Smart Beta Made Smart: Synthetic Risk Factors for Institutional and Retail Investors*

Andreas Johansson†
Stockholm School of Economics

Riccardo Sabbatucci‡
Stockholm School of Economics

Andrea Tamoni§
Rutgers Business School

ABSTRACT

We construct synthetic, tradable risk factors using optimal combinations of large and liquid mutual funds and ETFs. We find that investors are not able to harvest the unconditional factor risk premia, although the synthetic portfolios of institutional investors outperform those of retail investors. We also propose a methodology to identify market funds. Lastly, we show that (i) daily flows to naive smart beta strategies are more predictable than those to our synthetic strategies, and (ii) our synthetic HML outperforms a naive one based on fund names. Our results have implications for the evaluations of portfolio managers and cross-sectional return anomalies.

This version: March 7, 2021

Keywords: Smart beta, factor investing, tradable risk premia, daily flows to smart beta strategies

*We are grateful to Adam Farago (discussant), Kris Jacobs, Juhani Linnainmaa, Dong Lou, Per Strömberg, Allan Timmermann, Sheridan Titman for many valuable suggestions on our paper. We also received helpful comments from conference and seminar participants at the NFN Young Scholar conference, Stockholm School of Economics, University of Houston, and Warwick Business School. The paper received the ICPM Award Honourable Mention 2020.

†Ph.D. Student, Department of Finance, Stockholm School of Economics. **Email:** [redacted]

‡Department of Finance, Stockholm School of Economics and Swedish House of Finance (SHoF). **Email:** [redacted]

§Department of Finance, Rutgers Business School. **Email:** [redacted]

1 Introduction

Smart beta, or factor investing, has become a fundamental theme in the asset management industry over the last decade. Many providers have started offering mutual funds and exchange-traded funds (ETFs) whose objective is to track the strategies underlying the risk-factors discovered in academic research, such as value (Fama and French, 1992, 2006) or momentum (Jegadeesh and Titman, 1993).

The top panel of Figure 1 shows a clear trend in assets managed by smart beta funds, while the bottom panel highlights that both institutional and retail investors' demand for such assets has been steadily increasing over time. One explanation for this trend is that many institutional investors increase their exposure to smart beta assets to reconcile their need to control costs - most of these funds passively track factors and hence charge relatively low management fees - with the necessity to boost returns in a low interest rates environment, by earning the (unconditional) risk premia of the various factors reported in academic studies.[1] Indeed, over the last thirty years, academic research has uncovered the factor structure present in equity returns and highlighted a strong link between exposures to characteristics like book-to-market and risk premia.

However, how well investors can track the academic long-short risk factors remains an open question. Most importantly, exploring how and to what extent different types of investors can optimally replicate these factors in practice using portfolios of smart beta funds has not been studied. In fact, while the term "factor risk premia" refers to the excess returns obtained from spread (long-short) portfolios sorted on some underlying characteristic(s), (i) mutual funds and ETFs usually track only one "leg" of these factors (e.g., the value or growth legs of the HML factor), and (ii) investors cannot short-sell mutual funds.[2]

This evidence highlights several fundamental, and yet unanswered, research questions for both academics and practitioners: what is the "true"(e.g., tradable) set of factor strategies available to institutional and retail investors? Do institutional and retail investors differ in the way they can implement tradable risk strategies? Do these different implementations

[1]For example, the value premium (HML) has been around 3% per year over the period from July 1927 to October 2019.

[2]Institutional and retail investors cannot short-sell mutual funds, but they can take short positions in large, liquid ETFs. For our paper, it is irrelevant whether mutual funds or ETFs can engage in short-selling activities, since the constraint is on the mutual fund/ETF investors, not on the fund/ETF.

result in heterogeneous performance? What is the relation between the performance of these tradable strategies and investors' flows in the smart beta funds? Should investors evaluate fund managers against their true opportunity cost of investment? Do the anomalies discovered in the academic literature survive if benchmarked against tradable risk factors?

To provide insight on some of these questions, we propose a formal framework to construct tradable long-short risk factors. We first use a *minimum distance* statistical methodology to identify those funds that, regardless of their names and claimed strategy, are simply market funds. Second, we construct *tradable* factors after classifying the remaining, non-market funds into smart beta strategies based on their holdings. Specifically, we rely on a combination of smart beta mutual funds and ETFs to synthetically replicate the individual legs of the (non-tradable) risk factors based on their holdings. We then construct synthetic, long-short factors (e.g., HML) that are tradable by *both* retail and institutional investors by value-weighting the funds most exposed to the respective underlying factor characteristic. Our methodology is general and can be used to replicate *any* non-tradable smart beta index/benchmark with publicly disclosed holdings.[3]

We document several novel facts. First, we find that around one-half of the total existing funds cannot be distinguished from the market, both in real-time and over the full sample. Second, both retail and institutional investors are unable to harvest the various factor risk premia by investing in the synthetic, tradable proxy of the factors, with the exception of the size premium (SMB). In particular, the replication of momentum (MOM) and profitability (RMW) proves particularly challenging. Third, institutional investors are able to synthetically replicate the factors better than retail investors, as indicated by the lower alphas from regressions of the benchmarks on our synthetic portfolios, and by higher time series fit (R^2). This is particularly evident for the value (HML) and size (SMB) risk premia. Related to this point, we show that (i) the characteristic score of the synthetic value leg (i.e., the long leg of HML) of institutional investors is closer to that of the Fama-French benchmark relative to the one of retail investors (in line with institutions being better than retails in replicating HML); and (ii) the retail and institutional implementations of value and growth often differ in terms of the number of ETFs used in the synthetic portfolios, despite a non-trivial

[3]Some mutual funds are marketed as "active" (e.g., display an active tilt), but this has no impact on our methodology since we rely only on funds' holdings to construct synthetic, optimal portfolios with characteristic scores that are as close as possible to the non-tradable factors' ones. In other words, what matters is how close – in terms of holdings – funds are to the non-tradable factors, regardless of their type.

overlap. Fourth, the replication of SMB and its legs improves substantially by just using ETFs in the synthetic portfolio. The fact that some factors like SMB can be replicated more precisely using a smaller universe of assets (ETFs only vs. ETFs and mutual funds) might be related to the different benchmarking and incentives (i.e., greater leeway) of mutual fund managers. Fifth, given our focus on the tradability of the factors, we provide estimates on the capacity of smart beta strategies. A back-of-the-envelope calculation shows that most factors, with the exception of momentum, have more than $1 trillion available capacity, in line with the results in Ratcliffe et al. (2017). Lastly, we study the daily flows into synthetic and naive (e.g., based on fund names) smart beta strategies, and find that flows to the latter are more predictable than flows to sophisticated funds that optimally track the underlying characteristic of a risk factor but do not necessarily explicitly mention that characteristic in their names. We conclude that investors seem to allocate money into smart beta strategies based on fund names rather than their true factor exposure.

Overall, our results are consistent with a form of market incompleteness, meaning that investors cannot get a large exposure to the long leg of HML, or the short legs of MOM and RMW, when trading ETFs and mutual funds, as evidenced by characteristic scores of the synthetic factor legs that are substantially different from those implied by the Fama-French factors.

Our result that the space of tradable risk factors is quite different from that typically described by multi-factor asset pricing models has implications for evaluating fund managers' skill (i.e., α). Indeed, portfolio managers should be evaluated against the true opportunity cost of factor investing of institutional and retail investors. In other words, their alpha should be estimated with respect to what institutional and retail investors could have achieved by trading in a publicly available, optimal combination of smart beta funds.[4]

To summarize, the key contribution of this paper is to propose a methodology to construct tradable risk factors using large, liquid investment funds, and to show how different types of investors can trade them. We argue that these tradable risk factors should be used as benchmarks (e.g., to evaluate mutual funds), since they truly represent the opportunity cost of factor investing for retail and institutional investors.

[4]This is especially true in light of the several spurious anomalies discovered in the literature (e.g., Harvey et al., 2016). To this point, Hunter et al. (2014) augment the standard Carhart four-factor model with an active peer benchmark "tradable" factor, while Berk and van Binsbergen (2015) use comparable Vanguard index funds to evaluate the value added by portfolio managers.

Our paper proceeds as follows. In Section 2, we review the literature. In Section 3 we describe the data, while in Section 4 we present some descriptive evidence. In Section 5 we present our framework to categorize funds. In Section 6 we report the main results, while in Section 7 we compare our results to those obtained from naive smart beta strategies. Concluding remarks are in Section 8. Internet Appendix A reports several robustness tests.

2 Literature Review

Factor investing has become one of the most important topics in asset management over the last decade. Our paper contributes to several streams of research related to it.

Anomalies and Trading Frictions. Risk factors have been documented in various asset classes (e.g., Asness et al. (2013), Koijen et al. (2018)), although most smart beta funds tend to track equity factors, which is the focus of our paper. Momentum, value, size, or quality are just few examples of factor strategies available to investors through smart beta funds. Interestingly, a growing literature on cross-sectional predictability and data mining has pointed out several pitfalls of factor investing. Hou et al. (2019) considers almost 450 academically-reported anomalies in equity markets and finds that most of them (64%) fail to hold up. McLean and Pontiff (2016) and Linnainmaa and Roberts (2018) show that few anomalies persist out-of-sample. Harvey et al. (2016) examine 315 (macro and tradeable) factors, and find that most of these fail to survive statistical procedures that adjust for multiple testing. However, even in the multiple testing framework of Harvey et al. (2016), size, value, momentum, and volatility are found to be significant. Novy-Marx and Velikov (2016) document that strategies based on past returns like momentum tend to have high turnover and may thus be relatively expensive to trade. However, using proprietary data, Frazzini et al. (2015) find that the actual trading costs faced by a large institutional trader are an order of magnitude smaller than those estimated for the average trader, and state that "this is because a large institutional trader [...] often trades within the spread, using limit orders and tries to supply rather than demand liquidity" (p. 20). Consistent with our results, Patton and Weller (2020) find that real-world implementation of the momentum and value factors strongly underperform the relative on-paper strategies due to all-in implementation costs, while Arnott et al. (2017) find that some of the theoretical long-short factor returns are

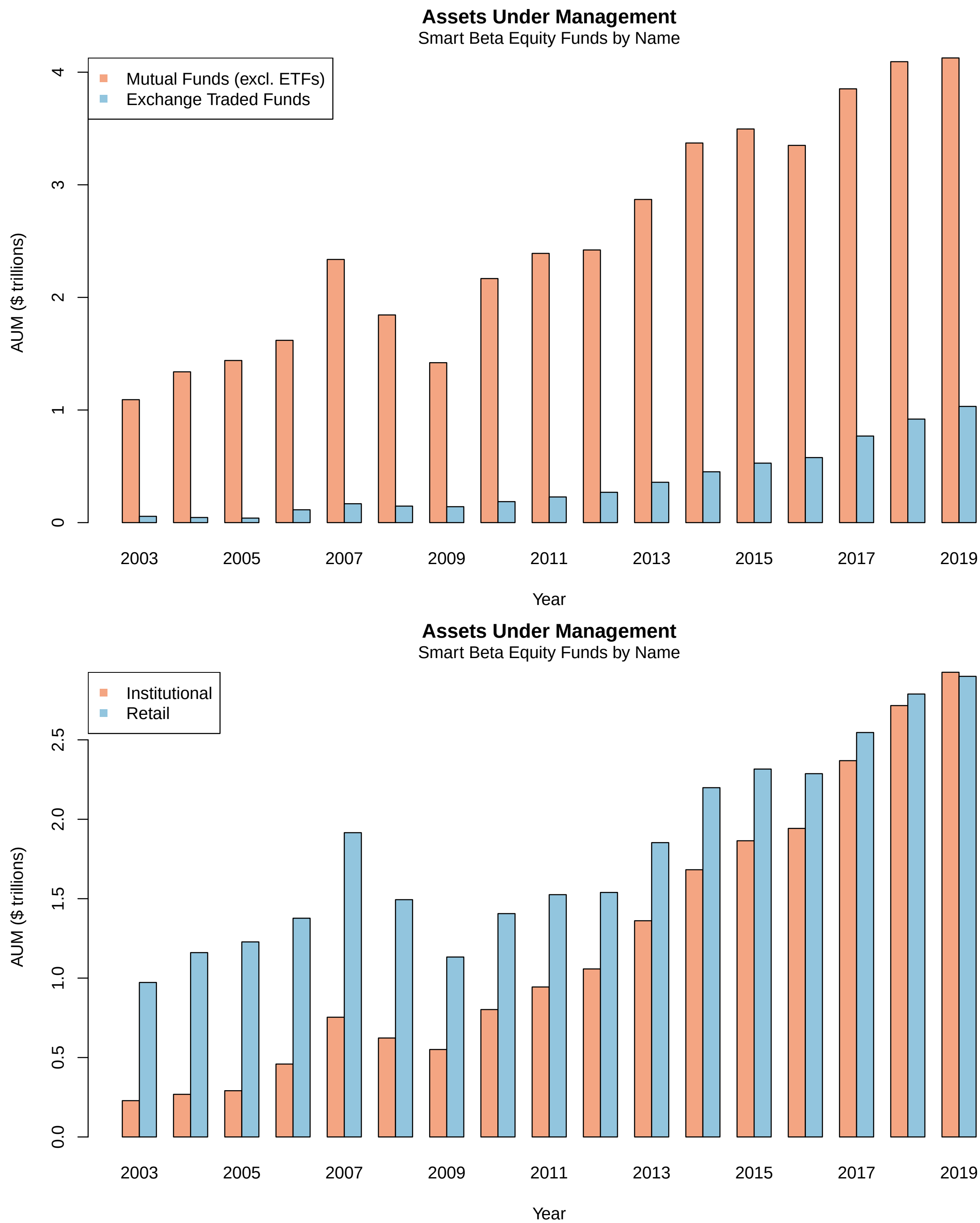


Figure 1: Smart Beta Equity Mutual Funds and ETFs. The top panel plots the assets under management of smart beta funds starting in 2003. The bottom panel plots the assets under management of mutual funds and ETFs available to institutional and retail investors. Funds are labeled as smart beta if their name contains words related to factor investing (e.g., momentum, value). Every year, the sum of the bars in the bottom panel is larger than the corresponding ones in the top panel because ETFs are available to both retail and institutional investors. See Appendix B.3 for a detailed description of the classification. The sample includes funds with assets under management greater than $50 millions.

hard to replicate in practice, mainly due to trading costs and expenses of shorting.[5] Mindful of the studies above, we investigate whether using highly liquid, cost-effective instruments available to both retail and institutional investors, it is possible to replicate robust factors that have survived out-of-sample tests and that are often included in state-of-the-art multi-factor models like the five-factor Fama and French (2015) model or the q-factor of Hou et al. (2015).[6]

Fund Performance and Closet Indexing. Our paper is also related to the literature on fund managers' performance with respect to (non-tradable) risk factors (i.e., the α often interpreted as the fund manager's skill).[7] The literature on fund managers' skill tends to benchmark fund performance on factor models that often cannot be traded, complicating the interpretation of the results.

Gerakos et al. (2020) provide evidence that institutional investors would earn higher returns by delegating their capital to asset managers rather than implementing mean-variance efficient portfolios using index funds and mutual funds available to them. Our research question is different: we do not evaluate the performance of institutional asset managers, but rather construct equity smart beta long-short benchmarks that are tradable by both retail and institutional investors given the set of assets that is publicly available (i.e., without the need of using intermediaries like consultants) and taking into account the capacity constraints of these strategies (e.g., Ratcliffe et al., 2017). This explains our focus on U.S. equities, the main asset class underlying smart beta strategies, which is also where the alpha generated by institutional asset managers seems limited (see Gerakos et al., 2020).

Closet indexing is another widely studied, closely related topic (e.g., Cremers et al., 2016).

[5] Both papers use the classical two-pass Fama and MacBeth (1973) procedure; we instead construct tradable factor portfolios using mutual funds and ETFs holdings.

[6] Evidence against replicability of these factors using mutual funds and ETFs can be interpreted in one of two ways. On the one hand, capital markets may be more efficient than previously recognized, so that effectively the CAPM holds and the market is the only factor. On the other hand, there may be frictions leading to a misalignment of incentives between managers and investors (e.g., Ma et al. (2019) find that portfolio manager compensation contracts are designed to mitigate agency conflicts); these frictions prevent investors from extracting the risk premia, even if these factors are not spurious.

[7] Several papers (Pastor and Stambaugh, 2002; Kacperczyk et al., 2005; Kosowski et al., 2006; Kacperczyk and Seru, 2007; Cremers and Petajisto, 2009; Koijen, 2014) document that (some) mutual fund managers have skill. With fund-level or industry-level decreasing returns to scale, Berk and van Binsbergen (2015) and Pastor et al. (2015) show that skill does not equate to average performance. Kacperczyk et al. (2014, 2016) study the stock picking and market timing abilities of mutual fund manages and Pastor et al. (2017) investigate the time series relation between fund performance and turnover.

Huang et al. (2020) document a sharp performance deterioration of smart beta indices after the corresponding ETFs are listed. Their evidence is consistent with the use of data mining when constructing smart beta benchmarks to attract flows into the related ETFs.

Flows, Returns, Names, and Styles. There is a vast literature on the relationship between fund flows and returns, with the majority of the studies looking at mutual funds (Warther, 1995; Sirri and Tufano, 1998; Coval and Stafford, 2007; Lou, 2012; Ben-Rephael et al., 2012). More recently, Brown et al. (2019) document that ETF flows signal non-fundamental demand shocks, which in turn leads to return predictability. Ben-David et al. (2018) focus on ETFs that track equity indices and document that ETFs lead to an increase in the non-fundamental volatility of the securities in their baskets.

Gruber (1996) and Zheng (1999) show that the short-term performance of funds that experience inflows is significantly better than those that experience outflows, suggesting that mutual fund investors have selection ability, a fact known as the "smart money" effect. However, Frazzini and Lamont (2008) find that this smart money effect is confined to short horizons of about one quarter, and at longer horizons mutual fund investors are "dumb" in the sense that their reallocations reduce their wealth on average. In particular, they find that money flows into mutual funds that own growth stocks, and flows out of mutual funds that own value stocks. Teo and Woo (2004) also find evidence for a dumb money effect at the style level.

A few studies also find that funds' names do not reflect their holdings. Lakonishok et al. (1992) show that the market betas of "value" funds are close to one and their returns are not correlated with the value portfolio. Cooper et al. (2005) find that flows to funds increase dramatically when funds change their names toward the current "hot" style or away from the current "cold" style. This relation holds even when the name change is cosmetic, in the sense that the fund's investing style, as reflected by the fund's new name, does not reflect its portfolio holdings. Similarly, Chen et al. (2020) provide evidence that bond fund managers misclassify their holdings to influence investor capital flows. Similarly to these studies, we also find that investors' flows into smart beta strategies tend to follow the fund names rather than the actual fund holdings.

Rakowski and Wang (2009) analyze the daily flows of individual mutual funds to study the behavior of fund investors, and find that it is more consistent with contrarian rather than momentum characteristics. They also highlight how the dynamics of daily and monthly mutual fund flows differ substantially. Finally, Lettau et al. (2019) provide a comprehensive analysis of the portfolios of active mutual funds, ETFs, and hedge funds through the lens of risk (anomaly) factors. They show that these funds do not systematically tilt their portfolios towards profitable factors, such as high book-to-market (BM) ratios, high momentum, small size, high profitability, and low investment growth. Similarly to Lettau et al. (2019), our paper focuses on the funds' exposures to the various risk factors by looking at their holdings.

3 Data

We use daily, monthly, and quarterly data on the universe of U.S. equity mutual funds, ETFs and individual U.S. stocks from January 2003 up to December 2019. We merge data from several sources to obtain our final dataset. Fund returns, assets under management (AUM), and quarterly fund holdings are from the CRSP Mutual Fund dataset, which also includes data on ETFs. We use the CRSP flag to identify institutional/retail mutual funds, as in Etula et al. (2019) and Cooper et al. (2020). On the other hand, ETFs can be traded by both institutional and retail investors. Individual stock characteristics are from the Compustat (quarterly) dataset. We obtain *daily* flows of mutual funds and ETFs from the EPFR dataset, the most comprehensive dataset on daily fund flows to date.

In addition to fund and individual stock characteristics, we require data on the most important risk factors tracked by smart beta funds, namely the aggregate market, value, size, profitability, and momentum factors and their individual legs.

We proxy these risk factors with the corresponding (non-tradable) Fama-French risk factors: the market factor (MKT-Rf), the value factor (HML), the size factor (SMB), the profitability factor (RMW), and the momentum factor (MOM). In the robustness section (see Appendix A.2), we also replicate the CMA factor of Fama and French (2015), the ROE and I/A factors of Hou et al. (2015), and the quality-minus-junk (QMJ) factor of Asness et al. (2019) as additional "benchmark" factors. We choose these benchmarks because these risk premia form the basis, among other things, of the literature on mutual fund performance evaluation and asset pricing anomalies.[8]

[8]We also considered including additional risk factors, such as the idiosyncratic volatility factor (IVOL) of Ang et al. (2006), and the betting-against-beta factor (BAB) of Frazzini and Pedersen (2014). However, fund names and prospectuses indicate that very few funds, if any, are marketing themselves as tracking these factors. Thus, we exclude these factors from the analysis.

Mutual fund returns are calculated by CRSP as the change in net asset value (NAV), including reinvested dividends from one period to the next. Mutual fund returns are net of all fees except for front and rear loads,[9] since our objective is to construct synthetic, tradable factors that can be thought of as the true opportunity cost of factor investing for retail and institutional investors. In other words, the returns of our synthetic tradable factors should be what investors actually earn (before taxes).[10]

To be consistent with the assets used to construct the non-tradable risk factors (e.g., stocks of U.S. companies), we restrict our main sample to mutual funds and ETFs whose mandate is to invest in U.S. equities. We also require each fund to have at least twelve months of returns, AUM greater than \$1 billion (in real terms as of 2018) and 50% of the observations in the year before inclusion above this cut-off.[11] Our choice of the \$1 billion cutoff is consistent with our goal of creating risk factors tradable by both institutional and retail investors. Many investors might not feel confident trading small mutual funds or illiquid ETFs.[12] Moreover, small funds are more likely to invest in illiquid securities given their size, hence their strategies might not be as scalable as those of larger funds.[13] Overall, we feel confident that by looking at funds with AUM above \$1 billion that we are ensuring the tradability of the underlying funds. In Internet Appendix B, we describe all the standard filters used to obtain the final dataset.[14] It is important to note that no funds of funds are left in our final dataset, as they would not be easily tradable.

[9] These are often zero for equity funds when investing large amounts, which is the case in our study.

[10] We do not explicitly account for the cost of trading the funds, which is nowadays very low. While trading costs are extremely important when trading several hundred individual stocks (e.g., Novy-Marx and Velikov (2016); Patton and Weller (2020)), this is not an issue for us, since our synthetic, optimal portfolios only consist of a few funds, which display a very low monthly turnover, with the sole exception of momentum (see Table 5).

[11] The latter requirement is only needed to avoid funds to jump in and out of the available sample on any given day – given the daily rebalance – should the funds' AUM hover around the threshold. This is consistent with the behavior of institutional investors who would not trade in and out of a fund simply because on any given day its AUM drops slightly below the threshold. The results are robust to percentages other than 50%.

[12] Brown et al. (2019) limit their sample to ETFs with at least \$50 million in assets to mitigate the impact of illiquidity and possible non-synchronous prices due to infrequent trading. See also Elton et al. (1996, 2001, among others) for a discussion on the biases induced by funds with a low assets under management value.

[13] Appendix A.1.1 reports results using funds with AUM larger than \$50 million. We also ran the empirical analysis using other fund size cutoffs, e.g., \$100 million and \$500 million. In all these cases, the results are qualitatively very similar.

[14] We use the standard filters reported in the WRDS documentation available here.

Daily fund flows, from the EPFR dataset, are constructed for individual mutual funds and ETFs as follows:

$$\%flow_{i,t+1} = \frac{TNA_{i,t+1} - TNA_{i,t} \cdot R_{i,t+1}}{TNA_{i,t}} - 1, \qquad \text{(Percentage Flows)}$$
$$\$flow_{i,t+1} = TNA_{i,t+1} - TNA_{i,t} \cdot R_{i,t+1}, \qquad \text{(Absolute Flows)}$$

where TNA is the total net asset value, and R_{t+1} is the gross return of the fund over the period $t : t + 1$.[15]

When we calculate the aggregate percentage flows of the various smart beta strategies in Section 6.4, we use the following "by-strategy" definition:

$$\%flow^s_{t+1} = \frac{\sum_{j \in S_{t+1,s}} TNA_{j,t+1} - \sum_{j \in S_{t+1,s}} (1 + r_{j,t+1}) TNA_{j,t}}{\sum_{j \in S_{t+1,s}} TNA_{j,t}}, \qquad (1)$$

where $S_{t+1,s}$ is the set of funds belonging to a particular smart beta strategy s (e.g., value) at time $t + 1$.

4 Descriptive Statistics

Table 1 reports the descriptive statistics of our final sample. We split the sample in two subperiods (2004-2011 and 2012-2019) to highlight the role that smart beta investing played following the Global Financial Crisis and the Great Recession. The total number of unique funds increases steadily over time. We have 1,456 (1,080) mutual funds in the latter (former) period, an increase of 35%. The pattern is even more striking when looking at ETFs (+195%), which became extremely popular over the last decade. The average (median) fund size is around \$5 billion (\$2.1 billion) in our sample but the size distribution is heavily right-skewed. In the fourth quarter of 2019, for example, the NAV of 1,029 funds exceeded \$1

[15] Equivalently, the percentage change in flows can be calculated as:

$$\%flow_{i,t+1} = \text{SharesOutstanding}_{i,t+1}/\text{SharesOutstanding}_{i,t} - 1.$$

billion and 157 funds exceeded $10 billion. It is also striking how the average ETF in our final sample is very large compared to the average mutual fund ($9.3 billion vs. $4.6 billion); this is in line with the large ETF inflows over the last decade. The median age of mutual funds and ETFs in the most recent sample period is around five years. The number of stocks in mutual fund portfolios also varies substantially across funds. The median number of stocks is around 120, while ETFs hold on average 375 stocks in their portfolio. This suggests that large mutual funds cherry-pick their holdings, while ETFs are passively tracking indices. Looking at the returns of mutual funds and ETFs, we note that the median mutual fund (ETF) has outperformed the S&P 500 in the first half of our sample by 25 bps (169 bps) per year, while it has underperformed by 88 bps (93 bps) in the second half. Overall, the median five-factor alphas are negative over both samples for both mutual funds and ETFs.[16]

Our focus is on equity smart beta funds. For exposition purposes, in this section we classify funds in smart beta strategies based on their names (Appendix B.3 provides details on the procedure) and describe their time series properties. Focusing on smart beta funds' names allows us to show the growing importance of factor investing in the asset management industry, at least from a marketing perspective. However, as we discuss in the next section, classifying funds by name is not optimal since many funds are not consistently tracking the strategy (i.e., factor) mentioned in their names (Cooper et al. (2005), Chen et al. (2020)).

The top panel of Figure 1 shows the growth in the smart beta industry over time. As of the fourth quarter of 2019, smart beta equity mutual funds and ETFs command around $5 trillion in total AUM. This represents about 50% of the total AUM of equity mutual funds and ETFs, with the relative importance of smart beta funds steadily increasing over the last decade.

The recent growth in smart beta investing is largely due to ETFs. These funds have become a popular instrument to passively invest in equities and risk factors. Over the last decade, the number of ETFs has skyrocketed, with a large fraction of them (supposedly) tracking risk factors. In 2018, for example, 50% of all ETFs that crossed $1 billion AUM were smart beta.

Some examples of large smart-beta ETFs included in our sample are the iShares Edge MSCI USA Quality Factor ETF (QUAL, $19.6 billion AUM), the iShares Edge MSCI USA Momentum Factor ETF (MTUM, $14.4 billion AUM), the iShares Edge MSCI USA Value

[16]The only exception is the alpha for ETFs before 2012, although it is not statistically different from zero.

Factor ETF (VLUE, \$10.5 billion AUM), and the Goldman Sachs ActiveBeta U.S. Large Cap Equity ETF (GSLC, \$11.7 billion AUM).[17] To further highlight the importance of ETFs in the competitive asset management industry, the SEC recently relaxed the requirements for launching ETFs by lowering the barriers to entry.[18]

To understand the importance of smart beta for institutional and retail investors, the bottom panel of Figure 1 shows the total AUM of institutional and retail funds investing in smart beta strategies. The relative share of institutional AUM relative to the total smart beta assets has increased over time from 10% to around 50%, suggesting that smart beta strategies have become extremely relevant for institutional investors.

Our sample does not include in-house smart beta strategies implemented by large, buy-side institutions such as the Norwegian Oil Fund, as these data are not publicly available. As a consequence, we underestimate the total assets invested in smart beta strategies by institutional investors. However, our sample *does* include mutual funds listed by smart beta providers such as AQR or Robeco, which cater to institutional investors; hence, we can still capture (part of) the institutional side of the market.

To get a more precise understanding of the distribution of smart beta funds across strategies, Figure 2 provides a bar plot of the growth in AUM and number of funds tracking the individual factor legs in our final sample. Growth funds have attracted substantial inflows, which is reflected both in the number of growth funds launched in the market and on the performance of growth stocks over the last decade (e.g., FAANG stocks becoming the most valuable companies in the world). Quality has also been a smart beta strategy sought after by investors, while surprisingly few funds and assets chased momentum strategies, probably due to scalability and capacity issues.[19]

Overall, it is evident that smart beta strategies have become increasingly important over the last decade for both institutional and retail investors, to the point of being coined 2019 as the "new kings of Wall Street."[20] As discussed earlier, one of the key contributions of this paper is to properly identify smart beta funds. In this section, we have looked at a classification of funds based on their names. In the next section, we discuss why such a naive classification is problematic, as many funds' names and prospectuses are often misleading

[17] AUM as of January 2021.

[18] See US regulator overhauls requirements for launching ETFs (Financial Times, 9/26/2019).

[19] We discuss this issue in Section 6.3.

[20] Index Funds Are the New Kings of Wall Street, WSJ September 2019.

(Cooper et al. (2005), Chen et al. (2020)), and we propose an alternative, arguably better way to classify funds into smart beta strategies.

5 Fund classification

A key issue for smart beta fund investors is how well the fund tracks the underlying "characteristic." In other words, what is the fund's tracking error with respect to the benchmark, where the benchmark for most smart beta funds is the long or short leg of a given factor (e.g., the growth leg of the HML factor).

Indeed, many academic studies show that several long-short factors earn consistently positive risk premia, while "hedging away" aggregate market risk. Investors want to harvest these risk premia. However, as discussed in Section 1, most funds tend to specialize and track only individual factor legs, so that no direct and easy way to replicate the performance of long-short factors (e.g., HML) is currently available to investors. The large number of available funds with similar names and objectives constitutes an additional challenge for retail and institutional investors. For example, there are well over 100 value and high-dividend ETFs in the U.S. alone, that track large, small or midsize stocks, based on different definitions of value. Moreover, the existence of multi-strategy funds simultaneously tracking multiple factors, such as momentum, quality or low volatility and different managers' incentives, makes the classification of funds a non-trivial task.

In this section, we propose a novel, general methodology to identify true smart beta funds that can be used both over the full sample (e.g., by policy makers) or in real-time (e.g., by investors). It is based on a two-step procedure. It identifies index funds based on a minimum distance statistical approach, and then it categorizes the remaining (non-index) funds into smart beta strategies according to their holdings. Before describing the procedure, we discuss why using fund names or regressions to categorize smart beta funds might result in misclassification.

5.1 Fund names and beta approach

The most naive way to classify funds is to look at their names. However, as already emphasized in the literature (e.g., Cooper et al. (2005), Chen et al. (2020)), fund names might be

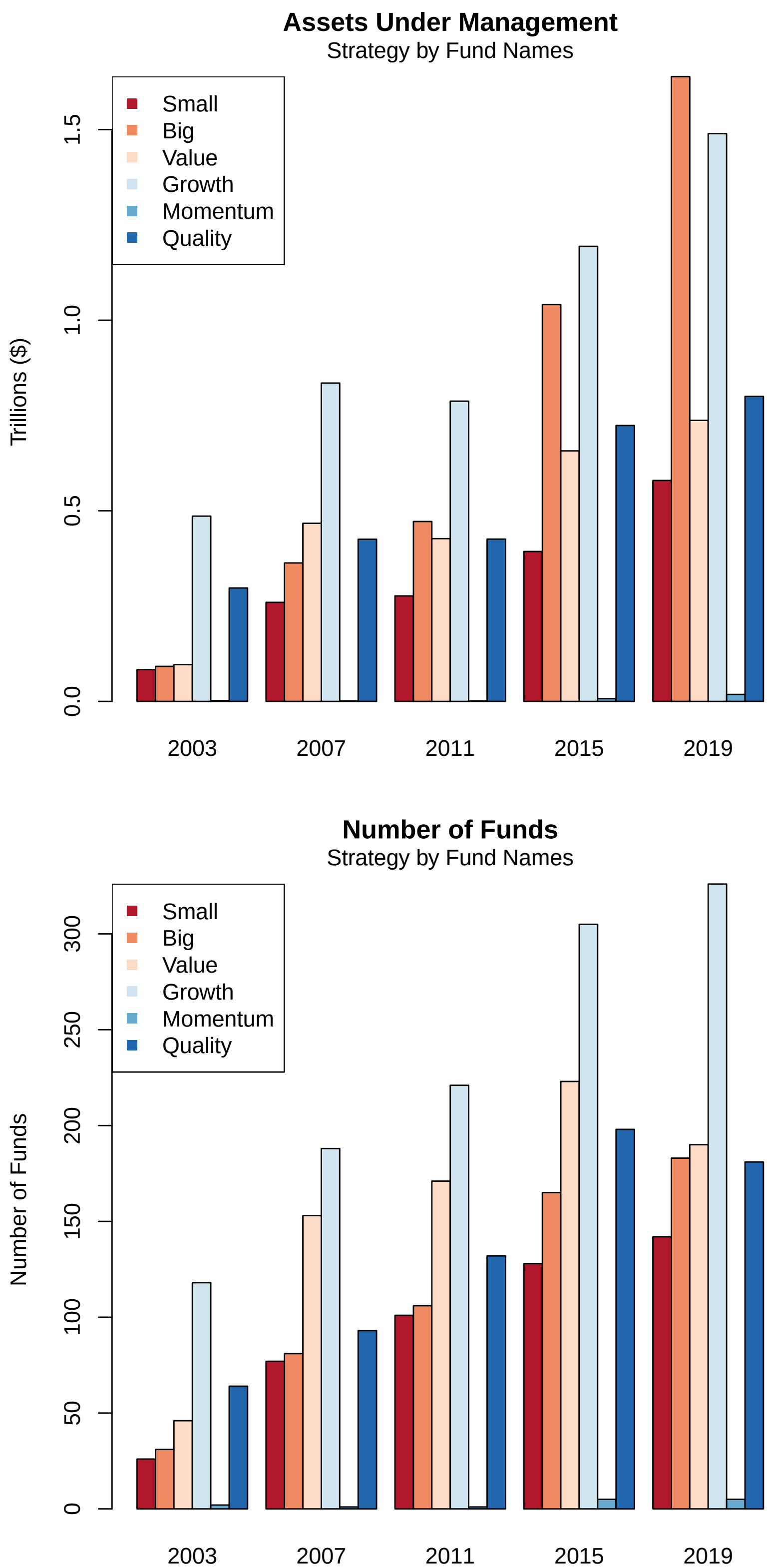


Figure 2: Net asset values and number of funds by strategy over time. The top panel plots the average net asset values of smart beta funds, by strategy, over time, while the bottom panel plots the time series of the number of funds available. Funds are categorized by name and must have more than $1bln AUM. The sample period is from 2003 to 2019.

misleading. The Securities and Exchange Commission (SEC) states[21] that investors should not "*assume that a mutual fund called the "ZYX Stock Fund" invests only in stocks or that the "Martian High-Yield Fund" invests only in the securities of companies headquartered on the planet Mars. The SEC generally requires that any mutual fund or ETF with a name suggesting that it focuses on a particular type of investment must invest at least 80% of its assets in the type of investment suggested by its name. But mutual funds and ETFs can still invest up to one-fifth of their holdings in other types of securities—including securities that a particular investor might consider too risky or perhaps not aggressive enough.*"

To emphasize the relevance of this issue, the SEC is requesting (as of March 2, 2020) public suggestions on how to eliminate misleading fund names.[22] Our hope is that our methodology can help address this problem, which severely distorts investors' allocation decisions.

In Section 7 we provide a formal analysis on whether it is feasible to replicate the long-short (non-tradable) factors using a naive strategy based on fund names for both institutional and retail investors. As we will see, a strategy based on fund names can only be implemented for the SMB and HML factors, and it fails to deliver a satisfactory replication, in particular, of HML.

A simple alternative to looking at fund names is to compute the correlation between a fund's historical returns and the benchmark of the fund (or the Fama-French leg return), and argue that the fund is tracking the benchmark well when such correlation is high (e.g., greater than 95%). This approach, which we dub the "beta approach," however, could be misleading for several reasons. First, even when a portfolio of funds displays a high correlation with the benchmark, it may still display a substantial alpha in a simple univariate return regression of the benchmark on the portfolio. Second, statistically speaking, some correlations might be spurious, with some funds tracking the benchmark(s) "by luck." Finally, another limitation of the "beta approach" is that many funds are relatively new and therefore there is not enough statistical power to test their relative performance with respect to the benchmark, and there might be time-variation in the beta exposure itself.

As an example, the Goldman Sachs Mid Cap Value Fund (GCMAX) can be considered a "value" fund given its name; this is also consistent with its claimed investment strategy.[23]

[21]See here: Looking Beyond A Mutual Fund or ETF Name.

[22]See here.

[23]We randomly pick this fund among the set of funds with similar inconsistencies. This should only be

However, this fund features, over the last decade, a partial correlation (i.e., after netting out the market) with the Fama and French Growth portfolio that is higher (0.89) than its correlation with the Value portfolio (0.45). Even when using rolling estimates, the fund has *always* a higher partial correlation with the growth factor.

To conclude, neither the (naive) approach based on fund names nor the beta approach provides a satisfactory answer as to whether investors can harvest the risk premia of well known factor strategies.

5.2 Identify Market Funds: Minimum Distance Approach

A common problem in identifying smart beta funds is related to closet indexing[24] (e.g., Cremers et al., 2016). Many mutual funds (and some ETFs) claim to be actively managing their portfolio, while displaying only a small tracking error with respect to their benchmark. In other words, these funds charge an active management fee while passively tracking the index. Therefore, an investor who wants to harvest factor risk premia would like to eliminate these funds. Unfortunately, the CRSP mutual fund database does not have an indicator variable that allows us to identify market funds (i.e., funds simply tracking the market).

The first step of our methodology addresses this issue by identifying market funds using a "minimum distance regression approach." We proxy the market by the MKT factor of Fama and French, although our results are robust to using other index proxies.

First, we run a time series regression of the individual factor legs (e.g., the S of SMB, the B of SMB, the H of HML, etc.) on the long-short risk factors

$$r_{l,t} = \alpha_l + \beta_{l,MKT} r_{MKT,t} + \beta_{l,SMB} r_{SMB,t} + \beta_{l,HML} r_{HML,t} + \beta_{l,MOM} r_{MOM,t} + \beta_{l,RMW} r_{RMW,t} + \varepsilon_{l,t} \tag{2}$$

where the factor leg $l \in \{$MKT, Small, Big, Value, Growth, Up, Down, Robust, Weak$\}$.

Table 2 reports the results for daily (Panel A) and monthly (Panel B) returns. The estimated loadings should be considered the "benchmark" coefficients for funds that track the various individual factor legs, as discussed by Lettau et al. (2019). Note that, as expected,

viewed as an illustrative example, not as an evaluation of the fund's performance.

[24]Recently the Financial Conduct Authority has for the first time publicly named and fined an asset manager for being involved in the controversial practice of closet tracking in a fund marketed to retail investors, who were overcharged by almost £1.8 million in fees.

the sum of the loadings of the short and long legs on their own risk factor sum to one. The estimated benchmark loadings appear very robust to the data frequency, suggesting that our methodology is not affected by it.

Any fund that tries to replicate the short or long leg of a specific risk factor should have loadings "close" to the ones of the individual relative legs estimated in (2). As a consequence, in the second step, we re-estimate regression (2) after replacing the dependent variable with the actual returns of fund i:

$$r_{i,t} = \alpha_i + \beta_{i,MKT} r_{MKT,t} + \beta_{i,SMB} r_{SMB,t} + \beta_{i,HML} r_{HML,t} + \beta_{i,MOM} r_{MOM,t} + \beta_{i,RMW} r_{RMW,t} + \varepsilon_{i,t}.$$

Under the hypothesis that fund i is replicating factor leg l, one expects all $\hat{\beta}_{i,f}$ to be equal to $\hat{\beta}_{l,f}$, which can be tested with an F-test. Then, we can directly compare the proximity of fund i to leg l using the statistical distance (i.e., the value of the F-test). As an example, if we want to test whether a fund return series is consistent with the "Growth" portfolio, the null would be set to H_0: $\beta_{mkt} = 1.02, \beta_{SMB} = 0.41, \beta_{HML} = -0.30, \beta_{MOM} = -0.02, \beta_{RMW} = -0.15$ (using daily data). More generally, we compute a total of ten F-tests (e.g., one for each of the regressions in (2)), and we assign a fund i to the strategy f corresponding to the lowest F-statistic. We label this approach the "minimum distance regression approach." Those funds that have the lowest F-statistic associated with the market are then removed from the results labeled "excluding MKT funds" in the empirical results section.

Using our approach, about one-half of the funds get classified as market index funds over the full sample.[25]

In Figure 3, we use an expanding window estimation to classify funds as market funds in real-time. Except for a short period of time around 2010, the percentage of the funds categorized as market funds is quite stable at around 50%. Our methodology is quite general and can be implemented with any preferred factor model. Most importantly, as discussed above, it can be used by policy makers and regulators to guarantee that funds are not misleading investors by following an investment strategy that is different from the one stated in their names or prospectuses.

The main argument in favor of such a two-step procedure (i.e., first compute betas, and then use a F-test to classify market funds) is motivated by the fact that the loadings on

[25] There are slight variations depending on whether we focus on retail/institutional funds, or ETFs.

the individual legs of the long-short risk factor (e.g., the "small" leg S on the SMB factor) are not symmetric, and the more volatile leg will display a larger beta (in absolute value). In other words, comparing two funds based only on the magnitudes of their HML betas, for example, is misleading. For illustration purposes, let us consider two funds, A and B. Fund A has a $\beta_{A,HML} = 0.25$, while Fund B has a $\beta_{B,HML} = -0.25$. In absolute value, they are the same, suggesting that they are "equally distant" from being pure value and growth funds. However, this inference is incorrect. Panel A of Table 2 shows that the "benchmark" loadings equal 0.7 for value (H) and -0.3 for growth (L). This implies that Fund B is much closer to be a growth fund (e.g., -0.25 vs. -0.30), than Fund A is a value fund (e.g., 0.25 vs. 0.70). Hence inference based solely on the absolute magnitude of the betas would result in fund misclassification.

Lettau et al. (2019) discuss three additional reasons why using simple betas to determine the fund type might be problematic. First, risk exposures are estimated using historical data and are thus subject to estimation error. Second, historical data might also not reflect the current portfolio of an active fund. This is especially true for firm characteristics that substantially change over time (e.g., the strategy has a high turnover), such as momentum. Third, the factor loadings may be time-varying, as often it is the case with growth firms that become value firms once their cash flows stabilize.[26]

Overall, the minimum distance approach (MDA) presented in this section addresses the issue of (mis)classifying market funds as smart beta ones. We use this approach only to identify market funds from the universe of available funds. If a fund is classified as a market fund using the statistical distance against the MKT, SMB, HML, MOM, and RMW factors, it will not be included in our synthetic, tradable factors when we report results labeled as "excluding market funds."

It is important to highlight that using holdings, rather than our methodology, to identify market funds is not meaningful because the "market," differently from equity risk factors, is not sorted on any firm characteristic. Similarly, using our minimum distance approach to classify smart beta funds would not be feasible since it requires a long time series of the funds' returns but most smart beta funds are relatively new. Instead, using holdings as in recent studies (e.g., Lettau et al., 2019) allows us to classify funds from inception in real

[26]However, using daily data and rolling regressions make these issues potentially less relevant in understanding the fund type.

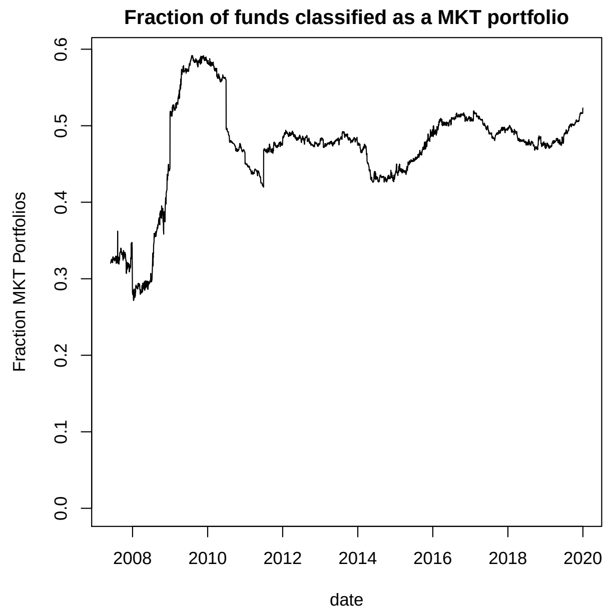


Figure 3: Fraction of funds classified as a market portfolio. This figure shows the fraction of funds classified as market funds by our methodology in real-time using an expanding window. A fund is classified as a market fund if it has the minimum F-stat with respect to the market portfolio. The sample is from June 2007 to December 2019.

time.

Whether a fund is classified as a market fund will depend on the choice of risk-factors used in regression (2). However, from the point of view of a regulator, or investor, this is not an issue since they can determine what risk factors they deem relevant or want to get exposure to, and only include those in the regressions. In other words, our methodology is applicable under any chosen set of risk factors.

The next subsection shows how we classify funds into specific smart beta strategies.

5.3 Fund Holdings

Following Leippold and Ruegg (2019) and Lettau et al. (2019), we categorize funds by looking at their holdings. For each fund, we construct the factor scores, that is, a number that describes how each fund is exposed to the different factors. To obtain these scores, we proceed as follows. Every period t, we sort each stock i based on a factor characteristic (e.g., book-to-market). That is, we obtain a factor score $s_{i,f,t}$ for each security $i = 1, 2, ..., N$, on each factor $f = HML, SMB, MOM, RMW$. We normalize the score for each characteristic from minus one (worst) to one (best). For example, if the characteristic is the book-to-market, the stock with the highest book-to-market (value stock) obtains a score of one and the company with the lowest book-to-market (growth stock) obtains a score of minus one. We then aggregate these scores in a time-varying factor score matrix $S_t \in R^{N \times F}$, with F denoting the number of factors. Lastly, we aggregate these scores at the fund level as follows:

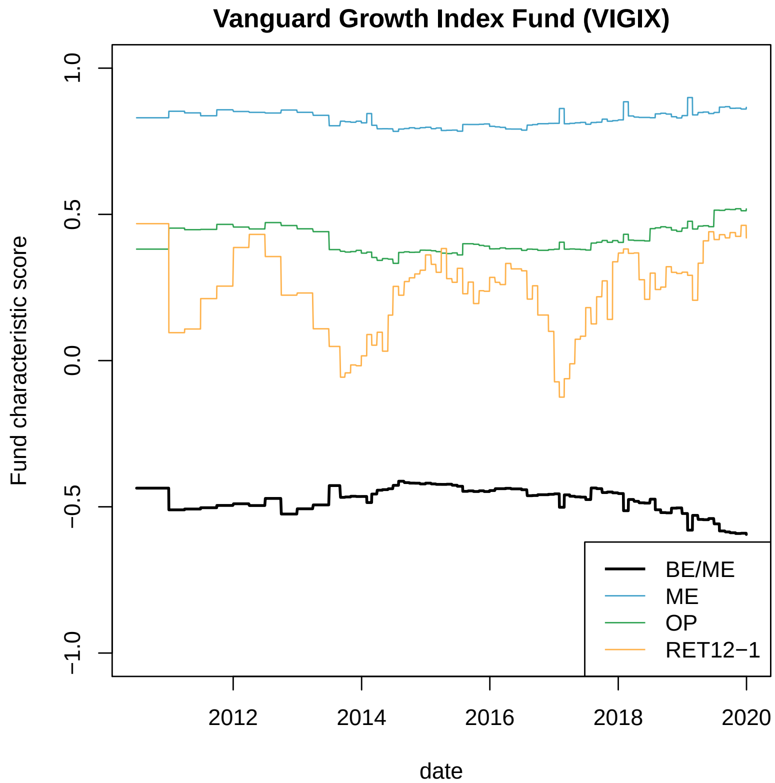
$$X_{m,t} = \omega_{m,t} \times S_t, \tag{3}$$

where $\omega_{m,t} \in R^N$ is the weight of the individual stock in the fund m at time t. Economically, $X_{m,t} \in R^F$ is the relative "strength" of the fund with respect to each factor.

This methodology is informative about what factors the funds are exposed to, including multi-factor funds[27] (e.g., if the book-to-market and momentum fund scores are both large and similar, our methodology will categorize the fund as "value-momentum").

Figure 4 displays the time series of the characteristics' score for two funds in our sample, Vanguard Growth Index Fund (top panel) and iShares Russell 2000 Value ETF (bottom panel). We focus on these two funds because they have a long time series, and they enter the construction of our synthetic portfolios (see Section 5.4) at some point in time. Several comments are in order. First, the book-to-market score is higher for the fund that markets itself as Value. We observe that the book-to-market score for the iShares Russel 2000 Value is less extreme (in absolute value) than that of the Vanguard Growth Index Fund. This result is reminiscent of Lettau et al. (2019), who document that, despite a large number of "value" ETFs, very few have consistently high book-to-market ratios. Second, funds are not neutral to those characteristics that are absent from their name; for example, the Vanguard Growth Index Fund features a high exposure to profitability as well. Third, the scores are quite stable over time for both funds, suggesting that our synthetic portfolios will not feature extreme turnover. Finally, note that the Vanguard Growth Index Fund is included in our synthetic "Growth" portfolio only during 2010 (when the book-to-market is at its lowest level). This latter fact suggests that the performance of the tradable factors obtained through our procedure can differ substantially from that of a simpler allocation based solely on names.

[27] According to a FTSE Russell survey done in 2019, multi-factor strategies are becoming the most popular among institutional investors globally. Multi-factor products captured 11% of new net flows into Europe-domiciled funds in 2019, compared to just 2% in 2015, according to BlackRock.



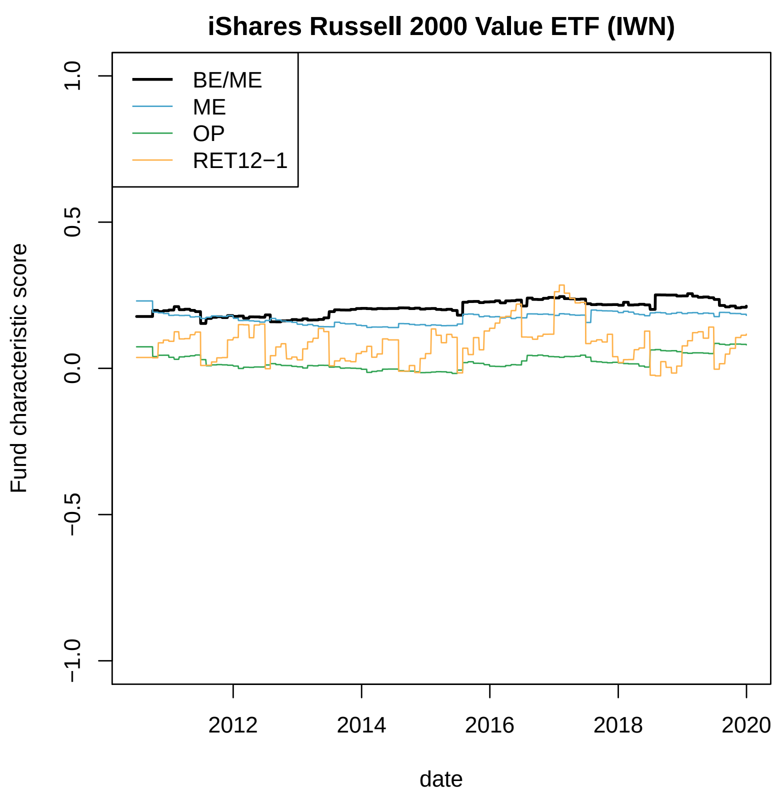


Figure 4: Time series of the characteristics' score for two funds in our sample. The top figure plots the time series of the characteristic scores for the Vanguard Growth Index fund which is classified as "Large-Growth" according to Morningstar. The bottom figure plots the time series of the characteristic scores for the iShares Russell 2000 Value fund which is classified as "Small-Value" according to Morningstar. The sample period is from July 2010 to December 2019.

5.4 Synthetic Smart Beta Strategies and Factors

The final step of our procedure involves constructing synthetic, tradable long-short factors, which are the basis of the empirical analysis in Section 6. We generate synthetic factors for two types of investors: retail and institutional. This allows us to study the properties of the *actual* portfolios tradable by a different set of investors. Perhaps, the set of trading strategies available to institutional investors is larger than the one available to retail investors; or, maybe, they all have access to the same factor trading strategies, but institutions are better able than retails to earn the unconditional factor risk premia (i.e., lower tracking error).

At each time t, and for each characteristic, we rank funds using the score at the fund level (see equation (3)). Then, for each characteristic, we construct the tradable long leg of the factor-mimicking portfolio by value-weighting the top ten mutual funds and ETFs. Analogously, we construct the short leg using the bottom ten funds, but restricting the funds to be ETFs. Then, we hold the synthetic portfolio until the next rebalancing time-period. Since we follow the construction method of Fama and French (1996), the Small/Big, Value/Growth, and Profitability legs are updated at the end of June each year, using firm characteristics from the end of last year. Therefore, even though we "rebalance daily," in practice, most portfolios will only be rebalanced once a year, with two exceptions. First, if a fund exits the sample during the year, a new fund will replace it in the synthetic portfolios at the next daily rebalancing. Second, when constructing the synthetic momentum factor, we rebalance daily using the previous year's return excluding last month's return as is standard in the literature. We repeat this procedure for both institutional and retail investors.[28]

Our choice of a fixed number of funds in the synthetic portfolios, namely ten, strikes a compromise between having a well-diversified portfolio (large number of funds), and making the synthetic portfolio easy to track and monitor for an investor (low number of funds).[29] Furthermore, fixing the number of funds allows us to to properly test the performance of the tradable factors across samples and specifications. Indeed, assume the selection was based on the top 10% of all available funds, rather than a predetermined number. This procedure would imply that the number of funds used to replicate the factor changes over time (due

[28]In Appendix A.2.1 we also construct the factors using "live" data.

[29]The results are robust to alternative choices (e.g., 25 funds) suggesting that a relatively small number of funds (10 in our case) ensures both tradability and diversification.

to the steady increase in funds throughout our sample, as seen in Table 1), and across styles. In turn, this procedure would imply that the synthetic strategy holds a different amount of idiosyncratic risk, making any comparison across samples and specifications hard to interpret. On the other hand, our choice of a fixed number of funds makes the estimates comparable.[30]

As mentioned above, we construct the long leg of the factors using both mutual funds and ETFs, while only ETFs in the short leg to ensure complete tradability of the long-short factors. The reason behind this choice is that it is not feasible to short mutual funds in practice, while it is relatively easy to short (liquid) ETFs.[31] We also report conservative results using only ETFs in both factor legs.[32]

The next section describes the performance of these synthetic, tradable long-short factors.

6 Empirical Results

6.1 Long-short Factors

In this section, we investigate the ability of our synthetic, tradable long-short portfolios to track the underlying (non-tradable) Fama-French risk factors. We run the following regression:

$$ret_{FF_{L-S},t} = \alpha + \beta ret_{synth_{L-S},t} + \varepsilon_t, \tag{4}$$

where $ret_{FF_{L-S},t}$ and $ret_{synth_{L-S},t}$ denote the returns of the long-short Fama-French factor, and their synthetic counterpart, respectively.

[30]We find a significant overlap between the value and growth funds selected by our procedure and those reported in (Lettau et al., 2019, Table 4), suggesting our methodology correctly identifies the "true" underlying source of risk. Examples include the Dimensional Fund Advisors (DFA) US Small Cap Value, and the DFA US Large Cap Value funds.

[31]In Appendix A.3, we construct and analyze the synthetic individual factor legs separately. They can be constructed using both mutual funds and ETFs since they are long-only.

[32]Sorting funds in (3x2) portfolios in a similar fashion to the double-sorted procedure used by Fama and French in constructing their long-short HML and SMB factors is also not feasible in practice. First, it would include hundreds of funds in the individual factor legs, making the portfolio costly to trade. Second, and more critical, the short leg would require only ETFs, making the total number of funds in the long and short portfolios different. In turn, this would imply different levels of idiosyncratic risk in the two legs. Third, the overall characteristic exposure of the final portfolio will be attenuated.

Table 3 reports the results. Panel A shows the estimates using the universe of funds available to retail investors, while Panel B shows the same results using funds available to institutional investors.

In Panel A, we observe economically large alphas for all risk factors.[33] The failure of SMB is particularly noticeable. Indeed, since shorting large and liquid stocks is relatively easy, one would expect SMB_{synth} to have good tracking ability. This is confirmed by a large (in fact, the largest) R^2 of 67%. Despite this, the alpha is economically large at 1.34%. The alphas on MOM_{synth} and RMW_{synth} are even larger at 2.37%, and 3.36%, respectively. Differently from SMB_{synth}, the R^2s for MOM_{synth} and RMW_{synth} are also quite low (30% and 23%, respectively). Retail investors also fail to replicate the HML factor: the alpha is 1.48% and the R^2 is only 41%. Excluding market funds from the analysis does not change the picture: despite an average improvement in terms of R^2, which implies that our real-time identification of market funds works as expected, we still observe large alphas. These results highlight the striking discrepancy between the performance of an optimal, factor-replicating portfolio of funds available to retail investors and the non-tradable "benchmark" factors commonly used in the literature.

Turning to institutional investors (Panel B), the replication of HML and SMB improves relative to retail investors: in particular, the alpha reduces to 0.94% for HML_{synth} and 0.52% for SMB_{synth}. The decent performance of SMB_{synth} by institutional investors is perhaps not surprising, since it is relatively easy to go long index funds tracking small cap stocks (e.g., Russell 2000 funds), and short liquid, large cap funds.

Despite an improvement relative to retail investors, the replication of the MOM and RMW factors continues to prove challenging even for institutional investors. Comparing Panel A to Panel B, with the exception of RMW_{synth}, we observe larger R^2s for institutional investors. In particular, the R^2 of HML_{synth} for institutional investors is 50% larger than that attained by the retail investors' HML_{synth}. Also, as in the case of retail investors, excluding market funds does not lead to a significant improvement in terms of alphas despite a generally better time series fit. Figure 5 provides a graphical representation of the results in Panels A and B of Table 3. Specifically, the figure displays the cumulative returns of the synthetic strategies (e.g., SMB_{synth}, HML_{synth}, MOM_{synth}, and RMW_{synth}), for both retail

[33]Despite being economically large, alphas are often statistically insignificant due to the large volatility of daily returns, and the relatively short sample.

and institutional investors, together with the non-tradable factor that each strategy tries to replicate. The top panels of Figure 5 confirm that the synthetic versions of SMB and HML tradable by institutional investors track the benchmarks much better than those by retail investors. The bottom two panels, consistent with the regression results, show that neither retail nor institutional investors are able to track the momentum and profitability factors. Overall, our results suggest that the tracking ability of institutions, and their capacity to earn the unconditional factor risk premia, is better than that of retail investors. We discuss potential explanations of this evidence in Section 6.2.

Finally, turning to synthetic strategies implemented using ETFs only (Panel C in Table 3), we see a substantial improvement for SMB$_{synth}$, whose alpha is further reduced to 12 bps. On the other hand, using only ETFs yields an inferior replication of HML relative to both retail and institutional investors: the HML$_{synth}$ alpha is quite large (1.94%) relative to the ones reported for retail investors (1.48%) and institutional investors (0.94%). This suggests that mutual funds in the Value leg are quite important in the synthetic replication.

The results highlight the complexity of synthetically replicating the (unconditional) factor risk premia identified in the literature. Institutional investors seem to be able to reap the value premium (HML), while both types of investors can successfully earn the size premium using ETFs. Also, the momentum and profitability factors cannot be replicated using synthetic portfolios of funds.

In Table 3 we focus on the replicability of the long-short factors based on the first moment (e.g., the average return differentials). Table 4 reports the summary statistics of the Fama- French factors, together with those of our synthetic tradable portfolios. The synthetic SMB$_{synth}$ and HML$_{synth}$ factors have similar volatility with respect to their non-tradeable counterparts, while MOM$_{synth}$ has approximately 40% lower volatility for both retail and institutional investors. This is due to the well-known fact that the short leg of MOM is not easily tradable and is subject to momentum crashes (Daniel and Moskowitz (2016)). Also, the synthetic versions of SMB, HML, and MOM have thinner tails relative to the non-tradable version of the factors: as an example, MOM has an excess kurtosis of 9.289, while our synthetic factors have an excess kurtosis between one-third and one-half of that (e.g., 3.031 and 4.222), suggesting our synthetic momentum factor has substantially less tail risk.

Table 4 also presents, for each factor, the ratio of the scores of the long and short portfolios on four characteristics (Size, B/M, Past returns, and Profitability). With the sole exception

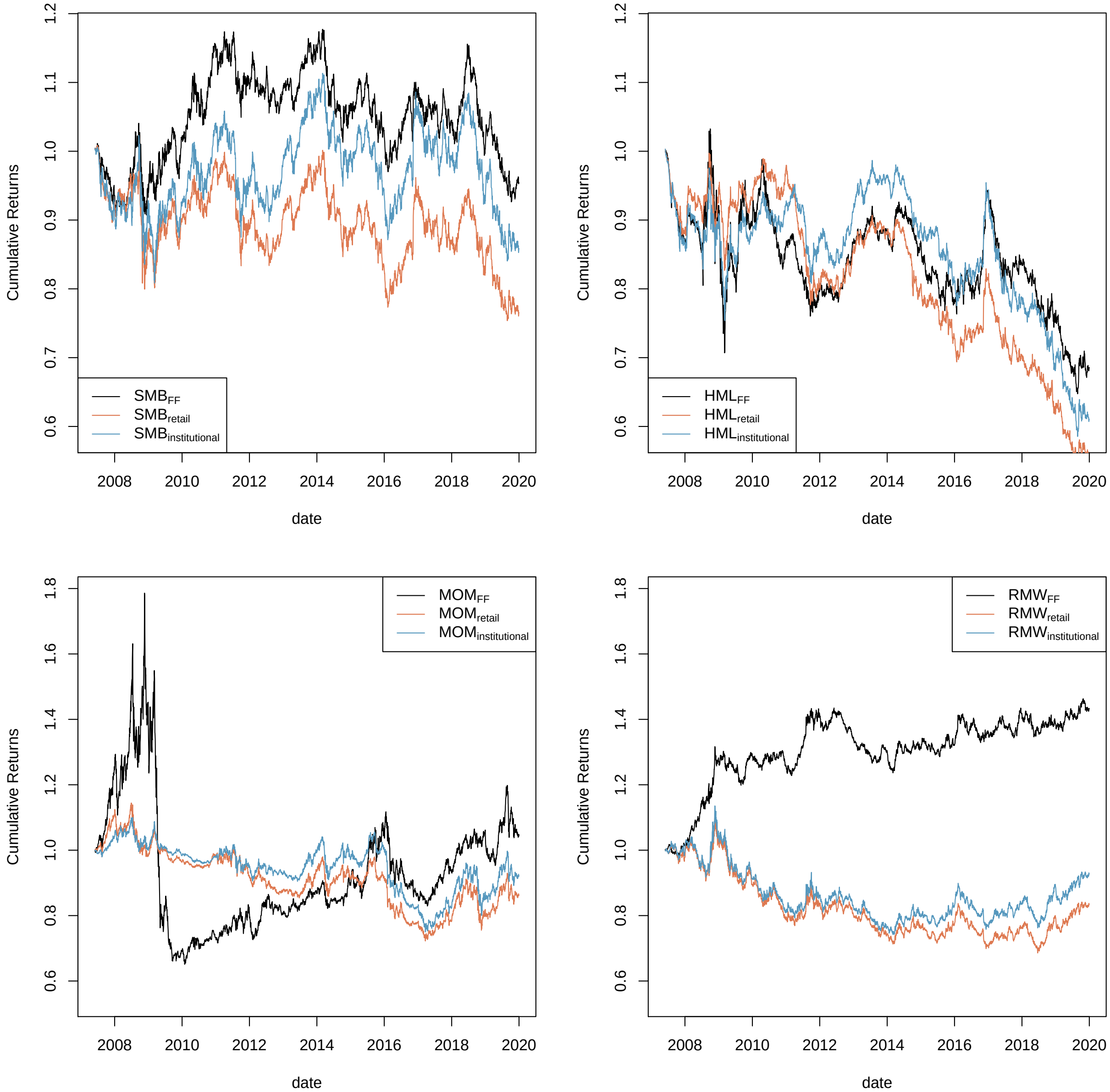


Figure 5: Cumulative returns of the synthetic long-short factors against the Fama French benchmarks. The figure shows in each panel three cumulative return series: the benchmark Fama-French long-short factor, and both the institutional (blue line) and retail (orange line) synthetic long-short portfolios. The sample period is from June 2007 to December 2019.

of SMB, we observe that our synthetic factor scores on characteristics other than the one used in the sorting procedure (cross-exposures) are extremely similar to those of the Fama-French benchmarks. This observation is important for two reasons. First, it provides additional evidence that our procedure is successful in attaining an accurate replication of the Fama-French benchmarks. Second, the fact that the cross-exposures are close to one implies that our synthetic factors are (close to) neutral with respect to other possible sources of risk. As a case in point, the score of the synthetic institutional HML on past returns is 0.94, which is also the same value for the Fama-French factor. This result underscores that the presence of multifactor funds in our sample is not an issue, since our methodology is able to generate synthetic portfolios with exposures to other factors of similar magnitude to those of the benchmark portfolios, and, hence, are close to neutral.

6.2 Drivers of Retail and Institutional Investors' Performance

An important economic question underpinning our results in Section 6.1 is what drives the different performance of retail and institutional investors in tracking the long-short factors.

To answer this question, we first investigate whether such a difference can be related to a differential exposure to the underlying factor characteristics. To this end, we show in Figure 6 the characteristic scores of the synthetic portfolios for institutional and retail investors, along with the ones of the Fama-French benchmark factors.[34]

For the long leg of SMB and HML, we observe a large score difference between the synthetic and the non-tradable benchmarks. This is less the case for the short legs, Big and Growth. Interestingly, institutional and retail investors' portfolios are often indistinguishable in terms of characteristic score, with the sole exception of Value, in which case institutional investors attain a higher score throughout the sample.

The evidence in Figure 6 suggests that the better performance of the institutional synthetic HML relative to its retail implementation is largely attributable to a better proxy for the long Value leg. This is indeed confirmed in Figure 7. The left panel compares the retail and institutional performance of the synthetic Value leg, along with the performance of the

[34]Figure 3 in Lettau et al. (2019) shows a similar analysis, although with some notable differences, namely (i) they select the funds based on their names, while we calculate the characteristic score of the synthetic, *optimal* factor replicating portfolio; (ii) they show the unconditional distribution of the characteristic scores across the funds, while we plot the time-series of the characteristic score of the synthetic portfolios; (iii) they only use mutual funds, while we include ETFs in the construction of the synthetic replicating portfolios.

respective non-tradable factor leg. The left panel confirms that the institutional investors do better than retail investors when it comes to replicating a Value strategy. The right panel shows the synthetic Growth leg, and there is no difference across types of investor.[35]

Turning to the Momentum and Profitability factors, we note that, differently from SMB and HML, it is the synthetic short leg that has a characteristic score far from that of the non-tradable factor leg. This is true for both types of investors. We conclude that the failure to replicate Momentum and Profitability is to a large extent attributable to the performance of the short legs since funds investing in "losers" or firms with weak profitability are not readily available.

Overall, we find that the characteristic scores of the institutional investors are, on average and across strategies, closer to the Fama-French ones. This, in turn, may explain the better performance of the synthetic institutional portfolios relative to the retail ones. More broadly, our results are consistent with a form of market incompleteness, meaning that investors trading ETFs and mutual funds cannot achieve a large exposure to the long leg of HML, or the short legs of MOM and RMW, as indicated by characteristic scores of the synthetic factor legs that are substantially different from those of the Fama-French factors.

6.3 Capacity of Smart Beta Strategies

Given the recent massive increase in smart beta mutual funds and ETFs, it is important to understand how much capital can flow into these smart beta strategies without distorting market prices, and whether the factor premia (e.g., HML) can be easily accessed through trading, even by large investors with several billions under management (e.g., Ratcliffe et al., 2017). In other words, how much capacity do these smart beta strategies *actually* have? In this section, we provide some insight on this important question.

Figure 8 shows the median net asset value of the synthetic institutional and retail portfolios over time.[36] Our synthetic portfolios only include, in each leg, ten large (i.e., AUM

[35] Of course, when we replicate the short leg of HML, we only use ETFs that are accessible to both retail and institutional investors. Therefore, mechanically any difference can only come from the long leg. However, even when the Growth leg is implemented using MF and ETFs, the figure shows no significant difference across investors.

[36] We plot the six-months smoothed median as a proxy of the stability of the strategies. The reason is that when a fund goes just below the $1 billion threshold, or a new fund enters into the synthetic portfolio, the median happens to jump, although this is rare.

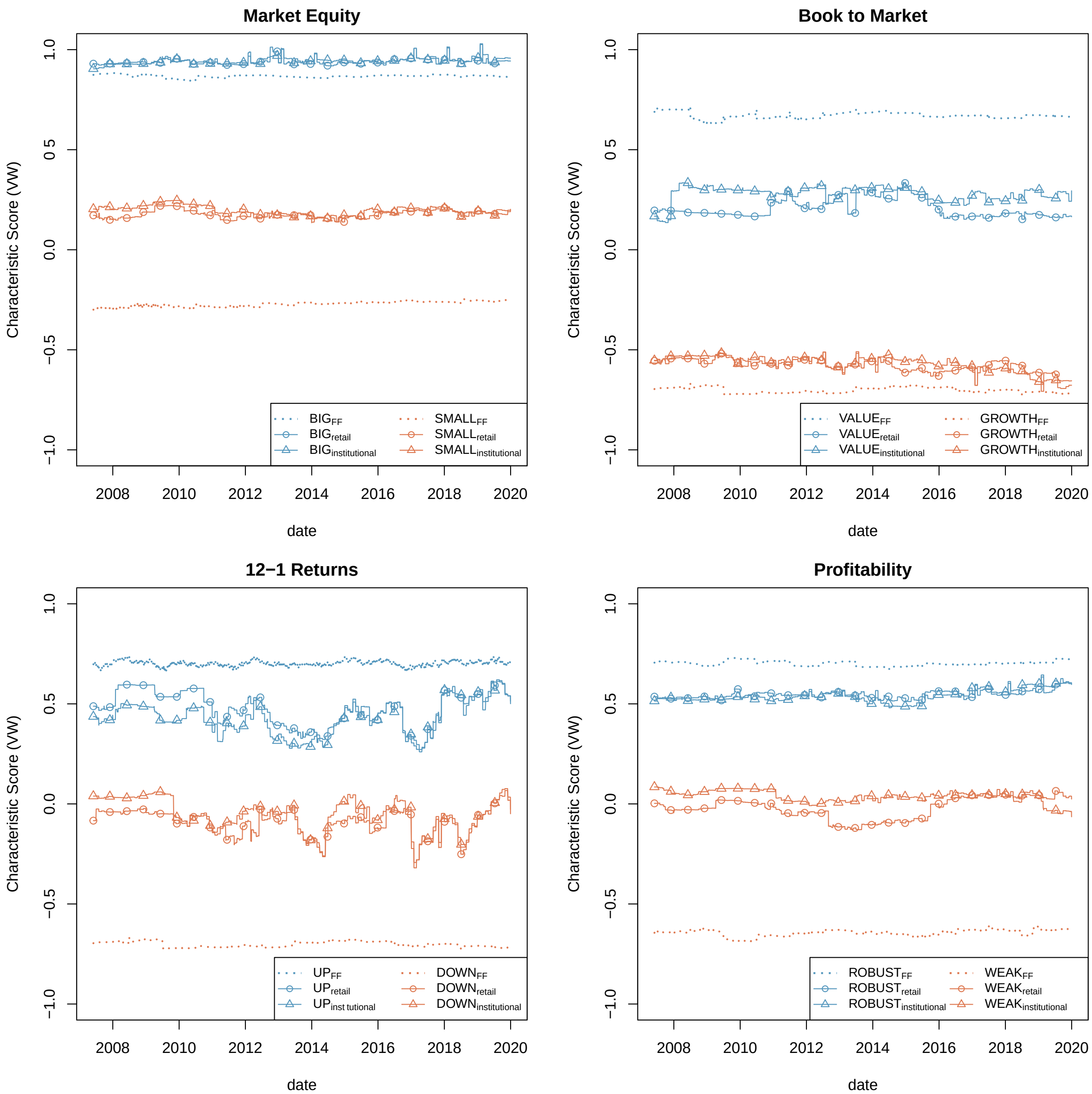


Figure 6: Characteristic scores of the synthetic portfolios. This figure plots the value-weighted characteristic scores of the synthetic portfolios (solid lines), for both retail and institutional investors, together with the Fama-French benchmark factor mimicking portfolio's characteristic scores (dashed line). The top (bottom) left quadrant reports the scores for the legs of SMB (MOM) on the left. The top (bottom) right quadrant reports the score for the legs of HML (RMW).The long leg of the synthetic portfolios is constructed using both mutual funds and ETFs, while the short leg only ETFs. The sample is from June 2007 to December 2019.

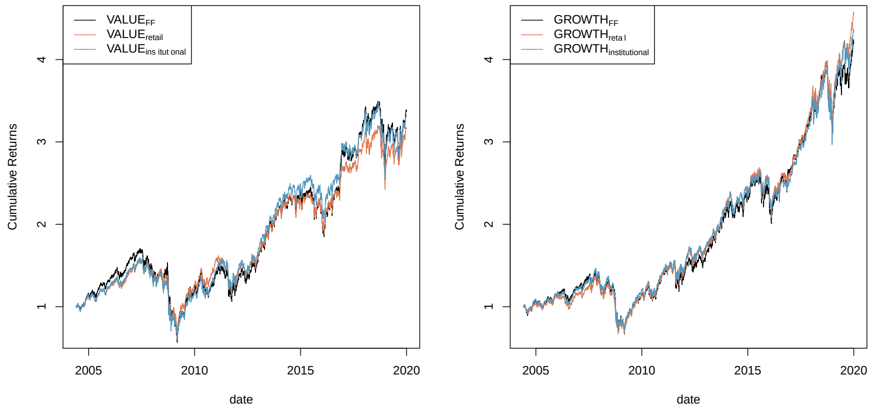


Figure 7: Replicating Value and Growth. The figure shows how institutional and retail investors can replicate the Value and Growth factor legs using both mutual funds and ETFs. The left panel plots three cumulative return series: the benchmark Fama-French factor leg, the institutional Value synthetic portfolio, and the retail Value synthetic portfolio. The right panel plots the same cumulative return series for Growth. The sample period is from June 2004 to December 2019.

greater than \$1 billion) funds with the largest exposure to the various characteristic scores. As a consequence, the figure does not provide a time-varying proxy for the overall capacity of the main smart beta strategies, but it is nonetheless informative. In fact, it is clear that for all the various smart beta strategies, institutional investors have access to larger optimal funds, with the median fund having a size of around \$4-5 billion, in contrast to the \$2-3 billion median fund size available to retail investors. This suggests that on any given day, a minimum of \$40-50 billion capital can be quite easily deployed by institutional investors to chase the risk premia through an optimal, synthetic combination of funds.

To better understand how trading in funds can indirectly affect these smart beta strategies, Table 5 reports the average monthly turnover of our synthetic portfolios, while in Figure 9 we plot it over time. The turnover of the funds in the optimal synthetic portfolios is quite low. On average, across strategies and time, the monthly turnover is around 20% (e.g., two new funds out of the original ten). For the Small and Value strategies, the retail turnover is higher, suggesting the "good" institutional funds exploiting these strategies tend to be quite stable.

Lastly, we provide some estimates of the total capacity of smart beta strategies. In principle, this requires a trading model that takes into account transaction costs and trading impact, so that one can calculate the amount of AUM that makes the risk premia drop to zero. An approximate solution is to take the number of large funds investing in a specific

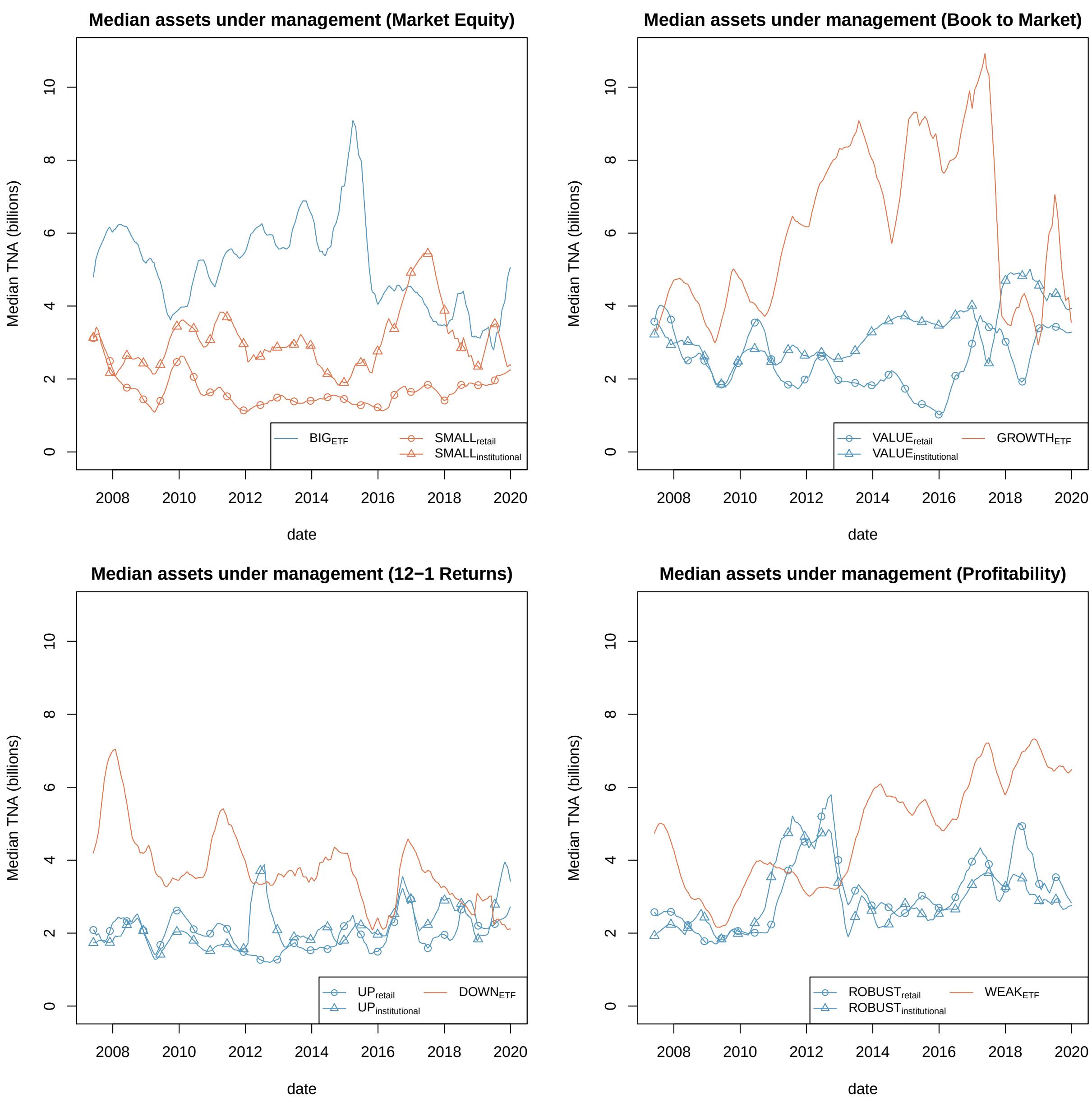


Figure 8: Median net asset value of the funds in the synthetic portfolios. The figure shows the time series of the (six-months smoothed) median NAV of the various synthetic portfolios for both retail and institutional investors. The short leg is constructed using only ETFs, available to both retail and institutional investors. The sample is from June 2007 to December 2019.

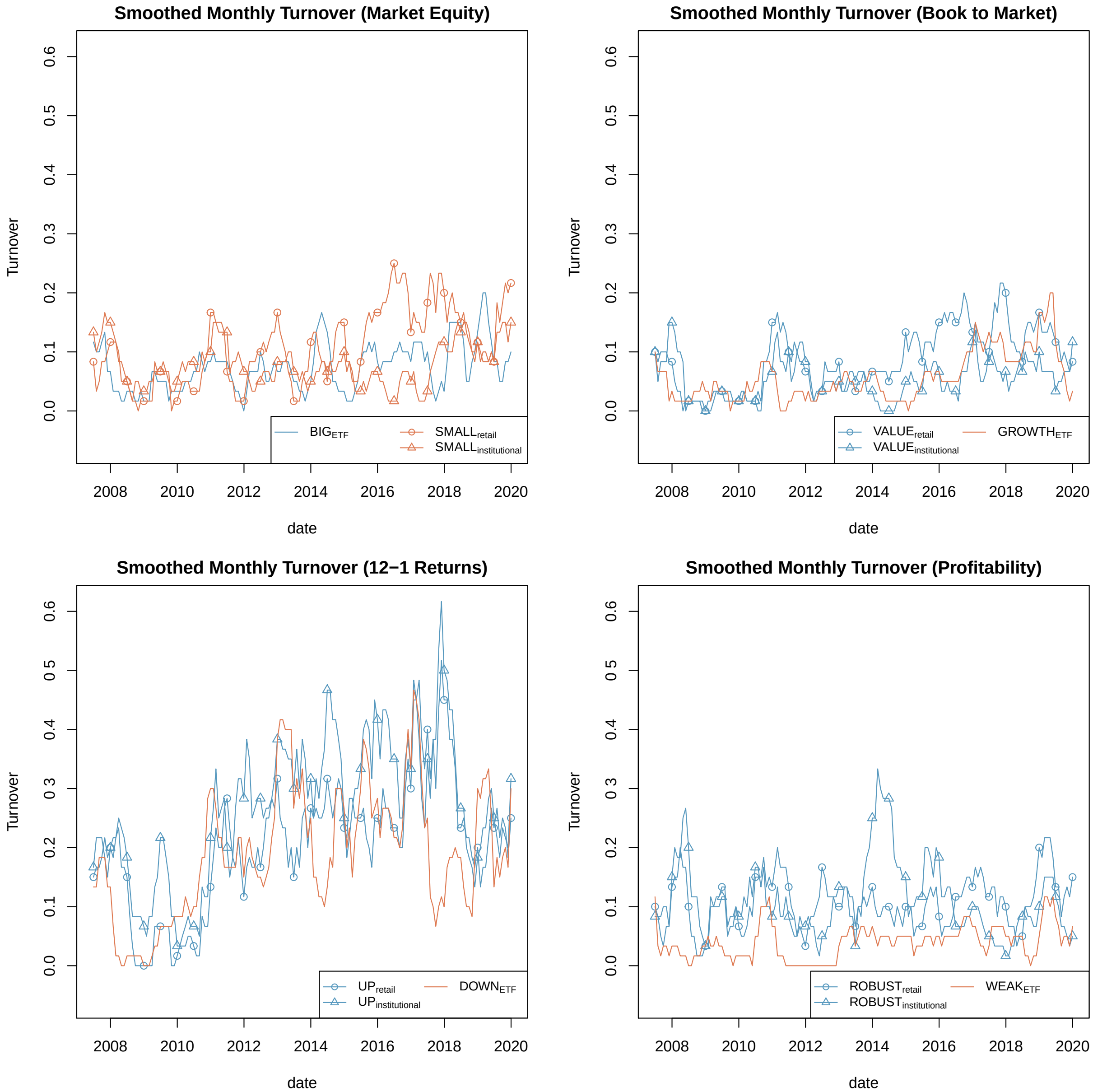


Figure 9: Monthly turnover of the synthetic portfolios. This figure shows the six-month smoothed turnover of the synthetic portfolios. The turnover is calculated as $TO = N_{new}/N_{tot}$. For example, if we replace three of the funds in our synthetic portfolio over the last month, the turnover will be 0.3 (since the synthetic portfolios are made up of ten funds in total). The sample is from June 2007 to December 2019.

strategy (e.g., around 300 funds for growth, 200 for value) from Figure 2 and assume that these funds will achieve the current median fund size in the optimal synthetic portfolios (e.g., around \$4-5 billion for Size, Value, and Profitability, and 2-3 billion for Momentum). A back-of-the-envelope calculation shows an overall capacity of around \$1.7 trillion for Size, \$2 trillion for Value and Growth, \$1 trillion for Profitability, and \$100 billion for Momentum, in line with the estimates of Ratcliffe et al. (2017).

6.4 Flows of Smart Beta Strategies

In this section, we examine whether the flows to the smart beta strategies are coming through funds that *in theory* should be tracking a strategy, versus funds that *in practice* track a strategy. In order to do so, we focus on the *daily* flows to smart beta strategies based on a naive classification (i.e., by looking at the funds' names) and smart classification (i.e., using the funds selected in our synthetic portfolios). Having access to *daily* flow data, in contrast to the standard monthly or quarterly flow data that is frequently used in the literature, allows us to highlight the impact of noise (or "technical analysis") traders from more professional, long-term investors.

We run the following regression:

$$F^{(1)}_{S,i,t+1} = \gamma_0 + \gamma_1 r^{(1)}_{S,i,t} + \gamma_5 r^{(5)}_{S,i,t} + \gamma_{22} r^{(22)}_{S,i,t} + \phi_1 F^{(1)}_{S,i,t} + \phi_5 F^{(5)}_{S,i,t} + \phi_{22} F^{(22)}_{S,i,t} + \theta_{1,i} r^{(1)}_{bench,t} + \theta_{5,i} r^{(5)}_{bench,t} + \theta_{22,i} r^{(22)}_{bench,t} + \varepsilon_{S,i,t+1} \quad (5)$$

where S = {Value, Growth, Small cap, Large cap, Momentum (long)}, i = {synthetic, by-names}, $r_{S,i,t}$ are the strategy S returns using the i construction, $F_{S,i,t}$ are the strategy S flows using the i construction, $r_{bench,t}$ is the return on a public smart beta strategy index (e.g., MSCI Value index[37]), and $r^{(k)}_t = \frac{1}{k}\sum_{l=0}^{k-1} r_{t-l}$ and $F^{(k)}_t = \frac{1}{k}\sum_{l=0}^{k-1} F_{t-l}$,

This specification is a generalization of the one in Corsi (2009), and allows us to identify behavioral effects in the trading of smart beta strategies by decomposing past returns and flows into daily, weekly and monthly blocks (1-, 5-, and 22-days). Given this specification, we test two economic hypotheses.

Our first hypothesis is that previous short-run fund returns within a strategy should lead to more positive inflows to the funds, regardless of whether the smart beta strategy portfolio

[37] We use the Fama-French strategy as a proxy, similarly to Cooper et al. (2005). The two are very highly correlated.

is constructed using fund names or synthetically:

Hypothesis 1. ***(Positive return-flows relationship.)*** *Positive short-run returns of smart beta funds should generate positive inflows into the same smart beta funds, regardless of whether the smart beta strategy is constructed naively (e.g., by names), or synthetically:*

$$H_1 : \gamma_1 > 0 \textit{ and } \gamma_5 > 0 \textit{ and } \gamma_{22} > 0.$$

Our second hypothesis is that there is a large set of investors who are unsophisticated. They observe the market returns of public proxies for smart beta strategies (e.g., MSCI Value Index) and trade as short-run momentum traders. In other words, this type of investors would naively trade funds based on their names after observing a positive return to the public smart beta strategy index. Contrary to this, the same effect for synthetic smart funds should be smaller, because unsophisticated (e.g., noise) traders would not select the "true" smart beta funds:

Hypothesis 2. ***(Naive investors invest in funds based on their names.)*** *Naive investors behave as short-run momentum traders based on the past returns of a proxy for smart beta strategies and invest in smart beta funds based on their names. This effect will be smaller for funds which synthetically replicate a strategy, traded by smart investors:*

$$H_{2a} : \theta_{1,by\_names} > 0; \theta_{5,by\_names} > 0$$

$$H_{2b} : \theta_{1,by\_names} > \theta_{1,synth}; \theta_{5,by\_names} > \theta_{5,synth}; \theta_{22,by\_names} > \theta_{22,synth}$$

Table 6 and Table 7 report the results for retail and institutional investors, respectively. Each pair of columns refers to a particular leg: Small and Big, Value and Growth, and Up and Robust.[38] Even columns report the results for a by-name replication, while odd columns report results for our synthetic smart beta strategies based on the methodology presented in Section 5.4.[39] Even columns tend to display higher R^2 than odd columns. In other

[38]Momentum "by name" has fewer than ten funds in the sample during the first few years. However, this makes the identification of flows into the momentum strategy even clearer since naive investors only have fewer funds to get exposed to the strategy.

[39]We do not separate retail and institutional investors' flows because ETFs are traded by both, and there is no clear way that we are aware of to identify their relative flows.

words, flows to "naive" smart beta funds are more predictable than flows to "sophisticated" funds that track well a certain risk factor but do not necessarily mention that factor in their name.[40]

The first hypothesis seems to hold, perhaps unsurprisingly, for momentum strategies (column 10), consistent with a momentum effect itself (e.g., investor flows follow momentum in returns). What is more surprising is that the effect is confined to flows into naive, not synthetic, momentum strategies for retail investors, and it is similarly strong for the naive momentum portfolio (22-days momentum) in the case of institutional investors. There is no clear return-flow pattern for growth and value strategies for either class of investors, while there is some evidence of momentum in the profitability strategies for retail investors.

Focusing on the second hypothesis, we see that the performance of the benchmark tends to be associated with greater inflows for the funds in the *by names* portfolio for Small and Big (independently from the class of investors), and also for Growth in the case of retail investors. In line with our hypothesis, for these strategies we find that $\theta_{22,by\_names} > \theta_{22,synth}$.[41]

We also note that, for both types of investors, the last month flows are strongly related to the future flows for most strategies, especially those based on the fund names. In fact, the effect is always larger for the "by-names" strategies, with the exception of Growth, for both types of investors. This suggests that investors tend to pay attention to smart beta funds according to their names, look at their past flows, and tend to invest in funds that have the highest flows over the previous month.

Lastly, the average daily flows are positive for most naive ("by name") strategies, and slightly negative for the true synthetic ones, suggesting that investors pour money into the former rather than investing in the optimal factor tradable portfolios.

Overall, we find some support for the second hypothesis: both types of investors react, at least partially, to the performance of the smart beta benchmark, but mainly use fund names as an indication of which fund strategy to buy after observing the return of the benchmark. Our conclusions are in line with the findings in Cooper et al. (2005), Chen et al. (2020) and Rakowski and Wang (2009).

[40]The notable exceptions are the Value and Growth legs.

[41]The loadings on Small for retail investors, and on Growth for institutional investors are however not statistically significant.

7 Replicating Factors "by Name"

A natural question is whether investors can simply replicate the main non-tradable smart beta indices using a portfolio of funds selected based on their names. The main issue with the "by name" approach is that it is feasible only for the Small/Big, Value/Growth, and Quality legs, separately, and only using mutual funds at the beginning of the sample, since (i) there are not enough funds whose names suggest the tracking of other smart beta strategies (e.g., momentum) and (ii) smart beta ETFs become available mainly after 2013. These two points together imply that implementing long-short factors using funds' names over our full sample is feasible only for SMB and HML factors.

In contrast, our methodology allows us to construct replicating tradable strategies for *any* non-tradable long-short factor or benchmark (e.g., momentum, idiosyncratic volatility) over *any* sample period, insofar the underlying assets of the benchmark can be retrieved. Moreover, by construction, the characteristic scores of our synthetic portfolios are always closer to the ones of the underlying benchmarks, meaning that if investors want to get exposed to a certain characteristic through a smart beta benchmark, our approach provides the best statistical viable solution (e.g., Section 6.1, and Table 4 in particular).

Despite these drawbacks, we next evaluate the performance of the "by Name" replicating portfolios, and compare them to the performance attained using our synthetic approach based on characteristic scores. Specifically, for each strategy, we classify funds using the set of keywords presented in Appendix B.3 (e.g., "value", "book", and "low p/e" for value funds). Then, we form the value-weighted portfolio of the largest ten (or fewer) funds[42] that are classified as members of a specific strategy to be consistent with our synthetic portfolio. The outcome is the "by Name" portfolio. Table 8 reports the results for the SMB and HML factors.

Two facts emerge by comparing the results of our synthetic approach (Table 3) to the "by Name" replication in Table 8. First, with respect to SMB, we find that the "by Name" replication works very satisfactorily: the alphas are below 50 bps and the betas close to one, similar to what is reported in Table 3 for institutional investors and using ETFs. The performance of SMB_{byName} is not surprising, since the strategy entails going long index

[42] In some periods at the beginning of our sample, there are fewer than ten funds available in some of the legs.

funds tracking small cap stocks (e.g., Russell 2000 funds), and short liquid, large cap funds. Second, and more importantly, we see that the implementation HML_{byName} leaves a much larger alpha relative to our synthetic replication. This is independent from the investor type: for example, institutional investors can reduce their alpha from 2.83% to 0.94% when moving from a naive replication "by Name" to a synthetic replication based on the characteristic score. We also observe a beta that is close to one in Table 3 but significantly larger than one in Table 8, suggesting a leveraged position.

Figure 10 provides graphical evidence. In particular, each panel shows the returns to our synthetic portfolios plotted against those of the Fama-French benchmark. Each panel also overlays the scatter plot of the naive, "by Name", version against the benchmark factor returns. The left (right) panel refers to SMB (HML). The better performance of our synthetic replication for HML is apparent. Also, the variance of the "by Name" portfolios is 50% smaller than the one for the synthetic portfolios, resulting in betas for the "by Name" strategies that are large at approximately 1.5, as shown in Table 8.

Overall, the difference between HML_{synth} (Table 3) and HML_{byName} (Table 8) is economically relevant. Thus, our procedure to construct synthetic factors is not a mere repackaging of the simpler "by Name" fund strategy. More aggressively, the failure of the "by Name" classification for HML suggests that funds' names and prospectuses may be misleading (Cooper et al. (2005),Chen et al. (2020)), and warrants a better way to classify funds in smart beta strategies. One of the contributions of our paper is to provide such a methodology.

8 Conclusion

Smart beta has become one of the most important investment topics over the last decade, as evidenced by the amount of assets under management flowing towards these strategies and the number of new funds tracking risk factors discovered in academic studies.

In this paper, we examine some fundamental, yet unanswered, research questions related to smart beta investing. First, do smart beta funds truly exist, or are they merely market funds? To address this question, we provide a formal framework to categorize funds into smart beta strategies. Funds are often categorized by investors according to their names and prospectuses, and many funds merely track the market (e.g., closet-indexing), regardless of what they claim. We propose a new methodology that relies on a *minimum distance*

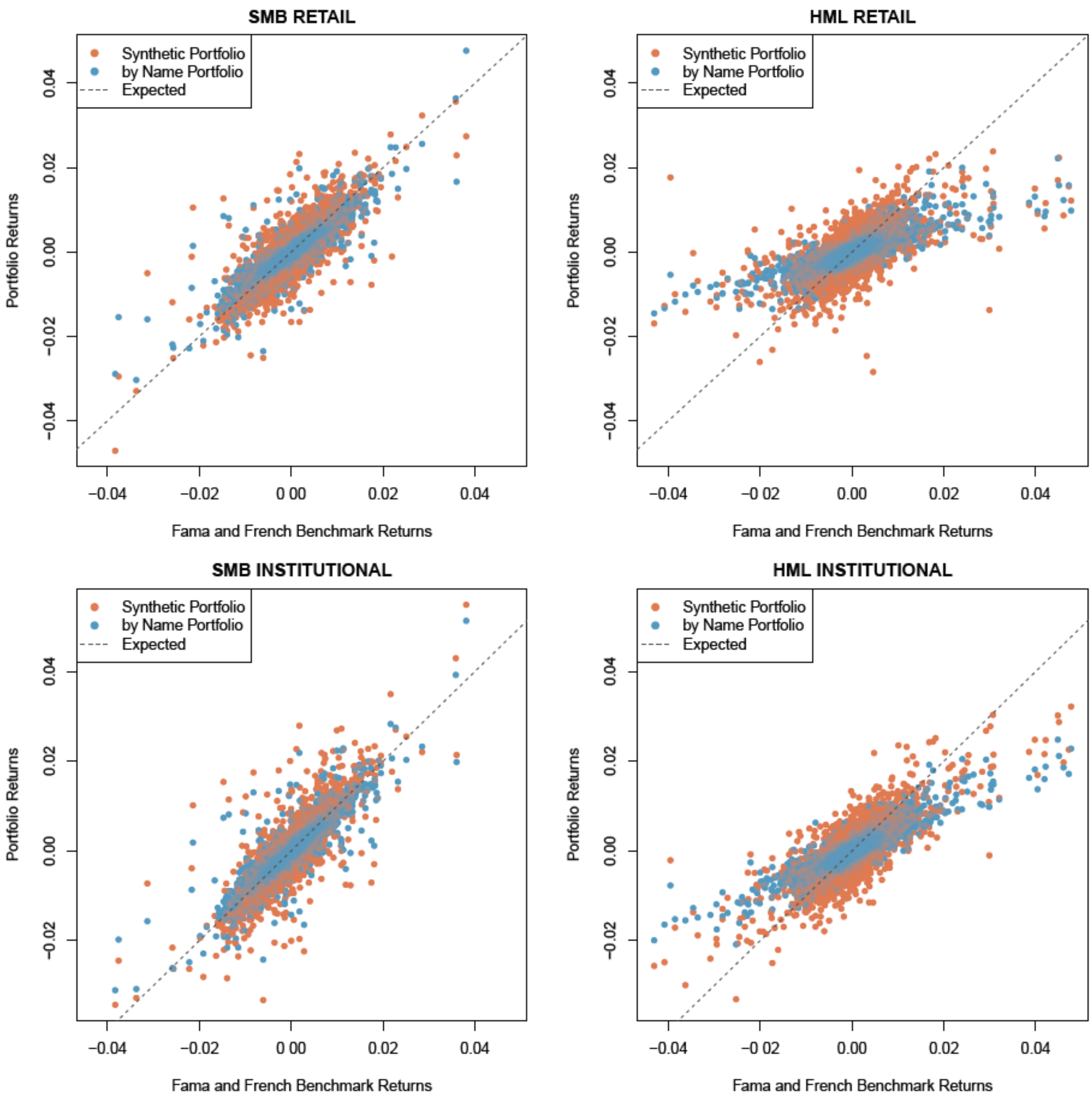


Figure 10: Comparison between the synthetic and "by Name" SMB and HML. The figure shows the scatterplot of the returns of our synthetic SMB (left panels) and HML (right panels) portfolios against the respective "by Name" version in replicating the Fama-French factors (the expected 45 degree line). The top (bottom) panels shows the results for the retail (institutional) investors. The sample period is from June 2007 to December 2019.

approach to identify market funds, and then we use the holdings of the remaining funds to sort them into smart beta strategies. We find that around one-half of the funds in our sample are simply market funds. Our results suggest that using fund names as a screening device is a suboptimal way to classify funds.

Given our screening and selection procedures, we can then study to what extent risk factors presented in academic studies are actually tradable, taking into account capacity constraints. To this end, we construct synthetic, *tradable* factors using an optimal combination (based on characteristic scores) of smart beta mutual funds and ETFs. We show that both retail and institutional investors are not able to synthetically trade standard risk factors, with the exception of the size premium (SMB). Interestingly, although, institutional investors are able to replicate the factors better than retail investors, as indicated by lower alphas from regressions of the benchmarks on our synthetic portfolios, but also by higher time series fit (R^2). This is particularly evident for the value (HML) premium. A potential explanation for this finding is that the characteristic score of the synthetic long leg of HML for institutional investors is closer to that of the Fama-French benchmark than the one synthesized by retail investors.

Lastly, we study the *daily* flows into synthetic and naive (e.g., based on fund names) smart beta strategies, and find that flows to the latter are more predictable than flows to sophisticated funds that optimally track the underlying characteristic of a risk factor but do not necessarily explicitly mention that characteristic in their names. We conclude that investors seem to allocate money into smart beta strategies based on fund names rather than their true factor exposure.

Overall, our results are consistent with a form of market incompleteness, meaning that investors cannot get exposure to the long leg of HML, or the short legs of MOM and RMW, when trading liquid financial instruments like mutual funds and ETFs. This implies that the investable set of strategies available to both retail and institutional investors may be smaller than previously thought.

Finally, our analysis is complementary to and has implications for the literature on the evaluation of portfolio managers (e.g., Berk and van Binsbergen (2015), Gerakos et al. (2020)) and cross-sectional return anomalies. In future work, we plan to use our methodology to provide additional insight on these topics.

References

Ang, Andrew, Robert J. Hodrick, Yuhang Xing, and Xiaoyan Zhang, 2006, The cross-section of volatility and expected returns, *The Journal of Finance* 61, 259–299.

Arnott, Rob, Vitali Kalesnik, and Lillian Wu, 2017, The incredible shrinking factor return, *Unpublished Manuscript* .

Asness, Clifford, and Andrea Frazzini, 2013, The devil in hml's details, *The Journal of Portfolio Management* 39, 49–68.

Asness, Clifford S., Andrea Frazzini, and Lasse Heje Pedersen, 2019, Quality minus junk, *Review of Accounting Studies* 24, 34–112.

Asness, Clifford S., Tobias J. Moskowitz, and Lasse H. Pedersen, 2013, Value and momentum everywhere, *The Journal of Finance* 68(3), 929–985.

Ben-Rephael, Azi, Shmuel Kandel, and Avi Wohl, 2012, Measuring investor sentiment with mutual fund flows, *Journal of Financial Economics* 104, 363–382.

Ben-David, Itzhak, Francesco Franzoni, and Rabih Moussawi, 2018, Do ETFs Increase Volatility?, *Journal of Finance* 73, 2471–2535.

Berk, Jonathan B., and Jules van Binsbergen, 2015, Measuring skill in the mutual fund industry, *Journal of Financial Economics* 118, 1–20.

Betermier, Sebastien, Laurent E. Calvet, and Paolo Sodini, 2017, Who Are the Value and Growth Investors?, *Journal of Finance* 72, 5–46.

Brown, David C., Shaun William Davies, and Matthew C. Ringgenberg, 2019, Etf flows, non-fundamental demand, and return predictability, *Working Paper* .

Chen, Huaizhi, Lauren Cohen, and Umit Gurun, 2020, Don't take their word for it: The misclassification of bond mutual funds, *Working Paper* .

Cooper, Michael J., Huseyin Gulen, and P. Raghavendra Rau, 2005, Changing names with style: Mutual fund name changes and their effects on fund flows, *The Journal of Finance* 60, 2825–2858.

Cooper, Michael J., Michael Halling, and Wenhao Yang, 2020, The persistence of fee dispersion among mutual funds, *Working Paper* .

Corsi, Fulvio, 2009, A Simple Approximate Long-Memory Model of Realized Volatility, *Journal of Financial Econometrics* 7, 174–196.

Coval, Joshua, and Erik Stafford, 2007, Asset fire sales (and purchases) in equity markets, *Journal of Financial Economics* 86, 479 – 512.

Cremers, K. J. Martijn, and Antti Petajisto, 2009, How Active Is Your Fund Manager? A New Measure That Predicts Performance, *The Review of Financial Studies* 22, 3329–3365.

Cremers, Martijn, Miguel A. Ferreira, Pedro Matos, and Laura Starks, 2016, Indexing and active fund management: International evidence, *Journal of Financial Economics* 120, 539 – 560.

Daniel, Kent, and Tobias J. Moskowitz, 2016, Momentum crashes, *Journal of Financial Economics* 122, 221 – 247.

Elton, Edwin J, Martin J Gruber, and Christopher R Blake, 1996, Survivorship Bias and Mutual Fund Performance, *Review of Financial Studies* 9, 1097–1120.

Elton, Edwin J., Martin J. Gruber, and Christopher R. Blake, 2001, A first look at the accuracy of the crsp mutual fund database and a comparison of the crsp and morningstar mutual fund databases, *The Journal of Finance* 56, 2415–2430.

Etula, Erkko, Kalle Rinne, Matti Suominen, and Lauri Vaittinen, 2019, Dash for Cash: Monthly Market Impact of Institutional Liquidity Needs, *The Review of Financial Studies* 33, 75–111.

Evans, Richard B., 2010, Mutual fund incubation, *Journal of Finance* 65, 1581–1611.

Fama, Eugene, and Kenneth French, 1992, The cross-section of expected stock returns, *The Journal of Finance* 47, 427–465.

Fama, Eugene F, and Kenneth R French, 1996, Multifactor Explanations of Asset Pricing Anomalies, *Journal of Finance* 51, 55–84.

Fama, Eugene F., and Kenneth R. French, 2006, The value premium and the capm, *The Journal of Finance* 61(5), 2163–2185.

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

Fama, Eugene F., and James D. MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of Political Economy* 81(3), 607–636.

Frazzini, Andrea, Ronen Israel, and Tobias Moskowitz, 2015, Trading costs of asset pricing anomalies, *Working Paper* .

Frazzini, Andrea, and Owen Lamont, 2008, Dumb money: Mutual fund flows and the cross-section of stock returns, *Journal of Financial Economics* 88, 299–322.

Frazzini, Andrea, and Lasse Heje Pedersen, 2014, Betting against beta, *Journal of Financial Economics* 111, 1 – 25.

Gerakos, Joseph, Juhani T. Linnainmaa, and Adair Morse, 2020, Asset managers: Institutional performance and factor exposures, *Journal of Finance* Forthcoming.

Gruber, Martin J, 1996, Another puzzle: The growth in activity managed mutual funds, *Journal of Finance* 51, 783–810.

Harvey, Campbell R., Yan Liu, and Heqing Zhu, 2016, ... and the cross-section of expected returns, *The Review of Financial Studies* 29(1), 5–68.

Hou, Kewei, Chen Xue, and Lu Zhang, 2015, Digesting anomalies: An investment approach, *The Review of Financial Studies* 28(3), 650–705.

Hou, Kewei, Chen Xue, and Lu Zhang, 2019, Replicating anomalies, *The Review of Financial Studies* Forthcoming.

Huang, Shiyang, Yang Song, and Hong Xiang, 2020, The smart beta mirage, *Working Paper* .

Hunter, David, Eugene Kandel, Shmuel Kandel, and Russ Wermers, 2014, Mutual fund performance evaluation with active peer benchmarks, *Journal of Financial Economics* 112(1), 1–29.

Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, *The Journal of Finance* 48(1), 65–91.

Kacperczyk, Marcin, Stijn Van Nieuwerburgh, and Laura Veldkamp, 2014, Time-Varying Fund Manager Skill, *Journal of Finance* 69, 1455–1484.

Kacperczyk, Marcin, and Amit Seru, 2007, Fund Manager Use of Public Information: New Evidence on Managerial Skills, *Journal of Finance* 62, 485–528.

Kacperczyk, Marcin, Clemens Sialm, and Lu Zheng, 2005, On the Industry Concentration of Actively Managed Equity Mutual Funds, *Journal of Finance* 60, 1983–2011.

Kacperczyk, Marcin, Stijn Van Nieuwerburgh, and Laura Veldkamp, 2016, A rational theory of mutual funds' attention allocation, *Econometrica* 84, 571–626.

Koijen, Ralph S. J., Tobias J. Moskowitz, Lasse H. Pedersen, and Evert B. Vrugt, 2018, Carry, *Journal of Financial Economics* 127(2), 197–225.

Koijen, Ralph S.J., 2014, The Cross-Section of Managerial Ability, Incentives, and Risk Preferences, *Journal of Finance* 69, 1051–1098.

Kosowski, Robert, Allan Timmermann, Russ Wermers, and Hal White, 2006, Can mutual fund "stars" really pick stocks? new evidence from a bootstrap analysis, *The Journal of Finance* 61, 2551–2595.

Lakonishok, Josef, Andrei Shleifer, and Robert W. Vishny, 1992, The impact of institutional trading on stock prices, *Journal of Financial Economics* 32, 23 – 43.

Leippold, Markus, and Roger Ruegg, 2019, Fama–french factor timing: The long-only integrated approach, *Working Paper* .

Lettau, Martin, Sydney Ludvigson, and Paulo Mandel, 2019, Chactarestics of mutual fund portfolios: where are the value funds?, *Working Paper* .

Linnainmaa, Juhani T, and Michael R Roberts, 2018, The History of the Cross-Section of Stock Returns, *The Review of Financial Studies* 31, 2606–2649.

Lou, Dong, 2012, A Flow-Based Explanation for Return Predictability, *The Review of Financial Studies* 25, 3457–3489.

Ma, Linlin, Yuehua Tang, and Juan-Pedro Gomez, 2019, Portfolio Manager Compensation in the U.S. Mutual Fund Industry, *Journal of Finance* 74, 587–638.

McLean, R. David, and Jeffrey Pontiff, 2016, Does academic research destroy stock return predictability?, *Journal of Finance* 71, 5–32.

Novy-Marx, Robert, and Mihail Velikov, 2016, A Taxonomy of Anomalies and Their Trading Costs, *Review of Financial Studies* 29, 104–147.

Pastor, Lubos, and Robert Stambaugh, 2002, Investing in equity mutual funds, *Journal of Financial Economics* 63, 351–380.

Pastor, Lubos, Robert F. Stambaugh, and Lucian Taylor, 2015, Scale and skill in active management, *Journal of Financial Economics* 116, 23–45.

Pastor, Lubos, Robert F. Stambaugh, and Lucian Taylor, 2017, Do funds make more when they trade more?, *The Journal of Finance* 72, 1483–1528.

Patton, Andrew J., and Brian M. Weller, 2020, What you see is not what you get: The costs of trading market anomalies, *Journal of Financial Economics* 137, 515 – 549.

Rakowski, David, and Xiaoxin Wang, 2009, The dynamics of short-term mutual fund flows and returns: a time-series and cross-sectional investigation, *Journal of Banking and Finance* 33, 2102–2109.

Ratcliffe, Ronald, Paolo Miranda, and Andrew Ang, 2017, Capacity of smart beta strategies from a transaction cost perspective, *The Journal of Index Investing* 8(3), 39–50.

Sirri, Erik R., and Peter Tufano, 1998, Costly search and mutual fund flows, *Journal of Finance* 1589–1622.

Teo, Melvyn, and Sung-Jun Woo, 2004, Style effects in the cross-section of stock returns, *Journal of Financial Economics* 74, 367 – 398.

Warther, Vincent A., 1995, Aggregate mutual fund flows and security returns, *Journal of Financial Economics* 39, 209–235.

Zheng, Lu, 1999, Is Money Smart? A Study of Mutual Fund Investors' Fund Selection Ability, *Journal of Finance* 54, 901–933.

	MF + ETF		MF		ETF	
	2004-2011	2012-2019	2004-2011	2012-2019	2004-2011	2012-2019
Number of Funds	1,121	1,577	1,080	1,456	41	121
Average AUM ($ millions)	3,940	4,934	3,884	4,578	5,237	9,259
Median AUM ($ millions)	1,757	2,118	1,708	2,024	3,543	4,772
Median fund age (in sample period)	3.643	5.151	3.516	5.141	4.849	5.175
Median number of holdings	114	124	112	119	424	375
Median return over S&P500 p.a. (%)	0.328	-0.907	0.252	-0.882	1.688	-0.932
Median FF5-factor α p.a. (%)	-0.461	-0.640	-0.497	-0.716	0.220	-0.348

Table 1: Descriptive statistics of the final sample. This table reports summary statistics of our final sample by fund type: mutual funds + ETFs (columns 1-2), mutual funds only (columns 3-4), and ETFs only (columns 5-6). The first (second) column in each set reports results from 2004 to 2011 (2012 to 2019). All funds have AUM greater than $1bln.

Panel A: Daily Estimates									
	"Mkt"	Small	Big	Value	Growth	Up	Down	Robust	Weak
α	0	0.00	0.00	0.00	0.00	0.01	0.01	-0.00	-0.00
MKT	1	1.01	1.01	1.02	1.02	1.06	1.06	1.02	1.02
SMB	0	**0.91**	**-0.09**	0.41	0.41	0.46	0.46	0.42	0.42
HML	0	0.09	0.09	**0.70**	**-0.30**	0.04	0.04	0.03	0.03
MOM	0	-0.01	-0.01	-0.02	-0.02	**0.35**	**-0.65**	-0.01	-0.01
RMW	0	-0.07	-0.07	-0.15	-0.15	-0.13	-0.13	**0.36**	**-0.64**

Panel B: Monthly Estimates									
	"Mkt"	Small	Big	Value	Growth	Up	Down	Robust	Weak
α	0	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
MKT	1	1.00	1.00	1.01	1.01	1.07	1.07	1.02	1.02
SMB	0	**0.92**	**-0.08**	0.43	0.43	0.48	0.48	0.45	0.45
HML	0	0.09	0.09	**0.67**	**-0.33**	-0.02	-0.02	-0.003	-0.003
MOM	0	-0.01	-0.01	-0.01	-0.01	**0.32**	**-0.68**	-0.02	-0.02
RMW	0	-0.06	-0.06	-0.12	-0.12	-0.09	-0.09	**0.40**	**-0.60**

Table 2: Population parameter estimates. This table reports full sample estimates of the time series regression

$$r_i^e = \alpha_i + \beta_{i,mkt} r_{mkt}^e + \beta_{i,smb} r_{smb} + \beta_{i,hml} r_{hml} + \beta_{i,mom} r_{mom} + \beta_{i,rmw} r_{rmw} + \varepsilon_i$$

where i denotes the individual legs of the factors. Panel A (Panel B) reports estimates using daily (monthly) returns. By construction, the total loading of each leg $(\beta_{i,long} - \beta_{i,short})$ on its own factor (highlighted in bold) must sum to one. The "MKT" column displays the theoretical loadings of a market portfolio on the different risk-factors. The sample period is from 2003 to 2019.

Panel A: Retail investors

				excl. "MKT" Funds		
	α	β	R^2	α	β	R^2
SMB_{synt}	1.34	0.77^{***}	0.67	1.55	0.88^{***}	0.77
HML_{synth}	1.48	0.88^{**}	0.41	1.98	0.96	0.60
MOM_{synth}	2.37	0.96	0.30	4.64	0.79^{***}	0.34
RMW_{synth}	3.36^{**}	0.30^{***}	0.23	0.85	0.28^{***}	0.17

Panel B: Institutional investors

				excl. "MKT" Funds		
	α	β	R^2	α	β	R^2
SMB_{synth}	0.52	0.70^{***}	0.70	0.80	0.87^{***}	0.82
HML_{synth}	0.94	0.94	0.60	1.14	0.95^{**}	0.64
MOM_{synth}	1.88	0.98	0.30	4.71	0.72^{***}	0.31
RMW_{synth}	3.11^{**}	0.30^{***}	0.22	0.81	0.29^{***}	0.18

Panel C: ETFs only

				excl. "MKT" Funds		
	α	β	R^2	α	β	R^2
SMB_{synth}	0.12	0.74^{***}	0.70	0.59	0.92^{***}	0.84
HML_{synth}	1.94	1.19^{***}	0.63	1.01	0.93^{***}	0.54
MOM_{synth}	1.29	0.95	0.21	4.49	0.89^{*}	0.29
RMW_{synth}	3.05^{**}	0.31^{***}	0.22	0.75	0.28^{***}	0.14

Table 3: Synthetic tradable risk factors against factor mimicking portfolios. This table reports the performance of long-short synthetic (tradable) factors from the regressions

$$ret_{FF_{LS},t} = \alpha + \beta ret_{synthetic_{LS},t} + \varepsilon_t$$

Panel A (Panel B) reports results for retail (institutional) investors using mutual funds and ETFs on the long leg, and ETFs on the short leg. The quadrant *excl. "MKT" funds* excludes from the analysis those funds classified as "market" funds by our minimum distance approach, i.e., funds with the lowest F-test statistic when compared to the expected MKT parameters are excluded. Panel C shows the results only using ETFs (available to both retail and institutional investors) on both legs. The tests on the coefficients are H_0: $\alpha_i = 0$ and H_0: $\beta_i = 1$. The left quadrant uses data from June 2007 to December 2019, while the right quadrant - where we exclude the market funds - uses data starting in August 2010. *p<0.1; **p<0.05; ***p<0.01.

Panel A: Retail investors

	MKT_{FF}	SMB_{FF}	HML_{FF}	MOM_{FF}	RMW_{FF}	SMB_{synt}	HML_{synt}	MOM_{synt}	RMW_{synt}
E(r)	0.095	0.001	-0.024	0.016	0.031	-0.017	-0.044	-0.008	-0.010
Std(r)	0.197	0.092	0.111	0.160	0.060	0.099	0.080	0.092	0.096
Skewness(r)	-0.182	0.067	0.625	-0.770	-0.174	0.021	0.186	-0.558	-0.134
Exc. Kurtosis(r)	9.573	4.044	8.733	9.289	2.975	2.807	2.384	3.031	3.144
SR(r)	0.481	0.006	-0.217	0.102	0.508	-0.170	-0.547	-0.084	-0.103
Size		0.390	0.943	1.022	1.131	0.617	0.774	1.034	1.436
B/M		1.037	5.616	0.930	0.766	1.627	2.390	0.792	0.678
Momentum		1.011	0.944	4.530	1.030	0.969	0.897	1.445	1.017
Profitability		0.651	0.810	1.053	4.795	0.739	0.760	1.022	1.389

Panel B: Institutional investors

	MKT_{FF}	SMB_{FF}	HML_{FF}	MOM_{FF}	RMW_{FF}	SMB_{synt}	HML_{synt}	MOM_{synt}	RMW_{synt}
E(r)	0.095	0.001	-0.024	0.016	0.031	-0.007	-0.035	-0.003	-0.002
Std(r)	0.197	0.092	0.111	0.160	0.060	0.110	0.091	0.090	0.095
Skewness(r)	-0.182	0.067	0.625	-0.770	-0.174	0.252	0.356	-0.615	-0.135
Exc. Kurtosis(r)	9.573	4.044	8.733	9.289	2.975	3.312	3.755	4.222	3.292
SR(r)	0.481	0.006	-0.217	0.102	0.508	-0.061	-0.389	-0.029	-0.017
Size		0.390	0.943	1.022	1.131	0.625	0.813	1.063	1.439
B/M		1.037	5.616	0.930	0.766	1.608	2.533	0.782	0.670
Momentum		1.011	0.944	4.530	1.030	0.993	0.939	1.402	1.039
Profitability		0.651	0.810	1.053	4.795	0.747	0.786	1.051	1.383

Table 4: Moments of long-short factors. This table reports summary statistics of the Fama French factors together with those of our synthetic tradable portfolios. Our synthetic portfolios include both mutual funds and ETFs in the long leg and only ETFs in the short leg to ensure tradability. Panel A (Panel B) reports the results for synthetic portfolios available to retail (institutional) investors. The bottom half of each panel shows, for each factor, the ratio of the characteristic scores of the long and short legs, normalized between zero and one, with respect to the various characteristics for both the Fama and French and our synthetic portfolios. The sample period is from June 2007 to December 2019.

Panel A: Retail investors								
	Small	Big	Value	Growth	Up	Down	Robust	Weak
Mean	0.11	0.14	0.09	0.12	0.21	0.24	0.11	0.09
Std	0.12	0.15	0.10	0.12	0.21	0.24	0.11	0.11
Min	0	0	0	0	0	0	0	0
q0.25	0	0	0	0	0	0	0	0
Median	0.10	0.10	0.10	0.10	0.20	0.20	0.10	0.10
q0.75	0.20	0.20	0.10	0.20	0.40	0.40	0.20	0.10
Max	0.70	0.60	0.40	0.40	0.90	1	0.40	0.60

Panel B: Institutional investors								
	Small	Big	Value	Growth	Up	Down	Robust	Weak
Mean	0.08	0.13	0.06	0.14	0.27	0.25	0.11	0.09
Std	0.09	0.14	0.09	0.14	0.25	0.24	0.13	0.10
Min	0	0	0	0	0	0	0	0
q0.25	0	0	0	0	0	0.05	0	0
Median	0.10	0.10	0	0.10	0.20	0.20	0.10	0.10
q0.75	0.10	0.20	0.10	0.20	0.40	0.40	0.20	0.10
Max	0.50	0.50	0.50	0.50	0.90	1	0.50	0.40

Table 5: Monthly Turnover. This table reports summary statistics of the monthly turnover of the long and short legs of our synthetic portfolios. Panel A (Panel B) shows the summary statistics for retail investors (institutional investors). The turnover is calculated, each month, as $TO = N_{new}/N_{tot}$. For example, an average turnover of 20% implies that two funds are replaced, on average, each month.

	Small (S)	Small (N)	Big (S)	Big (N)	High (S)	High (N)	Low (S)	Low (N)	Up (S)	Up (N)	Robust (S)	Robust(N)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(S) 1-day ret	0.122		0.151***		0.015		0.025		0.020		0.054**	
	(0.186)		(0.031)		(0.022)		(0.022)		(0.019)		(0.024)	
(S) 5-days avg ret	−0.710		0.042		0.072		−0.032		0.029		0.015	
	(0.504)		(0.085)		(0.059)		(0.058)		(0.051)		(0.065)	
(S) 22-days avg ret	−0.277		0.441**		−0.133		0.145		−0.027		0.271**	
	(1.053)		(0.186)		(0.124)		(0.122)		(0.100)		(0.133)	
(S) 1-day flow	0.051**		0.048**		−0.097***		−0.105***		−0.139***		0.060***	
	(0.020)		(0.020)		(0.019)		(0.020)		(0.020)		(0.020)	
(S) 5-days flow	−0.131***		−0.026		−0.394***		−0.093		−0.111*		−0.181***	
	(0.048)		(0.047)		(0.062)		(0.057)		(0.057)		(0.049)	
(S) 22-days flow	−1.015***		−0.006		0.615***		0.511***		0.348***		0.044	
	(0.143)		(0.094)		(0.111)		(0.091)		(0.099)		(0.103)	
(N) 1-day ret		−0.237		0.022		0.026		−0.001		−0.109		0.011
		(0.148)		(0.028)		(0.020)		(0.018)		(0.138)		(0.019)
(N) 5-days avg ret		−0.908**		0.066		0.053		0.035		0.072		-0.036
		(0.394)		(0.075)		(0.054)		(0.049)		(0.359)		(0.051)
(N) 22-days avg ret		−0.679		−0.334**		−0.126		−0.191*		2.847***		0.172*
		(0.792)		(0.159)		(0.105)		(0.102)		(0.784)		(0.097)
(N) 1-day flow		0.030		0.031		−0.031		−0.020		−0.230***		0.148***
		(0.020)		(0.020)		(0.019)		(0.020)		(0.034)		(0.019)
(N) 5-days flow		−0.172***		0.102**		−0.457***		−0.155***		−0.446***		−0.451***
		(0.048)		(0.047)		(0.059)		(0.054)		(0.126)		(0.050)
(N) 22-days flow		−1.199***		0.297***		0.652***		0.469***		0.605***		0.294***
		(0.153)		(0.073)		(0.108)		(0.089)		(0.202)		(0.112)
benchmark 1-day ret	−0.118	0.237	−0.116***	−0.018	−0.011	−0.018	−0.021	0.004	−0.015	0.070	−0.041**	−0.015
	(0.186)	(0.147)	(0.027)	(0.025)	(0.018)	(0.016)	(0.021)	(0.017)	(0.018)	(0.124)	(0.020)	(0.017)
benchmark 5-days ret	0.564	0.769**	−0.005	−0.021	−0.007	−0.017	0.057	0.010	−0.0004	0.226	0.030	0.060
	(0.491)	(0.383)	(0.071)	(0.068)	(0.049)	(0.043)	(0.054)	(0.046)	(0.049)	(0.317)	(0.054)	(0.044)
benchmark 22-days ret	0.425	0.814	−0.405***	0.257*	0.145	0.096	−0.169	0.158*	0.070	−1.497**	−0.242**	−0.130
	(1.010)	(0.757)	(0.157)	(0.146)	(0.104)	(0.083)	(0.113)	(0.095)	(0.096)	(0.639)	(0.107)	(0.082)
Constant	0.001*	0.001**	0.0003***	0.0002***	−0.0001**	0.0001	−0.0001**	0.0001	−0.0002***	0.001	0.0002**	0.0001
	(0.0003)	(0.0002)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.00004)	(0.0001)	(0.001)	(0.0001)	(0.0001)
Observations	3,122	3,122	3,122	3,122	3,122	3,122	3,122	3,122	3,122	982	3,122	3,122
Adjusted R^2	0.027	0.038	0.016	0.024	0.037	0.026	0.018	0.017	0.026	0.096	0.011	0.035

Table 6: Flows to smart beta strategies: retail investors. This table reports estimates of the regression

$$F^{(1)}_{S,i,t+1} = \gamma_0+\gamma_1 r^{(1)}_{S,i,t}+\gamma_5 r^{(5)}_{S,i,t}+\gamma_{22} r^{(22)}_{S,i,t}+\phi_1 F^{(1)}_{S,i,t}+\phi_5 F^{(5)}_{S,i,t}+\phi_{22} F^{(22)}_{S,i,t}+\theta_{1,i} r^{(1)}_{bench,t}+\theta_{5,i} r^{(5)}_{bench,t}+\theta_{22,i} r^{(22)}_{bench,t}+\varepsilon_{S,i,t+1}$$

where S = {value, growth, small cap, large cap, momentum (long)} denotes the smart beta strategies, i = {synthetic, by-names} denotes the synthetic (S) or naive (N) – based on fund names – strategy, $r_{S,i,t}$ are the strategy S returns using the i construction, $F_{S,i,t}$ are the strategy S flows using the i construction, $\mathrm{r}_{bench,t}$ is the return on a public smart beta strategy index, and $r^{(k)}_t = \frac{1}{k}\sum_{l=0}^{k-1} r_{t-l}$, and $F^{(k)}_t = \frac{1}{k}\sum_{l=0}^{k-1} F_{t-l}$. The synthetic strategies are constructed as in Section 5, while the naive (N) strategies include the largest ten funds. The sample period is from June 2007 to October 2019. The momentum UP "by name" strategy starts in July 2015 and it never includes ten funds. *p<0.1; **p<0.05; ***p<0.01.

	Small (S)	Small (N)	Big (S)	Big (N)	High (S)	High (N)	Low (S)	Low (N)	Up (S)	Up (N)	Robust (S)	Robust(N)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(S) 1-day ret	0.198		0.151***		0.003		0.017		0.006		0.043*	
	(0.221)		(0.033)		(0.030)		(0.024)		(0.027)		(0.024)	
(S) 5-days avg ret	−1.180**		0.008		0.019		0.010		0.007		0.017	
	(0.587)		(0.090)		(0.078)		(0.064)		(0.072)		(0.067)	
(S) 22-days avg ret	−1.250		0.237		−0.110		0.134		0.324**		0.176	
	(1.211)		(0.194)		(0.159)		(0.137)		(0.154)		(0.138)	
(S) 1-day flow	0.058***		0.040**		−0.074***		−0.228***		−0.159***		0.070***	
	(0.020)		(0.020)		(0.019)		(0.019)		(0.020)		(0.020)	
(S) 5-days flow	−0.123***		−0.062		−0.391***		−0.378***		0.001		−0.131***	
	(0.047)		(0.048)		(0.062)		(0.065)		(0.056)		(0.048)	
(S) 22-days flow	−0.999***		0.081		0.553***		0.998***		0.292***		0.109	
	(0.142)		(0.095)		(0.109)		(0.104)		(0.096)		(0.097)	
(N) 1-day ret		−0.004		0.013		0.031*		0.024		−0.122		0.012
		(0.208)		(0.027)		(0.019)		(0.020)		(0.144)		(0.022)
(N) 5-days avg ret		−0.899		0.069		0.054		0.065		0.114		−0.089
		(0.549)		(0.071)		(0.052)		(0.052)		(0.367)		(0.062)
(N) 22-days avg ret		−2.271**		−0.329**		−0.124		−0.197*		2.711***		0.177
		(1.153)		(0.152)		(0.102)		(0.107)		(0.795)		(0.115)
(N) 1-day flow		0.042**		0.023		0.002		−0.065***		−0.232***		0.204***
		(0.020)		(0.020)		(0.019)		(0.020)		(0.035)		(0.019)
(N) 5-days flow		−0.167***		0.089*		−0.353***		0.024		−0.441***		−0.315***
		(0.048)		(0.047)		(0.056)		(0.052)		(0.128)		(0.048)
(N) 22-days flow		−1.118***		0.277***		0.660***		0.469***		0.735***		0.260***
		(0.149)		(0.075)		(0.095)		(0.081)		(0.196)		(0.096)
benchmark 1-day ret	−0.195	0.010	−0.122***	−0.011	0.0002	−0.017	−0.018	−0.018	−0.012	0.081	−0.024	−0.004
	(0.223)	(0.208)	(0.029)	(0.025)	(0.026)	(0.015)	(0.023)	(0.019)	(0.026)	(0.130)	(0.020)	(0.019)
benchmark 5-days ret	1.030*	0.741	0.030	−0.021	0.034	−0.022	0.059	−0.015	0.051	0.234	0.032	0.067
	(0.575)	(0.535)	(0.077)	(0.065)	(0.066)	(0.041)	(0.059)	(0.049)	(0.069)	(0.325)	(0.055)	(0.049)
benchmark 22-days ret	1.333	2.282**	−0.232	0.244*	0.099	0.072	−0.189	0.141	−0.269*	−1.438**	−0.188*	−0.135
	(1.166)	(1.100)	(0.166)	(0.141)	(0.134)	(0.078)	(0.125)	(0.100)	(0.145)	(0.657)	(0.109)	(0.088)
Constant	0.001***	0.001***	0.0003***	0.0002***	0.00003	0.0002***	−0.00000	0.0001*	0.00004	0.0004	0.0002***	0.0003***
	(0.0003)	(0.0003)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.00005)	(0.0001)	(0.001)	(0.0001)	(0.0001)
Observations	3,122	3,122	3,122	3,122	3,122	3,122	3,122	3,122	3,122	960	3,122	3,122
Adjusted R^2	0.028	0.035	0.010	0.020	0.027	0.020	0.082	0.024	0.027	0.097	0.011	0.038

Table 7: Flows to smart beta strategies: institutional investors. This table reports estimates of the regression

$$F^{(1)}_{S,i,t+1} = \gamma_0 + \gamma_1 r^{(1)}_{S,i,t} + \gamma_5 r^{(5)}_{S,i,t} + \gamma_{22} r^{(22)}_{S,i,t} + \phi_1 F^{(1)}_{S,i,t} + \phi_5 F^{(5)}_{S,i,t} + \phi_{22} F^{(22)}_{S,i,t} + \theta_{1,i} r^{(1)}_{bench,t} + \theta_{5,i} r^{(5)}_{bench,t} + \theta_{22,i} r^{(22)}_{bench,t} + \varepsilon_{S,i,t+1}$$

where $S = \{$value, growth, small cap, large cap, momentum (long)$\}$ denotes the smart beta strategies, $i = \{$synthetic, by-names$\}$ denotes the synthetic (S) or naive (N) – based on fund names – strategy, $r_{S,i,t}$ are the strategy S returns using the i construction, $F_{S,i,t}$ are the strategy S flows using the i construction, $\mathrm{r}_{bench,t}$ is the return on a public smart beta strategy index, and $r^{(k)}_t = \frac{1}{k}\sum_{l=0}^{k-1} r_{t-l}$, and $F^{(k)}_t = \frac{1}{k}\sum_{l=0}^{k-1} F_{t-l}$. The synthetic strategies are constructed as in Section 5, while the naive (N) strategies include the largest ten funds. The sample period is from June 2007 to October 2019. The momentum UP "by name" strategy starts in July 2015 and it never includes ten funds. The Big (Robust)"by Name" strategy reaches ten funds at the end of October 2007 (mid-December 2009). *p<0.1; **p<0.05; ***p<0.01.

Panel A: Retail investors			
	α	β	R^2
SMB_{byName}	0.52	1.02	0.79
HML_{byName}	3.77**	1.68***	0.66

Panel B: Institutional investors			
	α	β	R^2
SMB_{byName}	0.38	0.92***	0.80
HML_{byName}	2.83*	1.55***	0.78

Panel C: ETFs only			
	α	β	R^2
SMB_{byName}	0.15	0.91***	0.81
HML_{byName}	1.90	1.51***	0.76

Table 8: Tradable risk factors constructed using fund names. This table reports the results of the regression

$$ret_{factor_{LS},t} = \alpha + \beta ret_{byname_{LS},t} + \varepsilon_t$$

using tradable factors constructed using fund names. Panel A (Panel B) reports results for retail (institutional) investors using mutual funds and ETFs on the long leg, and ETFs on the short leg. Panel C shows the results only using ETFs (available to both retail and institutional investors) on both legs. We use any number of funds available for constructing the long-short legs, up to a maximum of ten, to be consistent with the synthetic strategies. The tests on the coefficients are H_0: $\alpha_i = 0$ and H_0: $\beta_i = 1$. The sample is from June 2007 to December 2019. *p<0.1; **p<0.05; ***p<0.01.

Internet Appendix A Additional Results and Robustness Checks

A.1 Sample Restrictions

A.1.1 $50 million minimum fund size

In our main analysis, we restrict the sample to funds with over $1 billion (in real terms) of assets under management. Table A.1 reports the results using funds with a minimum of $50 million of AUM. The original results in Panel A and B of Table 3 are confirmed to a large extent. Institutional investors are able to track SMB and HML better than retail investors. In fact, with respect to the synthetic SMB_{synth} and HML_{synth}, the gap between institutional and retail investors widens. This is mainly due to a worsening performance for retail investors with alphas that are larger. In general, momentum and profitability are not replicatable by either class of investors. Turning to Panel C of Table A.1, we confirm that the alpha on SMB_{synth} is economically small when we restrict our focus to ETFs only (as was the case in Table 3). However, differently from Table 3, in Table A.1 we observe that using ETFs to replicate HML_{synth} produces the lowest alpha. We conclude that the synthetic replication based on ETFs only is more sensitive to the sample restriction. Importantly though, the results suggest that retail investors may gain exposure to SMB and HML by trading ETFs.

We conclude with one observation. By expanding the set of available funds, the score of the optimal synthetic leg obtained with the $50 million restriction is higher than the score implied by the $1 billion restriction and, hence, closer to the Fama-French one. However, the results in Table A.1 suggest that a higher synthetic portfolio score does *not* necessarily imply a return closer to the Fama-French ones (e.g., a fund potentially tries to minimize the tracking error with respect to the underlying non-tradable Fama-French factor by choosing other correlated stocks subject to feasibility constraints). In other words, it is still possible for the optimal synthetic portfolio to have a higher characteristic score, but a worse tracking error and larger alpha.

A.1.2 Excluding the Great Recession

Table A.2 reports our main result using a post-Great Recession sample, starting in July 2009.[1] The main results reported in Table 3 hold even in this shorter sample. In particular, we note that institutional investors continue to replicate all long-short factors better than retail investors, as confirmed by the lower alphas and larger R^2 for the SMB_{synth} and HML_{synth} factors. Interestingly, the replication of the MOM factor improves substantially when we only use ETFs, with an alpha of just 0.67%.

A.2 Alternative Factor Definitions

A.2.1 "Live" Factors

Table A.3 reports results for the synthetic tradable factors formed by sorting stocks (and hence funds) on the "live" characteristic, along the line of the factor implementation proposed by Asness and Frazzini (2013). In other words, we use the latest available accounting or market-based quantity available as of the day before the portfolio formation. The results broadly confirm the picture depicted by Table 3: the synthetic HML_{live} and SMB_{live} leave large alphas of more than 1.5% for retail investors. On the other hand, institutional investors are better able to get exposure to these factors; for example, the HML_{live} alpha halves to 70 bps and the R^2 increases from 40% to 60%. Finally, the results in Panel C confirm that the replication of SMB appears feasible using ETFs only, as confirmed by an alpha as low as 32 bps. Nevertheless, using ETFs result in an inferior replication of HML, suggesting that mutual funds in the Value leg are important for the synthetic replication that relies on funds with AUM greater than $1 billion.

A.2.2 CMA and Q-factors

Table A.4 reports results for synthetic tradable portfolios sorted on the CMA factor of Fama and French (2015), and two well known academic factors inspired by the neo-classical q-theory of investment (Hou et al., 2015): the return on a portfolio of high/low profitability (proxied by return on equity, ROE) stocks and the return on a portfolio of high/low in-

[1] The panels on the right "excl. MKT Funds" in Table 3 and Table A.2 are identical since we require three years of data to implement our minimum distance approach in real-time.

vestment stocks (I/A). Specifically, we follow Hou et al. (2015) in constructing the legs of these factors, the only difference being that our sorting procedure is univariate rather than trivariate as originally proposed by these authors.

The structure of Table A.4 is identical to Table 3. First, the results show that using using an alternative definition of profitability (ROE rather than Operating Profitability scaled by book equity as in Table 3) does not change our conclusion. In particular, we confirm that ROE cannot be replicated, as indicated by the large alphas of about 3% and by R^2s at most of 26%. Second, and contrary to profitability, the replication of the investment factor of Hou et al. (2015) proves (relatively more) feasible for institutional investors after removing the market funds: in this case, the alpha is only 54 bps. Using ETFs only and removing the market funds also provides a viable solution to synthesize the investment factor and results in an alpha of 57 bps. The CMA factor replication produces a slightly lower alpha but larger R^2 than the IA_{synth} factor for both types of investors.

A.2.3 QMJ

Another commonly used proxy for profitability is the quality-minus-junk (QMJ) factor of Asness et al. (2019). The last row of each panel in Table A.4 reports the results of our replicating exercise for this factor. The replication of the quality factor proves difficult with alphas at least as large as 3.50% and low R^2s, both for retail and institutional investors, and independently from the set of tradable assets considered. Note that our sample includes the AQR funds publicly available to retail and institutional investors, but it does not include the in-house managed ones; this may help explain the difference in performance when using our set of funds in the construction of the synthetic, tradable quality portfolios.

A.3 Individual Factor Legs

In addition to creating tradable factor risk premia for institutional and retail investors using mutual funds and ETFs, it might be important to understand whether and how well the individual long and short legs of the factors can be traded by the two types of investors.

Table A.5 addresses this question, by showing the performance of our synthetic factor legs, constructed using the approach described in Section 5.

Panel A (Panel B) displays the results when we employ only funds available to retail (institutional) investors. Each quadrant refers to a different universe of available assets. From left to right, we consider the set of mutual funds and ETFs, the set of ETFs only, and the set of mutual funds and ETFs purged of market funds (c.f., Section 5.2).

As in Section 6.1, we measure how close our synthetic legs mimic the Fama-French non-tradable ones by running a regression of the latter on the former:

$$ret_{FF_{leg},t} = \alpha + \beta ret_{synth_{leg},t} + \varepsilon_t \ .$$

Focusing on the universe of mutual funds and ETFs available to retail investors, we observe high R^2s and βs that are all close to one. Most importantly, we fail to reject the null hypothesis $H_0 : \alpha = 0$ for most strategies. The long leg of MOM (i.e., Up) is the sole exception: its alpha is statistically different from zero.

Despite the fact that (annualized) alphas are statistically insignificant for most legs, some of them are economically large (e.g., more than 1% per year), and generate substantial differences in cumulative returns.

Removing market funds improves the tracking ability of the Up synthetic leg, as shown by the decrease in the magnitude of alphas from 3.04% to 2.47% per year, although this leg remains hard to replicate synthetically.

Surprisingly, the replication of Small and Up improves substantially when we only use ETFs. The alpha of Small drops from 1.23% to 0.24%, while that of Up drops from 3.04% to 1.34%. We observe a similar improvement for the long leg of profitability: the alpha of Robust drops from 1.85% to 0.19%.[2] The fact that some legs such as Small and Robust can be replicated more precisely using a smaller universe of assets (ETFs only vs. ETFs and mutual funds) might be related to the different benchmarking and incentives (i.e., greater leeway) of mutual fund managers.[3]

Comparing Panel A to Panel B, we observe that the synthetic legs available to institutional investors leave, in general, lower alphas on the table: for example, the alphas of Up and Robust decrease from 3.04% to 2.28%, and from 1.85% to 0.97%, respectively. Whereas

[2]This is not due to the shorter sample available to study "ETFs only." When we restrict the mutual funds and ETFs to the same sample period (e.g., 2007 to 2019), the alphas of Small and Up from the combination of mutual funds and ETFs continue to be larger than those obtained using ETFs only.

[3]Small cap and quality investors have more discretion in the set of stocks to be purchased, given the large number of small caps and the different definitions of quality that various managers adopt as benchmarks, relative to passive, value-weighted ETFs.

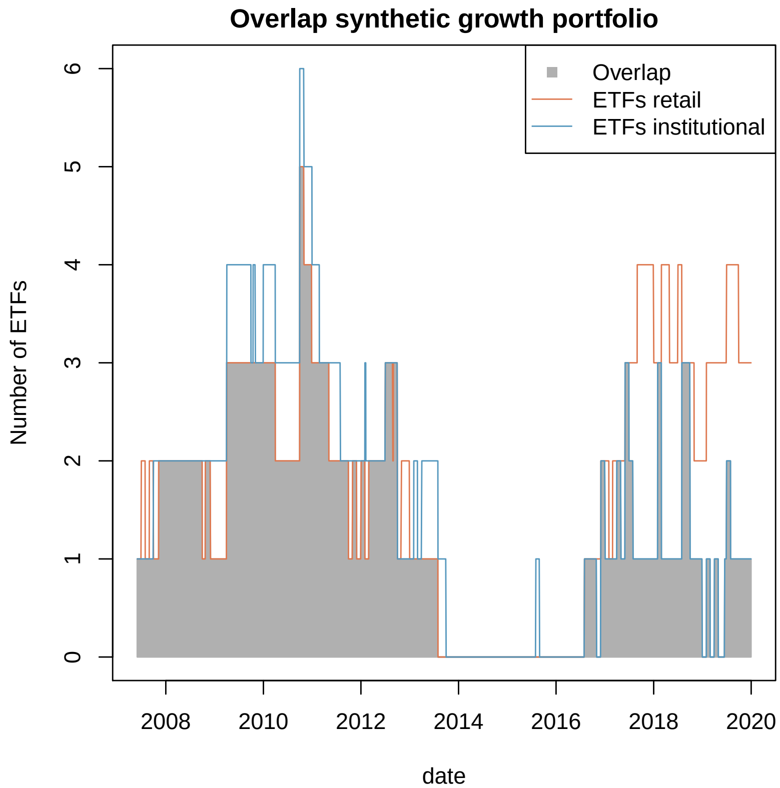
the long leg of HML has statistically insignificant alphas both for retail and institutional investors, we observe a larger R^2 of 96% for institutional investors. Finally, removing market funds has a slightly larger benefit for retail investors than for institutional investors.

A natural question is whether the different performance between institutional and retail legs replication is driven by the relative number of mutual funds and ETFs selected in the optimal synthetic portfolios. To investigate this point, Figure A.1 shows the number of ETFs in the synthetic Growth and Value portfolios.

Two observations stand out. First, there is an intriguing pattern: independently from the strategy considered, the number of ETFs selected in the synthetic portfolio is larger for institutional investors than for retail investors in the sample up to 2012; this pattern is then reversed in the latter part of the sample with more ETFs entering the synthetic retail portfolio relative to institutions. Second, there is a large overlap of ETFs chosen (the gray area coincides almost always with the minimum number of ETFs held by institution or retail investors). Overall, we conclude that the retail and institutional implementations of growth and value agree most of the time on the type but not on the number of ETFs chosen by the methodology.[4]

Lastly, from a theoretical standpoint, one might be interested in relating the underperformance of value (e.g., a negative value premium) over the last decade, with the corresponding surge in the amount of growth funds (e.g., the short leg of HML) available to investors. There could be several potential explanations for the latter, both behavioral and rational. Behavioral explanations include "herding behavior" by investors, e.g., *keeping up with the Joneses* – if all other investors are exposed to growth, you should be too – or "FAANG" hype, e.g., it is desirable to invest in public growth stocks because with some positive probability some of them will become the future Google or Amazon. Rational explanations might include (i) the high private market valuations of growth firms pushing the valuation of all public growth firms high up; (ii) the low interest rate environment of the last decade (e.g., when it is easy and cheap to borrow, desperate-for-yield investors will shift their asset allocation towards riskier equity); and (iii) *life-cycle dynamics*, e.g., it is optimal for young individuals to invest in growth firms, as discussed Betermier et al. (2017).

[4] This holds true for the other factors as well.



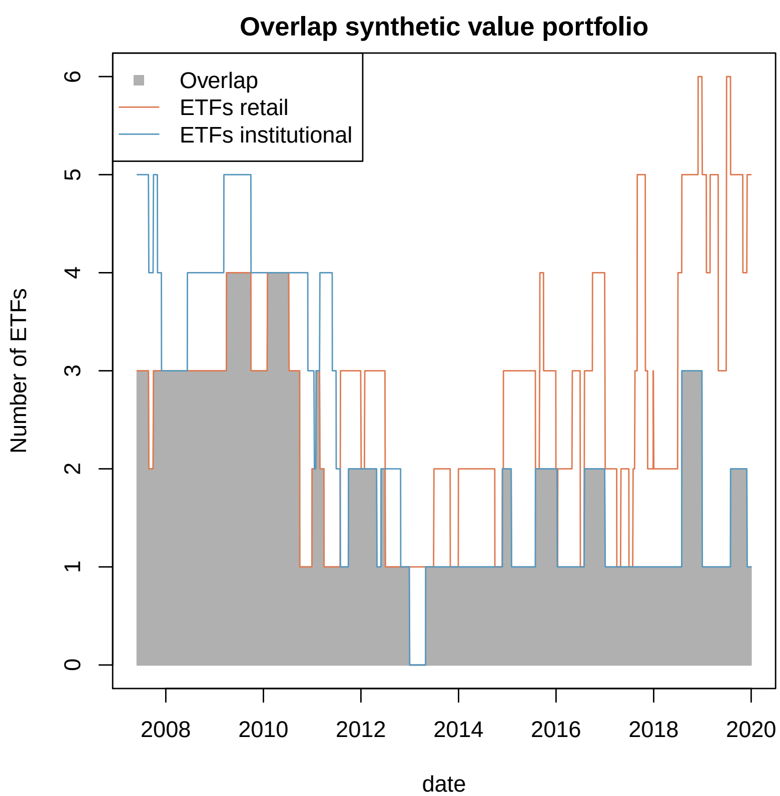


Figure A.1: ETF holdings comparison. The top figure plots the number of ETFs used in the synthetic portfolio when replicating the growth (factor) leg, split between funds available to retail and institutional investors. The bottom figure plots the number of ETFs used in the synthetic portfolio when replicating the value (factor) leg. The sample period is from June 2007 to December 2019.

Panel A: Retail investors						
				excl. "MKT" Funds		
	α	β	R^2	α	β	R^2
SMB_{synt}	1.77	0.70***	0.73	1.10	0.71***	0.75
HML_{synt}	1.64	0.62***	0.42	2.01	0.63***	0.59
MOM_{synt}	2.85	0.59***	0.22	5.35*	0.61***	0.40
RMW_{synt}	2.92**	0.23***	0.20	0.86	0.26***	0.20

Panel B: Institutional investors						
				excl. "MKT" Funds		
	α	β	R^2	α	β	R^2
SMB_{synth}	0.53	0.64***	0.76	0.26	0.71***	0.79
HML_{synth}	0.90	0.75***	0.54	0.65	0.58***	0.51
MOM_{synth}	3.63	0.70***	0.27	6.17**	0.57***	0.38
RMW_{synth}	2.91**	0.22***	0.19	0.74	0.27***	0.21

Panel C: ETFs only						
				excl. "MKT" Funds		
	α	β	R^2	α	β	R^2
SMB_{synth}	0.39	0.62***	0.69	-0.001	0.73***	0.78
HML_{synth}	0.38	0.72***	0.51	0.71	0.60***	0.50
MOM_{synth}	3.50	0.83***	0.30	5.30**	0.72***	0.41
RMW_{synth}	2.84*	0.23***	0.20	0.89	0.23***	0.16

Table A.1: Synthetic tradable risk factors - \$50-mln minimum. This table reports the performance of long-short synthetic "live" factors from the regression

$$ret_{FF_{LS},t} = \alpha + \beta ret_{synthetic_{LS},t} + \varepsilon_t$$

Panel A (Panel B) reports results for retail (institutional) investors using mutual funds and ETFs on the long leg, and ETFs on the short leg with at least \$50-mln AUM. The quadrant *excl. "MKT" funds* excludes from the analysis those funds classified as "market" funds by our minimum distance approach, i.e., funds with the lowest F-test statistic when compared to the expected MKT parameters are excluded. Panel C shows the results only using ETFs (available to both retail and institutional investors) on both legs. The tests on the coefficients are H_0: $\alpha_i = 0$ and H_0: $\beta_i = 1$. The left quadrant uses data from June 2007 to December 2019, while the right quadrant - where we exclude the market funds - uses data starting in August 2010. *p<0.1; **p<0.05; ***p<0.01.

Panel A: Retail investors						
				excl.	*"MKT" Funds*	
	α	β	R^2	α	β	R^2
SMB_{synt}	1.53	0.73***	0.56	1.55	0.88***	0.77
HML_{synt}	1.74	0.79***	0.38	1.98	0.96	0.60
MOM_{synt}	1.82	0.95	0.31	4.64	0.79***	0.34
RMW_{synt}	2.73**	0.35***	0.27	0.85	0.28***	0.17

Panel B: Institutional investors						
				excl.	*"MKT" Funds*	
	α	β	R^2	α	β	R^2
SMB_{synth}	0.89	0.71***	0.64	0.80	0.87***	0.82
HML_{synth}	1.15	0.87**	0.55	1.14	0.95**	0.64
MOM_{synth}	1.08	0.94	0.28	4.71	0.72***	0.31
RMW_{synth}	2.40*	0.34***	0.24	0.81	0.29***	0.18

Panel C: ETFs only						
				excl.	*"MKT" Funds*	
	α	β	R^2	α	β	R^2
SMB_{synth}	1.24	0.74***	0.58	0.59	0.92***	0.84
HML_{synth}	3.04**	1.18**	0.60	1.01	0.93***	0.54
MOM_{synth}	0.67	0.93	0.20	4.49	0.89*	0.29
RMW_{synth}	2.09	0.31***	0.16	0.75	0.28***	0.14

Table A.2: Synthetic tradable risk factors - post- Global Financial Crisis. This table reports the performance of long-short synthetic factors from the regression

$$ret_{FF_{LS},t} = \alpha + \beta ret_{synthetic_{LS},t} + \varepsilon_t$$

Panel A (Panel B) reports results for retail (institutional) investors using mutual funds and ETFs on the long leg, and ETFs on the short leg. The quadrant *excl. "MKT" funds* excludes from the analysis those funds classified as "market" funds by our minimum distance approach, i.e., funds with the lowest F-test statistic when compared to the expected MKT parameters are excluded. Panel C shows the results only using ETFs (available to both retail and institutional investors) on both legs. The tests on the coefficients are H_0: $\alpha_i = 0$ and H_0: $\beta_i = 1$. The left quadrant uses data after the GFC from July 2009 to December 2019, while the right quadrant - where we exclude the market funds - uses data starting in August 2010. *p<0.1; **p<0.05; ***p<0.01.

Panel A: Retail investors						
				excl.	*"MKT" Funds*	
	α	β	R^2	α	β	R^2
SMB_{live}	1.53	0.76^{***}	0.64	1.82	0.85^{***}	0.71
HML_{live}	1.52	0.79^{***}	0.40	1.61	0.84^{***}	0.55
RMW_{live}	3.27^{**}	0.26^{***}	0.14	1.18	0.30^{***}	0.14

Panel B: Institutional investors						
				excl.	*"MKT" Funds*	
	α	β	R^2	α	β	R^2
SMB_{live}	0.61	0.68^{***}	0.66	0.85	0.86^{***}	0.80
HML_{live}	0.70	0.93	0.60	0.72	0.92^{***}	0.64
RMW_{live}	3.33^{**}	0.26^{***}	0.16	0.95	0.32^{***}	0.18

Panel C: ETFs only						
				excl.	*"MKT" Funds*	
	α	β	R^2	α	β	R^2
SMB_{live}	0.32	0.73^{***}	0.68	0.58	0.92^{***}	0.83
HML_{live}	1.29	1.13^{**}	0.59	0.94	0.95^{**}	0.56
RMW_{live}	3.14^{**}	0.29^{***}	0.18	0.65	0.29^{***}	0.14

Table A.3: Live Factors. This table reports the performance of long-short synthetic "live" factors from the regression

$$ret_{FF_{LS},t} = \alpha + \beta ret_{synthetic_{LS},t} + \varepsilon_t$$

Panel A (Panel B) reports results for retail (institutional) investors using mutual funds and ETFs on the long leg, and ETFs on the short leg. The quadrant *excl. "MKT" funds* excludes from the analysis those funds classified as "market" funds by our minimum distance approach, i.e., funds with the lowest F-test statistic when compared to the expected MKT parameters are excluded. Panel C shows the results only using ETFs (available to both retail and institutional investors) on both legs. The synthetic portfolios use "live" characteristics (e.g., we use the latest available accounting or market-based quantity available as of the day before the portfolio construction, 'rdq' in Compustat) to sort funds, similarly to the "HML in the Devil" factor by AQR. The tests on the coefficients are H_0: $\alpha_i = 0$ and H_0: $\beta_i = 1$. The left quadrant uses data from June 2007 to December 2019, while the right quadrant - where we exclude the market funds - uses data starting in August 2010. *p<0.1; **p<0.05; ***p<0.01.

Panel A: Retail investors						
				excl. "MKT" Funds		
	α	β	R^2	α	β	R^2
CMA_{synt}	1.64	0.37***	0.31	1.23	0.47***	0.41
ROE_{synt}	3.03	0.38***	0.25	2.64	0.27***	0.11
IA_{synt}	1.39	0.37***	0.28	1.17	0.48***	0.36
QMJ_{synt}	3.84*	0.51***	0.33	3.88*	0.25***	0.09

Panel B: Institutional investors						
				excl. "MKT" Funds		
	α	β	R^2	α	β	R^2
CMA_{synt}	1.40	0.37***	0.32	0.61	0.48***	0.43
ROE_{synt}	3.02	0.37***	0.26	2.74	0.27***	0.11
IA_{synt}	1.14	0.36***	0.28	0.54	0.48***	0.37
QMJ_{synt}	3.73*	0.52***	0.32	3.78	0.26***	0.09

Panel C: ETFs only						
				excl. "MKT" Funds		
	α	β	R^2	α	β	R^2
CMA_{synt}	1.21	0.37***	0.32	0.64	0.47***	0.42
ROE_{synt}	2.87	0.38***	0.24	2.59	0.31***	0.12
IA_{synt}	0.98	0.38***	0.30	0.57	0.47***	0.37
QMJ_{synt}	3.90*	0.51***	0.26	3.71	0.38***	0.12

Table A.4: Additional factors. This table reports the performance of long-short synthetic factors from the regression

$$ret_{factor_{LS},t} = \alpha + \beta ret_{synthetic_{LS},t} + \varepsilon_t$$

for additional factors: the CMA of Fama and French (2015), the profitability factors (ROE and I/A) of Hou et al. (2015), and the QMJ of Asness et al. (2019). Panel A (Panel B) reports results for retail (institutional) investors using mutual funds and ETFs on the long leg, and ETFs on the short leg. The quadrant *excl. "MKT" funds* excludes from the analysis those funds classified as "market" funds by our minimum distance approach, i.e., funds with the lowest F-test statistic when compared to the expected MKT parameters are excluded. Panel C shows the results only using ETFs (available to both retail and institutional investors) on both legs. The tests on the coefficients are H_0: $\alpha_i = 0$ and H_0: $\beta_i = 1$. The left quadrant uses data from June 2007 to December 2019, while the right quadrant - where we exclude the market funds - uses data starting in August 2010. *p<0.1; **p<0.05; ***p<0.01.

Panel A: Retail investors									
	MFs+ETFs			ETFs only			MFs+ETFs, no "mkt" funds		
	$\hat{\alpha}$	$\hat{\beta}$	R^2	$\hat{\alpha}$	$\hat{\beta}$	R^2	$\hat{\alpha}$	$\hat{\beta}$	R^2
Small	1.23	1.05***	0.98	0.24	0.99	0.99	0.31	1.04***	0.98
Big	0.56	1.12***	0.94	-0.85	1.10***	0.94	0.52	1.05***	0.96
Value	-0.06	1.17***	0.93	-1.04	1.17***	0.95	0.23	1.13***	0.95
Growth	-0.28	0.98	0.93	-1.43	1.04***	0.94	-0.36	0.94***	0.94
Up	3.04**	0.98	0.94	1.34	0.99	0.93	2.47*	0.97**	0.94
Down	-1.24	1.32***	0.90	-2.60	1.29***	0.87	-1.71	1.29***	0.90
Robust	1.85	1.12***	0.92	0.19	1.09***	0.93	0.19	1.02	0.92
Weak	-1.17	1.08***	0.95	-1.88	1.00	0.96	-0.51	1.02**	0.95

Panel B: Institutional investors									
	MFs+ETFs			.			MFs+ETFs, no "mkt" funds		
	$\hat{\alpha}$	$\hat{\beta}$	R^2				$\hat{\alpha}$	$\hat{\beta}$	R^2
Small	0.91*	0.98***	0.99				0.51	0.98**	0.99
Big	0.72	1.13***	0.95				0.35	1.01	0.95
Value	-0.01	1.06***	0.96				-0.15	1.07***	0.96
Growth	0.08	0.98*	0.93				-0.46	0.94***	0.94
Up	2.28*	1.02	0.94				3.06**	0.94**	0.93
Down	-1.73	1.28***	0.87				-2.43	1.23***	0.87
Robust	0.97	1.11***	0.91				1.46	0.98	0.92
Weak	-0.12	0.94***	0.95				-0.29	0.94***	0.96

Table A.5: Tradable factor legs. This table reports the performance of individual factor legs from the regression

$$ret_{FF_{leg},t} = \alpha + \beta ret_{synthetic_{leg},t} + \varepsilon_t$$

Panel A (Panel B) reports results for retail (institutional) investors using mutual funds and ETFs (left quadrant), ETFs only (middle quadrant) and mutual funds and ETFs excluding "MKT" funds (right quadrant). The right quadrant excludes from the analysis those funds classified as "market" funds by our minimum distance approach, i.e., funds with the lowest F-test statistic when compared to the expected MKT parameters are excluded. The results in the quadrant "ETFs only" are available to both retail and institutional investors. The tests on the coefficients are H_0: $\alpha_i = 0$ and H_0: $\beta_i = 1$. The sample periods are the longest available for each class of investors/funds: from June 2004 for retail and institutional MFs+ETFs and from June 2007 for ETFs only. *p<0.1; **p<0.05; ***p<0.01.

Internet Appendix B Dataset Construction

Our data is comprised of daily, monthly, and quarterly data on the universe of U.S. equity mutual funds and ETFs from April 2003 up to December 2019. We also use the factor mimicking portfolios constructed by Fama and French (2015), Hou et al. (2015), and Asness et al. (2019).

B.1 Wharton Research Data Service

We construct the main mutual fund dataset by merging several datasets available on the Wharton Research Data Service (WRDS)

1. CRSP Daily Stock File
2. CRSP Mutual Funds Daily Returns
3. CRSP Mutual Funds Monthly Returns
4. CRSP Mutual Funds Summary (Quarterly)
5. CRSP Mutual Funds Portfolio Holdings (Quarterly)
6. Compustat Fundamentals Annual
7. Compustat Fundamentals Quarterly
8. CRSP/Compustat Linking Table

First, the daily, monthly, and quarterly CRSP mutual fund data is merged on 'crsp_fundno' and 'date', creating a dataset that contains daily fund returns, monthly fund flows, and quarterly fund characteristics. When we rebalance our synthetic portfolios daily, the funds' weights are calculated using the total net asset value of each fund from CRSP as of the end of the latest calendar month, and adjusting their NAV, every day, based on their daily returns.[5] Note that all analysis using daily flows are based on the EPFR data, which contains

[5] This will not take into account the daily flows to the funds, which are not available on CRSP, in updating the relative fund weights in our synthetic portfolios. However, looking at the daily flows from EPFR available for a subset of funds suggests that the relative weights are almost identical.

the reported daily flows and changes in the total net asset values. Second, the Compustat fundamentals are used to calculate the book equity of each 'gvkey', where the Compustat quarterly is used to fill in the gaps where available.

Third, the Compustat data is merged with the CRSP daily stock file by the CRSP/Compustat linking table. As there are multiple 'permno' (different share classes) for each 'permco' and 'gvkey', the Compustat 'gvkey' is linked to the CRSP 'permno' and 'permco' using the CRSP/Compustat linking table. The final merge is done on 'permno' and 'date', which implies that different 'permno' can have the same company characteristics if they belong to the same 'permco'. Therefore, depending on which analysis we are performing, the market equity, 'me', will sometimes be based on the share class, 'permno', and sometimes on the whole company's total market capitalization, 'permco'. For example, when we sort companies based on their characteristics, we use the entire company value, 'permco', but when we create the value-weighted returns, we use the weight on the individual share classes, i.e., the 'permno' (in the end it will get the same total value-weight, but different share-classes can exhibit different returns, which will be missed otherwise).

The merged CRSP/Compustat data set is used to calculate each firm's characteristic score in a similar vein to Asness and Frazzini (2013), Hou et al. (2015), Fama and French (2015), and Asness et al. (2019). At each 't', each firm is rank sorted from lowest to highest cross-sectional characteristic score. Since some funds are short-selling equities, the rank score is rescaled to -1 to 1, ensuring that the sign of the characteristic score will be consistent even if the fund has a negative exposure to a particular asset.

Fourth, the CRSP mutual fund portfolio holdings are merged with the CRSP+Compustat database. The addition of the size, value, momentum, profitability, IA, ROE, and quality characteristics of the fund holdings is used to calculate the fund specific exposure to the different strategies at each date. Fifth, at each date, the value-weighted strategy exposure of each fund is merged with the combined mutual fund database.

Finally, we merge the CRSP+Compustat mutual fund data with the EPFR data, which adds each fund's daily flows. An issue is that the EPFR data does not contain a 1-to-1 link table to the CRSP mutual fund data. Hence we match on CUSIP codes and ticker symbols. Both the CUSIP and the ticker symbol can change over time. However, using both CUSIP and ticker symbol yields a great match between the two datasets.

Additionally, the dates are sometimes misaligned, e.g., while Compustat and CRSP mutual fund characteristics are added at the end of each month, the CRSP firm returns are reported on each month's last trading day. Hence, whenever two data sets are merged with misaligned dates, the information is moved forward to the next available date, which prevents any look-ahead bias. Similarly, when we create the strategy exposures, each characteristic is lagged an additional day, which ensures that each strategy is tradable in practice.

B.2 Restricting our funds' population

Overall, there is no consensus in the literature on how to restrict the sample to only include domestic equity funds. We follow the WRDS guidelines, which have a step-by-step guide on how to restrict the fund sample.

1. Keep all equity funds using `'lipper_asset_cd'` equal to `'EQ'`.

2. There exist several alternatives for keeping domestic funds. Lettau et al. (2019), keeps all funds where the `'crsp_obj_cd'` starts with `'ED'`, which denotes domestic equity funds. In contrast, Evans (2010) only keep funds where `'per_com'` $> 90\%$ throughout the fund life. Following the 7-step procedure recommended in WRDS, we keep funds where `'lipper_class'` $\in$ { `'EIEI'`, `'G'`, `'LCCE'`, `'LCVE'`, `'MCCE'`, `'MCGE'`, `'MCVE'`, `'MLCE'`, `'MLGE'`, `'MLVE'`, `'SCCE'`, `'SCGE'`, `'SCVE'`}. However, after this step, there are still foreign sector funds left in the sample. Therefore, we remove funds where `'crsp_obj_cd'` doesn't start with `'ED'`. Note that we do a final restriction in step 5, where we require funds to hold a majority in US equities.

3. A common restriction is to also remove micro funds, which often consists of dead funds, or funds that are not actually traded in the market. Similar to the other restrictions, there is no consensus on how to do this. For tradability purposes, we keep funds with a CPI adjusted[6] AUM greater than $1 billion. We also use a $50 million restriction for robustness. To reduce the probability of funds going in and out of the sample each trading day, we only include the fund if it fulfills the restriction for at least half of the last years days of trading.

4. We also require funds to have at least one year of observations.

[6]Downloaded from Robert J. Shiller website.

5. Finally, we require the fund to hold 50%-130% in US equities (CRSP `'shrcd'` $\in \{10, 11\}$), which is estimated quarterly using the reported holdings of each fund.

B.3 Categorizing funds by name

Each fund is classified by partial matching of keywords in their name:

1. SMALL if it contains the phrase "small", "micro", "low-priced", or "russell 2000".
2. BIG if it contains the phrase "large", "mega cap", "nasdaq", "russell 1000", "russell 3000", "broad market", "total stock", "total market", "nyse", "dow jones industrial".
3. VALUE if it contains the phrase "value", "book", or "low p/e".
4. GROWTH if it contains the phrase "growth".
5. MOMENTUM if it contains the phrase "mom" or "trend".
6. QUALITY if it contains the phrase "quality", "dividend", "income", or "appreciation".

Each of the above keywords has been manually inspected.

We have also tried other keywords potentially associated with factors known in the academic literature. These include beta, idio, multi, fact, liq, smart, min, robust, aggressive, conservative, skew, tail, crash, tail risk, big, high, low, and junk. Unfortunately, these keywords either yield no, or very few funds, or a mixture of funds of different strategies, so we do not include them. We also excluded additional keywords that are already covered by the other keywords, e.g., 'MSCI USA' or 'S&P 600'.