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

This version: August 10, 2021

ABSTRACT

We construct tradable proxies of "on-paper" long-short risk factors using combinations of large and liquid mutual funds and ETFs, based on their holdings, for both retail and institutional investors. Using a novel proprietary dataset, we are able to account for the ETF shorting fees in constructing the short leg of the factors. In contrast with the recent literature, we find that investors are able to harvest the SMB and part of the HML risk premia, although MOM and RMW remain hard to replicate, with institutional investors outperforming retails. Our results have implications for the benchmarks that should be used to evaluate portfolio managers.

Keywords: Smart beta, factor investing, tradable risk premia, shorting costs, borrowing fees

*We are grateful to Adam Farago (discussant), Kris Jacobs, Minsoo Kim (discussant), 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 Eastern Finance Conference, NFN Young Scholar conference, Stockholm School of Economics, University of Houston, and Warwick Business School. We thank S3 partners for the ETF shorting fee data. The paper received the 2020 ICPM Award Honourable Mention.

†Ph.D. Student, Department of Finance, Stockholm School of Economics. **Email:** andreas.johansson@phdstudent.hhs.se

‡Department of Finance, Stockholm School of Economics and Swedish House of Finance (SHoF). **Email:** riccardo.sabbatucci@hhs.se

§Department of Finance, Rutgers Business School. **Email:** atamoni@business.rutgers.edu

1 Introduction

Smart beta, also known as 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 how both institutional and retail investors' demand for such assets has been steadily increasing over time. One explanation for this trend is that institutional investors have been increasing 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 real-world investors can track "on-paper" academic 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 stock characteristic, (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: how can institutional and retail investors optimally invest in "on-paper" risk factors? Do they differ in the way they can implement these strategies

[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 investors in mutual funds/ETFs, not on the funds themselves.

and obtain heterogeneous performances? Should investors evaluate fund managers against their opportunity cost of investing in tradable risk factors versus "on-paper" risk factors?

A few recent studies have attempted to quantify real-world implementation costs for academic "on-paper" factors. Existing approaches either use proprietary trading data to study trading costs for a single investment firm (e.g., Frazzini et al., 2015), use market-wide trading data such as NYSE Trade and Quote (TAQ) to estimate the trading costs of individual stocks selected by dynamic factor strategies (e.g., Novy-Marx and Velikov, 2016; Detzel et al., 2021), or estimate directly factor implementation costs based on cross-sectional Fama-MacBeth specifications augmented with mutual fund returns (Patton and Weller, 2020).

We tackle this issue in a novel and conceptually different way. We use mutual funds and ETFs to construct synthetic versions of the "on-paper" factors tradable by both institutional and retail investors in real-time, explicitly accounting for real-world shorting fees. More in detail, we select a portolio of smart beta mutual funds and ETFs based on their holdings to synthetically replicate the individual legs of the "on-paper" risk factors. We then construct synthetic, long-short factors (e.g., HML_{synth}) that are tradable by *both* retail and institutional investors by value-weighting the funds most exposed to the respective underlying factor characteristic. The short leg of our synthetic tradable factors only include ETFs and accounts for the daily shorting fees in constructing the factor return. Our methodology is general and can be used to replicate *any* "on-paper" factor or smart beta index with publicly disclosed holdings.[3]

Our approach relies on two main insights. First, in terms of transaction costs, trading a small number of funds is cheaper than trading hundreds of stocks. Indeed, "on-paper" academic factors include hundreds of stocks in both the long and short legs, and often have large turnover. Second, and most importantly, shorting illiquid, hard-to-borrow stocks, included in the short leg of the "on-paper" factors, is often infeasible. Several stocks have limited float and hence cannot be easily borrowed, and/or have extremely high shorting fees. Using ETFs in the short leg of the synthetic factors avoids this problem, since ETF shares can always be created by authorized participants (AP). As we will show in Section 6.1, and

[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 fund holdings to construct synthetic, optimal portfolios with characteristic scores that are as close as possible to the "on-paper" factors' ones. In other words, what matters is how close – in terms of holdings – funds are to the "on-paper" factors, regardless of their type.

one of the key contributions of this paper, ETFs have usually low shorting fees and larger short interest than individual stocks.[4] Thus, the conclusions of the previous literature, which uses individual stocks, do not automatically extend to our settings which instead relies on *portfolios* of stocks (e.g., funds). Indeed, we show that although investors face challenges in harvesting the various "on-paper" factor risk premia, there are important differences across strategies, type of investors (retail and institutional), as well as type of funds (mutual funds and ETFs).

We document several novel facts. First, looking across strategies, the replication of momentum (MOM) and profitability (RMW) proves particularly challenging for both types of investors, with 3%-4% lower annual returns relative to their relative "on-paper" risk premia. On the other hand, both types of investors can harvest the SMB premium using only ETFs in both legs: in this case, the synthetic tradable SMB leaves only a small alpha of 38 bps relative to the "on-paper" factor. The replication of HML falls in between these two extremes, and depends on the type of investors.

Second, institutional investors are able to synthetically replicate the "on-paper" factors better than retail investors, as indicated by lower alphas from regressions of the benchmarks on our synthetic portfolios, and by higher time series fit (R^2). This is true for all the factors we analyze, but especially for HML: even accounting for shorting costs, the synthetic HML available to institutional investors attains an alpha which is around 1% lower than the retail one. Consistent with this finding, we show that 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.

Third, the replication of SMB and MOM improve substantially by only using ETFs in the synthetic portfolios. 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. Taken together, our first three facts suggest that the replication of "on-paper" factors depends not only on what factor is the object of interest (e.g., SMB or MOM), but also on the type of investor (retail or institutional) and assets available to them (individual stocks, mutual funds and ETFs, or ETFs only.)

[4] Consistent with our finding, Evans et al. (2021) find that ETFs constitute 10% of the U.S. equity market capitalization but make up for over 20% of the aggregate short interest.

Fourth, using a proprietary dataset of daily ETF shorting fees, we calculate the daily financing costs (e.g., shorting fee) of our synthetic factors and subtract them to obtain the final long-short factor daily returns. This is an important but often overlooked aspect in the existing literature. Shorting individual stocks or ETFs is costly, and this cost should be explicitly accounted for in obtaining true, tradable factor returns. These factor shorting costs are time-varying and factor-dependent, with the short legs of HML and RMW averaging around 1% per year, and those of SMB and MOM being very volatile with spikes above 4% per year. We also show that the short interest of our synthetic portfolios are, on average, across factors and over time, larger than those implied by the "on-paper" Fama-French factors.

Fifth, given our focus on the tradability of the factors, we provide estimates on the capacity of tradable risk factors. A back-of-the-envelope calculation shows that SMB (HML) has \$1tn (\$2tn), while MOM and RMW have about \$450bn and \$60bn available capacity, respectively, in line with the results in Ratcliffe et al. (2017).[5]

Sixth, we show that our procedure to construct synthetic factors is not a mere repackaging of a more naive "by name" strategy, e.g., selecting funds using keywords like "value" or "low p/e" to construct a "by name" HML leaves a large alpha of 5% relative to the 1.5% obtained by institutional investors in our synthetic HML. This suggests that fund 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.

Our contributions to the existing literature are both methodological and in terms of data. To the best of our knowledge, our paper is the first to construct synthetic risk factors using funds and to explicitly model the short leg of the risk factors accounting for daily shorting fees. This is necessary in order to create truly tradable implementations of the "on-paper" risk factors. In order to do so, we rely on a unique, novel dataset containing the time series of the daily ETF shorting costs. Our paper is also the first to construct tradable "on-paper" factors for both retail and institutional investors. We show empirically

[5]In the Appendix, we also 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.

that these two sets of investors have access to different investment opportunity sets, which lead to differential performances, with institutional investors being better able to replicate the "on-paper" factors.

Our paper differs from prior work in several fundamental aspects. Differently from Frazzini et al. (2015), Novy-Marx and Velikov (2016) and Detzel et al. (2021), who focus on the transaction costs involved in replicating the risk factors using individual stocks, we provide estimates of truly tradable factor returns for both retail and institutional investors using publicly available investment instruments (e.g., funds). Relative to these studies, our paper has a key advantage. While transaction costs are extremely important when trading hundreds of stocks, and severely impact the real-world factor performance, our approach avoids such friction. In fact, our synthetic portfolios are only composed of a few large, extremely liquid funds in each leg, and the fund turnover in the optimal synthetic portfolios is extremely low, resulting in trading costs for our synthetic portfolios that are negligible. Differently from Patton and Weller (2020), we synthetically replicate the "on-paper" factors, using both mutual funds and ETFs. This allows us to construct real-time, tradable proxies of the "on-paper" factors, not only estimates of their implementation costs. Most importantly, by using ETFs instead of individual stocks in the short leg, and exploiting daily data on their short interest and shorting fees, our methodology generates time series returns of *truly* tradable factor short legs.

Our analysis shows that investors cannot obtain – through ETFs and mutual funds – large exposures to the long leg of HML (Value), or to the short legs of MOM and RMW, as evidenced by characteristic scores of the synthetic factor legs that are substantially different from those implied by the Fama-French "on-paper" factors. This finding is in line with Lettau et al. (2019) who document the "lack" of Value funds but also consistent with Patton and Weller (2020) who document that the short leg of the "on-paper" factors is what makes them hard to trade in practice. However, while in the work of Patton and Weller (2020) the limitation stems from the mutual funds' shorting constraints of individual stocks, this is not a concern in our setting, since all ETFs in our synthetic portfolios are highly liquid and can be easily sold short. The issue, in our case, is related to the structure of the fund industry: there is a scarcity of funds investing in "losers" or "weak profitability" stocks, which are required to replicate the "on-paper" factors. It is then not surprising that the SMB "on-paper" factor can be replicated quite well since several large-cap stocks ETFs exist.

More broadly, the fact 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). This raises the question of whether portfolio managers should be evaluated against the true opportunity cost of factor investing of institutional and retail investors rather than "on-paper" factors. In other words, their alpha should be estimated with respect to what institutional and retail investors could have achieved by trading themselves in a publicly available, optimal combination of smart beta funds.[6]

Our paper proceeds as follows. In Section 2, we review the literature. Section 3 describes the data used in the paper, while Section 4 presents some descriptive evidence on the smart beta industry. Section 5 describes how we construct synthetic, tradable risk factors using funds, while Section 6 reports the main empirical results. Section 7 concludes.

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, Shorting Costs 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. 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). Our results are consistent with Arnott et al. (2017) and Patton and Weller (2020) who find that real-world implementation of the momentum and value fac-

[6]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.

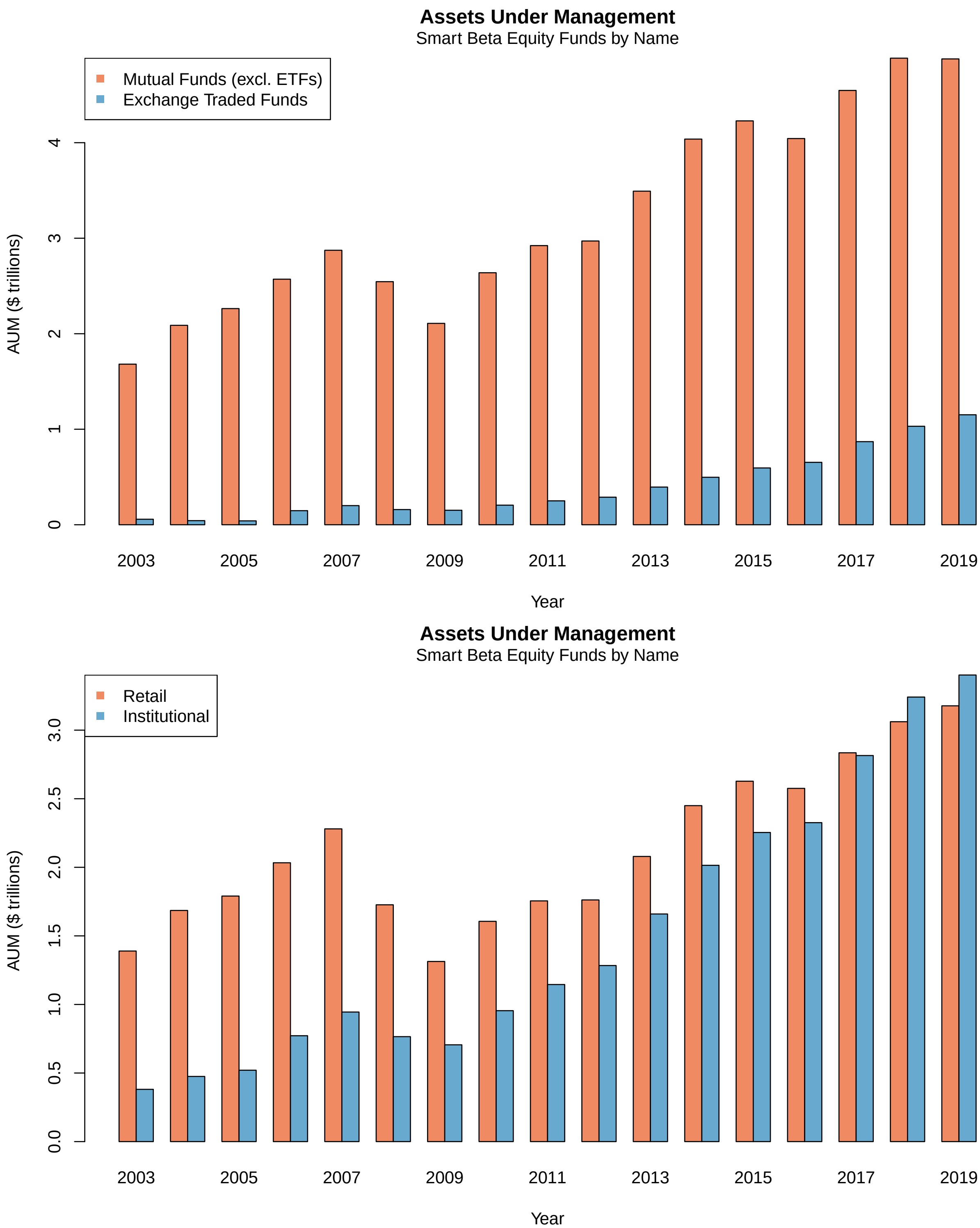

Figure 1: Smart Beta Equity Mutual Funds and ETFs. The top panel plots the assets under management of smart beta funds starting in 2003. The bottom panel plots the assets under management of mutual funds and ETFs available to institutional and retail investors. Funds are labeled as smart beta if their 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 C.3 for a detailed description of the classification. The sample includes funds with assets under management greater than $50 millions.

tors strongly underperform the relative "on-paper" strategies due to all-in implementation costs[7], and with Detzel et al. (2021), who find that accounting for individual stocks' transaction costs substantially reduces factor risk premia. Differently from the studies above, we investigate whether using highly liquid, cost-effective instruments available to both retail and institutional investors, and explicitly accounting for *shorting costs*, it is possible to replicate 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).[8]

Our work is complementary to Drechsler and Drechsler (2014), Engelberg et al. (2018), and Muravyev et al. (2021). Whereas these papers focus on the shorting costs of individual stocks, we exploit a novel dataset to account for the borrowing costs of ETFs when constructing tradable risk factors. Li and Zhu (2019) provide evidence that ETFs allow to circumvent short-selling constraints at the individual stock level.

Fund Performance and Closet Indexing. Our paper is also related to the literature on fund managers' performance with respect to "on-paper" risk factors (i.e., the α often interpreted as the fund manager's skill).[9] The literature on fund managers' skill tends to benchmark fund performance on factor models that cannot be easily 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

[7]Both papers use the classical two-pass Fama and MacBeth (1973) procedure to reach these conclusion; we instead construct tradable factor strategies using mutual funds and ETFs.

[8]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.

[9]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.

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). 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 fund 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 and fees, 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.

We get daily data on ETFs' shorting fees starting for the period 2015-2019 from S3 Partners. We interpolate the shorting cost for years before 2015 by taking the full sample daily average shorting costs of each ETF. The shorting fee data is very detailed, and include, among other variables, a bid, an ask and a last rate for each ETF. The bid rate is the average cost prime brokers are paying beneficial owners to borrow ETF shares, while the offer rate is the weighted average as to what existing shorts are paying to borrow the ETF shares.

In practice, this is the actual cost that investors, on average, pay to short ETF shares at a specific point in time. The last rate is the marginal borrowing cost for investors opening new short positions, and it is a proxy for where the average market borrowing fee (e.g., the offer rate) will move.[10] Data on ETF short interest is from the CRSP MF database.

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 "on-paper" 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 Appendix A.2, we also construct tradeable version of 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.[11]

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 in the main analysis of the paper are net of all fees except for front and rear loads,[12] 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).[13]

To be consistent with the assets used to construct the "on-paper" 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, and AUM greater than $1bn (in real terms as of 2019).

[10]When constructing our synthetic portfolios (Section 6.2), if the shorting fee for any fund in the optimal portfolio is not available on any day, we value-weight the shorting fees of the other funds in the portfolio.

[11]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.

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

[13]In Appendix A.3 we report results using gross fund returns, and verify that fund fees are not driving our results.

The choice of a \$1bn fund size cutoff is consistent with our goal of creating risk factors tradable by both institutional and retail investors. Investors might not feel confident trading small mutual funds or illiquid ETFs.[14] Moreover, small funds are more likely to invest in illiquid securities given their size, hence their strategies might not be as scalable as those available to larger funds.[15] Overall, by focusing on funds with more than \$1bn of AUM we are ensuring the synthetic factors are indeed tradable. Following the literature (e.g., Patton and Weller, 2020), for funds with multiple share classes, we eliminate the duplicated funds and compute the fund-level variables by aggregating across the different share classes.[16] It is important to note that no funds of funds are left in our final dataset, as they would not be easily tradable. In Internet Appendix C we describe all the standard filters used to obtain the final dataset.[17] Section 5.2 provides details on the construction of the long-short synthetic factors.

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. We also split the funds into those available to retail investors (Panel A) and institutional investors (Panel B).

The total number of unique funds increases steadily over time. Institutional investors have 692 (391) mutual funds and ETFs available in the latter (former) period, an increase of 77%. The increase for funds available to retail investors is a moderate 5%. The pattern is even more striking when looking at ETFs (+90%), which became extremely popular over the last decade.

The average (median) fund size is around \$6.8bn (\$2.6bn) for retail investors, and slightly less for institutional investors, but the size distribution is heavily right-skewed. In the fourth

[14] 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.

[15] 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.

[16] To aggregate returns within a fund group, we take total net asset (TNA) weighted returns.

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

quarter of 2019, for example, the AUM of 962 funds exceeded \$1bn and 174 funds exceeded \$10bn.

The average ETF in the most recent sample is very large compared to the average mutual fund (\$9bn vs. \$6bn); this is consistent with the large ETF inflows observed 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 held by mutual funds is 110, while the median ETF holds 371 stocks in their portfolio over the most recent sample period. This suggests that large mutual funds cherry-pick their holdings, while ETFs are more diversified. Looking at the returns of mutual funds and ETFs, we note that the median retail (institutional) fund has outperformed the S&P 500 in the first half of our sample by 84 bps (68 bps) per year, while it has underperformed by 93 bps (72 bps) in the second half. Overall, the median five-factor alphas are negative for both retail and institutional investors.[18]

Our focus is on equity smart beta funds. For exposition purposes, in this section we classify funds into smart beta strategies based on their names (Appendix C.3 provides details on the procedure) and describe their time series properties. Focusing on smart beta fund 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 Internet Appendix B, 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 \$5tn 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 \$1bn AUM were smart beta.

[18] The only exception is the 0.53% alpha for ETFs before 2012, although it is not statistically different from zero.

Some examples of large smart-beta ETFs included in our sample are the iShares Edge MSCI USA Quality Factor ETF (QUAL, $19.6bn AUM), the iShares Edge MSCI USA Momentum Factor ETF (MTUM, $14.4bn AUM), the iShares Edge MSCI USA Value Factor ETF (VLUE, $10.5bn AUM), and the Goldman Sachs ActiveBeta U.S. Large Cap Equity ETF (GSLC, $11.7bn AUM).[19] 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.[20]

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.[21]

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 in 2019 as the "new kings of Wall Street."[22]

[19] AUM as of January 2021.

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

[21] We discuss this issue in Section 6.1.

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

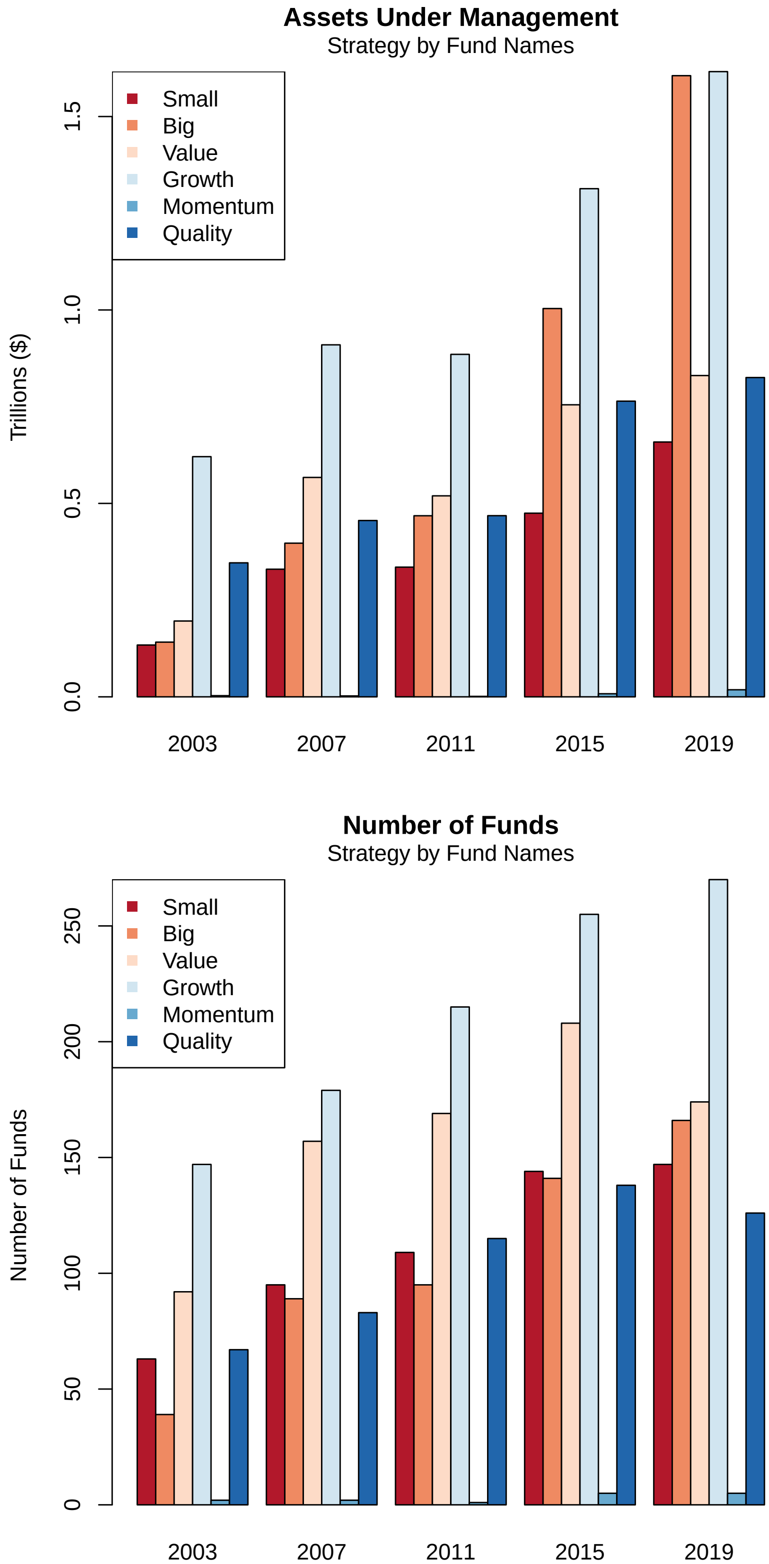

Figure 2: Net asset values and number of funds by strategy over time. The top panel plots the average net asset values of smart beta funds, by strategy, over time, while the bottom panel plots the time series of the number of funds available. Funds are categorized by name and must have more than $1bn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

5 Constructing Synthetic Factors Using Funds

5.1 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 $i = 1, 2, \ldots, N$, 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 denote with $S_t \in R^{N \times F}$ the time-varying factor score matrix, where F is the number of factors. We aggregate these scores at the fund level as follows:

$$X_{m,t} = \omega_{m,t} \times S_t, \tag{1}$$

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[23] (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").[24]

As an example, Figure 3 displays the time series of the characteristic scores for two funds, the 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 at some point in time. Several comments are in order. First, the book-to-market score is higher for the fund that markets itself

[23] 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.

[24] Of its own interest is the question whether smart beta funds truly exist, or are simply market funds. Thus, in Appendix B.2, we describe a classification methodology to categorize funds as market fund. We find that around one-half of the total existing funds cannot be distinguished from market funds, both in real-time and over the full sample.

as Value. However, we observe that the book-to-market score for the iShares Russell 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" funds, very few have high book-to-market ratios. Second, funds are not neutral to characteristics that are absent from their name; for example, the Vanguard Growth Index Fund features a high exposure to profitability. 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.

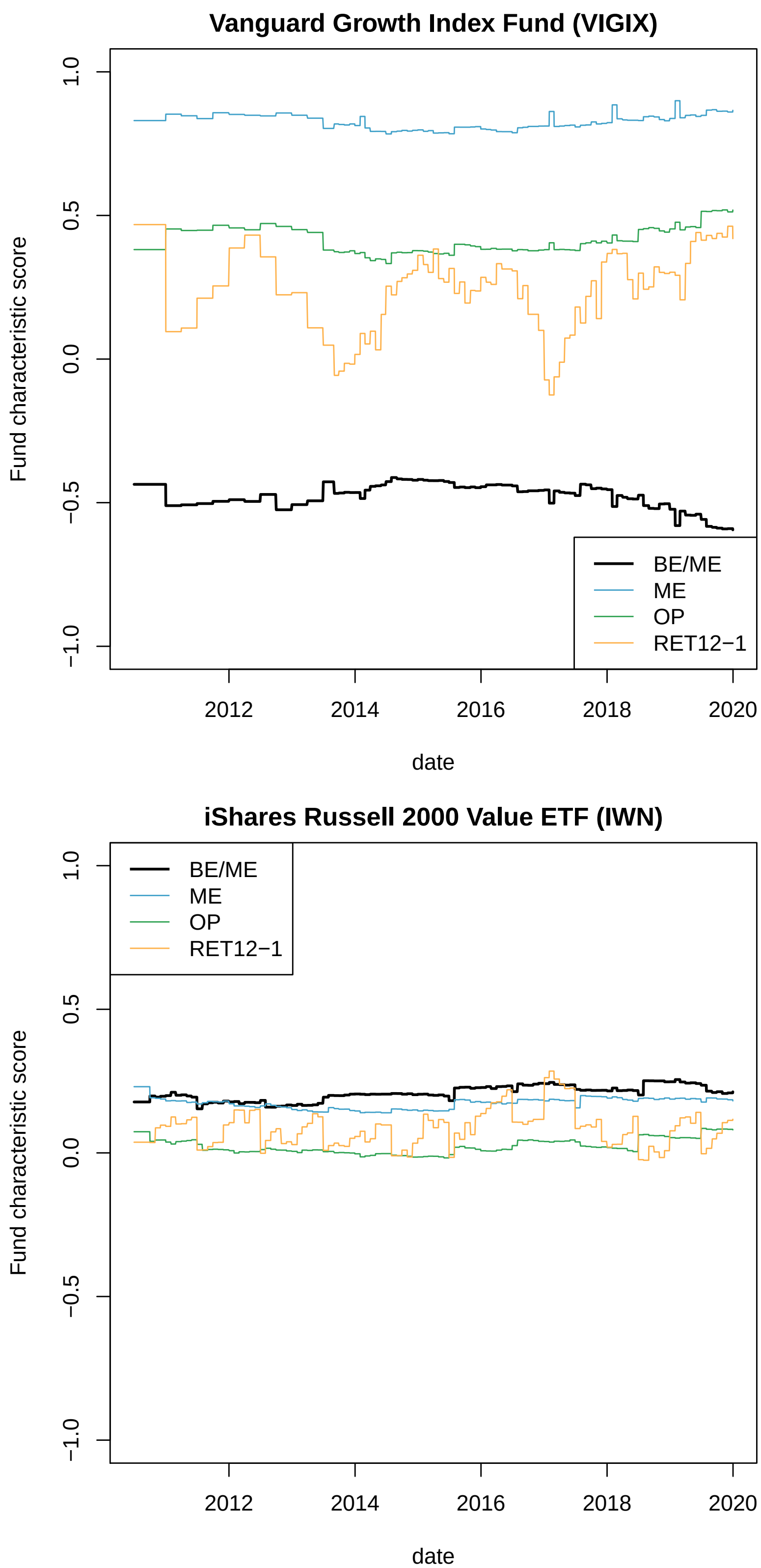

Figure 3: Time series of the characteristic scores for two funds in our sample. The top figure plots the time series of the characteristic scores for the Vanguard Growth Index fund which is classified as "Large-Growth" according to Morningstar. The bottom figure plots the time series of the characteristic 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.

5.2 Synthetic "Smart Beta" Risk Factors

Our ultimate goal is to construct synthetic, tradable long-short factors, and evaluate their performance in replicating traditional "on-paper" risk factors. By using portfolios of funds allows us to construct synthetic risk factors for two types of investors: *retail* and *institutional*. In turn, this allows us to study the properties of the portfolios available to *different* sets 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). It is worth mentioning that analyzing factor strategies for different investor types is not feasible using individual stocks, since they are theoretically tradable by both sets of investors, although institutional investors might have access to additional trading venues (e.g., OTC, dark pools). This explains why, to the best of our knowledge, no previous study has looked into this aspect.

At each time t, and for each characteristic, we rank funds using the score at the fund level (see equation (1)). 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. We then hold the synthetic portfolio until the next rebalancing date. 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 the prior 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 using the universe of funds available to institutional and retail investors, separately.[25,26]

Our choice of a fixed number of funds in the synthetic portfolios, namely ten, strikes a

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

[26] 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.

compromise between having a well-diversified portfolio (e.g., a large number of funds), and making the synthetic portfolio easy to track and monitor for an investor (e.g., a low number of funds).[27]

Furthermore, fixing the number of funds allows us to to properly test the performance of synthetic, 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 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 test estimates comparable.[28]

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 as will be discussed in the next section.

6 Empirical Results

6.1 Shorting Costs, Trading Costs and Capacity of our Synthetic Portfolios

A few recent papers have shown how accounting for trading costs affects the return performance of "on-paper" risk factors (Patton and Weller (2020), Detzel et al. (2021)). These papers focus only on one type of friction, e.g., transaction costs, that is the costs associated with the buying and selling of individual securities in a given factor portfolio. These factor portfolios include hundreds of stocks and have large turnover, hence incurring substantial trading costs. Detzel et al. (2021) estimate trading costs for standard factors ranging between

[27]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.

[28]We find that our results are qualitatively similar to using more funds (e.g., 25) in the synthetic portfolios.

0.36% (SMB) to 5.64% (MOM) per year.

However, transaction costs are not a major concern in our setting. In fact, by synthetically replicating "on-paper" factors using a limited number of funds most exposed to the characteristic underlying the factor, we practically avoid this friction. This is the case because funds internalize the trading costs of individual securities themselves, and the realized fund returns implicitly account for them. Most importantly, the transaction costs for most funds are negligible, sometimes zero, even for retail investors.[29]

To further highlight why trading costs are not an issue in our setting – i.e., when funds are used instead of individual stocks to replicate "on-paper" factors – Table 2 reports the average monthly turnover of our synthetic portfolios, while Figure 4 displays its six-months moving average over time. The fund turnover in the optimal synthetic portfolios is low. On average, across strategies and over time, the monthly turnover is around 14%. In other words, our synthetic portfolios trade only a handful of funds, some of which can be traded without any fee. Finally, whereas Table 2 suggests that retail and institutional investors have similar turnover on average, Figure 4 displays several instances where one type of investor has much higher turnover than the other, suggesting that institutional and retail investors may have access to potentially different strategies. More broadly, these turnover numbers should be compared to the much larger ones of "on-paper" factors.[30]

A more important friction concerns the implementation costs of the *short* legs of the risk factors, i.e., the costs incurred in borrowing the securities to be shorted to construct the factors' short leg. To the best of our knowledge, this friction has not been explicitly studied in the literature. For most securities, the borrowing cost, or shorting fee, is an order of magnitude larger than the transaction costs to trade them. This suggest that the "net of trading costs" factor returns discussed in the recent literature are, if anything, an upper bound, and that the actual implementation of "on-paper" factors using individual stocks might be quite different were the shorting fees of individual stocks taken into account.

A possible explanation for the little attention devoted to the borrowing costs of individual stocks is that they are quite time-varying, firm-characteristic dependent, and not always available at daily frequency. For example, the shorting fees of Microsoft (ticker: MSFT)

[29] A famous example is Fidelity.

[30] As an example, Detzel et al. (2021) show that turnover of MOM is around 25% per month, which is equivalent to trading hundreds of stocks every month.

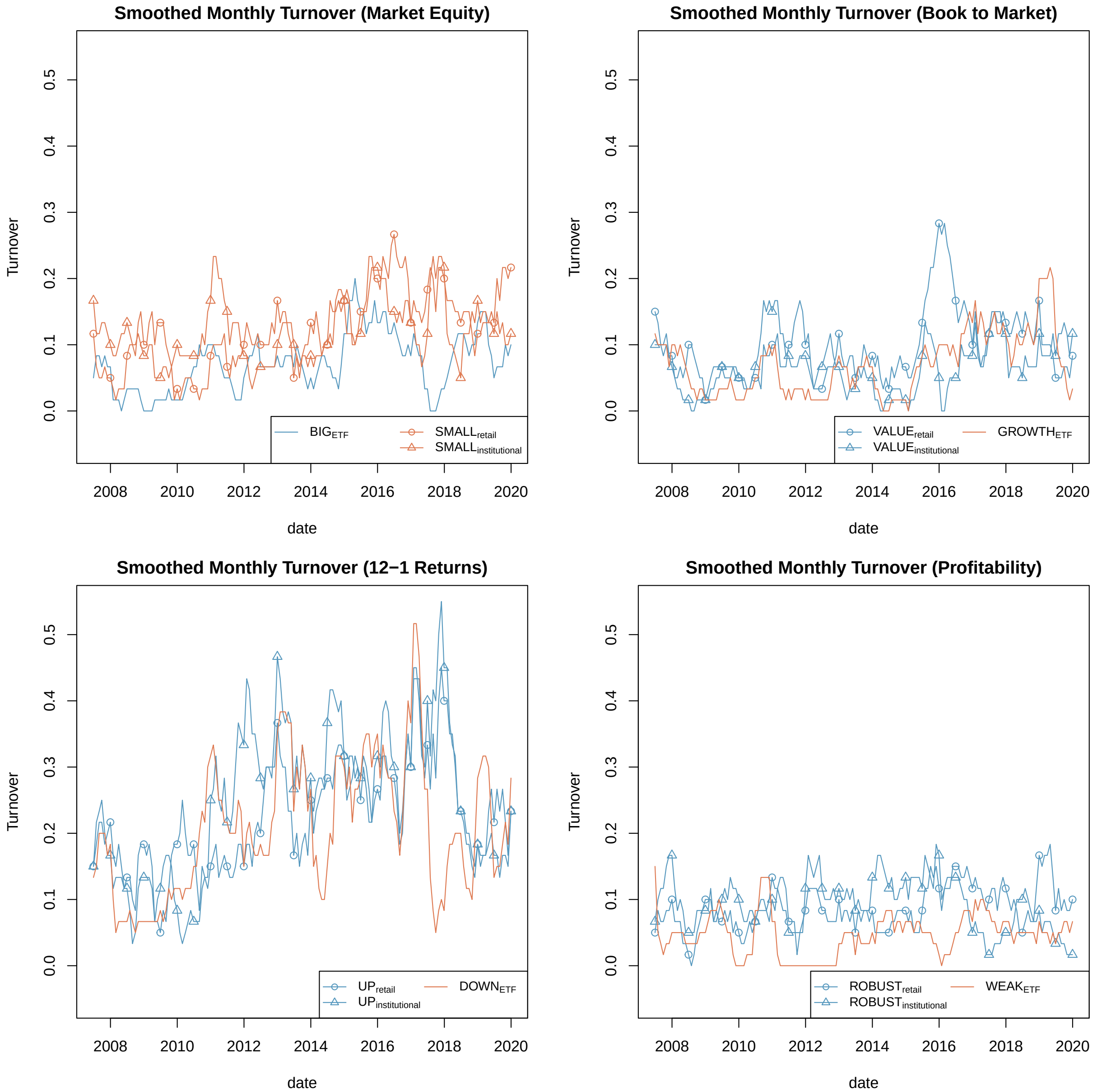

Figure 4: Monthly turnover of the synthetic portfolio legs. This figure shows the six-months smoothed turnover of the synthetic portfolios. The turnover is calculated as $TO = N_{new}/N_{tot}$. For example, if we replace three of the funds in our synthetic portfolio over the last month, the turnover 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.

in the short-leg of SMB or Gamestop (ticker: GME) in the short-leg of RMW were very different over the first two months of 2021, during the meme-stocks turmoil. MSFT had an average shorting fee of 0.25%, peaking at around 0.30% in early February, while GME had a much larger borrowing cost of 17.26%, on average, with a maximum of 33.9% and a daily annualized volatility of more than 9%. Hence, taking into account the shorting fees of individual stocks in constructing or evaluating risk factors, especially at lower frequencies such as monthly or quarterly, is challenging.

Our focus on funds-based factor replication is ideal in this respect. Our synthetic short-leg portfolios include only a limited number (namely ten) of ETFs instead of hundreds of potentially hard-to-borrow stocks, and the daily shorting fees and interest of large ETFs, that we use for tradability purposes, are almost always available. Most importantly, Li and Zhu (2019), Evans et al. (2021) and Ben-David et al. (2018) have documented that short-selling ETFs is relatively easier, since it is always possible to create new ETF shares to be shorted (e.g. naked shorting). As an example, during the recent meme-stocks turmoil of January 2021, while shorting GameStop might have been infeasible given the 30% shorting fee, it was still possible to short-sell ETFs owning GameStop shares. Quoting an extract from Blackrock ETF primer, "ETFs have functioned well in times of stressed markets, with ETF shares being at least as liquid as underlying portfolio assets and serving as an important vehicle of price discovery."[31]

This is confirmed in Figure 5, which shows that the short interest of the short legs of our synthetic tradable factors is often much larger than the one implied by the stock holdings of the Fama and French factor legs. The only exception is the Growth portfolio, for which the short interest of the stocks included in the Fama-French portfolio is greater starting from 2014. A plausible explanation for this finding relates to the performance of the FAANG stocks over the last few years, which make up a large chunk of the market capitalization of the Growth portfolio and are easy to borrow.

Figure 6 plots the shorting fees incurred in replicating the short legs of the "on-paper" factors from 2015 until 2019. The time-variation in the financing costs is evident, and the shorting fee declines over time, with the exception of the short momentum leg (e.g., Down), which is subject to jumps. The average financing cost for the Big, Growth and Weak legs are 1.4%, 1.0% and 1.3%, respectively, while slightly higher for the Down leg (2.2%). The

[31] See Novick et al. (2017).

median shorting fee for all ETFs in our final sample is 1.9% over the last five years, while its volatility is 4.1%, implying that the ETFs used in our synthetic short-leg portfolios have lower shorting fees than the median ETF, with the exception of the momentum leg.

Lastly, given the recent massive increase in smart beta mutual funds and ETFs, we are interested in understanding how much capital can flow into smart beta strategies through funds without distorting market prices, and whether the factor premia (e.g., HML) can be accessed through trading, even by large investors with several billions under management (e.g., Ratcliffe et al., 2017). In other words, we try to answer the question: "How much capacity do these smart beta strategies *actually* have?"

Figure 7 shows the median net asset value of the synthetic institutional and retail long and short portfolios over time.[32] Our synthetic portfolios only include, in each leg, ten large (i.e., AUM greater than \$1bn) 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 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, on average, to larger funds, with the median fund in the synthetic portfolios having a size of around \$3 bn, in contrast to the \$2bn median fund size available to retail investors. It is also interesting to note how the difference in fund size between institutional and retail investors is close to zero towards the end of our sample for all smart beta strategies. Overall, this suggests that, on any given day, a minimum of \$40-80bn capital (e.g., 10 funds × \$2-4bn for each leg) could be deployed by institutional investors to chase the risk premia by investing in a synthetic combination of funds.

We now attempt at providing 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 (e.g., DeMiguel et al. (2021)). An empirical proxy for the capacity of the factor strategies can be estimated by taking the number of large funds investing in a specific strategy[33] (e.g., around 260 funds for Growth, 160 for Value) from Figure 2 and assume that

[32]We plot the six-months smoothed median for clarity purposes. The reason is that when a fund goes just below the \$1bn threshold, or a new fund enters into the synthetic portfolio, the median happens to jump, although this is rare.

[33]We assume ten funds for the short leg of Momentum and Profitability since no fund seems to target these strategies.

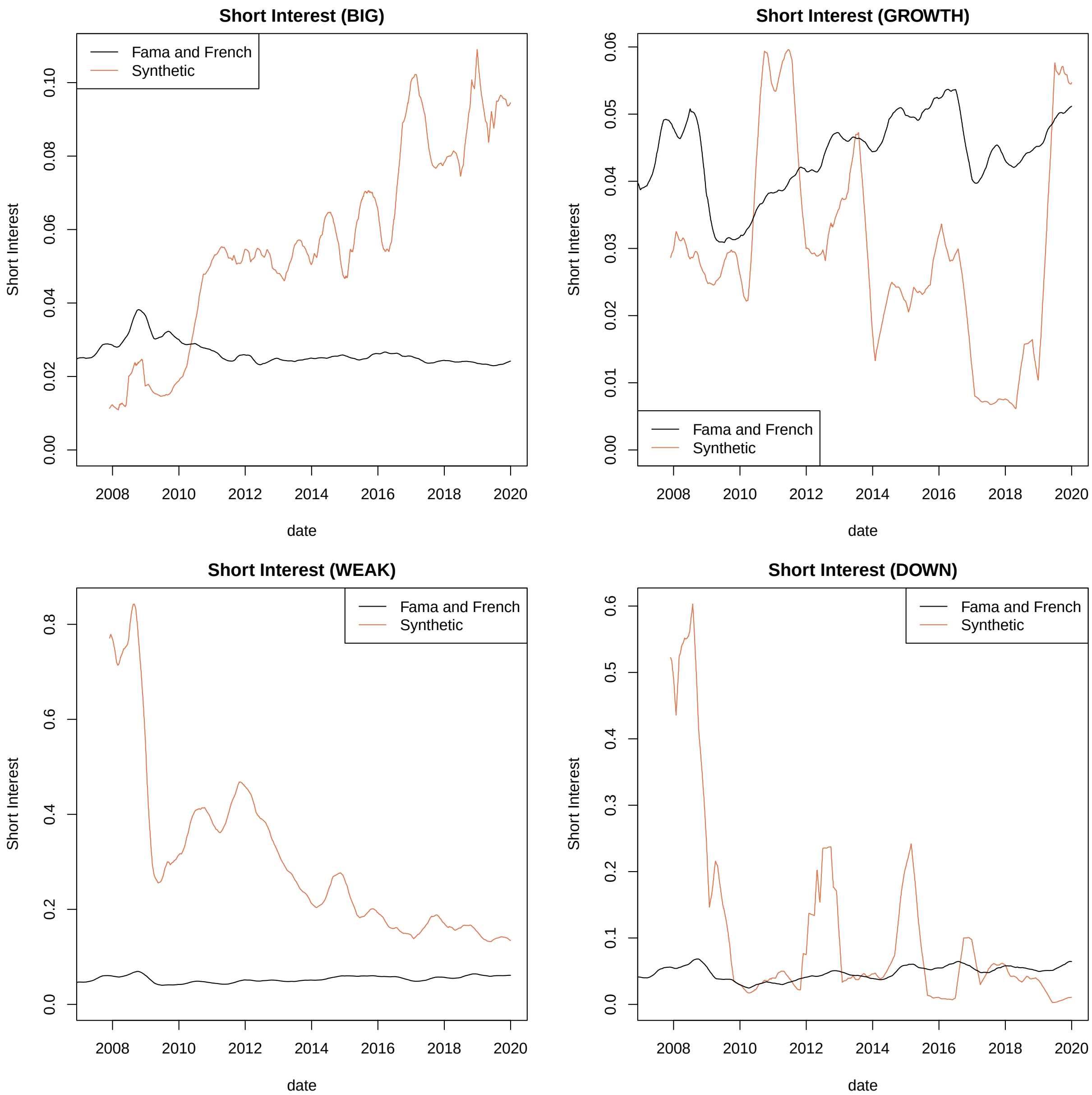

Figure 5: Short Interest of the Short Factor Legs. The figure shows the value-weighted short interest ($SHROUT_{shorted\ shares}/SHROUT_{total\ shares}$) of the short legs of our synthetic factors compared to the Fama-French one. The short leg of our synthetic factors is composed only of ETFs, so it will be the same for retail and institutional investors. The sample period is from June 2007 to December 2019.

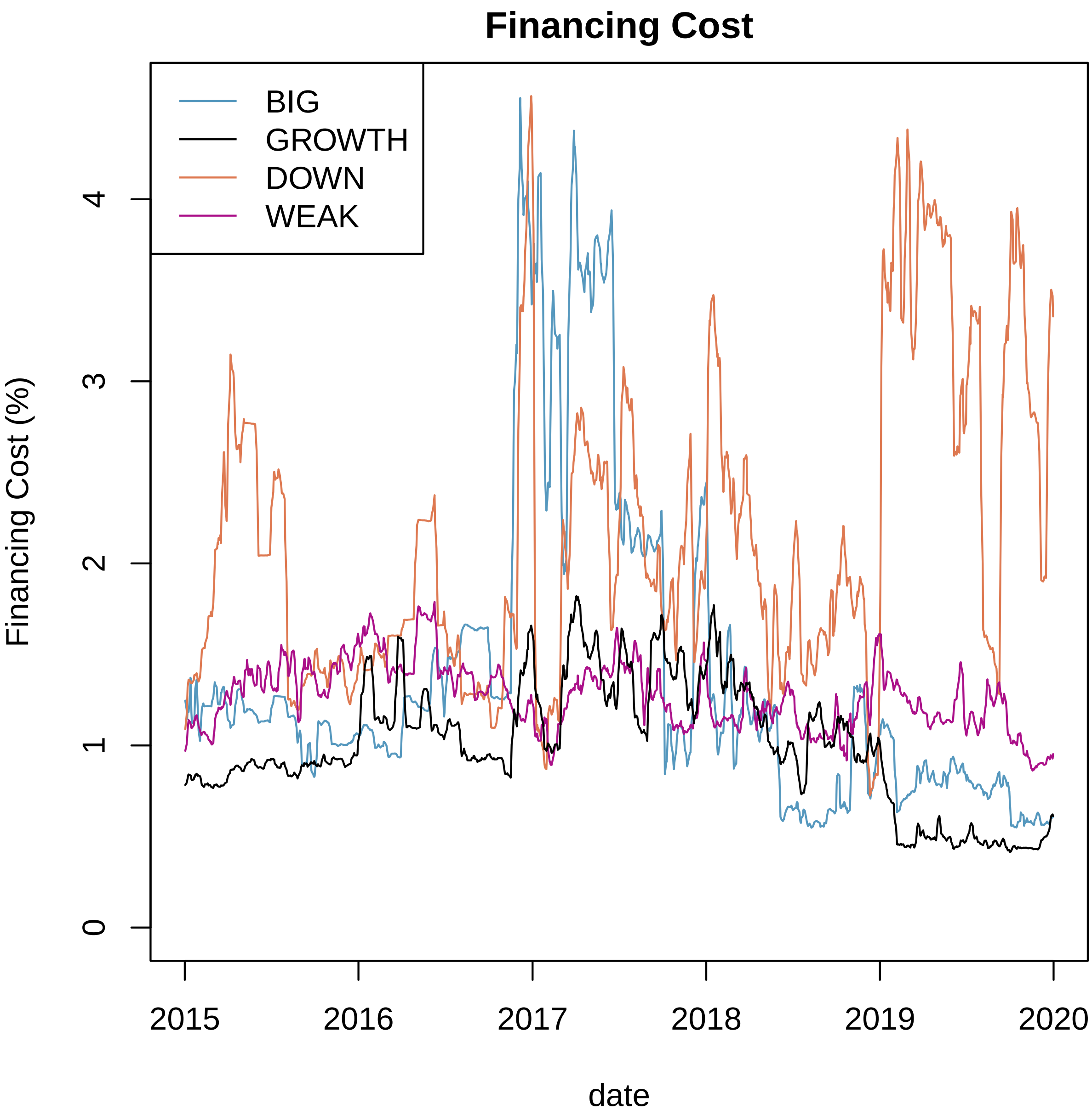

Figure 6: Financing cost of the factors' short legs. The figure shows the value-weighted shorting fees of the short legs of our synthetic tradable risk factors. The short leg of our synthetic factors is composed only of ETFs, so it is the same for retail and institutional investors. The sample period is from January 2015 to December 2019.

these funds will achieve the current median fund size in the synthetic portfolios (e.g., around \$2bn for Small and Momentum, \$4bn for Value, and \$3bn for Profitability). A back-of-the-envelope calculation shows an overall capacity of around \$1tn for SMB, \$2tn for HML, \$450bn for RMW, and \$60bn for MOM, in line with the estimates of Ratcliffe et al. (2017).

6.2 Long-short Factors

We now investigate the ability of our synthetic, tradable long-short portfolios to track the "on-paper" Fama-French risk factors. We run the following regression:

$$ret_{FF_{L-S},t} = \alpha + \beta ret_{synth_{L-S},t} + \varepsilon_t, \tag{2}$$

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.

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.[34] 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 66%. Despite this, the alpha is economically large at 1.30%. The alphas on MOM_{synth} and RMW_{synth} are even larger at 4.46%, and 3.84%, respectively. Differently from SMB_{synth}, the R^2s for MOM_{synth} and RMW_{synth} are also quite low (33% and 21%, respectively). Retail investors also fail to replicate the HML factor: the alpha is 2.41% and the R^2 is only 43%. These results highlight the striking discrepancy between the performance of an optimal, factor-replicating portfolio of funds available to retail investors and the "on-paper" 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 1.57% for HML_{synth} and 1.06% 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.,

[34]Despite being economically large, alphas appear statistically insignificant because of the volatility of daily returns, and the short sample available for our analysis (e.g., only 15 years).

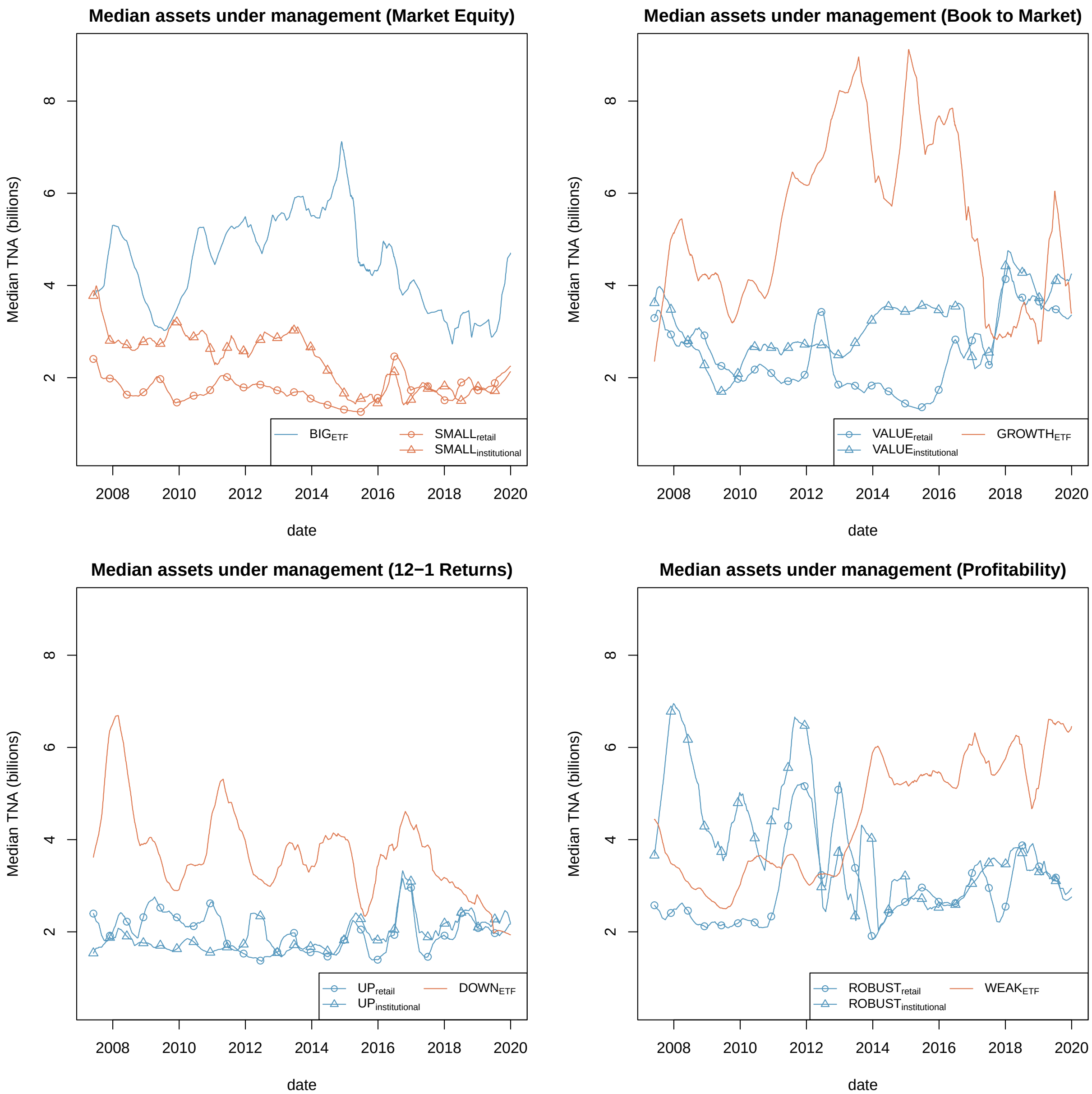

Figure 7: Median net asset value of the funds in the synthetic portfolios. The figure shows the time series of the (six-months smoothed) median NAV of the various synthetic portfolios for both retail and institutional investors. The short leg is constructed using only ETFs, available to both retail and institutional investors. The sample is from June 2007 to December 2019.

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, with alphas of more than 3% per year for both factors. Comparing Panel A to Panel B, we observe larger R^2s for institutional investors for the SMB and HML factors. In particular, the R^2 of HML_{synth} for institutional investors is 33% larger than that attained by the retail investors. Figure 8 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 and institutional investors, together with the cumulative returns of the "on-paper" Fama-French factor that each strategy tries to replicate. The top panels of Figure 8 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 6.3.

Finally, turning to synthetic strategies implemented using only ETFs (Panel C in Table 3), we see a substantial improvement for SMB_{synth}, whose alpha is further reduced to 38 bps. On the other hand, using only ETFs yields an inferior replication of HML relative to institutional investors, with a HML_{synth} alpha of 2.41%. This suggests that mutual funds in the Value leg are quite important in the synthetic replication of the "on-paper" HML factor.

The results of Table 3 confirm the complexity of obtaining in practice the "on-paper" factor risk premia identified in the literature, consistent with Patton and Weller (2020). Overall, institutional investors seem to be able to reap most of the value premium (HML), while both types of investors can successfully earn the size premium using ETFs. The momentum and profitability factors cannot be replicated by either type of investors.

Table 3 presents results on the replicability of long-short "on-paper" factors based on the first moment (e.g., the average return differentials), as it is common in the literature. However, average returns might not be the only valid metric to evaluate the replication performance of synthetic factors. Thus, Table 4 reports summary statistics of higher moments of the Fama-French factors, together with those of our synthetic tradable portfolios, for both

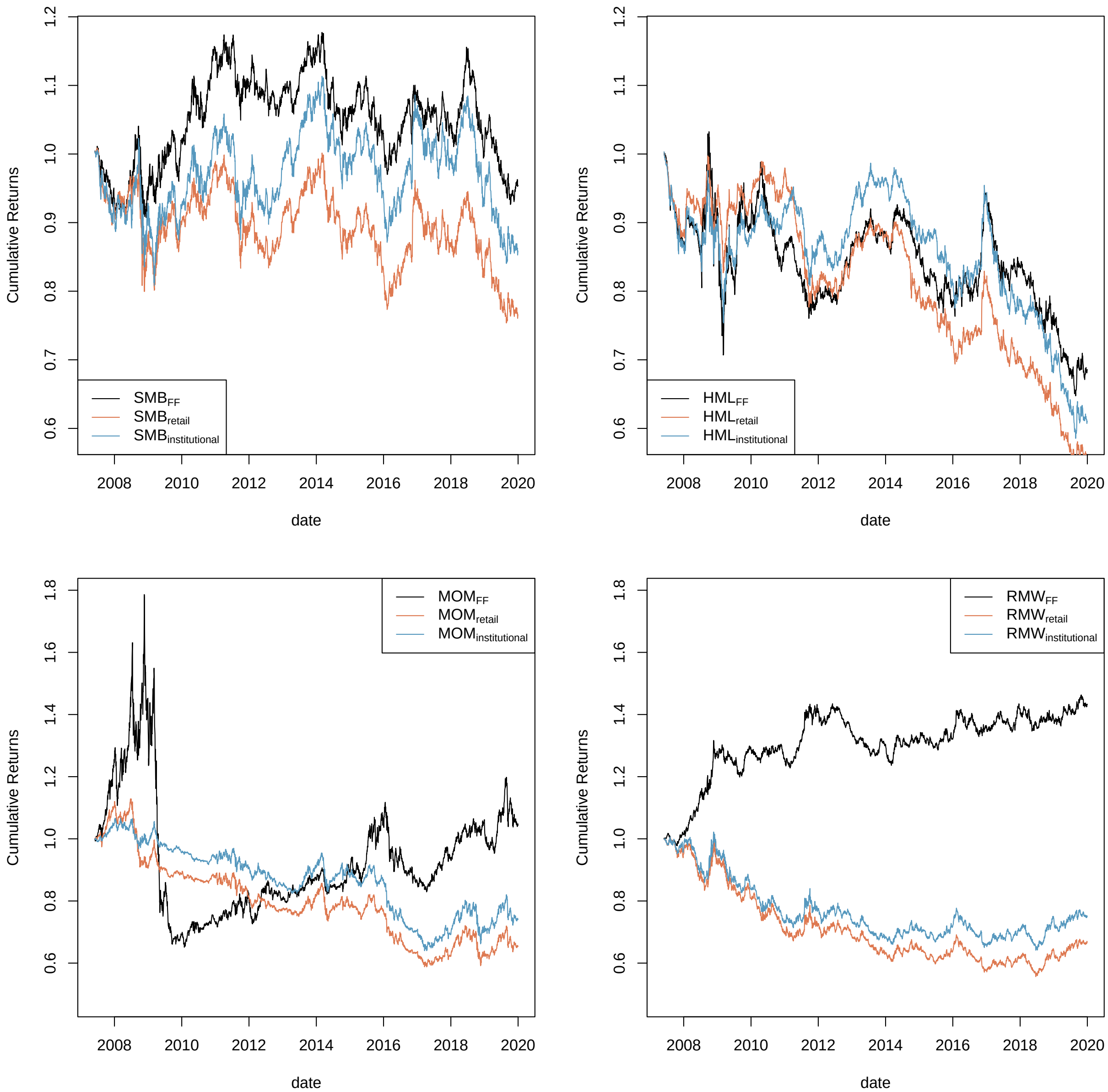

Figure 8: Cumulative returns of the synthetic long-short factors against the Fama-French benchmarks. The figure shows in each panel three cumulative return series: the benchmark Fama-French long-short factor, and both the retail (orange line) and institutional (blue line) synthetic long-short portfolios. The sample period is from June 2007 to December 2019.

types of investors. 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., 2.98 and 4.07), suggesting our synthetic momentum factor faces substantially less tail risk.

The bottom part of both panels in Table 4 presents, for each factor, the ratio of the scores of the long and short legs on four characteristics (size, B/M, past returns, and profitability). We observe that our synthetic factor scores on characteristics other than the one used in the sorting procedure (e.g., cross-exposures) are extremely similar to those of the Fama-French benchmarks.[35] 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 often 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 0.93, which is close to the 0.94 of 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, often close to neutral.

6.3 Drivers of Retail and Institutional Investors' Performance

An important economic question underpinning our results in Section 6.2 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, Figure 9 plots the characteristic scores of the synthetic portfolios for institutional and retail investors, along with the ones of the Fama-French benchmark factors.[36]

[35] The only exception is the exposure of the synthetic SMB to the book-to-market characteristic.

[36] Figure 3 in Lettau et al. (2019) shows a similar analysis, although with some notable differences, namely

For the long leg of SMB and HML, we observe a large score difference between the synthetic and the "on-paper" 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 9 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.

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 "on-paper" Fama-French benchmark. This is true for both types of investors. We conclude that the failure to replicate Momentum and Profitability is largely attributable to the performance of the short legs since funds investing in "losers" or firms with weak profitability are not readily available to investors.

More broadly, our results are consistent with a form of market incompleteness, resulting from investors not being able to achieve proper exposure to the long leg of SMB or HML, or the short legs of MOM and RMW, by trading ETFs and mutual funds.

6.4 Replicating Factors "by Name"

A natural question is whether investors can simply replicate the main "on-paper" factors 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) ETFs whose names include smart beta strategies become available mainly post-2013. These two points together imply that implementing long-short factors using fund names over our full sample is feasible only for the SMB and HML factors.

In contrast, our methodology allows us to construct replicating tradable strategies for

(i) they select funds based on their names, while we calculate the characteristic score of the synthetic factor replicating portfolio; (ii) they show the unconditional distribution of the characteristic scores across 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.

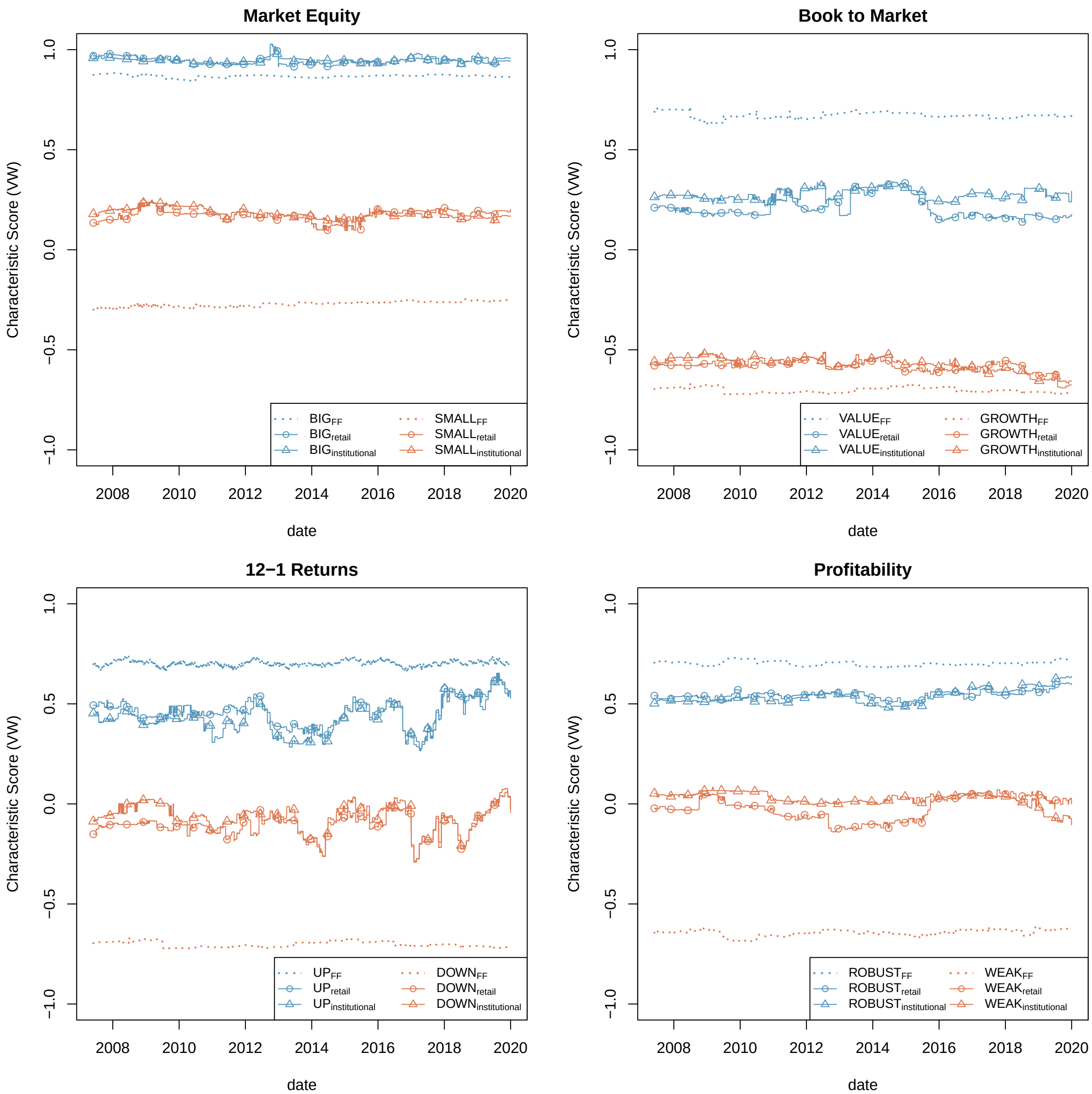

Figure 9: Characteristic scores of the synthetic portfolios. This figure plots the value-weighted characteristic scores of the synthetic portfolios (solid lines), for both retail and institutional investors, together with the Fama-French benchmark factor mimicking portfolio's characteristic scores (dashed 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.

any "on-paper" 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 close to the ones of the underlying benchmarks, so that investors indeed get exposed to a certain characteristic through a smart beta benchmark (e.g., Section 6.2, and Table 4 in particular).

Despite its 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 C.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[37] that are classified as members of a specific strategy to be consistent with our synthetic portfolio. Table 5 reports the tracking performance of the resulting "by name" SMB and HML factors.

Two facts emerge by comparing the results of our synthetic approach (Table 3) to the "by name" replication in Table 5. First, with respect to SMB, we find that the "by name" replication works satisfactorily: the alphas are slightly above 1%, the betas close to one, and the R^2 large at 0.8. When compared to Table 3, we see that SMB_{byName} offers an improvement for retail investors who can lower the alpha by 21 bps, while attaining a smaller tracking error; having said so, our synthetic SMB constructed using ETFs remains the best alternative to track the "on-paper" SMB with an alpha of 38 bps, about half of that obtained using the "by name" approach. Second, and most 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 5.14% to 1.57% when moving from a naive replication "by name" to our synthetic replication based on characteristic scores. We also observe a beta that is close to one in Table 3 but significantly larger than one in Table 5, suggesting a leveraged position.

Figure 10 provides graphical evidence. In particular, each panel shows the returns of 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) panels refer to SMB (HML). The better performance of our synthetic replication for HML is apparent. Also, the volatility of the "by name" HML portfolios is

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

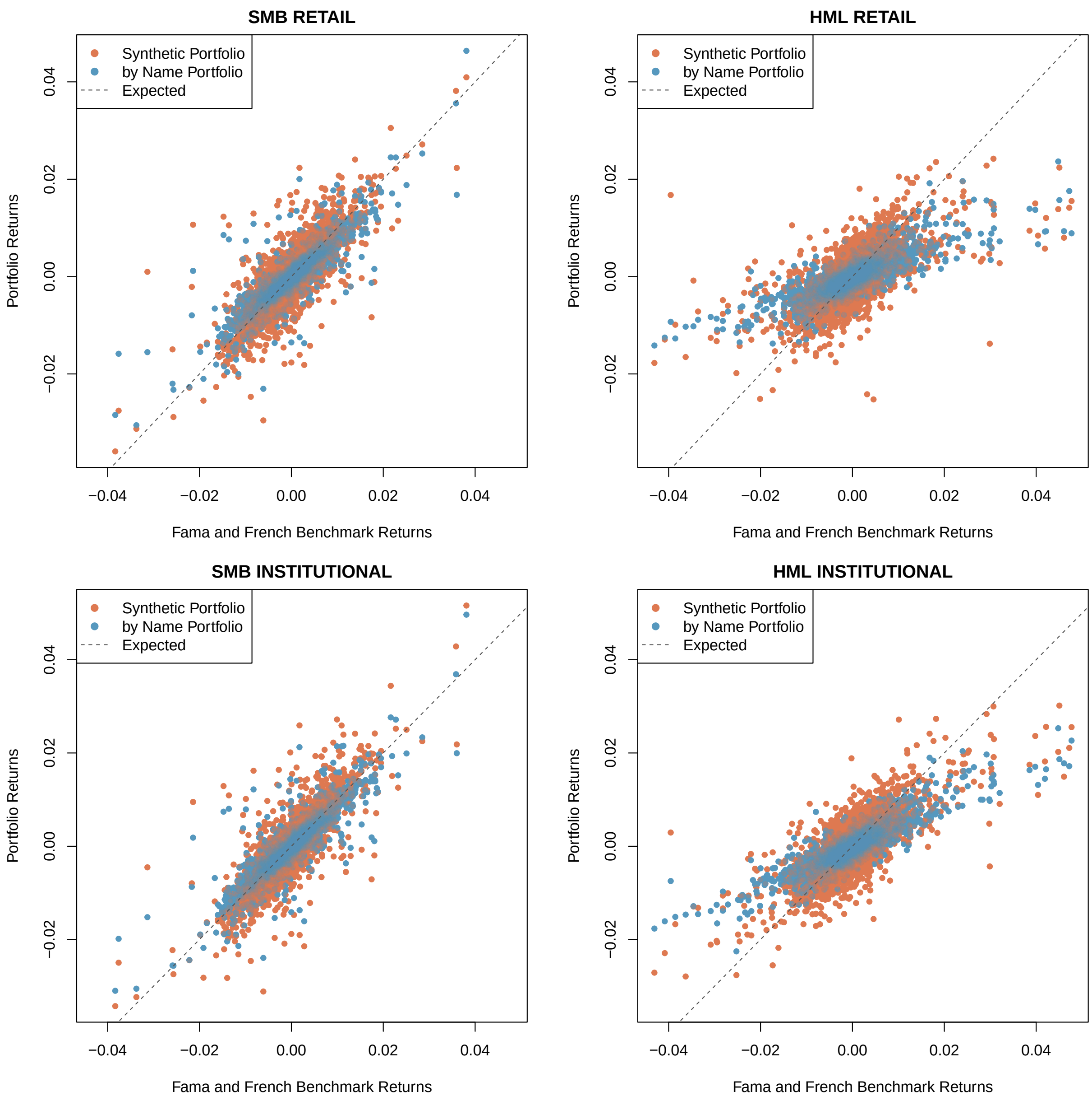

Figure 10: Comparison between the synthetic and "by name" SMB and HML. The figure shows the scatterplot of the returns of our synthetic SMB (left panels) and HML (right panels) portfolios against the respective "by name" version in replicating the Fama-French factors (the expected 45 degree line). The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

about 30% smaller than the one of our synthetic portfolios, resulting in betas for the "by name" strategies that are large at approximately 1.5, as shown in Table 5.

Overall, the difference between HML_{synth} (Table 3) and HML_{byName} (Table 5) is economically relevant. Thus, our procedure to construct synthetic factors is not a mere repackaging of the simpler "by name" strategy. More aggressively, the failure of the "by name" HML portfolio suggests that fund 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.

7 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 ("on-paper" factors). At the same time, a recent literature has documented that replicating "on-paper" factors using individual stocks is hindered by trading costs. Differently from this literature, in this paper we investigate the extent to which "on-paper" factors can be replicated directly by using *portfolio* of stocks (funds). Indeed, the use of mutual funds and ETFs offers several advantages. First, in terms of transaction costs, trading a small number of funds is cheaper than trading hundreds of stocks embedded in "on-paper" academic factors. Second, and most importantly, shorting illiquid, hard-to-borrow stocks, included in the short leg of the "on-paper" factors, is often infeasible. Using ETFs in the short leg of the synthetic factors avoids this problem, since – as we show – ETFs have usually low shorting fees and larger short interest than individual stocks. Therefore, conclusions of the previous literature relying on individual stocks do not automatically extend to our setting. Our analysis provides supportive evidence in favor of partial factor replicability, while also highlighting that the extent of replicability depends on the specific strategy considered, the type of investor (retail or institutional), and the type of assets (mutual funds and ETFs, or ETFs only).

More specifically, we construct synthetic, *tradable* factors using an optimal combination (based on characteristic scores) of smart beta mutual funds and ETFs. Mindful of the fact that open-ended mutual funds cannot be shorted, all our results are obtained by using ETFs in the short legs of the factors. Most importantly, and a key contribution of this paper, we exploit a novel dataset to properly account for ETF-level shorting fees (e.g., borrowing costs) in order to truly construct tradable proxies of "on-paper" risk factors, mimicking the frictions and costs faced by real-world investors.

Our results show that investors are unable to harvest the momentum and profitability risk premia, but differently from the previous literature, we find that it is quite feasible to synthetically replicate the size premium (SMB) using ETFs. Furthermore, institutional investors can harvest the value (HML) premium much 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). Further analysis shows our results to be consistent with a form of market incompleteness, meaning that investors cannot get exposure to the short legs of MOM and RMW, and only limited one to the Value leg of HML, 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, we show that it is important to select funds based on characteristic scores since equity fund names might not be indicative of the actual fund strategies, consistent with the results of Chen et al. (2020) for bond funds.

Overall, 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.

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.

DeMiguel, Victor, Alberto Martin-Utrera, and Raman Uppal, 2021, What alleviates crowding infactor investing? .

Detzel, Andrew, Robert Novy-Marx, and Mihail Velikov, 2021, Model selection with transaction costs, *Working Paper* .

Drechsler, Itamar, and Qingyi Freda Drechsler, 2014, The Shorting Premium and Asset Pricing Anomalies, NBER Working Papers 20282, National Bureau of Economic Research, Inc.

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.

Engelberg, Joseph E., Adam V. Reed, and Matthew C. Ringgenberg, 2018, Short-Selling Risk, *Journal of Finance* 73, 755–786.

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.

Evans, Richard B., Rabih Moussawi, Michael S. Pagano, and John Sedunov, 2021, Etf short interest and failures-to-deliver: Naked short-selling or operational shorting? .

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 J. 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* .

Li, Frank Weikai, and Qifei Zhu, 2019, Short selling etfs, *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.

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

Muravyev, Dmitriy, Neil D. Pearson, and Joshua Matthew Pollet, 2021, Is there a risk premium in the stock lending market? evidence from equity options, *The Journal of Finance* -, –.

Novick, Barbara, Ananth Madhavan, Samara Cohen, Sal Samandar, Sander Van Nugteren, and Alexis Rosenblum, 2017, A primer on etf primary trading and the role of authorized participants, *White Paper Blackrock* .

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.

Panel A: Retail investors				
	MF + ETF		**ETF**	
	2004-2011	2012-2019	2004-2011	2012-2019
Number of Funds	679	711	48	139
Average AUM ($ millions)	6,206	7,320	5,828	8,976
Median AUM ($ millions)	2,596	2,587	3,884	4,496
Median fund age (in sample period)	4.7	5.9	4.8	4.6
Median number of holdings	103	114	422	371
Median return over S&P500 p.a. (%)	0.84	−0.93	1.66	−0.95
Median FF5-factor α p.a. (%)	-0.22	−0.74	0.53	−0.30

Panel B: Institutional investors				
	MF + ETF		**ETF**	
	2004-2011	2012-2019	2004-2011	2012-2019
Number of Funds	391	692	48	139
Average AUM ($ millions)	3,800	6,058	5,828	8,976
Median AUM ($ millions)	1,943	2,434	3,884	4,496
Median fund age (in sample period)	2.8	4.6	4.8	4.6
Median number of holdings	132	130	422	371
Median return over S&P500 p.a. (%)	0.68	−0.72	1.66	−0.95
Median FF5-factor α p.a. (%)	−0.41	−0.53	0.53	−0.30

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

Panel A: Retail investors								
	Small	Big	Value	Growth	Up	Down	Robust	Weak
Mean	0.12	0.10	0.10	0.11	0.23	0.23	0.09	0.10
Std	0.13	0.10	0.11	0.11	0.19	0.23	0.09	0.10
Min	0	0	0	0	0	0	0	0
q0.25	0	0	0	0	0.10	0.05	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	0.30	0.40	0.10	0.10
Max	0.70	0.40	0.50	0.50	0.80	1	0.30	0.40

Panel B: Institutional investors								
	Small	Big	Value	Growth	Up	Down	Robust	Weak
Mean	0.12	0.09	0.07	0.12	0.25	0.25	0.09	0.11
Std	0.11	0.10	0.09	0.11	0.21	0.23	0.10	0.10
Min	0	0	0	0	0	0	0	0
q0.25	0	0	0	0	0.10	0.10	0	0
Median	0.10	0.10	0	0.10	0.20	0.20	0.10	0.10
q0.75	0.20	0.20	0.10	0.20	0.40	0.40	0.10	0.20
Max	0.60	0.40	0.50	0.50	0.90	1	0.40	0.40

Table 2: 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.

Panel A: Retail investors

	α	β	R^2
SMB_{synt}	1.30	0.75***	0.66
HML_{synth}	2.41	0.90*	0.43
MOM_{synth}	4.46	0.97	0.33
RMW_{synth}	3.84***	0.29***	0.21

Panel B: Institutional investors

	α	β	R^2
SMB_{synth}	1.06	0.71***	0.70
HML_{synth}	1.57	0.92	0.57
MOM_{synth}	3.48	0.94	0.28
RMW_{synth}	3.57**	0.29***	0.20

Panel C: ETFs only

	α	β	R^2
SMB_{synth}	0.38	0.74***	0.71
HML_{synth}	2.41	1.13**	0.59
MOM_{synth}	2.94	0.96	0.23
RMW_{synth}	3.39**	0.31***	0.22

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

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

after accounting for ETF shorting costs. 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. The tests on the coefficients are H_0: $\alpha_i = 0$ and H_0: $\beta_i = 1$. The sample period is from June 2007 to December 2019. *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)	0.094	0.001	-0.024	0.016	0.030	-0.017	-0.053	-0.029	-0.028
Std(r)	0.197	0.092	0.110	0.160	0.060	0.099	0.081	0.094	0.095
Skewness(r)	-0.182	0.067	0.625	-0.770	-0.174	0.108	0.244	-0.569	-0.091
Exc. Kurtosis(r)	9.573	4.044	8.733	9.289	2.975	2.722	2.167	2.980	2.578
SR(r)	0.479	0.006	-0.217	0.101	0.506	-0.167	-0.660	-0.311	-0.291
Size		0.390	0.943	1.022	1.131	0.612	0.780	1.033	1.446
B/M		1.037	5.616	0.930	0.766	1.655	2.415	0.796	0.675
Momentum		1.011	0.944	4.530	1.029	0.958	0.898	1.436	1.014
Profitability		0.651	0.810	1.053	4.795	0.734	0.765	1.012	1.391

Panel B: Institutional investors									
	MKT_{FF}	SMB_{FF}	HML_{FF}	MOM_{FF}	RMW_{FF}	SMB_{synt}	HML_{synt}	MOM_{synt}	RMW_{synt}
E(r)	0.094	0.001	-0.024	0.016	0.030	-0.014	-0.043	-0.020	-0.019
Std(r)	0.197	0.092	0.110	0.160	0.060	0.109	0.091	0.090	0.092
Skewness(r)	-0.182	0.067	0.625	-0.770	-0.174	0.223	0.335	-0.626	-0.111
Exc. Kurtosis(r)	9.573	4.044	8.733	9.289	2.975	2.952	2.966	4.069	2.744
SR(r)	0.479	0.006	-0.217	0.101	0.506	-0.132	-0.475	-0.220	-0.202
Size		0.390	0.943	1.022	1.131	0.617	0.806	1.065	1.448
B/M		1.037	5.616	0.930	0.766	1.655	2.556	0.765	0.669
Momentum		1.011	0.944	4.530	1.029	0.993	0.926	1.406	1.036
Profitability		0.651	0.810	1.053	4.795	0.737	0.775	1.048	1.384

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-French and our synthetic portfolios. The sample period is from June 2007 to December 2019.

Panel A: Retail investors			
	α	β	R^2
SMB_{byName}	1.09	1.03	0.80
HML_{byName}	5.01***	1.65***	0.64

Panel B: Institutional investors			
	α	β	R^2
SMB_{byName}	1.15	0.95**	0.81
HML_{byName}	5.14***	1.54***	0.75

Panel C: ETFs only			
	α	β	R^2
SMB_{byName}	0.70	0.91***	0.81
HML_{byName}	3.31**	1.51***	0.76

Table 5: 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 $1bn (in real terms) of assets under management. Table A.1 reports the results using funds with a minimum of $50 million of AUM. The main results in Panel A and B of Table 3 are confirmed to a large extent. First, either class of investors cannot harvest the momentum and profitability premia as confirmed by economically large alphas (greater than 2.8%) and R^2 that continue to stay low even after expanding the set of available funds. Second, 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 relative to the results in Table 3. Quantitatively, institutional investors are able to reap most of the value premium (HML); indeed they almost halve the HML alpha in Panel B of Table 3.

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 it 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.

A.2 Alternative Factor Definitions

A.2.1 "Live" Factors

Table A.2 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 2% for retail investors. On the other hand, institutional investors are better able to get exposure to these factors. Finally, the results in Panel C confirm that the replication of SMB appears somewhat feasible using ETFs only, as confirmed by an alpha of 79 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 $1bn.

A.2.2 CMA and Q-factors

Table A.3 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 investment 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.3 is identical to Table 3. First, the results show that 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 above 3% and by R^2 of at most 26%. Second, and contrary to profitability, the replication of the investment factor of Hou et al. (2015) proves (relatively more) feasible for institutional investors with an alpha of 1.65%. Using ETFs only also provides a viable solution to synthesize the investment factor and results in an alpha of 1.42%. Also, getting exposures to investment risk through replication of the CMA factor (CMA_{synth}) proves more difficult and produces alphas that are larger (and more statistically significant) 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.3 reports the replicating result for this factor. The replication of the quality factor proves difficult with alphas at least as large as 4.5% 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 Fund Fees

Table A.4 reports the replication results for all the benchmark factors discussed in the paper using before-fees fund returns (e.g., gross fund returns). All the alphas across the panels are slightly smaller, reflecting the impact of fund fees in the tradable replicating-factors. Betas and R^2 are unchanged, as the fund fees only represent a shift in the mean (e.g., α). Consistent with the main results in Table 3, the "on-paper" momentum and profitability factors continue to be the hardest to replicate (alphas are still large and economically significant), both for retail and institutional investors. Also, the alpha left over by HML_{synth} continues to be lower for institutional investors than for retail investors. The main difference relative to Table 3 is that the SMB factor appears replicatable by all investors if one is willing to abstract from fund fees.

The irrelevance of fund fees for the replicability of most "on-paper" factors can be better understood in light of the time series of fund fees in our sample. In fact, average mutual fund fees in our sample decline from 0.7% in 2003 to 0.5% in 2009 for institutional investors, and from 1.12% to 0.80% for retail investors, while ETFs fees are stable around 0.20% throughout the whole period. These fees are declining over time and are low compared to the alphas we find in our results, with the exception of the SMB and HML factors. Moreover, the fees on the long and short legs partially offset each other, implying that overall fund fees do not affect our conclusions.

It is important to highlight that, from an investor's perspective, the true opportunity cost of investment is the actual net return obtained from the synthetic factors, net of transaction and shorting costs, and hence net fund returns should be used in constructing tradable proxies of the "on-paper" factors. In conclusion, "on-paper" factors remain hard to replicate, regardless of whether we use net or gross fund returns, for both retail and institutional investors.

A.4 Flows of Smart Beta Strategies

An interesting question stemming from our previous results is whether smart beta investors are aware of the mismatch between fund names and underlying fund holdings. A formal test is to look at whether the flows to the smart beta strategies are coming through funds that should be tracking a smart beta strategy based on their *name*, versus funds that track a strategy *in practice* (e.g., have a characteristic score closer to the one of the "on-paper" factors). In order to do so, we focus on the *daily* flows to smart beta strategies based on a naive (i.e., by looking at the funds' names) and smart (i.e., using the funds selected in our synthetic portfolios) classification. Having access to *daily* flow data, in contrast to the standard monthly or quarterly flow data normally 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:

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

where $S = \{\text{Value, Growth, Small cap, Large cap, Momentum (long)}\}$, $i = \{\text{synthetic, by-names}\}$, $r_{S,i,t}$ 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[1]), 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 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:*

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

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

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:*

$$H_{2a} : \theta_{1,by\_names} > 0; \theta_{5,by\_names} > 0; \theta_{22,by\_names} > 0$$

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

Table A.5 and Table A.6 report the results for retail and institutional investors, respectively. Each pair of columns refers to a particular leg: Small-Big, Value-Growth, and Up and Robust.[2] Even columns report the results for a by-name replication, while odd columns report results for our synthetic smart beta strategies. On average, "naive" strategies (e.g., even columns) tend to display larger R^2 than those of "sophisticated" strategies (e.g., odd columns). In other 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.[3]

The first hypothesis seems to partially hold, perhaps surprisingly, for momentum strategies (column 10) of institutional investors, consistent with a momentum effect itself (e.g.,

[2]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.

[3]The main exception is the Value leg for retail investors, and, to a lesser extent, the Growth leg for institutional investors.

investor flows follow momentum in returns). What is more surprising is that the effect is confined to flows into naive, not synthetic, momentum strategies (22-days momentum) in the case of institutional investors.

There is also some clear return-flow pattern for the Big strategy for both retail and institutional investors, with positive returns inducing positive flows in the "sophisticated" Big but negative ones in the "naive" replication.

Focusing on the second hypothesis for institutional investors, we see that the performance of the benchmark tends to be associated with positive inflows for "sophisticated" Value, for Small (both sophisticated and naive), and "naive" Big, but, perhaps surprisingly, negative for the sophisticated Big.[4] As far as retail investors are concerned, we observe a positive relation for the "naive" Big but a negative one with the "sophisticated" Big strategy, as it was the case for institutional investors. Interestingly, both type of investors appear sophisticated with respect to Value, with a positive performance of the benchmark being associated with an inflow to "sophisticated" rather than "naive" Value.

We also note that, for both types of investors, the last month flows are strongly positively related to the future flows for most strategies, especially those based on the fund names. Also, the average daily flows are positive for most naive ("by name") strategies, and slightly negative for the true synthetic ones. This suggests that investors tend to pay attention to smart beta funds past (short and long-term) flows, and tend to invest in "by name" strateges that have experienced the highest flows.

Overall, we find some support for our 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, with Value being a notable exception. Our conclusions are in line with the findings in Cooper et al. (2005), Chen et al. (2020) and Rakowski and Wang (2009).

[4]Some of these loadings are however not statistically significant.

Panel A: Retail investors

	α	β	R^2
SMB_{synt}	1.48	0.71^{***}	0.74
HML_{synt}	1.74	0.62^{***}	0.43
MOM_{synt}	2.88	0.64^{***}	0.25
RMW_{synt}	2.87^{**}	0.23^{***}	0.21

Panel B: Institutional investors

	α	β	R^2
SMB_{synth}	0.49	0.64^{***}	0.75
HML_{synth}	0.88	0.74^{***}	0.53
MOM_{synth}	3.50	0.72^{***}	0.30
RMW_{synth}	3.10^{**}	0.22^{***}	0.17

Panel C: ETFs only

	α	β	R^2
SMB_{synth}	0.32	0.64^{***}	0.72
HML_{synth}	0.70	0.73^{***}	0.51
MOM_{synth}	3.40	0.87^{***}	0.37
RMW_{synth}	2.80^{*}	0.23^{***}	0.20

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. 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 sample period is from June 2007 to December 2019. *p<0.1; **p<0.05; ***p<0.01.

Panel A: Retail investors

	α	β	R^2
SMB_{live}	2.05	0.74***	0.63
HML_{live}	2.35	0.81***	0.41
RMW_{live}	3.50**	0.26***	0.14

Panel B: Institutional investors

	α	β	R^2
SMB_{live}	1.58	0.69***	0.66
HML_{live}	1.65	0.93	0.58
RMW_{live}	3.59**	0.27***	0.16

Panel C: ETFs only

	α	β	R^2
SMB_{live}	0.79	0.73***	0.70
HML_{live}	2.14	1.09	0.57
RMW_{live}	3.44**	0.29***	0.18

Table A.2: 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. 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 sample period is from June 2007 to December 2019. *p<0.1; **p<0.05; ***p<0.01.

Panel A: Retail investors

	α	β	R^2
CMA_{synt}	1.91^{*}	0.37^{***}	0.31
ROE_{synt}	3.45^{*}	0.38^{***}	0.25
IA_{synt}	1.71	0.38^{***}	0.29
QMJ_{synt}	4.61^{**}	0.50^{***}	0.32

Panel B: Institutional investors

	α	β	R^2
CMA_{synt}	1.87^{*}	0.40^{***}	0.35
ROE_{synt}	3.60^{*}	0.38^{***}	0.26
IA_{synt}	1.65	0.41^{***}	0.32
QMJ_{synt}	4.67^{**}	0.52^{***}	0.32

Panel C: ETFs only

	α	β	R^2
CMA_{synt}	1.59	0.36^{***}	0.31
ROE_{synt}	3.29^{*}	0.38^{***}	0.24
IA_{synt}	1.42	0.38^{***}	0.30
QMJ_{synt}	4.51^{**}	0.51^{***}	0.25

Table A.3: 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. 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 sample period is from June 2007 to December 2019. *p<0.1; **p<0.05; ***p<0.01.

Panel A: Retail investors			
	α	β	R^2
SMB_{synt}	0.16	0.75***	0.66
HML_{synt}	1.31	0.90**	0.43
MOM_{synt}	2.57	0.97	0.33
RMW_{synt}	3.46**	0.29***	0.21

Panel B: Institutional investors			
	α	β	R^2
SMB_{synt}	0.31	0.71***	0.70
HML_{synt}	0.73	0.92	0.57
MOM_{synt}	1.95	0.94	0.28
RMW_{synt}	3.26**	0.29***	0.20

Panel C: ETFs only			
	α	β	R^2
SMB_{synt}	-0.33	0.74***	0.71
HML_{synt}	1.54	1.13**	0.59
MOM_{synt}	1.69	0.96	0.23
RMW_{synt}	3.07**	0.31***	0.22

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

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

using fund gross returns (e.g., before fees). 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. The tests on the coefficients are H_0: $\alpha_i = 0$ and H_0: $\beta_i = 1$. The sample period is from June 2007 to December 2019. *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	0.005		0.182***		0.001		0.008		−0.001		0.033*	
	(0.042)		(0.028)		(0.011)		(0.011)		(0.011)		(0.017)	
(S) 5-days avg ret	−0.005		−0.027		0.079***		−0.024		0.010		0.041	
	(0.109)		(0.076)		(0.027)		(0.029)		(0.029)		(0.048)	
(S) 22-days avg ret	0.208		0.394**		−0.161***		0.004		0.073		0.037	
	(0.203)		(0.162)		(0.055)		(0.062)		(0.054)		(0.095)	
(S) 1-day flow	0.108***		0.012		−0.127***		−0.060***		−0.117***		0.061***	
	(0.020)		(0.020)		(0.020)		(0.020)		(0.020)		(0.020)	
(S) 5-days flow	−0.195***		−0.023		0.108**		−0.194***		−0.023		−0.095**	
	(0.048)		(0.049)		(0.053)		(0.055)		(0.056)		(0.047)	
(S) 22-days flow	−0.294**		0.038		0.540***		0.573***		0.382***		−0.152	
	(0.122)		(0.095)		(0.073)		(0.093)		(0.087)		(0.104)	
(N) 1-day ret		−0.027		0.007		0.005		0.002		0.011		0.005
		(0.046)		(0.009)		(0.005)		(0.003)		(0.021)		(0.004)
(N) 5-days avg ret		−0.167		0.018		0.006		−0.004		0.026		−0.014
		(0.119)		(0.022)		(0.015)		(0.007)		(0.054)		(0.010)
(N) 22-days avg ret		−0.203		−0.116**		−0.060**		−0.010		0.151		0.034*
		(0.238)		(0.047)		(0.028)		(0.014)		(0.103)		(0.020)
(N) 1-day flow		0.112***		0.022		0.045**		0.079***		0.011		0.220***
		(0.021)		(0.021)		(0.020)		(0.020)		(0.023)		(0.019)
(N) 5-days flow		−0.069		0.222***		−0.052		−0.043		0.193***		−0.312***
		(0.048)		(0.046)		(0.048)		(0.046)		(0.056)		(0.046)
(N) 22-days flow		−0.427***		0.276***		0.419***		0.197**		0.371***		0.440***
		(0.118)		(0.063)		(0.076)		(0.082)		(0.073)		(0.081)
benchmark 1-day ret	0.002	0.020	−0.152***	−0.006	0.002	−0.005	−0.010	−0.002	0.002	−0.008	−0.021	−0.006*
	(0.040)	(0.045)	(0.025)	(0.008)	(0.009)	(0.004)	(0.011)	(0.003)	(0.011)	(0.020)	(0.015)	(0.003)
benchmark 5-days ret	0.013	0.158	0.067	−0.003	−0.028	0.003	0.044	0.009	0.016	0.017	−0.009	0.015*
	(0.102)	(0.115)	(0.065)	(0.020)	(0.023)	(0.012)	(0.028)	(0.007)	(0.028)	(0.051)	(0.040)	(0.009)
benchmark 22-days ret	−0.136	0.246	−0.393***	0.090**	0.150***	0.036	−0.020	0.004	−0.052	−0.081	−0.036	−0.025
	(0.190)	(0.228)	(0.141)	(0.044)	(0.048)	(0.022)	(0.059)	(0.013)	(0.055)	(0.092)	(0.077)	(0.018)
Observations	3,169	3,169	3,169	3,169	3,169	3,169	3,169	3,169	3,169	2,597	3,169	3,169
Adjusted R^2	1.5%	1.2%	2.1%	5.1%	6.3%	1.7%	1.6%	1.2%	2.3%	5.5%	0.8%	5.1%

Table A.5: 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, $r_{bench,t}$ 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 Internet Appendix B, 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	0.188**		0.128***		0.001		0.006		0.006		0.026	
	(0.085)		(0.027)		(0.006)		(0.008)		(0.011)		(0.017)	
(S) 5-days avg ret	−0.608***		0.031		0.024		0.027		0.034		0.015	
	(0.216)		(0.073)		(0.016)		(0.022)		(0.028)		(0.046)	
(S) 22-days avg ret	−0.276		0.390**		−0.150***		−0.040		0.078		0.127	
	(0.416)		(0.159)		(0.033)		(0.047)		(0.061)		(0.094)	
(S) 1-day flow	0.110***		0.010		0.061***		−0.149***		−0.094***		0.040**	
	(0.021)		(0.020)		(0.020)		(0.020)		(0.020)		(0.020)	
(S) 5-days flow	−0.025		−0.009		0.083*		0.081		0.099*		−0.004	
	(0.047)		(0.049)		(0.046)		(0.054)		(0.052)		(0.047)	
(S) 22-days flow	−0.314***		0.059		0.259***		0.699***		0.372***		−0.030	
	(0.109)		(0.093)		(0.070)		(0.072)		(0.078)		(0.093)	
(N) 1-day ret		0.047		0.006		0.010		0.007*		0.019		0.005
		(0.063)		(0.011)		(0.007)		(0.004)		(0.035)		(0.009)
(N) 5-days avg ret		−0.366**		0.027		−0.004		−0.0004		0.086		−0.024
		(0.162)		(0.027)		(0.018)		(0.012)		(0.088)		(0.025)
(N) 22-days avg ret		−0.513		−0.168***		−0.048		−0.018		0.388**		0.044
		(0.329)		(0.057)		(0.034)		(0.024)		(0.181)		(0.049)
(N) 1-day flow		0.110***		0.056***		0.045**		0.049**		−0.027		0.113***
		(0.021)		(0.021)		(0.020)		(0.020)		(0.030)		(0.020)
(N) 5-days flow		−0.036		0.293***		0.158***		0.102**		0.208***		−0.152***
		(0.048)		(0.044)		(0.047)		(0.045)		(0.072)		(0.048)
(N) 22-days flow		−0.277**		0.320***		0.466***		0.434***		0.344***		0.707***
		(0.108)		(0.053)		(0.059)		(0.064)		(0.097)		(0.068)
benchmark 1-day ret	−0.197**	−0.053	−0.102***	−0.005	0.004	−0.004	−0.007	−0.005	−0.006	−0.011	−0.018	−0.003
	(0.085)	(0.063)	(0.023)	(0.010)	(0.006)	(0.005)	(0.008)	(0.004)	(0.010)	(0.032)	(0.015)	(0.008)
benchmark 5-days ret	0.572***	0.345**	0.013	−0.010	−0.009	0.011	0.004	0.011	−0.006	0.001	0.021	0.025
	(0.214)	(0.158)	(0.061)	(0.025)	(0.015)	(0.014)	(0.021)	(0.011)	(0.027)	(0.080)	(0.039)	(0.020)
benchmark 22-days ret	0.313	0.523	−0.383***	0.129**	0.135***	0.016	0.009	0.004	−0.050	−0.200	−0.115	−0.052
	(0.410)	(0.320)	(0.135)	(0.054)	(0.031)	(0.028)	(0.045)	(0.022)	(0.057)	(0.153)	(0.076)	(0.038)
Observations	3,169	3,169	3,169	3,169	3,169	3,169	3,169	3,169	3,169	1,490	3,169	3,169
Adjusted R^2	1.3%	1.2%	1.6%	11.8%	5.0%	8.5%	6.2%	5.6%	2.7%	6.3%	0.7%	6.0%

Table A.6: 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, $r_{bench,t}$ 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 Internet Appendix B, 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.

Internet Appendix B 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.

B.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 misleading. The Securities and Exchange Commission (SEC) states[5] 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*

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

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.[6] Our hope is that our methodology can help address this problem, which severely distorts investors' allocation decisions.

In Section 6.4 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.[7] 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

[6]See here.

[7]We randomly pick this fund among the set of funds with similar inconsistencies. This should only be viewed as an illustrative example, not as an evaluation of the fund's performance.

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.

B.2 Identify Market Funds: Minimum Distance Approach

A common problem in identifying smart beta funds is related to closet indexing[8] (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).

In this section, we show how one can identify 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

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

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

Table B.1 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, 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

[8] 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.

have loadings "close" to the ones of the individual relative legs estimated in (B.2). As a consequence, in the second step, we re-estimate regression (B.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 (B.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.[9]

In Figure B.1, 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 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

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

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 B.1 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.[10]

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 time.

Whether a fund is classified as a market fund will depend on the choice of risk-factors used in regression (B.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

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

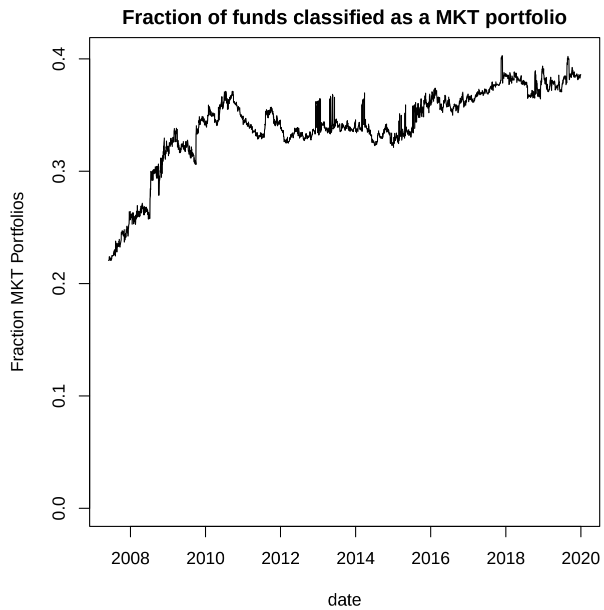

Figure B.1: Fraction of funds classified as a market portfolio. This figure shows the fraction of funds classified as market funds by our methodology in real-time using an expanding window. A fund is classified as a market fund if it has the minimum F-stat with respect to the market portfolio. The sample is from June 2007 to December 2019.

exposure to, and only include those in the regressions. In other words, our methodology is applicable under any chosen set of risk factors.

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 B.1: 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.

Internet Appendix C 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).

C.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.[11] Note that all analysis using daily flows are based on the EPFR data, which

[11] 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.

contains 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.

C.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[12] 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.

[12]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.

C.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'.