, and Japan. For non-nested models, we propose a new simulation methodology using a blocks bootstrap approach that takes into account factor dependencies. The resultant out-of-sample Sharpe ratios across all models and countries are lower than Fama and French's pairs bootstrap approach. While we confirm that the six-factor model with market, size, and small size spread factors for value, profitability, investment, and momentum produces the highest maximum squared Sharpe ratio in most economies, an exception is Asia excluding Japan. Additionally, spanning regressions reveal that size does not matter in any of the international equity markets, whereas value matters in Europe, Asia excluding Japan, and Japan.

JEL Classification: G12, G14

Keywords: Risk factors, maximum squared Sharpe ratio, asset pricing models, spanning regressions, international equity markets

---

[1] K. Grobys (corresponding author)
Department of Accounting and Finance, University of Vaasa, Wolffintie 34, 65200 Vaasa, Finland
e-mail: klaus.grobys@uwasa.fi; grobys.finance@gmail.com
[2] J. Kolari
Department of Finance, Texas A&M University, TAMU-4218, College Station, Texas, USA
e-mail: j-kolari@tamu.edu

[a]We are thankful to Wesley Gray (University of Chicago) for highlighting our paper in his popular investment blog alphaarchitect.com on December 26th, 2019 ("International Evidence on Factor Premiums"). Moreover, we received valuable comments from participants at the 2018 Annual International Conference on Macroeconomic Analysis and International Finance in Greece as well as the 2019 Finance Seminar at Hanken School of Economics and 2019 Accounting and Finance Seminar at the University of Vaasa, both in Vaasa, Finland. We would like to especially thank Janne Äijö, John Broussard, Pawel Fiedor, and Kam-Ming Wan for insightful comments. Finally, we would like to thank two anonymous reviewers for providing useful and helpful comments.

Choosing Factors: The International Evidence

1. Introduction

Almost 30 years after Jegadeesh and Titman (1993) documented the momentum effect in the U.S. equity market, Fama and French (FF) (2018) acquiesced to include momentum as a risk factor in their asset pricing models. Momentum is widely accepted by both academics and practitioners but has little or no theoretical foundation. FF (2018) test a six-factor model that adds the momentum factor to their five-factor model (FF 2015, 2017) comprised of market, size, value, profit, and investment factors.[1] Due to limitations of the left-hand-side approach for evaluating the performance of asset pricing models by means of time series regression intercept terms, they adopted a performance metric by Barillas and Shanken (2016) that focuses on the right-hand-side (RHS) factors of competing models. This approach selects the model with factors that produce the highest median or mean squared Sharpe ratio (MSSR). Moreover, FF (2018) implement pairs bootstrap simulations of full-sample (FS), in-sample (IS) and out-of-sample (OS) MSSR estimates for competing models. Overall, FF (2018) find that based on U.S. stock returns, the maximum MSSR was obtained for a model combining market and size factors with small stock spread factors for value, profitability, investment, and momentum.

Using pairs bootstrap is, however, a problematic issue if the data exhibit dependency structures because this approach treats data as independently distributed. Dependency structures can be manifested in the data in terms of switching regimes (Guidolin and Timmermann, 2007; Ang and Bekaert, 2002; Perez-Quiros and Timmermann, 2000) or simple autocorrelation structures as suggested by the well-known momentum effect, as documented first in Jegadeesh and Titman (1993). Furthermore, it also remains unclear how the FF (2018) six-factor model performs in other economies such as Europe, Asia Pacific or Japan.

Hence, the purpose of our study is to test various popular factor models in equity markets of North America, Europe, Asia Pacific excluding Japan and Japan. In doing so, we follow FF (2018) and use the MSSR metric to compare nested versions of FF's six-factor model, including the capital asset pricing model (CAPM) as well as their earlier three- and five-factor models (FF 1993, 2015). We implement bootstrap simulations of full-sample (FS), in-sample (IS) and out-of-sample (OS) MSSR estimates for competing models. Unlike FF (2018), who employed a version of pairs bootstrap, we utilize a blocks bootstrap simulation procedure that is appropriate for factor data with dependency structures. In their study, models are compared with different candidates for the

[1] Adding the momentum portfolio as risk factor to an asset pricing model is nothing new. Carhart (1997) proposed a four-factor model using Fama and French's (1993) three-factor model plus an additional factor capturing Jegadeesh and Titman's (1993) one-year momentum effect.

profitability factor as well as different portfolio spreads (i.e., spread portfolios using all stocks versus spread portfolios using either small or big stocks). Here we adhere to FF's (2018, p.247) advice that a parsimonious set of factors should be considered and therefore focus exclusively on portfolio spreads using all stocks or small stocks. While the factor definitions are similar to theirs, we use wider spreads in size in an effort to provide further empirical evidence on size factor effects.

Our study has some clear contributions. First, FF (2008, 2018) have recommended that sample-specific model comparisons would benefit from robustness tests of out-of-sample data in different periods and markets. Supporting this view, Hou, Xue, and Zhang (2020) have argued that the reliability of published results can be evaluated by replicating analyses with different sample data. Consistent with these studies, we contribute to the literature on choosing factors among many contenders by replicating FF's U.S. analyses in different regions of the world, including North America, Europe, Asia Pacific excluding Japan, and Japan. In line with replication requirements proposed by Hamermesh (2007), we employ different population samples and sample periods, a similar but not identical model, and alternative simulation procedures. Furthermore, we utilize a blocks bootstrap simulation procedure that is appropriate for factor data with dependency structures. Godfrey (2009) documents that pairs bootstrap allows for heteroskedasticity of unspecified form. However, this approach breaks the correlation structure in the data. Since it is generally accepted that regimes exist in financial returns (Guidolin and Timmermann, 2007; Ang and Bekaert, 2002; Perez-Quiros and Timmermann, 2000), another contribution of our study is that we address the dependency issue in factor returns by proposing a blocks bootstrap.

In summary, our international results corroborate U.S. findings that the six-factor model dominates other models. A surprising finding of our study is that the popular three-factor model, which has been used as a benchmark model in empirical asset pricing research over the past three decades, does not outperform the CAPM in terms of pricing equities in North America, Europe, and Japan. This finding is notable in that FF (1992, 1993, 1995) rejected the CAPM in the early 1990s due to its inability to describe the cross section of average U.S. stock returns and alternatively proposed the three-factor model. Early empirical evidence by FF (1992, 1993) showed that the three-factor model outperformed the CAPM. Our result suggests that FF's (1992, 1993) empirical observation could be specific to the sample and market at that time. In further analyses, we deconstruct MSSR by means of spanning regressions that explain each of FF's (2018) six factors based on the other five factors. Unlike FF (2018), we find that a commonality in international equity markets is the insignificance of the size factor. Another new international finding is that the value factor is significant and positive, whereas the investment factor is not significantly different from zero. Hence, value matters outside the U.S. but not investment. A further consistent finding across countries is that the profitability factor plays in important role in pricing stocks. This finding contradicts earlier evidence by FF (2017) in which the profitability factor was insignificant in Japan and Asia excluding Japan. Finally, the marginal contributions to MSSR of the momentum factor in Europe and AP are relatively high compared to other multifactors but approximately equal to those for the market factor, which suggests that momentum plays a more salient role in asset pricing than previously believed. Finally, we will see in forthcoming analyses that choosing factor models among non-nested models depends not only on FS, IS, and OS comparisons but the simulation approach chosen (viz., pairs bootstrap versus blocks bootstrap).

The next section discusses the data and factor definitions. Section 3 provides the RHS approach results for testing nested asset pricing models. Section 4 tests non-nested versions of the six-factor model using blocks bootstrap in FS, IS, and OS simulations. Section 5 investigates spanning regressions of the factors. Section 6 concludes.

2. Data and Factor Definitions

Local factors for the FF (2015) five-factor model as well as momentum for North American (NA), Europe, Asia excluding Japan (AP), and Japan are downloaded from Kenneth French's website. These factors are used for the local benchmark models. All data are denominated in U.S. dollars. Following FF (2018), the standard risk factors in the benchmark model are defined as follows: Mkt_t^{ex} is the value-weight (VW) market portfolio return minus the risk-free rate at time t, SMB_t (small minus big) is the size factor, HML_t (high minus low book-to-market equity) is the value factor, RMW_t (robust minus weak) is a profitability factor, CMA_t (conservative minus aggressive) is an investment factor, and MOM_t (winners minus minus losers) is a momentum factor. Additionally, for each region, we downloaded the following value-weighted local portfolios: 25 size and book-to-market ratio, 25 size and profitability, 25 size and investment, and 25 size and momentum. These portfolios are used to construct small size and big size factors by conditioning the long-short strategies on only the first (small size factors) and fifth (big size factors) quintiles.[2] The small (big) value, profitability, investment, and momentum factors are denoted as $HML_{S,t}$, $RMW_{S,t}$, $CMA_{S,t}, MOM_{S,t}$ ($HML_{B,t}, RMW_{B,t}, CMA_{B,t}, MOM_{B,t}$), respectively. Widening the spread should offer a larger contrast between risk factors based on different size groups of stocks. Our sample period is from November

[2] Note that the size breakpoints for a region are due to the distinct market environments defined as the 3rd, 7th, 13th, and 25th percentiles of the region's aggregate market capitalization. (See Kenneth's French website for further details.) This implies that our factors are defined using the 3/97 market cap percentile as opposed to the 10/90 market cap percentile as used in FF (2018) for the U.S. market. Since we operate with spreads that are slightly wider, our analysis should offer a larger potential contrast between risk factors based on different size groups of stocks.

1990 to August 2019. Hence, following FF (2018), the six-factor asset pricing model that we explore in our study is given by

$$R_{i,t}^{ex} = a_i + b_i Mkt_t^{ex} + s_i SMB_t + h_i HML_t + r_i RMW_t + c_i CMA_t + m_i MOM_t + u_{i,t}.$$

Table 1 provides summary statistics for the factors. First, with respect to the five-factor model in the NA region, only the market and profitability factors have average returns (premiums) that are significant at the 5% level. Unlike FF (2017), the premium for the value factor using small stocks is economically large at 72 basis points per month and statistically significant even when using the critical value of 2.78 based on the multiple testing hurdle of Harvey, Liu, and Zhu (2016). The difference HML_{S-B} is 84 basis points per month with t = 3.53 indicating significance at any level. The spread is almost twice than that reported in FF (2017) due most likely to using a wider spread. Similarly, the investment factor is 69 basis points per month when implemented among small stocks with a highly significant t = 4.89. Unlike their 2017 study, the profitability factor is only significant when implemented among big stocks. The reason for this discrepancy between our results and those of FF can be attributed to different sample periods. Their data ends in December 2015, but from January 2016 to August 2019 the profitability factor implemented among big stocks produce 77 basis points per month on average with t = 2.13 but only -8 basis points per month with t = -0.22 among small stocks. In Europe, the average returns for the risk factors of the five-factor model are similar to the U.S. The standard value and profitability factors produce 28 and 38 basis points per month with t = 2.21 and t = 4.56, respectively, both significant at the 5% level. These values are close to the 32 and 41 basis points per month reported by FF (2017). Moreover, the investment factor is significant when implemented among small stocks.

Turning to the AP region, the premiums are again similar to those in FF (2017). The standard value and investment factors had average returns of 58 and 35 basis point per month with t = 3.71 and t = 2.73, respectively, both highly significant. As in their study, none of the risk factors in the five-factor model generates premiums when implemented among big stocks. Moreover, while they found that the standard value factor in Japan produced a significant premium of 36 basis points per month with t = 2.19, our findings do not corroborate this result in the extended sample period. Indeed, none of the standard risk factors of the five-factor model produce significant premiums. As before, the reason for the discrepancy between our results and theirs is sample specific. The value factor in the January 2016 to August 2019 period produces an average return of –48 basis points per month with t = 1.03, thereby diminishing the overall premium in terms of economic magnitude and statistical significance. Moreover, we find that the value factor, when implemented among small stocks in Japan, produces 36 basis points per month with $t = 2.68$, which was insignificant in their study.

A commonality across different geographic regions is that the size factor does not produce any significant premium in the sample period as in FF (2017). Also, the standard momentum factor produces the highest premium among long/short portfolios in the NA, European, and AP regions ranging from 59 to 91 basis points per month with significant t-statistics ranging from 2.32 to 4.37. The momentum factor is considerably stronger when implemented among small stocks but does not reach significance among big stocks. Confirming Asness, Moskowitz, and Pedersen (2013), irrespective of size group, the momentum factor does not produce a significant premium in Japan.

3. Nested Models

Following FF (2018), we use spanning regression intercepts and GRS tests for choosing among nested models. We compare the FF (1993) three-factor model versus CAPM, five-factor versus three-factor models, and five-factor versus six-factor models. Using the local risk factors of the benchmark model, a new finding in Table 2 is that, for NA, Europe, and Japan, the intercepts associated with *SMB* and *HML* are statistically not different from zero. The corresponding GRS test of the intercepts from the spanning regressions for *SMB* and *HML* on Mkt^{ex} in Table 3 does not reject the hypothesis that the intercepts are jointly zero with p-values ranging from 0.13 to 0.28. These results imply that adding *SMB* and *HML* to Mkt^{ex} does not produce higher MSSR than Mkt^{ex} alone. By contrast, in the U.S. market, FF (2018) found that these combined factors produced a higher MSSR than Mkt^{ex} alone. The AP region is an exception, as adding *SMB* and *HML* to Mkt^{ex} yields a higher MSSR than the Mkt^{ex} alone. Given that the intercept of the regression of *SMB* on Mkt^{ex} is insignificant, we infer the value factor explains this result.

Based on spanning regressions of the profitability and investment factors on Mkt^{ex}, *SMB*, and *HML*, Table 2 indicates that at least one of these factors exhibits an intercept that is statistically significant. GRS tests on the intercepts from the spanning regressions for *RMW* and *CMA* on Mkt^{ex}, *SMB*, and *HML* in Table 3 reject the hypothesis that they are jointly zero with p-values ranging from 0.04 to 0.00. Hence, this combination of factors produces a higher MSSR than Mkt^{ex} alone, which is consistent with FF's U.S. evidence. Not surprisingly, adding the momentum factor to the five-factor model produces a higher MSSR in NA, Europe, and AP, which can be explained by the regression of the momentum factor on the risk factors of the five-factor model with intercepts between 68 and 108 basis points per month and significant t-statistics in the range of 2.76 to 4.64.

In sum, our international evidence confirms U.S. findings in FF (2018) that the six-factor model outperforms the other asset pricing models. However, contrary to FF's (1992, 1993, 1995) early studies, the popular three-factor model does not outperform the CAPM in most of the international markets investigated.

4. Non-Nested Models

FF (2018) split their sample period of 636 months into 318 adjacent pairs: months (1, 2), (3, 4) … (635, 636). Each simulation run draws (with replacement) a random sample of 318 pairs and randomly assigns the month for each pair to an in-sample (IS) sample month. They then use the IS months to compute IS MSSR values for all models in each run. Furthermore, each model's IS MSSR identifies weights for factors in the IS tangency portfolio in a simulation run. These weights are combined with the unused months of the chosen pairs to compute the out-of-sample estimate of the Sharpe ratio for each run. Even though IS MSSR are subject to upward bias according to FF, because monthly returns of factor portfolios are close to serially uncorrelated, out-of-sample (OS) Sharpe ratios are free of bias. In their study, FS, IS, and OS are compared, but IS and OS MSSRs are expected to rank models differently. Godfrey (2009) has referred to this methodology as pairs bootstrap, which allows for heteroskedasticity of unspecified form.

A potential problem in pairs boostrap is the assumption of serially uncorrelated factor portfolio returns. In this regard, it is generally accepted that regimes exist in stock or bond returns (e.g., see Guidolin and Timmermann, 2007; Ang and Bekaert, 2002; Perez-Quiros and Timmermann, 2000). Wahal (2019) has provided evidence of time dependency in factor returns by showing that the *CMA* factor is redundant prior to 1963. Similarly, Weber (2018) did not confirm FF's (2015) finding that the value factor is redundant. Also, the literature on momentum crashes suggests that momentum factor returns are subject to time dependency. As another empirical example, FF (2017) documented that the standard value factor in Japan produced a significant premium of 36 basis points per month with $t = 2.19$ in the sample period July 1990–December 2015. However, in the January 2016–August 2019 period, the payoff is –48 basis points per month, which is indicative of a regime shift. Finally, recent studies suggest that risk factors exhibit autocorrelation. In this regard, Arnott, Clements, Kalesnik and Linnainmaa (2019) employed 51 factors in the literature and showed that the momentum factor fully subsumes industry momentum. Forming momentum portfolios based on their one-month returns and holding them for a month ahead earned an annualized average return of 10.5% with significant $t = 5.01$ at any level including the multiple testing hurdle of 2.78 by Harvey, Liu, and Zhu (2016). These results imply first-order autocorrelation.

To address the dependency issue in factor returns, we propose a blocks bootstrap. This approach allows for autocorrelation of unknown form. To implement blocks bootstrap for 100,000 FS simulations, we divide our sample of 346 observations into 326 overlapping blocks. Each block has a length of 20 consecutive observations. Each run draws a random block with probability 1/326 and replacement. If the original sample size is achieved, the drawing ends and the sample is completed. For the IS and OS simulations, we divide our sample of 346 observations into 306 overlapping blocks. Each block has a length of 20 consecutive observations. Each run draws for the IS a random block with probability 1/306 and replacement. The following 20 observations in the time series are allocated in the OS. For example, if in the first iteration the block 306 (i.e., the last IS block) is picked, which comprises the observation series (306, 307, 308, …, 324, 325) and allocates them to the IS sample, the observation series (326, 327, 328, …, 345, 346) is allocated to the OS sample. This approach accounts for unknown dependencies in the time series of factor returns. In our analysis, we follow FF's advice "… that undisciplined search for the best model in a large set of potential factors can create an overwhelming multiple comparisons problem that preempts statistical inference. If we are to order models in a reliable way, the number of models considered must be limited." (FF, 2018, p. 236) Since we are interested in investigating the small factor model's superiority in international equity markets, we order two models on their MSSR: standard six-factor model using all stocks and a six-factor model that incorporates HML_S , RMW_S , CMA_S , MOM_S instead of the corresponding standard risk factors.

For NA, Europe, and Japan, our findings are similar to FF (2018). That is, the actual MSSR for the models that includes small spread risk factors produces the highest value. For instance, in Europe the actual MSSR for the model that includes small spread risk factors is 0.53 compared to the benchmark model at only 0.23. One exception is the AP region, for which the benchmark model exhibits the highest actual MSSR. In Appendix Table A.1, the corresponding results are reported using the FF pairs bootstrap approach. As shown there, full-sample (FS) estimates based on 100,000 simulation runs confirm that the model that includes small spread risk factors produces the highest value for MSSR in all markets.

In view of the upward bias problem discussed earlier, we employ the OS MSSR. As expected, OS estimates are consistently lower than IS estimates for all models. The OS approach confirms that the model with small spread risk factors produces the highest MSSR for NA, Europe, and Japan. However, in the AP region, the benchmark model is superior. One potential issue is that the MSSR values in the benchmark model and the small spread risk factor model are statistically indistinguishable from one another. To address this issue, Table 5 summarizes 100,000 FS, IS, and OS samples using blocks bootstrap. The table shows the average MSSR (Column) – MSSR (Row) (i.e., the difference between MSSR for a column model and a row model) as well as the percent of simulation runs in which the difference is negative. We observe from the FS, IS, and OS simulations that the benchmark model outperforms the small spread risk factor model in 1.62%, 0.11%, and 0.53%, respectively, of the simulation runs in NA. The corresponding results are all 0% for Europe and 21.01%, 15.87%, and 16.08% for Japan, which confirms that the first place model that combines Mkt^{ex}, SMB , and small stock spread factors performs better in terms of MSSR than the benchmark model in most simulation runs.

Unlike FF (2018), we propose the blocks bootstrap for simulation analyses of factors. For comparison purposes, we report the results using FF's pairs bootstrap approach in Appendix Table A.2. The results are similar for NA, Europe and Japan. In the AP region, the pairs bootstrap approach suggests that the standard benchmark model and the model that combines Mkt^{ex}, SMB, and small stock spread factors perform equally well (viz., the pairs bootstrap based on 100,000 suggests that the benchmark model beats the latter model in 44.52% and 49.76% of the respective IS and OS simulation runs). The corresponding results in Table 5 using blocks bootstrap are 64.45% and 82.63%, which indicates that the first place model for AP is again the standard benchmark model.

How much do the results between pairs and block bootstrap differ? To shed light on this question, we compound the differences between the IS and OS simulation runs. We find that the average OS MSSR is 0.0319 lower using pairs bootstrap than the average IS MSSR. Blocks bootstrap shows that the average OS MSSR is even worse at 0.0462, which implies that pairs bootstrap produces 44.95% higher MSSRs in OS simulation runs. This s*eemingly independence bias* occurs in situations wherein dependencies (perhaps long memory) in data are treated in simulations as if they are independent. In our paper, we use a heuristic of block length $\sqrt{T}$ rounded to the next decimal corresponding to 20 consecutive factor return observations. Increasing the block length tends to decrease in statistical power. An interesting topic would be to investigate the optimal block length. Because this empirical question is beyond the scope of the present paper, it is left for future research.

In sum, our findings indicate that, similar to FF's (2018) findings in the U.S. equity market, the model that combines Mkt^{ex}, SMB, and small stock spread factors outperforms the benchmark model in three-out-of-four international markets (viz., NA, Europe, and Japan). However, in the AP excluding Japan region, the benchmark model is the first place model. Also, departing from FF, we find that not taking into account dependency structures in financial markets can lead to biases in statistical inferences.

5. Spanning Regressions

Following FF (2018), Table 6 reports spanning regressions that explain each of the six factors in a model with the other five factors. To conserve space, we only report results for the standard factors. The marginal contribution of each factor is given by the squared intercept terms divided by the squared standard error of the regression equation. The *t*-statistic for the intercept in a factor's spanning regression measures the reliability of the factor's marginal contribution to the MSSR. The NA sample confirms FF's finding that *HML* is redundant in combination with *CMA*, but this result is not evident in the other international samples. In the latter samples, the value factor exhibits an economically high and statistically significant exposure to *CMA* ranging from 0.39 to 0.75, and the intercept terms of the value factor's spanning regressions are statistically significant even for Japan. Another finding contrary to FF is that, except for NA, *CMA* appears to be redundant in all other markets (viz., Europe, AP, and Japan). These findings are similar to FF (2017), who found that *HML* is significant in spanning regressions that control for the remaining risk factors in their five-factor model. Also, they documented that *CMA* is redundant in Europe and Japan. Our evidence shows that *CMA* is insignificant for AP after controlling for the remaining risk factors of the six-factor model, whereas in FF (2017) it is significant. Thus, we infer that factor spanning inferences are sensitive to sample specific data.

An important commonality that we observe from Table 6 is that, contrary to FF (2017, 2018), *SMB* does not matter irrespective of different samples. While small firms load negatively on *CMA* for NA and AP as in FF (2018) for the U.S. market, we do not find such evidence for Europe and Japan. Another commonality that we observe across our interational samples is that *RMW* plays an important role in explaining the cross section of stocks returns. In this regard, the intercepts of the spanning regressions are significantly different from zero at the 5% level or lower. This result also agrees with FF (2017), who reported positive and statistically significant intercepts for spanning regressions that regress the local profitability factor on the remaining local risk factors in their five-factor model. Finally, the momentum factor produces significant intercept terms in spanning regressions for NA, Europe, and AP but not Japan. The lack of momentum returns in Japan has been documented in earlier studies (Asness, Moskowitz, and Pedersen, 2013; FF, 2012). However, Asness (2011) found that Japanese equity results support momentum as a strong ex-ante strategy.

It is worth noting that the marginal contributions to MSSR of the momentum factor in Europe and AP equal 0.0393 and 0.0733, respectively, which are close to those for the market factor at 0.0480 and 0.0750, and noticeably higher than the other risk factors in the six-factor model. These results suggest that momentum plays a larger role in empirical asset pricing than previously recognized. Seminal work by Jegadeesh and Titman (1993) identified the momentum anomaly. Subsequently, Carhart (1997) augmented FF's three-factor model with the momentum factor. He found that this four-factor model substantially improved the average pricing errors of the CAPM and three-factor model. Also, Novy-Marx (2013) proposed a four-factor model incorporating the market factor in excess form plus industry-adjusted versions of the value, profitability, and momentum factors. Industry-adjusted factor portoflios were found to outperform standard risk factors in the Carhart four-factor model.[3]

6. Conclusions

Based on the sample period November 1990 to August 2019, we investigated the performance of FF's (2018) six-factor model in international equity markets. Employing tests for non-nested models, our results confirmed that the momentum factor played an important role for pricing equities in North America, Europe, and Asia excluding Japan but not Japan. Given that FF's winner model was a six-factor model combining Mkt^{ex} and SMB with the small stock spread factors HML_S , RMW_S , CMA_S , MOM_S , and their advice to focus on a short list of models, we compared three different six-factor model candidates. The six-factor model using the FF (2018) standard risk factors served as a benchmark model, whereas the model that combines Mkt^{ex} and SMB with the small stock spread factors denoted as HML_S , RMW_S , CMA_S , and MOM_S serves as alternative model.

In general, our findings confirmed FF (2018) for NA, Europe, and Japan wherein the six-factor model comprised of Mkt^{ex} and SMB with the small stock spread factors HML_S , RMW_S , CMA_S , and MOM_S produced the highest MSSR in 100,000 OS simulation runs. Since the empirical evidence suggested factor dependency structures, which is consistent with prior literature, we proposed simulations based on a blocks bootstrap approach instead of FF's pairs bootstrap approach. However, using this bootstrap approach in the AP region, the benchmark model performed better than the six-factor model combining Mkt^{ex} and SMB with the small stock spread factors HML_S , RMW_S , CMA_S , and MOM_S . As aptly phrased by Asness (2011), our empirical evidence for AP appears to be the "exception that proves the rule." Because Sharpe ratios in OS simulation runs using pairs bootstrap are systematically higher than in blocks bootstrap, we infer that pairs bootstrap can result in misleading statistical inferences due to an additional upward bias, dubbed s*eemingly independence bias*.

[3] Novy-Marx (2013) performed spanning regressions wherein each industry-adjusted risk factor is regressed on the risk factors of Carhart's four-factor model. Each regression produced intercepts that are significant with *t*-statistics equal to 4.17 (industry-adjusted momentum factor), 6.42 (industry-adjusted value factor), and 4.88 (industry-adjusted profitability factor). These results imply that industry-adjusted factors expand the mean-variance frontier. Moreover, his four-factor model was compared to Carhart's model in terms of ability to explain 15 investment anomalies. The root mean squared error of his four-factor model was only 0.22 compared to 0.54 for Carhart's model, which further supported his model.

An important commonality across economies covered in our study is that the local *RMW* factor plays an important role in pricing the cross section of international equity returns (i.e., the intercept in multivariate spanning regressions is statistically significant). This result confirms FF (2017) even after taking into account momentum factors in the set of control variables. Another commonality is that the *SMB* factor is insignificant across all countries. This result differs from FF (2018), who found that *SMB* generates a significant intercept when running multivariate spanning regressions. This divergent result could be attributable to using a shorter sample period than in their paper. Surprisingly, employing FF's (2018) test for nested models, the three-factor model did not outperform the CAPM in NA, Europe, and Japan. This finding is interesting in view of the fact that approximately 30 years ago FF proposed the three-factor model in lieu of the CAPM's inability to describe the cross section of average returns.

In sum, the main implications of our study can be summarized as follows: First, relying on asset pricing models for pricing equity portfolios can be a risky endeavor because factor models that appear to be valid in a given sample period are not necessarily relevant in extended samples. Second, investors engaging in factor investing need to be informed which factors are actually relevant in a given economy. Third, disregarding well-known dependency structures in factor data may result in biased statistical inference.

Our paper has a number of implications for future research. For example, Fleming, Kirby, and Ostdiek (2003), Barroso and Santa-Clara (2015), Daniel and Moskowitz (2016), and Moreira and Muir (2017) explored the effects of volatility timing on factor returns. Future research could investigate the effects of volatility timing on the superiority of small stock spread factors compared to big stock spread factors. Moreover, Asness, Moskowitz, and Pedersen (2013) documented the presence of momentum across various unrelated asset markets. While FF (2018) and our paper focused exclusively on the U.S. and international equity markets, future studies are encouraged to explore the implications of factor models accounting for momentum in other asset markets. In this regard, Okunev and White (2003) and Menkhoff, Sarno, Schmeling, and Schrimpf (2012) have examined the profitability of different momentum strategies in currency markets, and Lustig, Roussanov, Verdelhan (2011) have proposed a two-factor asset pricing model for currency markets consisting of dollar risk and carry risk factors. Finally, Racicot and Rentz (2016) and Racicot, Rentz, Tessier, and Théoret (2019) employ Racicot's (2015) methodology based on the Generalized Methods of Moments technique accounting for robust instruments to test the FF (2015) five-factor model extended by a liquidity factor. Their findings indicate that except for the market factor, all of the factors including liquidity are not significant. Addressing this issue is beyond the scope of this paper and therefore is left for future research.

References

Ang, A., G. Bekaert, 2002, International asset allocation with regime shifts, *Review of Financial Studies* 15, 1137–1187.

Ang, A., J. Chen, 2002, Asymmetric correlations of equity portfolios, *Journal of Financial Economics* 63, 443-494.

Arnott, R., M. Clements, V. Kalesnik, J. Linnainmaa, 2019, Factor momentum, Working paper, Dartmouth University.

Asness, C., 2011, Momentum in Japan: The exception that proves the rule, *Journal of Portfolio Management* 37, 67–75.

Asness, C.S., T. J. Moskowitz, L. H. Pedersen, 2013, Value and momentum everywhere*, Journal of Finance* 68, 929–985.

Barillas, F., J. Shanken, 2016, Which alpha?, *Review of Financial Studies* 30, 1316–1338.

Barroso, P., P. Santa-Clara, 2015, Momentum has its moments, *Journal of Financial Economics* 116, 111–120.

Carhart, M., 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57–82

Daniel, K., T. Moskowitz, 2016, Momentum crashes, *Journal of Financial Economics* 122, 221–247.

Fama, E. F., K. R. French, 1992, The cross-section of expected stock returns, *Journal of Finance* 47, 427–465.

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

Fama, E. F., K. R. French, 1995, Size and book-to-market factors in earnings and returns, *Journal of Finance* 50, 131–156.

Fama, E. F., K. R. French, 2008, Dissecting anomalies, *Journal of Finance* 63, 2971–2995.

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

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

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

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

Fleming, J., C. Kirby, B. Ostdiek, 2003, The economic value of volatility timing using "realized" volatility, *Journal of Financial Economics* 67, 473–509.

Godfrey, L. G., 2009. *Bootstrap Tests for Regression Models*. Palgrave MacMillian, New York.

Guidolin, M., A. Timmermann, 2007, Asset allocation under multivariate regime switching, *Journal of Economic Dynamics and Control* 31, 3503–3544.

Hamermesh, D. S., 2007, Viewpoint: Replication in economics, *Canadian Journal of Economics* 40, 715–733.

Harvey, C., Y. Liu, H. Zhu, 2015, … the cross-section of expected returns, *Review of Financial Studies* 29, 5–68.

Hou, K., Xue, C., and L. Zhang, 2020. Replicating Anomalies, *Review of Financial Studies* 33(5), 2019–2133.

Jegadeesh, N., S. Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, *Journal of Finance* 48, 65–91.

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

Menkhoff, L., L. Sarno, M. Schmeling, A. Schrimpf, 2012, Currency momentum strategies. *Journal of Financial Economics* 106, 660–684.

Moreira, A., T. Muir, 2017, Volatility managed portfolios, *Journal of Finance* 72, 1611-1644.

Novy-Marx, R., 2013, The other side of value: The gross profitability premium, *Journal of Financial Economics* 108, 1–28.

Okunev J., D. White, 2003, Do momentum-based strategies still work in foreign currency markets?, *Journal of Financial and Quantitative Analysis* 38, 425–447.

Perez-Quiros, G., A. Timmermann, 2000, Firm size and cyclical variations in stock returns, *Journal of Finance* 55, 1229–1262.

Racicot, F.E., 2015. Engineering robust instruments for GMM estimation of panel data regression models with errors in variables: a note. *Applied Economics* 47(10), 981-989.

Racicot, F.E. and Rentz, W.F., 2016. Testing Fama-French's New Five-Factor Asset Pricing Model: Evidence from Robust Instruments, *Applied Economics Letters* 23(6), 444-448.

Racicot, F.E., Rentz, W.F., Tessier, D. and Théoret, R. 2019. The Conditional Fama-French Model and Endogenous Illiquidity: A Robust Instrumental Variables Test. *PLOS ONE* 14(9), 1-26.

Wahal, S., 2019, The profitability and investment premium: Pre-1963 evidence. *Journal of Financial Economics* 131, 362–377.

Weber, M., 2018, Cash flow duration and the term structure of equity returns. *Journal of Financial Economics* 128, 486–503.

Table 1. Summary statistics for monthly factor returns

We download data from Kenneth French's website for the local FF (2015) five-factor models and the local momentum factors for North America (NA), Europe, Asia excluding Japan (AP), and Japan. These factors are used for the local benchmark models. Following FF (2018), the standard risk factors used in the benchmark model are defined as follows: Mkt_t^{ex} is the excess return on the value-weighted market portfolio at time t, SMB_t (small minus big) and HML_t (high minus low book-to-market equity) are the size and value factors of the FF (1993) three-factor model, RMW_t (robust minus weak) is a profitability factor, CMA_t (conservative minus aggressive) is an investment factor, and MOM_t (winners minus minus losers) is a momentum factor. Moreover, for each country, we also download the following local value-weighted portfolios: 25 size and book-to-market ratio, 25 size and profitability, 25 size and investment, and 25 size and momentum. These portfolios are used to construct small size and big size factors by conditioning the long-short strategies on only the first (small size factors) and fifth (big size factors) quintiles. The value, profitability, investment, and momentum factors are denoted as *HML*, *RMW*, *CMA*, and *MOM*, respectively. Panel A reports the results for North America, Panel B for Europe, Panel C for Asia excluding Japan, and Panel D for Japan. The sample period is from November 1990 to August 2019.

Factors	Mkt	SMB	HML	RMW	CMA	MOM
Panel A. North America						
Fama and French (2018)	0.66	0.10	0.15	0.33	0.24	0.59
	(2.91)	(0.66)	(0.86)	(2.60)	(1.73)	(2.32)
Small stocks factor			0.72	0.21	0.69	1.30
			(2.83)	(1.26)	(4.89)	(4.73)
Big stocks factor			-0.12	0.46	0.16	0.55
			(-0.56)	(2.54)	(0.79)	(1.59)
Small – Big stock factor			0.84	-0.24	0.53	0.75
			(3.53)	(-1.19)	(2.98)	(3.40)
Panel B. Europe						
Fama and French (2018)	0.46	0.04	0.28	0.38	0.18	0.91
	(1.79)	(0.34)	(2.21)	(4.56)	(1.93)	(4.37)
Small stocks factor			0.76	0.61	0.36	1.93
			(5.01)	(6.97)	(3.01)	(8.88)
Big stocks factor			0.09	0.45	0.09	0.56
			(0.42)	(2.34)	(0.63)	(1.73)
Small – Big stock factor			0.67	0.16	0.27	1.38
			(2.98)	(0.82)	(2.01)	(5.73)
Panel C. Asia excluding Japan						
Fama and French (2018)	0.66	-0.15	0.58	0.27	0.35	0.83
	(2.13)	(-0.98)	(3.71)	(1.86)	(2.73)	(3.59)
Small stocks factor			0.91	0.44	0.36	1.39
			(3.82)	(2.39)	(3.01)	(5.51)
Big stocks factor			0.37	0.11	0.09	0.17
			(1.30)	(0.40)	(0.63)	(0.45)
Small – Big stock factor			0.54	0.33	0.27	1.22
			(1.67)	(1.06)	(2.01)	(3.60)
Panel D. Japan						
Fama and French (2018)	0.06	0.11	0.27	0.13	0.06	0.09
	(0.20)	(0.66)	(1.74)	(1.16)	(0.46)	(0.39)
Small stocks factor			0.22	0.36	-0.12	0.03
			(1.03)	(2.68)	(-0.76)	(0.10)
Big stocks factor			0.42	0.12	-0.03	0.04
			(1.58)	(0.53)	(-0.14)	(0.11)
Small – Big stock factor			-0.20	-0.07	-0.09	-0.02
			(-0.61)	(-0.22)	(-0.43)	(-0.05)

Table 2. Spanning regressions for nested models all stocks

This table reports tests of whether: the excess market return (*Mkt*) spans the size spread factor (*SMB*) and value spread factor (*HML*); Mkt^{ex}, *SMB* , and *HML* span the investment spread factor (*CMA*,) and profitability spread factor (*RMW*); and Mkt^{ex}, *SMB*, *HML*, *CMA*, and *RMW* span the momentum spread factor (*MOM*). The tests focus on the intercept terms from spanning regressions of additional factors on base factors: *SMB* and *HML* regressed on Mkt^{ex}; *CMA* and *RMW* on Mkt, *SMB*, and *HML*; and *MOM* on Mkt^{ex}, *SMB*, *HML*, *CMA*, and *RMW*. The t -statistics for the coefficients are shown in parentheses. Panel A reports the results for North America, Panel B for Europe, Panel C for Asia excluding Japan, and Panel D for Japan.

Panel A. North America								
LHS	RHS							
	Intercept	*Mkt*	*SMB*	*HML*	*RMW*	*CMA*	R^2	s(e)
SMB	0.00	0.14					0.05	2.70
	(0.03)	(4.13)						
HML	0.27	-0.18					0.06	3.15
	(1.58)	(-4.59)						
RMW	0.41	-0.13	-0.29	0.23			0.35	1.95
	(3.91)	(-4.98)	(-7.58)	(6.80)				
CMA	0.26	-0.16	-0.00	0.58			0.67	1.51
	(3.19)	(-8.02)	(-0.13)	(22.62)				
MOM	0.68	-0.21	0.30	-0.69	0.13	0.36	0.16	4.37
	(2.76)	(-3.20)	(3.19)	(-5.59)	(1.04)	(2.31)		
Panel B. Europe								
SMB	0.07	-0.08					0.03	2.10
	(0.66)	(-3.26)						
HML	0.24	0.09					0.03	2.33
	(1.91)	(3.42)						
RMW	0.50	-0.06	-0.03	-0.33			0.34	1.27
	(7.33)	(-4.46)	(-1.04)	(-11.40)				
CMA	0.12	-0.15	-0.05	0.47			0.47	1.30
	(1.75)	(-10.29)	(-1.52)	(15.86)				
MOM	0.68	-0.12	0.11	-0.25	0.81	0.28	0.24	3.42
	(3.39)	(-2.65)	(1.29)	(-2.16)	(5.58)	(1.99)		

(Table 2 continued)

Panel C. Asia								
LHS				RHS				
	Intercept	*Mkt*	*SMB*	*HML*	*RMW*	*CMA*	R^2	s(e)
SMB	-0.14	-0.01					0.00	2.91
	(-0.92)	(-0.48)						
HML	0.54	0.06					0.02	2.92
	(3.45)	(2.36)						
RMW	0.63	-0.15	-0.17	-0.50			0.50	1.92
	(6.02)	(-8.27)	(-4.66)	(-14.27)				
CMA	0.36	-0.21	-0.07	0.21			0.29	2.03
	(3.24)	(-11.23)	(-1.94)	(5.51)				
MOM	1.08	-0.07	0.08	-0.52	-0.05	0.36	0.16	3.99
	(4.64)	(-1.44)	(1.09)	(-5.27)	(-0.43)	(3.40)		
Panel D. Japan								
SMB	0.11	0.06					0.01	3.19
	(0.64)	(1.92)						
HML	0.28	-0.09					0.03	2.85
	(1.81)	(-3.50)						
RMW	0.24	-0.13	-0.06	-0.34			0.29	1.80
	(2.45)	(-7.26)	(-1.92)	(-9.93)				
CMA	-0.08	0.04	0.10	0.46			0.34	1.93
	(-0.76)	(2.27)	(3.15)	(12.45)				
MOM	0.14	-0.12	-0.04	-0.33	0.42	0.04	0.15	4.05
	(0.62)	(-2.60)	(-0.55)	(-3.50)	(2.82)	(0.30)		

Table 3. Multifactor tests

This table shows GRS statistics from Gibbons et al. (1989) and associated *p* -values testing whether the additional factors jointly improve the maximum squared Sharpe ratio (MSSR) produced by the factors of the benchmark model (i.e., the capital asset pricing model (CAPM) or the FF three-factor model). Panel A reports the test results for North America, Panel B for Europe, Panel C for Asia excluding Japan, and Panel D for Japan.

Panel A. North America			
Model	LHS	GRS	p-value
CAPM	SMB, HML	2.53	0.2817
Three-factor model	RMW, CMA	27.16	0.0000
Panel B. Europe			
CAPM	SMB, HML	4.01	0.1348
Three-factor model	RMW, CMA	56.23	0.0000
Panel C. Asia excluding Japan			
CAPM	SMB, HML	13.46	0.0012
Three-factor model	RMW, CMA	41.59	0.0000
Panel D. Japan			
CAPM	SMB, HML	3.48	0.1758
Three-factor model	RMW, CMA	6.65	0.0359

Table 4. Comparison of six-factor models (blocks bootstrap)

This table provides the results for implementing blocks bootstrap for 100,000 full-sample (FS) simulations. We divide our sample of 346 observations into 326 overlapping blocks. Each block has a length of 20 consecutive observations. Each run draws a random block with probability 1/326 and replacement. If the original sample size is achieved, the drawing ends and the sample is completed. For in-sample (IS) and out-of-sample (OS) simulations, we divide our sample of 346 observations into 306 overlapping blocks. Each block has a length of 20 consecutive observations. Each run draws for the IS a random block with probability 1/306 and replacement. The following 20 observations in the time series are allocated to the OS. Average MSSRs are computed. Panel A reports the MSSR results for North America, Panel B for Europe, Panel C for Asia excluding Japan, and Panel D for Japan.

Model	Actual	FS	IS	OS	IS-OS
Panel A. North America					
Standard six-factor model (FF 2018)	0.1409	0.1679	0.1106	0.0860	0.0246
Small stock six-factor model	0.2297	0.2690	0.2871	0.2062	0.0809
Panel B. Europe					
Standard six-factor model (FF 2018)	0.2319	0.2711	0.1828	0.1488	0.0340
Small stock six-factor model	0.5307	0.6064	0.6059	0.4880	0.1179
Panel C. Asia excluding Japan					
Standard six-factor model (FF 2018)	0.2557	0.2793	0.3022	0.2642	0.0380
Small stock six-factor model	0.2362	0.2809	0.2750	0.1961	0.0789
Panel D. Japan					
Standard six-factor model (FF 2018)	0.0308	0.0546	0.0368	0.0186	0.0182
Small stock six-factor model	0.0464	0.0785	0.0684	0.0498	0.0186

Table 5. Distributions of differences between MSSR column model minus row model

This table shows average differences between MSSR for a column model and a row model from 100,000 full-sample, in-sample, and out-of-sample simulation runs, as well as the percent of simulation runs in which the row model has higher MSSR than the column model (% < 0). Panel A reports the test results for North America, Panel B for Europe, Panel C for Asia excluding Japan, and Panel D for Japan.

Panel A. North America (blocks bootstrap)

Model	$Mkt, SMB, HML, RMW, CMA, MOM$		$Mkt\ , SMB\ , HML_S , RMW_S\ , CMA_S , MOM_S$	
Full-sample	Mean	%<0	Mean	%<0
$Mkt, SMB, HML, RMW, CMA, MOM$			0.1011	1.62
$Mkt\ , SMB\ , HML_S , RMW_S\ , CMA_S , MOM_S$	- 0.1011	98.39		
In-sample				
$Mkt, SMB, HML, RMW, CMA, MOM$			0.1765	0.11
$Mkt\ , SMB\ , HML_S , RMW_S\ , CMA_S , MOM_S$	-0.1765	99.89		
Out-of-sample				
$Mkt, SMB, HML, RMW, CMA, MOM$			0.1202	0.53
$Mkt\ , SMB\ , HML_S , RMW_S\ , CMA_S , MOM_S$	-0.1202	99.47		

(Table 5 continued)

Panel B. Europe (blocks bootstrap)

Model	$Mkt, SMB, HML, RMW, CMA, MOM$		Mkt_S , SMB_S , HML_S , RMW_S , CMA_S , MOM_S	
Full-sample	Mean	%<0	Mean	%<0
$Mkt, SMB, HML, RMW, CMA, MOM$			0.3353	0.00
Mkt , SMB , HML_S , RMW_S , CMA_S , MOM_S	-0.3353	100.00		
In-sample				
$Mkt, SMB, HML, RMW, CMA, MOM$			0.4231	0.00
Mkt , SMB , HML_S , RMW_S , CMA_S , MOM_S	-0.4231	100.00		
Out-of-sample				
$Mkt, SMB, HML, RMW, CMA, MOM$			0.3392	0.00
Mkt , SMB , HML_S , RMW_S , CMA_S , MOM_S	-0.3392	99.98		

Panel C.. Asia excluding Japan (blocks bootstrap)

Model	$Mkt, SMB, HML, RMW, CMA, MOM$		Mkt_S , SMB_S , HML_S , RMW_S , CMA_S , MOM_S	
Full-sample	Mean	%<0	Mean	%<0
$Mkt, SMB, HML, RMW, CMA, MOM$			0.0016	50.68
Mkt , SMB , HML_S , RMW_S , CMA_S , MOM_S	-0.0016	49.32		
In-sample				
$Mkt, SMB, HML, RMW, CMA, MOM$			-0.0272	64.45
Mkt , SMB , HML_S , RMW_S , CMA_S , MOM_S	0.0272	35.55		
Out-of-sample				
$Mkt, SMB, HML, RMW, CMA, MOM$			-0.0681	82.63
Mkt , SMB , HML_S , RMW_S , CMA_S , MOM_S	0.0681	17.37		

(Table 5 continued)

Panel D. Japan (blocks bootstrap)

Model	$Mkt, SMB, HML, RMW, CMA, MOM$		$Mkt_S\ , SMB_S\ , HML_S\ , RMW_S\ , CMA_S\ , MOM_S$	
Full-sample	Mean	%<0	Mean	%<0
$Mkt, SMB, HML, RMW, CMA, MOM$			0.0239	21.01
$Mkt\ \ , SMB\ \ , HML_S\ , RMW_S\ \ , CMA_S\ , MOM_S$	-0.0239	78.99		
In-sample				
$Mkt, SMB, HML, RMW, CMA, MOM$			0.0316	15.87
$Mkt\ \ , SMB\ \ , HML_S\ , RMW_S\ \ , CMA_S\ , MOM_S$	-0.0316	84.14		
Out-of-sample				
$Mkt, SMB, HML, RMW, CMA, MOM$			0.0312	16.08
$Mkt\ \ , SMB\ \ , HML_S\ , RMW_S\ \ , CMA_S\ , MOM_S$	-0.0312	83.92		

Table 6. Spanning regressions

This table provides spanning regressions that explain each of the six factors in a model with the other five factors. To conserve space, we only report spanning regressions for the standard factors. The marginal contribution of each factor equals the squared intercept terms divided by the squared standard error of the regression equation. The *t*-statistic (in parentheses) for the intercept in a factor's spanning regression measures the statistical reliability of the factor's marginal contribution to the MSSR. Panel A reports the test results for North America, Panel B for Europe, Panel C for Asia excluding Japan, and Panel D for Japan.

Panel A. North America									
LHS	Int.	Mkt	SMB	HML	RMW	CMA	MOM	R^2	s(e)
Mkt	1.07		0.11	0.31	-0.46	-0.83	-0.14	0.30	3.53
	(5.49)		(1.41)	(3.05)	(-4.85)	(-6.93)	(-3.20)		
SMB	0.19	0.05		0.13	-0.49	-0.05	0.10	0.21	2.47
	(1.35)	(1.41)		(1.80)	(-7.64)	(-0.61)	(3.19)		
HML	-0.13	0.09	0.07		0.25	0.94	-0.12	0.68	1.85
	(-1.21)	(3.05)	(1.80)		(5.03)	(21.50)	(-5.59)		
RMW	0.42	-0.14	-0.30	0.28		-0.07	0.02	0.36	1.94
	(3.89)	(-4.85)	(-7.64)	(5.03)		(-0.96)	(1.04)		
CMA	0.22	-0.15	-0.02	0.62	-0.04		0.04	0.68	1.50
	(2.57)	(-6.93)	(-0.61)	(21.50)	(-0.96)		(2.31)		
MOM	0.68	-0.21	0.30	-0.69	0.13	0.36		0.16	4.37
	(2.76)	(-3.20)	(3.19)	(-5.59)	(1.04)	(2.31)			
Panel B. Europe									
Mkt	0.87		-0.34	0.70	-0.46	-1.35	-0.16	0.32	3.97
	(3.77)		(-3.39)	(5.40)	(-2.61)	(-9.06)	(-2.65)		
SMB	0.10	-0.10		0.08	-0.12	-0.14	0.04	0.04	2.11
	(0.76)	(-3.39)		(1.13)	(-1.26)	(-1.59)	(1.29)		
HML	0.34	0.11	0.05		-0.51	0.75	-0.05	0.56	1.58
	(3.70)	(5.40)	(1.13)		(-7.84)	(14.25)	(-2.16)		
RMW	0.39	-0.04	-0.04	-0.30		0.02	0.10	0.40	1.22
	(5.67)	(-2.61)	(-1.26)	(-7.84)		(0.37)	(5.58)		
CMA	0.06	-0.14	-0.05	0.50	0.02		0.04	0.48	1.30
	(0.77)	(-9.06)	(-1.59)	(14.25)	(0.37)		(1.99)		
MOM	0.68	-0.12	0.11	-0.25	0.81	0.28		0.24	3.43
	(3.39)	(-2.65)	(1.29)	(-2.16)	(5.58)	(1.99)			

(Table 6. continued)

LHS	Int.	Mkt	SMB	HML	RMW	CMA	MOM	R^2	s(e)
Panel C. Asia excluding Japan									
Mkt	1.29		-0.22	-0.02	-0.70	-1.00	-0.09	0.34	4.71
	(4.69)		(-2.41)	(-0.14)	(-5.43)	(-8.48)	(-1.44)		
SMB	0.02	-0.08		-0.06	-0.34	-0.10	0.04	0.07	2.81
	(0.14)	(-2.41)		(-0.80)	(-4.33)	(-1.29)	(1.09)		
HML	0.75	-0.00	-0.03		-0.68	0.39	-0.15	0.49	2.11
	(6.23)	(-0.14)	(-0.80)		(-14.07)	(7.14)	(-5.27)		
RMW	0.59	-0.11	-0.15	-0.54		0.17	-0.01	0.51	1.90
	(5.37)	(-5.44)	(-4.33)	(-14.07)		(3.29)	(-0.43)		
CMA	0.14	-0.17	-0.05	0.33	0.18		0.09	0.34	1.97
	(1.19)	(-8.48)	(-1.29)	(7.14)	(3.29)		(3.41)		
MOM	1.08	-0.07	0.08	-0.52	-0.05	0.37		0.16	3.99
	(4.64)	(-1.44)	(1.09)	(-5.27)	(-0.43)	(3.41)			
Panel D. Japan									
Mkt	0.41		0.13	-0.55	-1.26	-0.53	-0.17	0.22	4.82
	(1.58)		(1.52)	(-4.99)	(-7.53)	(-3.19)	(-2.60)		
SMB	0.10	0.05		-0.03	-0.00	0.28	-0.02	0.05	3.13
	(0.57)	(1.52)		(-0.35)	(-0.03)	(2.58)	(-0.55)		
HML	0.29	-0.12	-0.01		-0.18	0.54	-0.11	0.38	2.29
	(2.34)	(-4.99)	(-0.35)		(-2.12)	(7.25)	(-3.50)		
RMW	0.18	-0.11	-0.00	-0.07		-0.53	0.05	0.55	1.45
	(2.27)	(-7.53)	(-0.03)	(-2.12)		(-13.06)	(2.82)		
CMA	0.06	-0.06	0.07	0.25	-0.62		0.01	0.57	1.56
	(0.66)	(-3.19)	(2.58)	(7.25)	(-13.06)		(0.30)		
MOM	0.14	-0.12	-0.04	-0.33	0.42	0.04		0.15	4.05
	(0.62)	(-2.60)	(-0.55)	(-3.50)	(2.82)	(0.30)			

Appendix

Table A.1. Comparison of six-factor models (pairs bootstrap)

This table shows the average MSSRs based on 100,000 full-sample (FS), in-sample (IS), and out-of-sample (OS) simulation runs. FS simulations estimate MSSR using random samples (with replacement) of 346 months for the sample period November 1990–August 2019. IS and OS simulations split the 346 sample months into 173 adjacent pairs: months (1,2), (3,4), … (345, 346). A simulation run draws a random sample with replacement of 173 pairs. The IS simulation run chooses a month randomly from each pair in the respective run. IS MSSR is calculated for all models in each sample month. Panel A reports the test results for North America, Panel B for Europe, Panel C for Asia excluding Japan, and Panel D for Japan.

Model	Actual	FS	IS	OS
Panel A. North America				
Standard six-factor model (FF 2018)	0.1409	0.1616	0.1393	0.1151
Small stock six-factor model	0.2297	0.2524	0.2579	0.2187
Panel B. Europe				
Standard six-factor model (FF 2018)	0.2319	0.2568	0.2349	0.2111
Small stock six-factor model	0.5307	0.5726	0.5756	0.5217
Panel C. Asia excluding Japan				
Standard six-factor model (FF 2018)	0.2557	0.2814	0.2596	0.2292
Small stock six-factor model	0.2362	0.2676	0.2678	0.2287
Panel D. Japan				
Standard six-factor model (FF 2018)	0.0308	0.0469	0.0229	0.0116
Small stock six-factor model	0.0464	0.0634	0.0632	0.0359

A.2. Distributions of differences between MSSR column model minus row model (pairs bootstrap)

Using pairs bootstrap, the table shows average differences between MSSR for a column model and a row model based on 100,000 full-sample, in-sample, and out-of-sample simulation runs, as well as the percent of simulation runs in which the row model has higher MSSR than the column model (% < 0). Panel A reports the test results for North America, Panel B for Europe, Panel C for Asia excluding Japan, and Panel D for Japan.

Panel A. North America (pairs bootstrap)

Model	$Mkt, SMB, HML, RMW, CMA, MOM$		$Mkt_S, SMB_S, HML_S, RMW_S, CMA_S, MOM_S$	
Full-sample	Mean	%<0	Mean	%<0
$Mkt, SMB, HML, RMW, CMA, MOM$			0.0908	1.44
$Mkt\ , SMB\ , HML_S, RMW_S, CMA_S, MOM_S$	-0.0908	98.56		
In-sample				
$Mkt, SMB, HML, RMW, CMA, MOM$			0.1186	1.03
$Mkt\ , SMB\ , HML_S, RMW_S, CMA_S, MOM_S$	-0.1186	98.98		
Out-of-sample				
$Mkt, SMB, HML, RMW, CMA, MOM$			0.1036	1.11
$Mkt\ , SMB\ , HML_S, RMW_S, CMA_S, MOM_S$	-0.1036	98.89		

(Table A-2. continued)

Panel B. Europe (pairs bootstrap)

Model	$Mkt, SMB, HML, RMW, CMA, MOM$		$Mkt_S, SMB_S, HML_S, RMW_S, CMA_S, MOM_S$	
Full-sample	Mean	%<0	Mean	%<0
$Mkt, SMB, HML, RMW, CMA, MOM$			0.3158	0.00
$Mkt\ , SMB\ , HML_S, RMW_S, CMA_S, MOM_S$	-0.3158	100.00		
In-sample				
$Mkt, SMB, HML, RMW, CMA, MOM$			0.3407	0.00
$Mkt\ , SMB\ , HML_S, RMW_S, CMA_S, MOM_S$	-0.3407	99.97		
Out-of-sample				
$Mkt, SMB, HML, RMW, CMA, MOM$			0.3106	0.00
$Mkt\ , SMB\ , HML_S, RMW_S, CMA_S, MOM_S$	-0.3106	99.98		

(Table A-2. continued)

Panel C. Asia (pairs bootstrap)

Model	$Mkt, SMB, HML, RMW, CMA, MOM$		$Mkt_S, SMB_S, HML_S, RMW_S, CMA_S, MOM_S$	
Full-sample	Mean	%<0	Mean	%<0
$Mkt, SMB, HML, RMW, CMA, MOM$			-0.0138	58.99
$Mkt\ \ , SMB\ \ , HML_S, RMW_S, CMA_S, MOM_S$	0.0138	41.01		
In-sample				
$Mkt, SMB, HML, RMW, CMA, MOM$			0.0082	44.52
$Mkt\ \ , SMB\ \ , HML_S, RMW_S, CMA_S, MOM_S$	-0.0082	55.48		
Out-of-sample				
$Mkt, SMB, HML, RMW, CMA, MOM$			-0.0005	49.76
$Mkt\ \ , SMB\ \ , HML_S, RMW_S, CMA_S, MOM_S$	0.0005	50.24		

(Table A-2. continued)

Panel D. Japan (pairs bootstrap)				
Model	$Mkt, SMB,$ $HML, RMW,$ CMA, MOM		Mkt_S , SMB_S , HML_S , RMW_S , CMA_S , MOM_S	
Full-sample	Mean	%<0	Mean	%<0
$Mkt, SMB, HML, RMW, CMA, MOM$			0.0165	25.04
Mkt , SMB , HML_S , RMW_S , CMA_S , MOM_S	-0.0165	74.97		
In-sample				
$Mkt, SMB, HML, RMW, CMA, MOM$			0.0403	8.05
Mkt , SMB , HML_S , RMW_S , CMA_S , MOM_S	-0.0403	91.95		
Out-of-sample				
$Mkt, SMB, HML, RMW, CMA, MOM$			0.0243	10.42
Mkt , SMB , HML_S , RMW_S , CMA_S , MOM_S	-0.0243	89.58		