Journal Of Investment Management, Vol. 20, No. 1, (2022), pp. 48–69

www.joim.com

HOW WELL DO FACTOR ETFS CAPTURE THE FAMA–FRENCH FACTORS?

Nicholas Apergis[a], *Thomas Poufinas*[b,*], *Alexandros Panagakis*[c], *and Ioannis Ritsios*[d]

Institutional investors are investigating systematic, rule-based investment directions other than purely passive investing, such as factor-based investing. This study examines how well the factor-ETFs capture the Fama–French factors and attempts to explain their difference from the smart beta indexes applied in practice. The findings document that the market factor explains a substantial part of the expected returns, with the remaining factors, except momentum, posting smaller or no contribution. Style ETFs exhibit mixed results in capturing their referenced style, with almost all of them exhibiting non-neutral momentum. The findings are of interest to investment managers, investors, risk managers, and stock exchanges.

1 Introduction

Many asset managers and assets owners worldwide are seeking the (mix of) investment strategies that will be optimal for their portfolios and customers or beneficiaries to follow. Factor-investing is among the investment approaches followed by a series of institutional investors, among which are pension funds and sovereign wealth funds, which account potentially for the biggest portion of assets under management. Depending on the size of the portfolio, especially when the portfolio is relatively large, asset managers may choose to pursue factor-investing themselves. However, when the size of the portfolio

[a]University of Derby, Kedleston Road Campus, Derby, DE22 1GB, United Kingdom. Tel. +4407708051354, E-mail: n.apergis@derby.ac.uk
[b]Department of Economics, Democritus University of Thrace, Komotini, 69100, Greece. Tel: +302531039568, E-mail: tpoufina@econ.duth.gr
[c]Graduate Program in Business Mathematics, Athens University of Economics & Business and University of Athens, Athens, 15772, Greece. Tel: +306949557429, E-mail: alexandros.panagakis@gmail.com
[d]Discretionary Asset Management, National Bank of Greece Asset Management, 103 Syngrou Avenue, Athens 11745, Greece. Tel: +306973730660, E-mail: yritsios@nbgam.nbg.gr
*Corresponding author.

is smaller, then they may opt to rely on already available Exchange-traded funds (ETFs), in particular those that claim to follow the factors of interest.

Fama and French (1992, 1993, 2012, 2015) are the pioneers in introducing and testing such factors, extending essentially the Capital Asset Pricing Model (CAPM) in an attempt to better explain the return of assets and investment portfolios. They initially start with three-factor models and extend them to five-factor models; nowadays, six-factor models—extending the Fama—French five-factor model with the inclusion of momentum (MOM)—are employed in order to better describe the over-performance of securities and portfolios compared to the risk-free rate.

In addition to the research that has been conducted in the field both by academics and practitioners, we strongly believe that we can still contribute to the relevant literature, as is outlined in the following paragraphs. But why are ETFs fascinating to the investors and to the academics? And why did we choose to assess ETFs versus the Fama–French factors? We focus on ETFs because the investment industry has produced a big number of ETFs over the last 10 years that implement factor investing, i.e., select their securities based on certain factors of the Fama–French type. Why is there interest in factors within an ETF? As Vanguard (Grim *et al.*, 2017) puts it, *although stocks can be sorted in many different ways, attention is typically paid to those factors with an extensive academic literature and empirical evidence of historical positive risk-adjusted excess returns—in other words, certain factors that have "worked" in the past.* Consequently, the financial institutions have developed ETFs that follow specific factors. *Equity factor-based investing is a form of active management that aims to achieve specific risk or return objectives through systematic, rules-based strategies*. It has a wide number of applications and capitalizes on the choice of stocks that exhibit the relevant factor attributes. The use of factors in practice is exploited also by Dimensional Fund Advisors—DFA (Dimensional). Professor Fama is a Director of Dimensional. As a result, an experience of four decades is gradually implemented in the actual investment world. On June 26, 2020 Dimensional announced the launch of three ETFs. This makes our research more relevant than ever.

We do, therefore, believe that it is important to have a model—which we deliver in this paper—that would: (i) allow us to compare the performance of ETFs with the one that they would ideally have if they copied the Fama–French factor approach; (ii) enable us to predict the future performance for a given set of FF factor performance outcomes; (iii) assist us in explaining the contribution of each factor to the performance of the ETF; and (iv) following the anonymous reviewer, attempt to interpret why smart beta indexes applied in practice are different from the FF (research) factors.

More specifically, the analysis assesses a series of ETFs that (claim to) follow the Fama–French factors, as these ETFs may be used by investors in managing their portfolios through a factor investing approach. Therefore, the study attempts to provide evidence of the explanatory power of the six-factor model for these ETFs and to conclude whether they indeed follow their announced strategy. Moreover, we try to understand in what manner the relevant factors impact the performance of the ETFs over the risk-free rate. We do that by following two perspectives; one when the market return is measured by the performance of the specific underlying index that each of the ETFs follows and a second when the market return is measured by the performance of a common index, namely the S&P 500. The findings can be of substantial value to the asset owners, i.e., (primarily institutional) investors that wish to pursue factor-investing by employing these ETFs in order to take well-informed decisions with regard to the choice and usage of the most appropriate for their profile ETFs.

At the same time, the results highlight the points at which attention is required when the Fama–French factors are used, since there may be differences between the approach followed by Fama–French and the way it is implemented by the ETFs. Such differences are identified by our analysis. They are potentially due to the different definitions used by the ETFs compared with the Fama–French approach (possibly attributed to the difference between the smart beta indexes and the FF factors), the feasibility of the exact replication of the market or index, as well as the distribution of dividends. This is an additional novelty of this paper. Last, but not least, the study investigates the forecasting capacity of the five- and six-factor models, to realize that the addition of the momentum driver as the sixth factor improves the forecasting ability of the Fama–French modeling approach.

Summarizing, we identify our contribution to the existing literature in the following directions, as anticipated at the beginning of the introduction: (i) assessment of the (performance of the) ETFs that pursue factor investing—in total and as per their referenced style/smart beta index versus the Fama–French factors; (ii) explanation and evaluation of the contribution of each factor to the performance of the ETFs via the five- and six-factor models; (iii) prediction and comparison of the future performance of the ETFs through the five- and six-factor models to realize that momentum improves the forecasting capacity of the Fama–French models.

2 Background Discussion

There is a series of papers in the literature that present and analyze the characteristics, the behavior and performance of ETFs. We choose (among others) Elton *et al.* (2002), who examine the characteristics and performance of an exchange-traded index fund, namely the Standard & Poor's Depository Receipts or SPDR or Spiders to find that: (i) their NAV is kept close to market price due to the in-kind transactions, (ii) they underperform the S&P Index due to the lost income caused by holding dividends received on the underlying shares in cash, and (iii) they are the most actively traded stock and the instrument of choice for most hedging (at the time of their research).

Kostovetsky (2003) investigates the explicit and implicit costs incurred by ETFs and compares them with the costs of index mutual funds through a one-period and a multi-period model that depends on the investor trading preferences, tax implications, and other characteristics. He finds that the major differences between these two fund types are management fees, shareholder transaction fees, taxation efficiency, and other qualitative differences, which are important for decision-making. Guedj and Huang (2009) develop an equilibrium model to investigate whether an ETF is a more efficient indexing vehicle compared with an Open-Ended Mutual Fund (OEF) to show that: (i) OEFs provide cross-subsidization among investors by sharing the transaction costs for those investors experiencing large liquidity shocks, at the cost of lower average returns for the remaining investors; (ii) the OEFs provide partial insurance against future liquidity needs and is ex-ante beneficial for risk-averse investors, which can though cause moral hazard issues in the form of excessive trading and increase the cost of the insurance; and (iii) investors with higher liquidity needs benefit more from the liquidity insurance and hence prefer to invest via the OEF; and (iv) however, ETFs are better suited for narrower and less liquid underlying indexes, and for investors with longer investment horizons. Agapova (2011) examines conventional index mutual funds and exchange-traded funds (ETFs) by comparing aggregate fund flows into conventional open-ended index funds to those into ETFs for various underlying indexes to find that conventional funds and ETFs are substitutes, but not perfect substitutes for one another and that there is evidence that the coexistence of both instruments can be explained by a clientele effect that segregates the two vehicles into different market niches. Tse (2015) examines the profitability of momentum strategies with international iShares and US sector exchange-traded funds (ETFs) traded on the NYSE to find that: (i) momentum profits are insignificant for the late 1990s to 2014 period; (ii) few country and industry ETFs yield positive results using time series momentum; (iii) the overall performance is worse than the buy-and-hold strategy; and (iv) time series momentum offers significant profits during the 2008 global financial crisis, but the profits decline sharply for the post-crisis period.

Elton *et al.* (2019) examine the determinants that are important in explaining differences across funds following the same index and demonstrate how to select a passive vehicle that has a high probability of having the best performance in the following years. They find that: (i) for index funds: (a) cross-sectionally, the major determinants affecting pre-expense performance for index funds are turnover, the number of passive funds in the same family, and the return from security lending; and (b) by examining the standard deviation of return differences from the index they follow, the main determinant is the type of index followed, with emerging market indexes and foreign stock indexes having the largest deviations; and (ii) for ETFs: (a) the major determinants of the differential return are the number of passive funds in the same family and the amount of security lending they do; and (b) the standard deviation of deviations from the index is primarily determined by which index they follow, although security lending also plays a role. When they examine the determinants of the post-expense cross-sectional returns, they realize that the same factors matter, with the expense ratio becoming much more important in affecting differential return. None of these articles address the question that we investigate in our paper; they focus on the characteristics, the behavior, and performance of ETFs.

Fama and French (1992, 1993, 2012, 2015) have performed research on the inclusion of certain factors to extend the CAPM for many years, enhancing continuously their models, so as to incorporate factors that explain the excess returns of a portfolio over the risk-free rate. They did so by embodying factors that came out of the research of other individuals as well. Initially, Fama and French (1992) introduce two variables that are easily measured so as to capture the cross-sectional variation in average stock returns associated with the market beta (β), size, leverage, book-to-market equity, and earnings–price ratios. These are size and book-to-market equity. They find that when the tests allow for the variation in the β factor that is unrelated to size, the relation between the market β and average returns is flat, even when the β factor is the only explanatory variable.

Next, they identify five common risk factors in the returns of stock and bonds (Fama and French, 1993). Three of them are stock-market factors: namely, an overall market factor, and the factors related to firm size and book-to-market equity. Two of them are bond-market factors related to maturity and default risks. Stock returns are found to have a shared variation due to the stock market factors, whereas, they are linked to bond returns through shared variation in the bond-market factors. In addition, they find that the bond-market factors capture the common variation in bond returns, with the exception of low-grade corporate bonds. They conclude that the five factors seem to explain better the average returns on stocks and bonds.

Moreover, Fama and French (2015) explore a five-factor model that captures the size, value, profitability, and investment patterns in average stock returns. It behaves better than the three-factor model they developed in 1993. According to the authors, its limitation is that it fails to capture the low average returns on small stocks whose returns behave like those of firms that invest a lot despite low profitability. In addition, the performance of the model is not sensitive to the way the factors are defined. Moreover, with the addition of the profitability and investment factors, the value factor of the three-factor model seems to become redundant for describing the average returns in the sample that the authors examine.

Hou *et al.* (2014) provide supportive evidence that is quite close to the one by Fama and French (2015). In particular, they examine a (four-factor) q-model consisting of the market, the size, the investment, and the profitability factors to find that it largely summarizes the cross section of average stock returns, i.e., they do not include the High Minus Low (HML) factor. They focus on the returns associated with anomalies to see that about half of the anomalies are insignificant. In addition, they realize that the performance of the q-factor model is—with a few exceptions—at least comparable and in many cases better than that of the Fama–French (1993) three-factor model, as well as the Carhart (1997) four-factor model in capturing the remaining significant anomalies.

By contrast, Carhart (1997) attempts to capture momentum returns and thus proposes a four-factor model for returns in the U.S. This is essentially the three-factor model by Fama and French, enhanced with a fourth factor that accounts for the momentum returns, as measured by the difference between the monthly returns on diversified portfolios of the winners and losers of the past year. He demonstrates that common factors in stock returns and investment expenses almost completely explain the persistence in equity mutual funds' mean and risk-adjusted returns. He realizes the one-year momentum effect but finds that individual funds do not earn higher returns from following the momentum strategy in stocks. His explanation is based on the argument that the only persistence not explained is concentrated in a strong underperformance by the worst-return mutual funds. Consequently, the presence of skilled or informed mutual fund portfolio managers is not supported.

Fama and French (2012) continue to study the impact of size, value, and momentum in international stock returns. They examine four regions, namely, North America, Europe, Japan, and Asia Pacific. They find that there are value premia in average stock returns that decrease with size, with the exception of Japan. The same holds for the case of returns momentum; moreover, the spreads in average momentum returns decrease from smaller to bigger stocks. They test whether empirical asset pricing models can capture the value and momentum patterns in international average returns and whether asset pricing seems to be integrated across the four regions. They infer that integrated pricing across regions is not strongly supported in their tests. More specifically, for three regions (North America, Europe, and Japan), local models that use local explanatory returns provide possible descriptions of local average returns for portfolios formed on size and value versus growth. In addition, local models are less successful in tests on portfolios formed on size and momentum. Lilloe-Olsen (2016) tests the effect of the Fama–French factors on Nordic capital markets (Sweden, Denmark, Norway, and Finland). They investigate the three-, four-, five-, and six-factor models in two dimensions: (a) to price equity returns, and (b) to capture the returns on an individual and combinational basis. They find that there is evidence supporting the presence of factor effects in these markets.

Fama and French (2017) extend their five-factor model tests to international markets, namely, four regions, to find that average stock returns in North America, Europe, and Asia Pacific increase with the book-to-market ratio and profitability, whereas they decrease with investment. Japan has a different behavior; although the relation between average returns and the book-to-market ratio is strong, the average returns show a weak relation to profitability and investment. The five-factor model that adds profitability and investment to the three-factor model absorbs the patterns in average returns. The main limitation of this modeling approach is that it does not succeed in fully capturing the low average returns of small stocks whose returns behave like those of low profitability firms that invest aggressively.

A series of research papers assess the ability of the Fama–French models in explaining excess returns, as well as their predicting ability. The size effect, the book-to-market effect, and the ability of the Fama–French three-factor model to incorporate these effects and improve the asset pricing ability of the CAPM is studied for the Australian stock market by Gaunt (2004). He finds that the three-factor model provides significant improved explanatory power over the CAPM and evidence that the book-to-market factor plays a key role in asset pricing.

The power of multifactor models to explain mutual fund returns is studied by Huij and Verbeek (2009). They (i) identify the extent to which professional managers are able to capture the value, size, and momentum premia; and (ii) analyze the degree to which the use of these factor proxies systematically biases the performance estimates of mutual funds, as the factor proxies are based on hypothetical stock portfolios and as such do not take into account transaction costs, trade impact, and trading restrictions; hence, the factor premia may be overestimated or underestimated. They find that funds with a value- (growth-)oriented style and funds that hold stocks that do well (poorly) over the past year earning returns that are higher (lower) than those predicted by the CAPM. Their findings indicate though that the value and momentum premia are significant and persist beyond transaction costs and trading restrictions. However, the factor premia seem to be miscalculated; the performance estimates for value funds and past losers are biased downwards and those for growth funds and past winners are biased upwards. Consequently, the use of the standard three- and four-factor models to mutual funds needs to be made with caution. They recommend as an alternative the construction of the corresponding factors with the use of mutual funds returns instead of stock returns. Cakici *et al.* (2013) examine the value and momentum effects in 18 emerging stock markets to find strong evidence for the effect of the value factor in all emerging markets and for the effect of the momentum factor for all, but Eastern Europe. They also form portfolios sorted on size and book-to-market ratio and portfolios sorted in size and lagged momentum to explain their returns based on factors constructed using local, U.S., and aggregate global developed stock markets data to find that local factors perform much better. The latter indicates the presence of emerging market segmentation.

The validity and accuracy of the CAPM and the three-factor Fama–French models are tested for the Johannesburg Stock Exchange (Karp and van Vuuren, 2017) to realize that both models perform relatively poorly, most likely due to inadequate market proxy measures, market liquidity restrictions, risk factors that are not priced-in and volatility inherent in an emerging market environment. The value premium is found to explain a larger proportion of the variation in the excess returns compared with the size premium and is more pronounced in portfolios with relatively higher book-to-market. The effectiveness of the Fama–French three-factor model and the CAPM in explaining the excess return in the Dhaka Stock Exchange is examined by Sattar (2017). He analyzes five listed firms of the cement industry to realize that the Fama–French three-factor model has a better predicting ability of the variation in excess returns over the risk-free rate, compared with the CAPM, while the coefficients are all statistically significant. The forecasting ability of the Fama–French factors is investigated by Panopoulou and Plastira (2014). They test the HML, SMB, the long-term reversal, and momentum factors for the U.S. stock returns to realize that they exhibit a better predicting ability, which suggests that these factors contain significant information for future stock market returns. They also find that the default spread, and to a lesser extent the term spread, contains important information for the evolution of these factors. In addition, appropriate decompositions of the factors in their size and value components can enhance predictability.

A series of papers follow the actively managed ETFs. Robotis (2011) compared their risk and return with the risk and return of the market (S&P 500 Index) to find that there is no significant difference between them. With the use of CAPM, the Fama–French three- and four-factor models, he realized that the managers did not deliver any significant excess return (alpha – α). Moreover, the managers do not seem to exhibit any ability to time the market. Garyn-Tal (2014) found an investment strategy in active ETFs, with the use of the Fama–French–Carhart four-factor model, which delivers a positive risk-adjusted excess return. Finally, Schizas (2014) empirically assessed active ETFs in terms of risk, return, and incentives and compared them with passive ETFs, mutual funds, and hedge funds to realize that they are not as active as they are perceived.

When it comes, however, to the assessment of the contribution of the Fama–French factors in explaining the ETF returns not much has been found in the academic literature. Practitioners have been more productive though. Vanguard (Grim *et al.*, 2017) has produced a guide to the suitability and key considerations for the construction of equity factor-based portfolios. Morningstar (Boyadzhiev *et al.*, 2018) has developed a framework for analyzing multifactor funds. Consequently, the novelty of our approach is that we do regress the excess returns of the ETFs, not only with the five factors but also with all six factors, while we investigate the forecasting capability of the five- and six-factor models. We consider this important as it incorporates into one single model all the factors that are currently followed by ETFs.

3 Data, Variables, and Methodology

3.1 Data

For the Fama–French factors the analysis uses the notation, definitions, and language as introduced by French (2019) in his Data Library, where he explains how the Fama–French factors are constructed. More precisely, the five Fama–French factors are constructed using the six value-weighted portfolios formed on size and book-to-market, the six value-weighted portfolios formed on size and operating profitability, and the six value-weighted portfolios formed on size and investment: Small Minus Big (SMB) is the average return on the nine small stock portfolios minus the average return on the nine big stock portfolios, while HML (High Minus Low) is the average return on the two value portfolios minus the average return on the two growth portfolios. Robust Minus Weak (RMW) is the average return on the two robust operating profitability portfolios minus the average return on the two weak operating profitability portfolios, while Conservative Minus Aggressive (CMA) is the average return on the two conservative (low) investment portfolios minus the average return on the two aggressive (high) investment portfolios. $R_m - R_f$ measures the excess return of the market over the risk-free rate. The first component is the value-weighted return of all the CRSP (Center for Research in Security Prices) firms incorporated in the U.S. and listed on the NYSE, AMEX, or NASDAQ that have the required data as described in French (2019). The second component is the one-month Treasury-bill rate, as taken from Ibbotson Associates. Finally, WML (Winners Minus Losers), a.k.a. momentum (MOM), is the equal-weighted average of the returns for the two winner portfolios for a region minus the average of the returns for the two loser portfolios. To measure for momentum, the Fama–French portfolios are sorted on size and lagged momentum. For portfolios formed at the end of month $t-1$, the lagged momentum return is a stock's cumulative return for month $t-12$ to month $t-2$. The momentum break points for a region are the 30th and 70th percentiles of the lagged momentum return of the big stocks of the region. The time span period is from 2008 to 2018. The analysis used Bloomberg as the source of the ETF-related data, i.e., the ETFs themselves (name, ISIN number, underlying index), their share prices, as well as the level of their underlying indexes and the level of the S&P 500 Total Return Index. We calculated the monthly returns of the ETFs from the corresponding ETF monthly closing prices.

3.2 Variables

The independent variables are consequently the six Fama–French factors, i.e., $R_m - R_f$, SMB, HML, RMW, CMA, and MOM. R_m is either the return of the underlying index or of the S&P 500 Total Return Index, depending on the model under investigation. Our dependent variable is the excess return of each ETF over the risk-free rate $R_{\mathrm{ETF}} - R_f$.

3.3 Methodology

We attempt to link the excess performance of the ETFs over the risk-free rate with the Fama–French factors with the use of a linear regression. The general form of the regression equation yields:

$$\begin{aligned} R_{\mathrm{ETF}_t} - R_{f_t} &= a_t + \beta_1(R_{mt} - R_{f_t}) + \beta_2 SMB_t \\ &\quad + \beta_3 HML_t + \beta_4 RMW_t \\ &\quad + \beta_5 CMA_t \end{aligned}$$

or

$$\begin{aligned} R_{\mathrm{ETF}_t} - R_{f_t} &= a_t + \beta_1(R_{mt} - R_{f_t}) + \beta_2 SMB_t \\ &\quad + \beta_3 HML_t + \beta_4 RMW_t \\ &\quad + \beta_5 CMA_t + \beta_6 MOM_t \end{aligned}$$

depending on whether we use the MOM factor or not. The analysis uses OLS, along with White's test to detect potential heteroskedasticity, while using Robust Standard Errors to tackle it when present.

4 Empirical Analysis and Results

4.1 Regressions against the underlying index: The period 2008–2018

Exhibit 1 provides some descriptive statistics for the 34 ETFs for which data were available for the entire period of 2008 to 2018. SMB, HML, RMW, CMA, and MOM supporting descriptive statistics of size, value, operating profitability, investment sorted, and momentum factors for explaining portfolio returns. The results for the selected ETFs—which are marked with their ISIN (International Securities Identification Number)—indicate that the average market premium is negative, the average value premium is positive, and the average profitability premium is positive, whereas the average size premium is negative, the average investment premium is positive, and the average momentum premium is negative. The market premium is found less volatile, in contrast with the size, the value, the profitability, and the investment premia. The momentum premium seems to be the most volatile.

A negative SMB implies that the average returns of the big stocks outperform the average returns of the small ones, while a positive HML indicates that the average returns of the value stocks are larger than the average returns of the growth ones. A positive RMW signifies that the average returns of highly profitable stocks outperform the low profitable ones, while a positive CMA indicates that the average returns of high investment stocks are smaller than the average returns of the

Exhibit 1 Descriptive statistics.

Portfolios (ETFs)	Mean	SD	Minimum	Maximum	Portfolios (ETFs)	Mean	SD	Minimum	Maximum
Market-Rf	−0.27	0.025	−0.30	−0.24	US4642873099	−0.29	0.013	−0.30	−0.28
SMB	−1.77	1.782	−3.03	−0.51	US78464A4094	−0.29	0.009	−0.29	−0.28
HML	0.87	1.541	−0.22	1.96	US4642876068	−0.29	0.019	−0.31	−0.28
RMW	1.17	1.457	0.14	2.20	US46137V2667	−0.28	0.003	−0.29	−0.28
CMA	1.07	3.649	−1.51	3.65	US4642876712	−0.29	0.008	−0.29	−0.28
MOM	−3.08	6.809	−7.89	1.74	US4642878874	−0.29	0.028	−0.31	−0.27
US4642888774	−0.27	0.025	−0.29	−0.26	US78464A2015	−0.30	0.013	−0.31	−0.29
US4642888857	−0.27	0.027	−0.28	−0.25	US78464A8210	−0.29	0.027	−0.31	−0.27
US4642882736	−0.29	0.032	−0.31	−0.27	US46137V1750	−0.29	0.052	−0.32	−0.25
US4642865822	−0.26	0.037	−0.29	−0.24	US46137V2170	−0.30	0.025	−0.32	−0.28
US4642874089	−0.27	0.032	−0.29	−0.24	US4642878049	−0.29	0.037	−0.32	−0.26
US4642878791	−0.28	0.048	−0.32	−0.25	US78464A8137	−0.30	0.028	−0.32	−0.28
US4642877058	−0.28	0.035	−0.31	−0.26	US78463X8719	−0.31	0.024	−0.32	−0.29
US78464A5083	−0.27	0.028	−0.29	−0.25	US78467Y1073	−0.29	0.026	−0.31	−0.27
US4642876639	−0.27	0.028	−0.29	−0.25	US4642875078	−0.29	0.027	−0.31	−0.27
US46137V2584	−0.14	0.068	−0.19	−0.09	US46137V4564	−0.27	0.021	−0.29	−0.26
US78464A3005	−0.28	0.049	−0.32	−0.25	US46137V4804	−0.28	0.059	−0.32	−0.24
US78464A8392	−0.28	0.043	−0.31	−0.25	US46137V4648	−0.30	0.034	−0.32	−0.28
US46137V1917	−0.27	0.102	−0.34	−0.19	US46137V4986	−0.30	0.006	−0.31	−0.30
US46137V1677	−0.25	0.119	−0.34	−0.17	US46137V4721	−0.28	0.001	−0.28	−0.28

Note: ETFs are in excess returns, SD = standard deviation.

Exhibit 2 Correlation matrix of the Fama–French factors.

Variables	MKT	SMB	HML	RMW	CMA	MOM
MKT	1.000					
SMB	0.143	1.000				
HML	−0.355	−0.386	1.000			
RMW	0.068	0.102	0.052	1.000		
CMA	0.148	0.298	−0.246	0.544	1.000	
MOM	−0.346	−0.209	0.224	−0.102	−0.447	1.000

low investment stocks. A negative MOM displays that the average returns of higher lagged momentum return stocks are lower than the average returns of lower lagged momentum return stocks. Finally, all individual excess returns over the risk-free rate indicate the presence of a negative premium.

Exhibit 2 reports the correlations across the explanatory (aggregate) variables to explore the likelihood of multicollinearity. They illustrate that the correlation between the market risk factor with the size and value factors is low. The market premium is positively correlated with the size premium, but the strength of this correlation is low, while it is negatively correlated with the value premium, with this correlation also being weak. Overall, the presence of correlation is low and in that sense no multicollinearity can be detected across the explanatory (aggregate) variables.

Exhibit 3 presents a synopsis of the results of the regression analysis of the Fama and French five- and six-factor models for all 34 portfolios under study. All of the 34 ETFs are sorted on the basis of size, book-to-market ratio, operating profitability, investment in total assets, and momentum premia. One by one, all of the 34 portfolios are regressed on market, size, value, operating profitability, investment, and momentum premia to explain the relationship between the ETF excess returns and those premia.

The empirical findings clearly highlight that the coefficient of the market premium is positive and statistically significant at 1% across all 34 ETFs. This implies that the market premium has a significant linear relationship with the portfolio returns in all the regressed ETFs, which is consistent with the conventional assets pricing model (CAPM). Moreover, the results also document that the significance (in terms of the *R*-squared) of the overall model increases whenever the MOM factor is explicitly introduced. Such findings remain consistently similar across all categories of funds. Finally, looking at all the factors themselves it seems that they capture most of the over-performance of the ETFs against the risk-free rate.

In addition, the results document that the SMB factor is significantly (at varying levels) and negatively associated with the excess portfolio returns of all the regressed ETFs. The negative value of the SMB coefficient indicates that potentially the ETFs concentrate more on big caps as this secures easier implementation, due to liquidity, trading cost (less trading) and volume. However, the relevant coefficients are close to zero, implying that it does not contribute a lot to the excess returns of the ETFs over the risk-free rate.

As far as the HML factor is concerned, the findings display that the value premium is positive and statistically significant at all levels; however, the

Exhibit 3 Regression results of the Fama–French five- and six-factor models (2008–2018).

Model	A	B	AM	BM	Model	A	B	AM	BM
US4642888774	X		X		US4642876068	X		X	
US4642888857		X	X		US46137V2667	X		X	
US4642882736	X		X		US4642876712	X		X	
US4642865822	X		X		US4642878874	X		X	
US4642874089	X		X		US78464A2015	X		X	
US4642878791	X		X		US78464A8210	X		X	
US4642877058	X		X		US46137V1750	X		X	
US78464A5083	X		X		US46137V2170	X		X	
US4642876639	X		X		US4642878049	X		X	
US46137V2584	X		X		US78464A8137	X		X	
US78464A3005	X		X		US78463X8719	X		X	
US78464A8392	X		X		US78467Y1073	X		X	
US46137V1917	X		X		US4642875078	X		X	
US46137V1677	X		X		US46137V4564	X		X	
US4642873099	X		X		US46137V4804	X		X	
US78464A4094	X		X		US46137V4648	X		X	
US4642876068	X		X		US46137V4986	X		X	
US46137V2667	X		X		US46137V4721	X		X	

Model A: All variables except RMW are statistically significant – MOM not included. Model B: All variables are statistically significant – MOM not included. Model AM: All variables except RMW are statistically significant – MOM included. Model AM: All variables are statistically significant – MOM included. R^2: mean: 0.62; standard deviation: 0.05; max: 0.75; min: 0.50. Statistical significance: p-value $\leq$0.10.

coefficients are also very close to zero. This yields that the HML factor has also a small contribution to the excess return of the ETFs over the risk-free rate. The positive HML coefficient means that the ETFs focus more on value. In terms of the RMW factor, the results signify that it is not statistically significant across all portfolio returns, except in the case of one ETF, where it turns out to be significant at 10%. Therefore, these findings clearly support the insignificance of this factor. It could mean that none of the ETFs or their underlying indexes focuses on the operating profitability.

However, in terms of the CMA factor, the findings illustrate that it turns out to be significantly and negatively associated with excess portfolio returns, for the majority of the funds at the 1% level. In other words, the selected ETFs tend to be weighted more towards aggressive investment stocks, i.e., with lower dividend yield. This implies that there is somehow more aggressive investment, betting on growth and future profitability. Once again, the factor is relatively small; hence, a small contribution is attributed to this factor. The MOM factor turns out to be positively—for some ETFs—and negatively—for some other ETFs-related and statistically significant (at 1%) across all ETFs. This can be possibly interpreted by the fact that the selected with a positive (negative) coefficient ETFs tend to invest more in winners (losers) in terms of lagged momentum, i.e., stocks that exhibit higher (lower) cumulative returns over the past period as incorporated by the relevant factor. The coefficient of this factor is higher than those of SMB, HML, RMW, and CMA, which could mean that it has a higher contribution to the excess returns of the ETFs. Finally, the insignificance of the intercept, i.e., alpha (α), in the majority of cases potentially signifies that either little or no active management versus the underlying index is attempted by the ETF managers or that the active management did not yield any over-performance compared to the Fama–French model. As the adjusted R^2 ranges between 0.53 and 0.79 makes us wonder whether there is room is left to include further sources of the systematic risk factors in the six-factor model. We draft a potential direction in the future venues section.

4.2 *Robustness check*: *The period* 2014–2018

This part of the analysis repeats the baseline estimations reported in Exhibit 3, but this time over the period 2014–2018 where more (68) ETFs enter the sample. The new results are reported in Exhibit 4 in a concise format. All of the 68 ETFs are sorted again on the basis of size, book-to-market ratio, operating profitability, investment in total assets, and momentum premia. One by one, all of the 68 portfolios are regressed in a similar to the period 2008–2018 manner. The findings of these regressions are identical to the ones for the period 2008–2018, with two minor differences: (i) the RMW factor is not statistically significant across ALL new portfolio returns; and (ii) the CMA factor, turns out to be significantly and negatively associated with excess returns at the 1% level for ALL ETFs. These results verify the validity of our initial findings and along with the insignificance of the intercept—in most of the ETFs—indicate AGAIN that there is little room left for additional factors.

4.3 *Regressions against a common index*

As a consequence of the aforementioned observations we regress the ETFs against the S&P 500 Total Return Index, i.e., a common index, so as to investigate the contribution of other factors. We categorized the funds according to their descriptions (styles) in an attempt to evaluate how closely they follow their investment strategy. The results are summarized in a synoptic manner in Exhibit 5. We realize that: (a) the excess performance of the market contributes the most in the excess returns of the ETF over the risk-free rate; (b) HML, CMA, and MOM contribute the most among the remaining factors in the excess returns of the ETF over the risk-free rate; and (c) SMB and RMW have almost no contribution in explaining the excess performance of the ETF above the risk-free rate.

The market beta gets lower; this is probably attributed to the fact that the ETF underlying index deviates from the S&P 500. The other coefficients get higher, but still not very much higher, except for momentum that increases even further. Significance gets lower for CMA, as it posts at all cases low (10%) or no significance. Furthermore, CMA is negative, which indicates an aggressive investment, betting on growth and future profitability. RMW remains statistically insignificant, while momentum remains significant at all levels for all ETFs. Momentum is positive for the US-oriented Momentum ETF and negative for International, Emerging and US Tilt Momentum ETFs.

The remaining results are comparable, as (i) value seems to be appropriately captured by the Value ETFs; (ii) growth is not captured by the Growth ETFs, as the sign is the opposite of the anticipated one; and (iii) size is not captured by the small- and mid-cap ETFs, as the sign of the coefficient is the opposite of the expected one. The deviation of the sign of the SMB coefficient from the anticipated one has been observed by Chen and Bassett (2014). They present a very large portfolio with a positive SMB coefficient in the three-factor Fama–French model. Consequently, a positive SMB coefficient does not necessarily

Exhibit 4 Regression results of the Fama–French five- and six-factor models (2014–2018).

Model	A	C	AM	CM	Model	A	C	AM	CM
US9220427184		X		X	US46137V2667	X		X	
US4642888774		X		X	US4642876712	X		X	
US46432F3881		X		X	US4642878874	X			X
US46432F3964	X		X		US9219325050	X			X
US4642888857	X			X	US78464A2015	X			X
US4642882736	X		X		US9219328690	X		X	
US4642865822	X		X		US78464A8210	X			X
US46429B6149	X			X	US9219327940	X		X	
US4642864759	X		X		US46137V1750	X			X
US4642891315	X		X		US46137V2170	X		X	
US46429B4169	X		X		US4642878049	X		X	
US46429B2007	X		X		US46138G1022		X		X
US46429B4656	X		X		US46138E1495	X			X
US4642874089	X		X		US78464A8137	X			X
US4642878791	X		X		US78463X7562	X			X
US4642877058	X		X		US78463X8719	X			X
US78464A5083	X		X		US9219328286		X		X
US4642876639	X			X	US68386C3025		X		X
US46137V2584	X		X		US46138E1727		X		X
US78464A3005	X		X		US46138E1156	X			X
US78464A8392	X		X		US46138E1560	X		X	
US9219327031	X		X		US46138E1230	X			X
US9219327783	X		X		US46137V1834		X		X
US9219328443	X		X		US46138E1644	X			X
US46137V1917	X		X		US46138E1800	X			X
US46137V1677	X		X		US46138G2012	X			X
US78464A1280	X			X	US78467Y1073		X		X
US78468R7052	X			X	US4642875078		X		X
US46138E2899	X		X		US46137V4564	X		X	
US46138E2220	X		X		US46137V4804	X			X
US4642873099	X		X		US46137V4648	X		X	
US78464A4094	X		X		US46137V4986		X		X
US4642876068	X		X		US46137V4721		X		X

Model A: All variables except RMW are statistically significant – MOM not included. Model C: All variables except RMW and SMB are statistically significant – MOM not included. Model AM: All variables except RMW are statistically significant – MOM included. Model AM: All variables except RMW and SMB are statistically significant – MOM included. R^2: mean: 0.64; standard deviation: 0.06; max: 0.79; min: 0.53. Statistical significance: p-value $\leq$0.10.

Exhibit 5 Regression results of the Fama–French five- and six-factor models (2014–2018)—A common market index.

Model	A	C	D	E	AM	CM	DM	EM	Model	A	C	D	E	AM	CM	DM	EM
Small-cap value, Small-cap growth, Small-cap equal weight, Mid-cap-equal weight, Mid-cap value, Mid-cap growth, Large- and mid-cap -growth																	
US46137V1677	X				X				US4642877058	X				X			
US46137V1750	X						X		US4642878791	X				X			
US46137V1834	X						X		US4642878874	X				X			
US46137V1917	X						X		US78464A2015	X				X			
US46137V2170	X						X		US78464A3005	X				X			
US46137V2667	X				X				US78464A8210	X				X			
US46137V4564	X						X		US9219327783	X				X			
US46137V4648	X						X		US9219327940	X				X			
US46137V4721	X						X		US9219328443	X				X			
US4642876068	X				X				US9219328690	X				X			
US4642876639	X						X										
Value																	
US46137V2584		X				X			US78464A1280		X						X
US46137V4804				X				X	US78464A5083		X				X		
US4642874089				X				X	US78464A8392		X						X
US4642888774				X				X	US9219327031			X					X
US46432F3881		X						X									
Growth																	
US46137V4986		X						X	US4642888857		X				X		
US4642873099		X						X	US78464A4094	X					X		
US4642876712	X					X			US9219325050	X					X		
Small-cap																	
US46138E1156			X				X		US4642882736	X				X			
US46138E1230	X						X		US4642891315	X						X	
US46138E1495	X						X		US46429B2007	X						X	
US46138E1560	X						X		US46429B4169	X						X	
US46138E1644	X				X				US46429B4656	X						X	
US46138E1727	X						X		US46429B6149	X				X			
US46138E1800	X						X		US68386C3025	X						X	
US46138G1022	X						X		US78463X7562	X				X			
US46138G2012	X						X		US78463X8719	X				X			
US4642864759	X				X				US78464A8137	X						X	
US4642865822	X				X				US9219328286	X						X	
US4642878049	X						X		US9220427184	X						X	
Mid-cap																	
US4642875078	X				X				US78467Y1073	X				X			

Exhibit 5 (*Continued*)

Model	A	C	D	E	AM	CM	DM	EM	Model	A	C	D	E	AM	CM	DM	EM
Momentum																	
US46138E2220				X				X	US46432F3964		X						X
US46138E2899				X				X	US78468R7052	X					X		

Model A: All variables except RMW are statistically significant – MOM not included. Model C: All variables except RMW and SMB are statistically significant – MOM not included. Model D: All variables except RMW and CMA are statistically significant – MOM not included. Model E: All variables except RMW, SMB, and CMA are statistically significant – MOM not included. Model AM, CM, DM, EM: Same as A, C, D, E – MOM included. R^2: mean: 0.62; standard deviation: 0.06; max: 0.74; min: 0.51. Statistical significance: p-value ≤ 0.10.

mean small. We discover that the opposite is also possible; a negative SMB does not necessarily mean large. Overall, as some of the ETFs exhibit exposure in factors other than their style indicates, this implies that they are not neutral with respect to the remaining factors. As a matter of fact, most of the ETFs are not momentum neutral. Consequently, investors are exposed to momentum (or the other factors not falling within their investment mandate) unknowingly.

4.4 Out-of-sample forecasting

This section of the present paper investigates the validity of the above estimates by studying the forecasting power of the reached results. In particular, it contrasts the forecasting accuracy of the estimates with those of a model that does not contain the variable of MOM. Forecasts are generated using an expanding estimation window. Models are initially estimated with respect to the in-sample period 2008–2014 and evaluated on the out-of-sample period 2015–2018. This breakdown was chosen so as to have anomalies in both periods and thus be able to assess the predicting capacity of our model also in declining markets. In the first period 2008–2014, we had the downturn of 2008–2009, whereas in the second period, that is, 2015–2018, we had the downturn of 2015–2016, that both took place in the first years of the corresponding periods. In the second period, we also had the corrections of February and October 2018. Moreover, once models are estimated on the in-sample period, after which the portfolio returns are forecasted for one year ahead. The model coefficients are then re-estimated in the expanded sample 2008–2015, and a new set of forecasts is generated. This process is repeated for all forecasting horizons (i.e., four years). A ratio below one implies a better performance. The forecast error is measured using both the root mean squared error and the mean absolute error.

Exhibit 6 reports the average forecast performance ratios for horizon $k = 1$, 2 3, and 4 years. The findings indicate that the root mean square error (RMSE) of the model with MOM ranges well below one across all ETFs, indicating a substantial forecasting performance. Relatively, the model with MOM growth forecasts is more accurate with respect to the model without this variable. The performance ratios based on the mean absolute error (MAE) are also well below one across all ETFs.

4.5 Results interpretation

The baseline findings are most likely explained by the fact that the ETFs were regressed against their own underlying indexes. Consequently, we

Exhibit 6 Comparison of out-of-sample forecast errors (four-year average).

Horizon *k*	Model with MOM							
	RMSE	MAE		RMSE	MAE		RMSE	MAE
US4642888774	0.023	0.015	US4642888857	0.025	0.017	US46137V2170	0.042	0.024
US4642882736	0.026	0.016	US4642865822	0.027	0.018	US78464A8137	0.046	0.028
US4642874089	0.030	0.019	US4642878791	0.031	0.021	US78467Y1073	0.046	0.029
US4642877058	0.033	0.021	US78464A5083	0.034	0.023	US46137V4564	0.046	0.031
US4642876639	0.028	0.018	US46137V2584	0.032	0.021	US46137V4648	0.042	0.029
US46137V2584	0.032	0.022	US78464A3005	0.034	0.024	US46137V4721	0.043	0.032
US78464A8392	0.036	0.025	US46137V1917	0.036	0.028	US4642878049	0.037	0.029
US46137V1677	0.041	0.026	US4642873099	0.037	0.030	US78463X8719	0.040	0.031
US78464A4094	0.040	0.025	US4642876068	0.036	0.030	US4642875078	0.038	0.031
US46137V2667	0.043	0.026	US4642876712	0.039	0.032	US46137V4804	0.037	0.032
US4642878874	0.040	0.024	US78464A2015	0.037	0.029	US46137V4986	0.036	0.030
US78464A8210	0.043	0.026	US46137V1750	0.039	0.031			
			Relative errors compared to the model without MOM					
US4642888774	1.011	1.032	US4642888857	0.8358	0.8588	US46137V2170	0.929	0.966
US4642882736	0.952	0.974	US4642865822	0.8623	0.8810	US78464A8137	0.965	1.016
US4642874089	0.945	0.967	US4642878791	0.8573	0.8755	US78467Y1073	0.976	1.018
US4642877058	0.880	0.905	US78464A5083	0.8763	0.8968	US46137V4564	0.973	1.011
US4642876639	0.872	0.888	US46137V2584	0.8720	0.8900	US46137V4648	0.956	0.977
US46137V2584	0.863	0.886	US78464A3005	0.8683	0.8830	US46137V4721	0.926	0.952
US78464A8392	0.872	0.900	US46137V1917	0.8655	0.8815	US4642878049	0.934	0.955
US46137V1677	0.877	0.908	US4642873099	0.8700	0.8935	US78463X8719	0.967	0.995
US78464A4094	0.966	0.975	US4642876068	0.9595	0.9975	US4642875078	0.962	0.985
US46137V2667	0.944	0.966	US4642876712	0.9458	0.9730	US46137V4804	0.965	0.996
US4642878874	0.917	0.943	US78464A2015	0.9190	0.9553	US46137V4986	0.933	0.976
US78464A8210	0.900	0.927	US46137V1750	0.9070	0.9243			

Note: The relative errors are the ratio of the average forecast error of the model with the MOM variable to the average forecast error of the model without the MOM variable. Figures less than one (a ratio of <1) denote that the former model has a better forecasting Performance than the latter one. Results are shown as the average for four forecast horizons.

can infer that they indeed mimic these indexes, as market excess return clearly prevails and the relevant coefficient, i.e., beta, is between 0.9 and 1, i.e., not far from 1. The (relatively small) deviation of the excess performance of the ETFs from the excess performance of their underlying index, as measured by their beta (compared with the market beta) is probably attributed to the following reasons: (a) A smaller than one beta implies that we may be missing (i) the dividend that the fund distributes, (ii) or the cash that it possibly contains, (iii) or the smaller number of shares in which it invests; thus, not fully incorporating the very small-cap stocks which are the stocks that are anticipated to deliver growth, as a result of which the growth factor is most likely partially missed. (b) ETFs may allow for security lending and they may apply their own replication strategy as they cannot always invest in all the stocks of the underlying index. (c) ETFs and their underlying smart beta indexes may use a different definition of the factors compared to the research ones of Fama–French. In addition, we cannot be certain that the ETF manager security selection is based fully on the Fama–French factors. (d) ETFs may apply their own idiosyncratic process; normally the manager does not intervene. (e) ETFs are subject to transaction costs and potential trading restrictions.

At the same time, we can observe that the intercept (alpha – α) is very close to zero, which could indicate either lack of active management—as anticipated in ETFs that follow and index, or active management that did not deliver the expected results. Consequently, the ETFs seem to follow to a great extent their underlying index and thus the market excess performance is the prevailing factor. The other factors have a smaller contribution—judged by the size of their coefficient—which subscribes to this point of view.

Attempting to explain why smart beta indexes (along with the replicating ETFs) differ from the Fama–French factors—as suggested by an anonymous reviewer—we realize that the Fama–French factors are theoretical constructs which capture potential underlying sources of risk and/or return. While the ETFs under examination might follow style indexes that use potentially different construction rules, they should still load on the relevant Fama–French factors in a way that is consistent with the style they meant to capture. This is the underlying theory behind popular risk attribution models, such as MSCI BARRA and AXIOMA. It would be very strange indeed for an index (and, therefore, the ETF that replicates that index) that is meant to capture growth to load positively on the HML factor that was specifically designed to capture the value premium.

Now, there are potential differences that could explain away some of the exposures we observe in the analysis. Recall that smart beta indexes (or strategies) use factor-based investing methods to identify risk factors that can be exploited. The most common factors are value, size, and momentum, which are consistent with our findings. One potential explanation is associated with the criteria ETFs used to identify mispriced (cheap) equities even from the index. The literature seems to prefer the price-to-book (or book-to-price) ratio, especially in terms of the high-minus-low (HML) factor. However, in practice, there may be two potential deviations from this approach. The first pertains to the use of other metrics, such the price-to-earnings and price-to-cash flow ratios, the dividend yield, and the earnings yield. The second is relevant to the method used to calculate these metrics. Practitioners challenge the standard methods used to calculate them at portfolio formation, such as the one used for the book-to-price ratio (Asness and Frazzini, 2013). In this case, they recommend replacing the standard approach that uses lagged book data aligned with price data keeping these values unchanged until rebalancing, with alternatives that use more punctual price data and maintaining the necessary lags for book data. In the framework of a five-factor model including momentum, these alternative approaches seem to deliver superior returns. Consequently, if ETFs are formed in practice with such deviations from the research factors, then a difference in the coefficients may exist.

A second cause at which the difference may be attributed is the sector neutrality of the ETF. The ETFs may be formed to be sector neutral, which is not the case for the portfolios constructed in the theory by Fama and French, implying that in practice the cheapest stocks are chosen per sector and not for the entire universe of the index. This may also deliver returns that are different from the ones provisioned by the academic work and may lead to the ETF factors being different from the research ones.

Finally, throughout the analysis, we do not account for the presence of transaction costs. That should be a far smaller issue though, because first we try to understand the risk exposures to the underlying factors and gross returns should retain that information. Secondly, style indexes, and therefore the ETFs, are not rebalanced frequently (momentum indexes, might be the exception here), as the half-life of the underlying factors they try to capture is slow-moving.

The analysis attempts to address the following questions (using a value index as an example): (1) is the index loading with the right sign on the premium it has been designed to capture?, (2) is the average return of the ETF associated with the underlying index, consistent with the performance of the factor during the period of the study?, and (3) is the marginal contribution to risk associated with the relevant factor the dominant source of risk (other than the market)? For example, we would expect a value ETF to load positively to the HML factor. Given that HML and momentum are negatively correlated, we would also expect a negative loading to the momentum factor (one could try to mitigate the effect of the negative correlation during the portfolio construction process, but we leave this exercise as a separate line of research). Unless the value index the relevant ETF is tracking is explicitly stating that is a small-cap index, we would not expect the loading to SMB to be statistically significant. Moreover, continuing with the value ETF, from a risk decomposition perspective, we would expect that the value factor exposure to result in the largest marginal contribution to risks, other than the market, given that the ETFs are only long portfolios. Any significant economic and statistical deviation from these priors should provide evidence that the underlying style indexes are not constructed in a way that appropriately capture the sources of return they were meant to.

In order to measure the impact of each factor to the return and volatility of the ETF, we calculate the (percentage) incremental returns of each factor to the returns attributed to the market portfolio, the (percentage) contribution of each factor to the volatility of the ETF, and the incremental returns to the variance contribution. The descriptive statistics are displayed in Table 7. When the momentum is not taken into account, the incremental returns on top of the market returns are primarily determined by the SMB, CMA, and HML factors, in that order (in absolute value), across all, but two ETFs. As a matter of fact, the incremental returns attributed to SMB and HML are in the opposite direction with those of the market, whereas the incremental returns attributed to CMA are in the same direction. The order is not affected by the style of the ETF, i.e., value-growth, large–mid–small, or their combinations. A potential explanation is that even in the formation of value-growth funds, the size prevails, with growth stocks being potentially small-caps and value stocks being potentially large-caps (with mid-caps serving both directions). Looking at the volatility metric when momentum is not among the factors of the modeling approach, we realize that CMA has the largest contribution to the volatility of the ETFs in absolute values across all, but one, ETFs. SMB, and HML follow in that order. Furthermore, when we look at the incremental returns, adjusted for the contribution to variance, we realize that SMB systematically comes first and HML second (across all, but one ETFs) and CMA third.

Combining the observation that the incremental returns of SMB and HML are in the opposite direction with the realization that the average returns (both the actual and the model-derived)

Exhibit 7 Incremental return and variance contribution descriptive statistics (models without MOM and with MOM).

	MKT	SMB	HML	RMW	CMA	MKT	SMB	HML	RMW	CMA	MOM
						Percentage incremental return					
Min	99.4%	−56.8%	−24.4%	−1.4%	0.8%	−28020.0%	−3577.7%	−2036.2%	−0.3%	−3923.3%	−10321.3%
Max	154.6%	−0.7%	−0.6%	0.0%	25.0%	13758.5%	9636.7%	2900.0%	0.0%	2276.6%	19506.7%
Mean	132.5%	−35.1%	−15.6%	−0.4%	18.9%	483.2%	−43.7%	−103.0%	0.0%	81.0%	−317.6%
SD	11.4%	10.6%	5.7%	0.5%	5.8%	6231.4%	1957.2%	756.5%	0.1%	899.8%	4445.4%
						Percentage variance contribution					
Min	−1.9%	0.5%	−1.7%	−2.5%	6.1%	−2.2%	0.1%	0.1%	−0.1%	−3.1%	70.6%
Max	95.1%	37.9%	22.6%	0.0%	69.9%	2.6%	9.2%	4.2%	0.0%	19.7%	100.8%
Mean	1.9%	25.7%	15.7%	−0.2%	57.0%	0.3%	3.6%	1.9%	0.0%	5.7%	88.5%
SD	16.8%	6.7%	5.2%	0.4%	10.8%	1.8%	2.0%	1.3%	0.0%	8.5%	9.7%
						Incremental return/variance contribution					
Min	−79.7	1.3	−4.2	−3.0	−1.4	−39.5	2.2	1.5	−6.1	−3107.7	−0.5
Max	150.6	15.0	1.8	−1.9	−0.3	38.9	12.3	91.4	−6.1	31.0	0.5
Mean	71.4	1.9	0.9	−2.1	−0.4	−2.6	5.6	8.8	−6.1	−91.1	0.1
SD	39.9	2.3	0.9	0.3	0.2	35.4	3.3	15.8	0.0	533.2	0.5

are negative in the period under consideration, we may conclude that factor investing—as provisioned by the Fama–French model—would have improved the returns of the ETFs. CMA is in the same direction, potentially influenced by the fact that the re-investment (or dividend) policy of a firm may be affected by its earnings, which are reflected globally by market returns. Consequently, years of negative market performance could indicate reduced earnings, thus, affecting the investment strategy of the listed firms. This difference in the attribution of the incremental returns, i.e., the predominance of SMB, may have been observed at the formation (or rebalancing) of the ETFs and thus the smart beta factors selected are different from the research ones. In addition, the remark that it does not have the largest contribution to the variance among the significant factors, but it exhibits the highest incremental returns against its contribution to the variance, subscribes to this point of view. The same explanation may be offered for the case of the HML factors that comes second.

When momentum is considered, it takes the lead as far as the incremental returns are concerned (in absolute values), with SMB being second and HML and CMA following, with their positions interchanging, depending on the type of ETF. The incremental returns due to momentum are in the same direction with the market returns for the ETFs that have a positive coefficient and are in the opposite direction for the ETFs that have a negative coefficient. As a result, in the period under examination, during which average negative returns were posted, the ETFs with a negative momentum coefficient managed to have positive returns attributed to the momentum factor. The opposite held true for the ETFs that had a positive momentum coefficient. As none of these funds was a momentum ETF, it seems that they offer exposure to momentum as a side effect of the security selection. Turning now to volatility, we realize that when momentum is one of the factors under investigation in our model, it has the highest contribution to the volatility of the ETF, which is true across all ETFs. The CMA factor follows (in absolute values and across all, but one, ETFs) and always exhibits the opposite contribution from that of the market (across all, but one, ETFs). SMB comes next (in absolute values and across all, but five, ETFs) before HML and the market. HML exceeds the contribution of the market (in absolute values) when the latter is negative (across all, but three, ETFs) and lags when the contribution of the market is positive. The incremental returns adjusted for the contribution to the variance confirm the lead of the SMB factor across all, but six, ETFs at which another factor seems to come first. HML follows across all, but eight, ETFs.

There are two questions that need to be explained with regards to the momentum effect. The first is its presence and the second is its difference from the theoretical factors. The presence of momentum may be explained by the fact that ETFs select stocks that have persistently better performance among the value or size stocks they choose at the formation of the portfolio. It could be also explained by the fact that momentum is a factor that explains the returns of stocks and the ETFs are not neutralized for momentum. As such, the impact of momentum tends to increase as historically the factor is influenced by market crashes. Its difference from the corresponding research factor may be attributed to the fact that the ETFs may use different periods compared to the period that the Fama–French approach uses, i.e., 12–1 months, the impact of crashes, as well as investibility constraints (MSCI, 2017). Due to its contribution to volatility momentum, most likely may not be missed. The incremental returns attributed to the other three factors could be explained through the arguments presented above, when momentum was not one of the factors of the modeling approach. Consequently, when ETFs that follow smart beta strategies are formed, there may be a tilt towards SMB and HML stocks, no matter what the orientation of the ETF is as a result of the incremental returns they offer either in absolute terms, or relative to their contribution to variance. Momentum may be present due to its high incremental returns, as well as its high volatility, considering that the ETFs may not be neutralized for its presence.

5 Future Research Venues

We do realize that this paper has certain limitations that primarily stem from the fact that it uses the Fama–French factors as have been constructed to include all NYSE, AMEX, and NASDAQ stocks as per the Kenneth French Data Library (French, 2019), whereas the ETFs used follow sub-indexes that have been constructed to somehow follow the Fama–French factors. We plan in the future to attempt to construct our own portfolios so as to better reflect the particulars of the ETFs we used in our research. We also aim at looking deeper in the particulars of the ETFs so as to understand how to capture the deviation of the excess performance of the ETFs from the excess performance of their underlying index or the common market index, as explained by the reasons mentioned in the Results Interpretation subsection of the manuscript. As a next step we observe that ESG (Environmental, Social & Governance) investing is gaining grounds among the asset managers, as well as asset owners. We intent to study the effect of the ESG criteria and potentially mimic the creation of the Fama–French factors in creating a scale—metric of ESG investing so as to compare firms or ETFs meeting High ESG standards versus the ones succeeding Low ESG standards.

6 Conclusions

The analysis explored the performance of a series of ETFs against their underlying index so as to verify whether they implement their announced strategy. It seems that globally they did indeed follow their underlying index. This paper employed the five- and six-factor Fama–French models to find that the excess return of the market is the prevailing factor, with the other ones following. Momentum seemed to contribute the most compared to the other factors and RMW the least, even being insignificant. We also expanded the analysis to explain the excess performance of the ETFs compared with a common index, namely the S&P 500 Total Return Index. We found once and again that the excess return of the market is the prevailing factor, with a smaller coefficient though. Among the remaining factors, HML, CMA, and MOM contribute the most, whereas SMB and RMW have almost no explanatory power. Even more, MOM seems to improve in all cases the forecasting ability of the model. The findings of this paper can be used by investors in assessing whether the ETFs indeed follow the factor-investing approach they have opted to adopt, as well as in predicting their future performance.

Acknowledgments

Many thanks go to Dr. George Sakoulis of GMO for his valuable contribution and advice for making this a better paper and increase its interest from a practitioner's point of view.

The authors wish to thank Dr. Charalampos Agiropoulos for his kind advice in reshaping the tables that depict the results.

We are thankful to the anonymous reviewer for the recommendations and directions that helped us to improve our manuscript and address the difference between the Fama–French factors and smart beta indexes.

References

Agapova, A. (2011). "Conventional Mutual Index Funds versus Exchange-Traded Funds," *Journal of Financial Markets* **14**(2), 323–343.

Asness, C. and Frazzini, A. (2013). "The Devil in HML's Details," *The Journal of Portfolio Management* **39**(4), 1–21.

Boyadzhiev, D., Bryan, A., Dutt, M., Johnson, B., and McCullough, A. (2018). "A Framework for Analyzing Multifactor Funds," *Morningstar Manager Research* https://www.morningstar.com/content/dam/marketing/shared/research/methodology/869053-FrameworkAnalyzingMultifactorFunds.pdf. Accessed: 25.3.2019.

Carhart, M. M. (1997). "On the Persistence of Mutual Fund Performance," *The Journal of Finance* **52**(1), 57–82.

Cakici, N., Fabozzi, F. J., and Tan, S. (2013). "Size, Value, and Momentum in Emerging Marketstock Returns," *Emerging Markets Review* **16**, 46–65.

Chen, H-I. and Bassett, G. (2014). "What Does $\beta_{\mathrm{SMB}} > 0$ Really Mean?" *The Journal of Financial Research* **XXXVII**(4), 543–551.

Elton, E. J., Gruber, M. J., Comer, G., and Li, K. (2002). "Spiders: Where are the Bugs?" *Journal of Business* **75**(3), 453–473.

Elton, E. J., Gruber, M. J., and de Souza, A. (2019). "Passive Mutual Funds and ETFs: Performance and Comparison," *Journal of Banking & Finance* **106**, 265–275.

Fama, E. F. and French, K. R. (1992). "The Cross-section of Expected Stock Returns," *The Journal of Finance* **XLVII**(2), 427–465.

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

Fama, E. F. and French, K. R. (2012). "Size, Value, and Momentum in International Stock Returns," *Journal of Financial Economics* **105**, 457–472.

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

Fama, E. F. and French, K. R. (2017). "International Tests of a Five-Factor Asset Pricing Model," *Journal of Financial Economics* **123**, 441–463.

French, K. R. (2019). "Data Library. Tuck School of Business at Dartmouth," http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html. Accessed: 19.1.2019.

Garyn-Tal, S. (2014). "An Investment Strategy in Active ETFs," *The Journal of Index Investing* **4**(1), 12–22.

Gaunt, C. (2004). "Size and Book to Market Effects and the Fama French Three Factor Asset Pricing Model: Evidence from the Australian Stockmarket," *Accounting and Finance* **44**(1), 27–44.

Grim D. M., Pappas, S. N., Tolani, R. G. and Kassidis S. (2017). "Equity Factor-Based Investing: A practitioner's

Guide," *Vanguard Research* https://americas.vanguard.com/docs/literature/equity-factors-guide.pdf. Accessed: 25.3.2019.

Guedj, I. and Huang, J. (2009). "Are ETFs Replacing Index Mutual Funds?" Working Paper.

Hou, K., Xue. C. and Zhang, L. (2014). "Digesting Anomalies: An Investment Approach," *The Review of Financial Studies* **28**(3), 650–705.

Huij, J. and Verbeek, M. (2009). "On the Use of Multifactor Models to Evaluate Mutual Fund Performance," *Financial Management* **38**(1), 75–102.

Karp, A. and van Vuuren, G. (2017). "The Capital Asset Pricing Model and Fama-French Three Factor Model in an Emerging Market Environment," *International Business & Economics Research Journal* **16**(3), 231–256.

Kostovetsky, L. (2003). "Index Mutual Funds and Exchange-Traded Funds," *Journal of Portfolio Management* **29**(4), 80–93.

Lilloe-Olsen, E. (2016). "Factor Effects in Nordic Equities—A Theoretical and Practical Study. Master's thesis. M.Sc. Economics and Business Administration," *Copenhagen Business School* https://studenttheses.cbs.dk/bitstream/handle/10417/6213/einar_lilloe_olsen.pdf?sequence=1. Accessed: 19.1.2019.

MSCI (2017). "Factor Focus: Momentum," *Factor Investing* 1-8. msci.com. Accessed 4.10.2020.

Panopoulou, E. and Plastira S. (2014). "Fama French Factors and US Stock Return Predictability," *Journal of Asset Management* **15**(2), 110–128.

Robotis, G. G (2011). "The Performance of Actively Managed Exchange-Traded Funds," *The Journal of Index Investing* **1**(4), 53–65.

Sattar, M. (2017). "CAPM Vs Fama-French Three-Factor Model: An Evaluation of Effectiveness in Explaining Excess Return in Dhaka Stock Exchange," *International Journal of Business and Management* **12**(5), 119–129.

Schizas, P. (2014), "Active ETFs and Their Performance vis-a-vis Passive ETFs, Mutual Funds and Hedge Funds," *The Journal of Wealth Management* **17**(3), 84–98.

Tse, Y. (2015). "Momentum Strategies with Stock Index Exchange-Traded Funds," *The North American Journal of Economics and Finance* **33**, 134–148.

Keywords: Exchange-traded funds; Fama–French factors; factor-based investing; portfolio return

JEL Classification: G11, G12, G15