The Smart Beta Mirage

Shiyang Huang, Yang Song, and Hong Xiang*

Abstract

We document and explain the sharp performance deterioration of smart beta indexes after the corresponding smart beta ETFs are launched for investment. While smart beta claims to deliver excess returns through factor exposures, the market-adjusted return of smart beta indexes drops from 2.77% per year "on paper" before ETF listings to −0.44% after ETF listings. This performance decline cannot be explained by strategic timing, the declining trend in factor premia, or diminishing returns to scale. Instead, we find strong evidence of data overexploitation in constructing smart beta indexes, which helps ETFs attract flows as investors respond positively to stellar backtests.

*We thank Constantinos Antoniou, Hank Bessembinder, Itzhak Ben-David, Philip Bond, Darrell Duffie, Jiacui Li, Tyler Muir, Jeff Pontiff, Stephan Siegel, Yuehua Tang, Yao Zeng, Zhuo Zhong, several Blackrock and Vanguard professionals, and seminar participants at University of Amsterdam, Erasmus university Rotterdam, University of Warwick, University of Washington, The University of Hong Kong, Virtual Asset Management Seminar, Victoria University of Wellington, Renmin University of China, the joint seminar by University of Melbourne and the Financial Research Network (FIRN), and The University of International Business and Economics for helpful comments for helpful comments. Huang and Xiang are with HKU Business School of The University of Hong Kong, and Song is with University of Washington. Email: huangsy@hku.hk, songy18@uw.edu, and hongxiang18@hku.hk.

I. Introduction

Exchange-Traded Funds (ETFs) have experienced tremendous growth during the past two decades. In the ETF marketplace, the fastest-growing and possibly most discussed segment is "smart beta" ETFs, which ETF sponsors often refer to as one of the most important financial innovations supported by academic research.[1] Unlike traditional ETFs that track broad cap-weighted market indexes, smart beta ETFs are designed to track non-cap-weighted indexes that are built upon formulaic rules (known as smart beta indexes), claiming to deliver excess returns through systematic exposures to certain asset pricing factors/anomalies, e.g., value and momentum. Since their first appearance in 2000, the assets under management (AUM) of US equity smart beta ETFs have increased to about $700 billion as of 2018, accounting for more than 20% of the aggregate ETF market (Morningstar, 2019). Blackrock also forecasts that the AUM of equity smart beta ETFs will grow to more than $2 trillion by 2025 (Bloomberg, 2016).

One key attribute defining smart beta ETFs is the so-called "rules-based index construction."[2] Although smart beta indexes are claimed to be built upon pre-determined formulaic rules, the degree of freedom over smart beta index construction is enormous in practice. Given the large-scale data availability and tremendous amount of computing power one may possess, it is not difficult to develop a smart beta index with stellar backtested returns from thousands, if not millions, of trials.

Indeed, backtested index performance is widely used for promoting smart beta ETFs, and practitioners highlight that backtests play a key role in launching and marketing smart beta ETFs (Morningstar, 2017; Financial Times, 2017; ETF Stream, 2019).[3] Presumably, the better the backtested performance of smart beta indexes, the easier it is to sell ETFs

[1] For example, Investco (2018) cites more than ten academic papers to endorse its smart beta ETFs.

[2] While some early smart beta ETFs were built to track existing indexes, in the majority of cases, ETF sponsors either construct smart beta indexes by themselves or collaborate with third-party index providers to develop indexes. For example, JP Morgan US Momentum Factor ETF states in its prospectus,"The JP Morgan US Momentum Factor Index uses a rules-based risk allocation and factor selection process developed by J.P. Morgan Asset Management."

[3] For example, Morningstar (2017) notes that marketing of smart beta relies almost exclusively on backtested data possibly due to the lack of sufficient real-life track record. A senior market participant also writes in ETF Stream (2019),"The purpose of backtesting is to simulate the historical performance of a strategy and the results are commonly used for marketing investment products."

to investors.[4] This is not surprising as ETF investors have been shown to chase past performance (e.g., Dannhauser and Pontiff, 2020), yet they might fail to differentiate between hypothetical backtested returns and "real" returns.

The substantial discretion over smart beta index construction, together with data mining incentives for stellar backtests, raises several important questions: To what extent is the construction of smart beta indexes susceptible to data overexploitation? Does the claimed "smartness" of smart beta exist out of backtests? Has the rapid proliferation of smart beta ETFs created value for investors? Answers to these questions are urgent as hundreds of billions of dollars have been poured into smart beta ETFs, and trillions more are expected to follow in the coming years.[5] Moreover, ETF sponsors typically market their smart beta ETFs as "smart" financial innovations endorsed by academic research.

To address these questions, we compile a sample of US domestic equity smart beta ETFs listed between 2000 to 2018 and manually collect their underlying smart beta indexes and the index returns before and after ETF listing. Our sample is representative as it covers about 80% of the total AUM of the US equity smart beta ETFs classified by Morningstar. Note that before ETF listings, smart beta index returns are not accessible to investors. Thus, we refer to the index returns before the ETF listing as "on paper" or backtested returns. We refer to the index returns after listing as "real" or "live" returns.

We document several novel findings in this paper. Probably the most disappointing finding to smart beta investors is that the claimed "smart" performance of smart beta indexes exists only in backtests, and smart beta index performance deteriorates dramatically after ETF listings. For example, smart beta indexes earn a significantly positive "on paper" CAPM alpha of 2.77% (t = 4.59) per year over an average 13-year period before ETF listings.[6] Once smart beta ETFs are launched to investors, we observe a substantial decline

[4]Vanguard (2012), for example, claims that hypothetical performance data helps attract investment flows and contributes to the viability of new ETFs.

[5]For comparison, as of 2018, the total assets of US domestic equity active mutual funds and closed-end funds are $4.65 trillion and $68 billion, respectively. Moreover, the total size of active equity mutual funds and closed-end funds has been stable over the past decade (Investment Company Institute, 2019).

[6]We follow Fama and French (2015) and use the value-weighted returns of all CRSP US firms as the market returns. This choice is reasonable as investors can invest in this market index through the Vanguard Total Stock Market ETF (VTI) at a very low cost (i.e., 3 basis points (bps) per year). When we use returns of the SPDR S&P 500 ETF (SPY) as market returns, we can get even stronger results.

in the index performance. Specifically, after ETF listings, smart beta indexes have a negative CAPM alpha of -0.44% per year over an average 6-year post-listing period. The performance difference between the pre-listing and the post-listing periods (referred to as the performance decline hereafter) is 3.22% per year ($t = 4.63$). We also conduct a large set of robustness checks that further confirm this finding.

It is worth noting that we mainly use smart beta index returns rather than ETF returns to measure the post-ETF-listing performance. There are two advantages of using smart beta index returns: First, ETF returns do not exist before ETF listings, and focusing on index returns makes the pre- and post-listing performance directly comparable. Second, since we use smart beta index returns to gauge the post-listing performance, it is not the transaction costs in replicating indexes that drive the performance decline.[7] When we use ETF returns over the post-ETF-listing period, we even observe a slightly sharper performance decline.

We explore four plausible explanations for the sharp decline in smart beta index performance after ETFs are launched to investors: (*i*) strategic timing of ETF listings, (*ii*) declining trend in factor premia, (*iii*) diminishing returns to scale caused by ETF flows, and (*iv*) data overexploitation in index construction. Our results suggest that the first three explanations fail to explain the performance decline. Instead, we find strong support for data mining in index construction.

We first show that the "strategic timing" hypothesis cannot explain our findings. Under this hypothesis, ETF sponsors strategically launch products when the tracked factors have recently outperformed the aggregate market. Possibly because factor returns are mean-reverting, the performance of smart beta indexes is not sustained after ETF listings. To examine this explanation, we control for factor-wide return fluctuations, including the potential mean-reversion of factor returns. Specifically, we categorize smart beta indexes by the Morningstar-classified factor themes and then benchmark each index relative to the earliest-constructed index in its factor-theme category.[8] After controlling for factor-wide return fluctuations, the magnitude of the performance decline (3.27%) is similar to that measured

[7] In fact, smart beta ETFs track their underlying indexes very closely. The average correlation between monthly ETF returns and index returns is 0.99 in our sample.

[8] For example, value smart beta indexes are benchmarked to the first value smart beta index (the Russell 1000 Value Index), and growth smart beta indexes are benchmarked to the Russell 1000 Growth Index.

by CAPM alpha (3.22%), suggesting that the "factor timing" hypothesis cannot explain the sharp deterioration in smart beta index performance after ETF listings.

The second explanation is that factor profitability declines over time (e.g., McLean and Pontiff, 2016), and the post-ETF-listing return decline is a natural outcome of this general trend. We argue that this explanation also fails to work. First, by adjusting for factor-theme benchmark returns, we have effectively controlled the time trend in factor profitability, and we get similar results. Second, quantitatively, the smart beta index performance decline is much larger than the decline in factor premia documented in prior studies. For example, the average annualized smart beta performance drops by more than 300 bps after ETF listings. In contrast, Chordia, Subrahmanyam, and Tong (2014) estimate that the average profitability of twelve well-studied factors declines by about 30 bps per year.

Another possibility is that investment flows into smart beta ETFs could make factor-based strategies over-crowded, and consequently, hurt the performance of smart beta indexes after ETF listings due to decreasing returns to scale. We show that this is also not the case. Specifically, we find that ETFs with smaller average post-listing AUM either across all ETFs or within the factor-theme category experience an even larger performance decline, which is inconsistent with decreasing returns to scale.

As these seemingly plausible explanations fail to work, we explore the "data mining" hypothesis. We find strong support that the construction of smart beta indexes involves data overexploitation so that backtested performance is not sustained after ETFs go live. To illustrate data mining in action, Figure 1 plots cumulative returns of two smart beta indexes of the same factor theme constructed by one of the world's major index providers: the XYZ Factor-A Index and XYZ Enhanced Factor-A Index.[9] The XYZ Factor-A Index was released in 1997. After years of no excess returns relative to the market, the index provider constructed the XYZ Enhanced Factor-A Index in December 2014 to presumably "enhance" returns. A smart beta ETF tracking this "enhanced" index was listed shortly thereafter. Ironically, the "enhanced" performance only occurred in the backtesting period before the ETF listing. After the ETF tracking this "enhanced" index was launched, the

[9]We choose not to disclose the name of the index provider. The cumulative returns here are relative to an aggregate market index from XYZ.

"on paper" stellar performance of the "enhanced" index quickly disappeared and trailed the aggregate market.

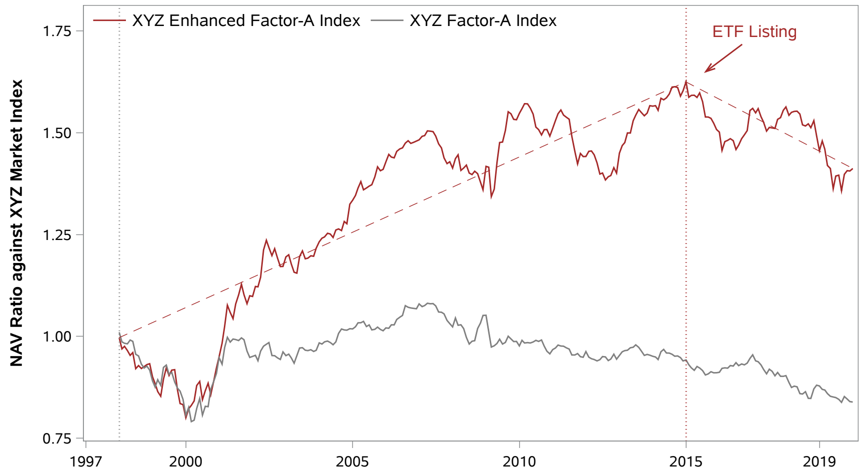

Figure 1: **An Indicative Case of Data Mining.** This figure plots the cumulative NAV ratios of two smart beta indexes of the same factor theme relative to an aggregate market index. All three indexes are from one of the world's largest index providers, whose name we choose not to disclose. The first smart beta index (in grey) was constructed in December 1997. The "enhanced" version of the first index (in red) was constructed in December 2014, and the ETF tracking this index was listed shortly. Note that the returns of the "enhanced" index before the ETF listing are "on paper" and are not investable.

To show that data mining largely drives the performance decline of smart beta indexes more generally, we provide three pieces of evidence.

First, we explore the opaqueness of smart beta indexes. We argue that the so-called "multi-factor" indexes are more susceptible to data mining than single-factor indexes. This is because multi-factor indexes can freely combine multiple factors so that the degree of freedom in building multi-factor indexes is higher by nature. Moreover, the category of multi-factor smart beta ETFs has the largest number of ETF offerings (Morningstar, 2019), implying intense competition and thus heightened incentives to data mine. Meanwhile, even among single-factor indexes, value and growth indexes should have relatively less space for data mining as they are built on the most-studied factor. Consistent with our conjecture, we find that before ETF listings, multi-factor indexes have the best backtests and the largest post-listing performance decline. In contrast, value and growth indexes have the smallest performance decline after ETF listings.

Second, we explore the discretion of ETF sponsors on index construction. As stellar backtests can help ETF sponsors attract flows, when ETF sponsors have more control over index construction, we should observe higher backtested returns and a sharper performance decline.[10] To examine this conjecture, we classify smart beta indexes into "high-ETF-discretion" and "low-ETF-discretion" groups. A "high-ETF-discretion" index should satisfy at least one of the two non-exclusive conditions: (*i*) the index is constructed "in-house" by the ETF sponsor, and (*ii*) the index name contains the ETF sponsor's name or the index release date is within six months of the ETF listing date.[11] Although the second criterion is crude, it suggests that a smart beta index is tailored for the ETF and thus indicates the ETF sponsor's influences on index construction. Align with our conjecture, the performance decline for "high-ETF-discretion" smart beta indexes is much larger than "low-ETF-discretion" indexes.

Third, we explore the heterogeneity in ETF sponsor size. We argue that because small ETF sponsors are keen to increase market share, they are likely to have stronger desires for stellar backtested performance than large ETF sponsors, who may care more about reputation. Indeed, we find evidence that the smart beta indexes used by small ETF sponsors display a larger performance decline after ETF listings than those used by large sponsors.

Thus far, our evidence is consistent with that data mining in indexation leads to a large performance decline after the launch of smart beta ETFs. One natural and important question is whether ETF investors are aware of the index data mining. To answer this question, we examine the response of ETF flows to pre-ETF-listing index returns. We find that ETF investors respond strongly to the backtested performance, consistent with the claim of Vanguard (2012) that backtests contribute to the viability of new ETFs. For example, a one-standard-deviation increase in pre-listing index returns is associated with an increase in monthly ETF flows of 6% over the first post-listing year. This effect is economically significant as the sample median is 11% per month over this period. The strong response of ETF flows to the "on paper" returns also justifies the practice of data overexploitation in

[10]When ETF sponsors collaborate with third-party index providers to construct smart beta indexes, ETF sponsors can involve in index construction extensively, and index providers are likely to cater to ETF sponsors' demand (Financial Times, 2015; Pensions & Investments, 2018).

[11]About 60% smart beta indexes are classified as high-ETF-discretion based on these criteria. Our results are not sensitive to the time window in condition (*ii*).

index construction, as this behavior gets rewarded through flows.[12]

Our finding that smart beta indexes trail the aggregate market when they become accessible to investors suggests that the majority of smart beta ETFs do not add value to investors in terms of returns. However, one might argue that the proliferation of smart beta ETFs is still valuable as new smart beta ETFs might help investors get higher or cheaper exposure to asset pricing factors. Unfortunately, our further analyses show that this is not the case because later-constructed smart beta indexes mostly have *lower* average exposure to the designated factor than the first-constructed index in the same factor-theme category. Moreover, investors need to pay significantly higher fees to access the later-constructed indexes.

In summary, our findings raise concern on the proliferation of smart beta ETFs because the majority of smart beta ETFs add little value to investors both in terms of excess returns beyond the broad market and in terms of desired factor exposures. The claimed "smart" performance of smart beta seems to be a mirage that only exists in backtests. We also extend our study to other developed markets, including Europe, the UK, Australia, and Canada, and we find that the post-listing decline in smart beta performance is ubiquitous.

Admittedly, we cannot entirely rule out the possibility that ETF sponsors and index builders are unaware that they overexploit data when developing smart beta indexes. That is, although ETF sponsors or index builders have set up intricate and arbitrary rules in smart beta indexation, they believe that backtests best represent future performance. If this is the case, investors might need to be even more cautious about smart beta ETFs, as failure to recognize data mining suggests incompetence.

The rest of the paper is organized as follows. Section II briefly discusses the literature. Section III introduces smart beta ETFs and our data collection process. Section IV shows that returns of smart beta indexes decline dramatically after ETFs listings and rules out several explanations for this performance dropoff. Section V provides evidence of data mining in smart beta index construction and shows that this behavior helps ETFs attract flows. Section VI provides supporting evidence from other developed markets. Section VII concludes. Additional results and robustness checks are provided in the appendix.

[12] In Appendix A, we further show that flows into smart beta ETFs significantly negatively predict future ETF performance, reinforcing that ETF investors are likely to be unsophisticated.

II. Literature Review

Our paper is related to several strands of literature. First, our paper contributes to the literature that studies asset managers' strategic behavior. For example, prior studies document that asset managers deviate from their claimed investment policies (e.g., Chan, Chen, and Lakonishok, 2002; Brown, Harlow, and Zhang, 2009; Wermers, 2012), misreport their portfolios (e.g., Bollen and Pool, 2009, 2012; Cici, Gibson, and Merrick Jr, 2011; Agarwal, Barber, Cheng, Hameed, and Yasuda, 2019; Chen, Cohen, and Gurun, 2020), take excess risk to boost performance (e.g., Carhart, Kaniel, Musto, and Reed, 2002), and obfuscate fee structures and disclosures (e.g., Barber, Odean, and Zheng, 2005; Edelen, Evans, and Kadlec, 2012). To the best of our knowledge, our study is the first to analyze the possibility of agency frictions in ETF index construction and the resulting implications for ETF investors.[13] While we focus on smart beta ETFs in this study, our findings also have important implications for other segments of ETFs, such as industry ETFs and thematic ETFs.

In another important contribution to this literature, Evans (2010) shows that mutual fund companies use an incubation strategy when launching new mutual funds, which harms mutual fund investors.[14] Although both strategies serve to attract investment flows, data mining in ETF indexation and mutual fund incubation are fundamentally different in several aspects. During a mutual fund incubation trial, a limited number of investors or the investment company deploy real capital into the tested mutual funds. Thus, the incubation strategy is based on real track records. In contrast, data mining is entirely "on paper," as it searches for "superior" returns from historical data and does not require any capital. Possibly because real capital is invested during mutual fund incubation, the incubation period is relatively short, mostly less than three years (Evans, 2010). However, the average length of "on paper" backtested index returns is about 13 years in our sample.

[13] Suhonen, Lennkh, and Perez (2017) analyze 215 rules-based investment strategies across five asset classes offered by 15 investment banks. Similar to our findings on smart beta ETFs, they observe significant performance dropoffs when these strategies go live from backtests. Several practitioners (Li and West, 2017; Pattabiraman, 2020) also observed that smart beta indexes experienced return declines after ETFs are listed and suspected the possibility of data mining. Neither of them examines the underlying mechanism nor explores the influences on investment flows.

[14] Mutual fund incubation is a strategy that a mutual fund company uses to test multiple new funds with only outperforming funds are open to the public.

Our paper also contributes to the growing literature on ETFs, particularly on the debates over the bright and dark sides of ETFs. On one hand, some argue that ETFs increase asset volatility and harm liquidity (e.g., Israeli, Lee, and Sridharan, 2017; Ben-David, Franzoni, and Moussawi, 2018; Da and Shive, 2018; Agarwal, Hanouna, Moussawi, and Stahel, 2018; Pan and Zeng, 2019) and ETF investors often make sub-optimal decisions (e.g., Brown, Cederburg, and Towner, 2020). A contemporary paper by Johansson, Sabbatucci, and Tamoni (2020) finds that investors cannot effectively harvest factor premia through smart beta ETFs. On the other hand, some argue that ETFs improve market efficiency. For example, Glosten, Nallareddy, and Zou (2020) show that ETFs ownership improves price efficiency. Box, Davis, Evans, and Lynch (2020) examine intraday returns and find little evidence that trading in ETFs impacts the underlying securities. Huang, O'Hara, and Zhong (2021) find that industry ETFs help investors hedge industry-wide risk. Our study complements the former strand of literature by highlighting one detrimental impact of ETFs on investors that results from overexploiting the data during index construction.

Our paper is also related to the recent literature that questions the out-of-sample reliability of the large set of asset pricing factors/anomalies documented by academic research (e.g., Harvey, Liu, and Zhu, 2016; Novy-Marx, 2016; McLean and Pontiff, 2016; Yan and Zheng, 2017; Arnott, Harvey, Kalesnik, and Linnainmaa, 2019; Hou, Xue, and Zhang, 2020). We show that smart beta ETFs, arguably the most feasible investment vehicle for investors to access asset pricing factors, are susceptible to data mining in index construction. As a result, the proliferation of smart beta ETFs seems to add little value to investors in terms of excess returns relative to the aggregate market and desired exposures to asset pricing factors.

III. Background and Data of Smart Beta ETFs

In this section, we briefly introduce the institutional background of smart beta ETFs and describe how we construct our data set.

A. *Smart Beta and the Soil of Data Mining*

In the booming ETF market, smart beta ETFs have experienced the largest growth and accounted for more than 20% of the overall ETF marketplace as of 2018 (Morningstar, 2019).[15] In the simplest terms, smart beta ETFs follow an indexation approach, but do so differently than traditional broad cap-weighted market indexes. Smart beta ETFs emphasize alternative index construction rules and claim to provide investors excess returns through exposures to asset pricing factors (anomalies). There are two broad categories of smart beta ETFs: single- and multi-factor smart beta ETFs. Popular themes of single-factor smart beta ETFs include value, growth, dividend, momentum, risk/volatility, and quality. Multi-factor smart beta ETFs can freely combine multiple factors altogether.

While some early smart beta ETFs track existing indexes, in the majority of cases, ETF sponsors either construct smart beta indexes by themselves or collaborate with third-party index providers to build indexes. In the latter case, ETF sponsors are often extensively involved in index construction (Financial Times, 2015; Pensions & Investments, 2018).[16]

Although smart beta claims to follow rules-based and formulaic indexation procedures, the discretion in constructing smart beta indexes is enormous in practice. We use arguably the most transparent and well-studied factor theme, value, to illustrate the large degree of freedom in index construction.

To build a value smart beta index, one may use any one or any combination of the various "value" measures, such as price-to-book ratio, price-to-earnings ratio, price-to-projected-earnings ratio, price-to-sales ratio, price-to-cash flow ratio, and enterprise value-to-operating cash ratio. The degree of freedom in choosing the number of stocks and the weight of each selected stock is even larger. For example, Appendix B presents the "rules" in determining the number of constituents of a value smart beta index with the ETF offered by a large ETF sponsor.[17] These index rules are quite intricate and hard to justify by economic or intuitive

[15] Smart beta is also known as strategic beta, alternative beta, or factor investing. Among smart beta ETF sponsors, Blackrock iShares and Vanguard are the largest in terms of AUM, while Invesco and First Trust are the largest in terms of the number of smart beta ETF products.

[16] For example, an executive of FTSE Russell, one of the largest smart beta index providers, noted in an interview (Pensions & Investments, 2018), "Our methodologies are developed in consultation with clients and external advisers. There's transparency throughout and a collaborative process when developing custom indexes for asset owners and asset managers."

[17] The index name and the ETF sponsor name are hidden on purpose. As of June 2020, this ETF has an

rationale. Given the enormous discretion one may possess, it is not difficult to build a "smart" value index that produces stellar backtests out of thousands or even millions of trials. As value is arguably the most-studied factor, the degree of freedom for other smart beta factor themes can only be higher, let alone multi-factor smart beta which can freely choose among multiple factors.

Furthermore, backtests also play a key role in launching and promoting smart beta products as stellar backtested returns help attract investment flows (Morningstar, 2017; Financial Times, 2017; ETF Stream, 2019).[18] The lure of increased flows is likely to induce ETF sponsors to overexploit historical data in index construction. Even if ETF sponsors collaborate with third-party index providers to develop smart beta indexes, index providers also have incentives to cater to ETF sponsors' demand for stellar backtests. Moreover, index providers are paid according to the amount managed against their benchmarks, known as index-licensing fees (Financial Times, 2019). In short, there are plenty of incentives and space for data mining in smart beta index construction.

B. Data Description

Now we describe how we construct the sample of smart beta ETFs and how we manually collect their index returns before and after ETF listings.

We take several steps to construct the sample. First, we obtain a list of U.S. equity smart beta ETFs identified by Morningstar.[19] For each ETF, we obtain the ETF name, the ETF listing date, and the factor theme from Morningstar Direct. Specifically, factor theme is the factor category of smart beta ETFs classified by Morningstar. We then use the CRSP database to cross-check these smart beta ETFs. After this step, we obtain 379 smart beta ETFs.

In the second step, for each smart beta ETF, we manually identify the underlying index

AUM of more than 50 billion US dollars.

[18]According to Vanguard (2012) and based on our conversations with several Blackrock professionals, prospective investors are often guided and pointed to the backtested performance of smart beta indexes, and index providers always offer backtested performance through public means, typically their websites and third-party data providers.

[19]Specifically, we extract the list of smart beta ETFs from all ETFs in Morningstar Direct by applying search criteria of (*i*) U.S. Category Group = "U.S. Equity," (*ii*) Strategic Beta = "Yes," and (*iii*) Strategic Beta Group not in "Commodity" or "Fixed-Income."

and collect the index performance start date, index release date, and return history of the index.[20] Specifically, for each smart beta ETF, we collect the name of its underlying index from its official website or from professional third-party websites (e.g., ETF.com). With index names, we collect index information and the entire history of index returns from Morningstar Direct or from the websites of index providers. To compare the smart beta index performance before and after ETF listings, we require the underlying index of a smart beta ETF to have non-missing returns twelve months before and twelve months after the ETF listing.

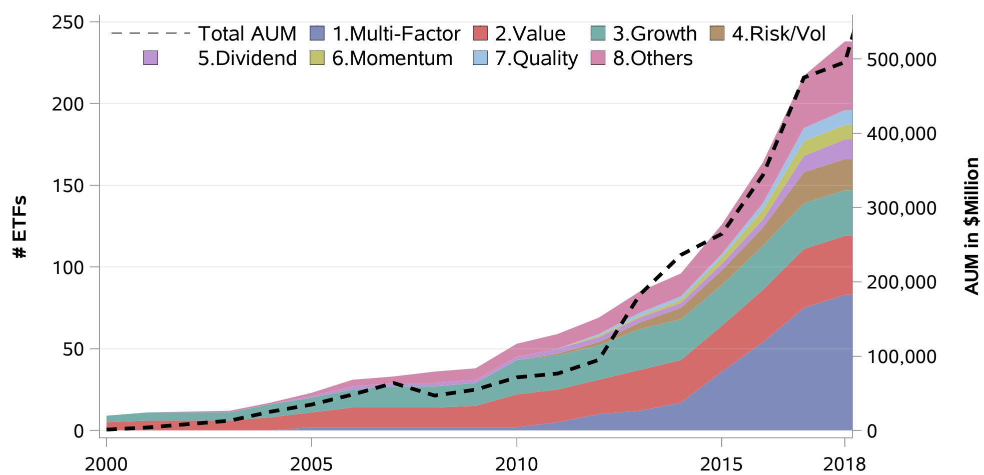

Figure 2: **Aggregate AUM and Number of Smart Beta ETFs by Factor Theme.** The left axis shows the number of smart beta ETFs listed each year under different smart beta factor themes based on our sample. The right axis shows the total AUM across all ETFs in our sample.

Our final sample has 238 US domestic equity smart beta ETFs listed between 2000 and 2018, tracking 223 unique smart beta indexes.[21] Based on the AUM at the end of 2019, these 238 smart beta ETFs account for 82.3% of the total AUM of the preliminary list of 379 ETFs (in the first step of data collection), with an average AUM of 2.8 billion. This suggests that our sample is representative of the U.S. equity smart beta ETF market and thus, there is

[20]The performance start date is the earliest time of index performance in backtests, and the index release date is the time when an index is released.

[21]There are ten indexes that are tracked by multiple smart beta ETFs. When evaluating smart beta index performance, we only count these indexes once.

little selection bias in our analyses. Figure 2 plots the total AUM of the ETFs in our sample and the number of ETFs by factor theme from 2000 to 2018. It is clear that the smart beta ETF market is rapidly expanding. Among the smart beta themes, the multi-factor category has the largest number of ETF offerings as of 2018, consistent with Morningstar (2019).

[Insert Table 1 here]

Table 1 also reports other summary statistics of our sample. About 16% of smart beta ETFs were launched before 2010, 24% were launched between 2010 to 2014, and 60% were launched between 2015 and 2018. The ETFs are offered by 41 ETF sponsors, with an average expense ratio of 33 bps per year. Moreover, after ETFs are listed, the average correlation between monthly ETF returns and index returns is 99%, suggesting that these ETFs track their underlying smart beta indexes very closely.

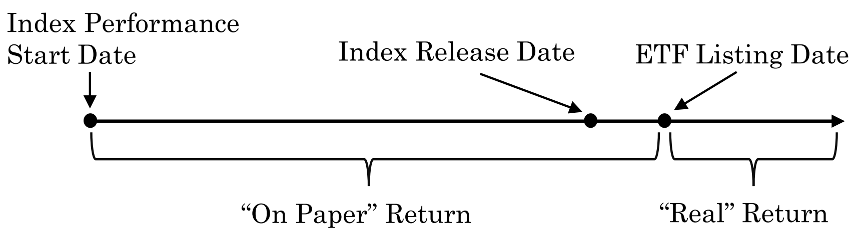

Figure 3: **An Illustration of "On Paper" and "Real" Returns.**

Note that each smart beta index has "on paper" returns and "live" returns, as illustrated in Figure 3. From index performance start date to ETF listing date, smart beta returns are not accessible to investors and thus are denoted by "on paper" or backtested returns. The "live" or "real" returns are those after the ETF listing date and are accessible to investors. As shown in Panel B of Table 1, a smart beta index has 228 months of returns available on average, with 157 months of "on paper" returns and 71 months of "real" returns. While we define "on paper" and "real" returns based on whether smart beta returns are investable to investors, in Appendix A, we also use index release date as the cutoff, and we get even sharper results.[22]

[22] As Table 1 reports, in more than half of our ETF sample, smart beta indexes were released within five months of the ETF listing dates.

IV. "Smart" Performance Only Exists in Backtests

In this section, we demonstrate that the performance of smart beta indexes declines significantly after the corresponding ETFs are launched to investors. We then examine possible explanations for such performance decline. We show that strategic timing in ETF listings, negative time trend in factor premia, and diminishing returns to ETF scale cannot explain this performance decline.

A. *Smart Beta Index Performance Before and After ETF Listings*

We start by analyzing the performance of smart beta indexes before and after ETF listings, the cutoff at which smart beta returns become accessible to investors.

Because smart beta ETFs often claim that their primary goal is to provide excess returns beyond the broad cap-weighted market index through factor exposures, we use the aggregate market as the benchmark for performance evaluation. Specifically, we use the value-weighted return of all CRSP US firms as the market return following Fama and French (2015). In fact, investors can easily invest in this broad market index through the Vanguard Total Stock Market ETF (VTI) at a very low cost (i.e., 3 bps per year).[23]

In Panel A of Table 2, we report the performance of smart beta indexes before and after ETF listings. We find that smart beta indexes only outperform the market index in backtests. For example, the average CAPM alpha of smart beta indexes is 2.77% per year (t = 4.59) over an average 13-year period before ETF listings. However, after ETFs are listed, smart beta indexes deliver an average negative CAPM alpha of −0.44% per year over an average 6-year post-listing period. The difference of CAPM alphas between the periods before and after ETF listings is 3.22% per year and is highly statistically significant.

[Insert Table 2 here]

It is worth noting that in our main analyses, we use smart beta index returns rather than the corresponding ETF returns over the post-ETF-listing period. There are two advantages of using smart beta index returns. First, ETF returns do not exist before ETF listings,

[23]Our results are unchanged if we use returns of VTI as the aggregate market returns.

and thus focusing on index returns can make the pre- and post-listing performance directly comparable. Second, since we use smart beta index returns to measure the post-ETF-listing performance, the performance decline cannot be attributed to transaction costs in replicating smart beta indexes. In Panel B of Table 2, we also gauge post-listing performance using ETF returns and find the performance decline to be slightly sharper. In our estimation, smart beta ETFs deliver a negative CAPM alpha of -0.59% ($t = -1.69$) per year before fees and -0.87% ($t = -2.40$) per year after fees.

To give a more direct sense of smart beta performance, we also use returns of the SPDR S&P 500 ETF (SPY), the first and most-traded ETF, as the benchmark, and we repeat the above analysis. Table 3 shows that before ETF listings, smart beta indexes outperform SPY by 3.07% ($t = 4.87$) per year "on paper." However, over the post-listing period, smart beta ETFs *underperform* SPY by 1.24% ($t = 3.33$) per year on average. If we further account for that smart beta ETFs charge higher fees than SPY, smart beta ETFs *underperform* SPY by 1.43% ($t = 3.69$) per year on average. That is, smart beta ETFs deliver strictly lower returns relative to SPY after listing.

We also conduct several robustness tests in Appendix A. First, we confirm the post-ETF-listing performance decline of smart beta through panel regressions. The results are reported in Table A.1. Second, when we compare index performance before and after the index release date rather than the ETF listing date, our results are even stronger (see Table A.2). Third, we conduct sub-sample analysis and show that the patterns in Table 2 are robust for different sample periods (see Table A.3). We also find that later-listed smart beta ETFs experience a larger performance decline in their underlying smart beta indexes. Last, we show that the results in Table 2 are robust when we require ETFs to have a longer history (at least three years) of non-missing index returns before ETF listings (see Table A.4).

In sum, the results in this section and in Appendix A show that the claimed outperformance of smart beta indexes only exists in backtests. Once the corresponding smart beta ETFs are launched to investors, smart beta performance declines sharply and is worse than broad market indexes.

B. Discussions of Potential Explanations

We proceed to explore explanations for the performance deterioration of smart beta indexes. There are four plausible explanations: (*i*) strategic timing of ETF listings, (*ii*) negative time trend in factor premia, (*iii*) diminishing returns to scale caused by ETF flows, and (*iv*) data overexploitation in index construction. In this section, we argue that none of the first three explanations can materially explain the sharp decline in index performance after ETF listings. We will provide supporting evidence of data mining in index construction in Section V.

B.1. Strategic Timing of ETF Listings

The first plausible explanation is that smart beta ETF sponsors strategically time their tracked factors when launching smart beta ETFs. That is, they choose to launch ETFs when the tracked factors have recently outperformed the aggregate market. Possibly because factor returns are mean-reverting, the post-ETF-listing performance is not as good as the pre-listing performance. We refer to this specific "strategic timing" explanation as "factor timing."

To examine this "factor timing" hypothesis, for each smart beta index, we use the earliest-constructed index of the same factor theme as the benchmark to control for factor-wide return fluctuations, including the potential factor return mean-reversion.[24] Specifically, we measure smart beta performance using both excess returns and alphas relative to the factor-theme benchmark (i.e., the earliest-constructed index in the same factor-theme category).

Table 4 reports the results. Controlling for factor-wide return fluctuations, smart beta indexes still experience a sharp performance decline after ETF listings, and the magnitude of the decline is similar to that in Table 2. For example, relative to the benchmark index in each factor-theme category, the excess return of smart beta indexes is, on average, 2.85% per year before ETF listings, and it drops to −0.42% per year post ETF listings. The performance decline is 3.27% per year and is highly statistically significant. When regressing index returns

[24] For example, we measure the performance of value smart beta indexes relative to the Russell 1000 Value Index, and we measure the performance of growth smart beta indexes relative to the Russell 1000 Growth Index. For smart beta indexes based on momentum, volatility, quality, and dividend, we use the MSCI USA Momentum Index, the MSCI USA Minimum Volatility Index, the MSCI Quality Index, and the Nasdaq Broad Dividend Achiever Index, respectively.

on the factor-theme benchmark returns, we observe similar performance deterioration after ETF listings (see Panel B of Table 4). Thus, factor timing fails to explain the post-ETF-listing performance decline of smart beta indexes.[25]

[Insert Table 4 here]

An alternative explanation related to 'factor timing" is the "index timing" hypothesis. That is, ETF sponsors launch ETFs to track existing smart beta indexes when the indexes have recently outperformed the aggregate market or other indexes of the same factor theme. We argue that this "index timing" explanation also fails to work. As shown in Tables 2 to 4, when we use a wider time window around ETF listing to measure performance, smart beta indexes have higher pre-listing performance. This pattern is inconsistent with the "index timing" hypothesis under which indexes should outperform more around ETF listings.

B.2. Time Trend in Factor Premia

Another plausible explanation is that factor premia are waning over time, possibly due to improving market efficiency or an increase in factor trading following the academic discovery of such factors (e.g., McLean and Pontiff, 2016). We argue that the time trend of factor premia cannot explain the performance decline documented in Tables 2 to 4 for the following two reasons.

First, since we adjust for factor-theme benchmark returns in Table 4, we effectively control for the time trend of factor premia, and we still get similar results. Second, quantitatively, the decline in smart beta index returns is much sharper than that in factor premia documented in prior studies. For example, Chordia et al. (2014) analyze twelve well-known asset pricing factors/anomalies and estimate the average "half-life" of these factors to be 12.8 years, suggesting an annualized decline of about 30 bps in factor returns. By comparison, smart index returns and alphas decline by more than 300 bps per year on average after ETF listings. Thus, the waning in factor premia does not materially explain the performance decline in smart beta indexes.

[25] Controlling for factor-theme benchmark returns also rules out the possibility that the performance decline is due to a drop in loadings on the designated factors.

B.3. Diminishing Returns to Scale

One might also argue that investment flows into smart beta ETFs could make certain factor-based strategies over-crowded, and consequently, impair the performance of smart beta indexes due to decreasing returns to scale. Because we use index returns rather than ETF returns to avoid transaction costs in replicating indexes (an important driving force for decreasing returns to scale), our findings are unlikely to be explained by this argument. We now conduct tests to formally rule out this explanation.

Specifically, we split smart beta ETFs into two groups based on their average post-listing AUM. We then analyze the pre-listing and post-listing index performance for these two groups as in Table 2. To alleviate concerns that different factor-based strategies are affected differently by this scale effect, we also group smart beta ETFs based on their average post-listing AUM within their factor-theme category.

Table 5 report the results. As one can see, whether we rank ETFs by their AUM within the whole sample (Panel A) or within the factor-theme category (Panel B), the indexes of smaller-sized ETFs experience even larger performance declines after ETF listings. Thus, we can conclude that the post-ETF-listing performance decline cannot be attributed to diminishing returns to scale.

[Insert Table 5 here]

In short, the results in this section suggest that strategic timing of ETF listings, the negative time trend in factor premia, and diminishing returns to scale cannot sufficiently explain the performance deterioration of smart beta indexes post ETF listings.

V. Data Mining and ETF Flows

In this section, we provide supporting evidence that data overexploitation in smart beta index construction drives the sharp performance decline after ETF listings. We further show that investors respond positively to backtested performance, which incentivizes the practice of boosting "on paper" returns. In addition, we find that later-constructed smart beta indexes/ETFs mostly fail to deliver higher or cheaper average exposure to the designated factor than the first-offered index of the same factor theme. Altogether, our findings raise concerns on the large wave of smart beta offerings.

A. Evidence of Data Mining in Index Construction

While smart beta indexes are claimed to be built under rules-based and formulaic indexation procedures, the discretion in constructing indexes is enormous in practice. This discretion comes from the multitude of ways of defining factors and the flexibility in selecting stocks and choosing their weights in the index. Meanwhile, ETF sponsors often advocate stellar backtested results (e.g., Morningstar, 2017). Presumably, the stellar performance in backtests could help attract investment flows, although the outperformance only exists "on paper."

To demonstrate that data mining is the key driving force for the smart beta performance decline in Tables 2 to 4, we conduct three tests.

In the first test, we explore the opaqueness of smart beta indexes. We argue that the discretion in constructing multi-factor smart beta indexes is larger than single-factor smart beta indexes. Intuitively, multi-factor indexes can freely choose among multiple factors, and the flexibility in combing factors leads to larger discretion in index construction by nature. Moreover, as the category of multi-factor smart beta ETFs has the largest number of ETF offerings, multi-factor smart beta ETFs are likely to face the most competitive product market and thus have stronger desires for stellar backtests to attract investment flows. While single-factor indexes are generally less susceptible to data mining than multi-factor indexes, value and growth indexes should have even less space for data mining than other single-factor indexes as they are built on the most-studied factor.

[Insert Table 6 here]

We indeed find strong supporting evidence for our argument. As shown in Panel A of Table 6, before ETF listings, multi-factor smart beta indexes have the highest average CAPM alpha of 4.11% per year, while value and growth smart beta indexes have the lowest average CAPM alpha of 1.16% per year. After ETF listings, multi-factor smart beta indexes have the largest performance drop of 4.91% per year, while value and growth smart beta indexes have the smallest performance decline of 1.76% per year on average.

To provide further support for data mining, in the second test, we explore the degree of control that ETF sponsors possess over smart beta index construction. Since backtests can help ETF sponsors attract flows and thus collect management fees (Vanguard, 2012), ETF sponsors are likely to have a strong desire for stellar backtests. Thus, in cases where ETF sponsors have more control over index construction, we should observe higher "on paper" returns before ETF listings and more substantial performance declines.

To test this, we classify smart beta indexes into "high-ETF-discretion" and "low-ETF-discretion" groups. To be classified as a high-ETF-discretion index, a smart beta index should satisfy at least one of the two non-exclusive conditions: (*i*) the index is constructed "in-house" by the ETF sponsor, and (*ii*) the index name contains the ETF sponsor's name or the index's release date is within six months of the ETF listing date. The second condition implies that the index is specifically tailored for the ETF, and thus, is indicative of the ETF sponsor's control over the index construction. Although this is a crude classification, we argue that ETF sponsors are likely to have more influence over high-ETF-discretion indexes than low-ETF-discretion indexes.

In Panel B of Table 6, we examine the performance of high-ETF-discretion and low-ETF-discretion smart beta indexes. Consistent with our hypothesis, high-ETF-discretion indexes indeed have better "on paper" performance and experience larger performance deterioration. For example, before ETF listings, high-ETF-discretion indexes can generate an "on paper" CAPM alpha of 3.64% per year, while low-ETF-discretion indexes have a CAPM alpha of 1.51% per year on average. After ETF listings, high-ETF-discretion indexes experience an average drop in CAPM alpha of 4.26% per year, and low-ETF-discretion indexes have much smaller performance deterioration, averaging at 1.69% per year. The difference in the performance decline between the two groups is also statistically significant.

In the third test, we explore the heterogeneity in ETF sponsor size. Since small ETF sponsors are keen to increase market share, they should have stronger incentives to boost index backtests to attract investment flows than large ETF sponsors, who are likely to care more about reputation. To show that this is indeed the case, we classify the top-three ETF sponsors by AUM (Blackrock iShares, Vanguard, and State Street Global Advisors) into one group and all other ETF sponsors into another group. We then examine the index performance for these two groups of ETF sponsors separately.

As shown in Panel C of Table 6, the smart beta indexes used by smaller ETF sponsors experience a larger performance decline than those of the top-three ETF sponsors. Specifically, the CAPM alpha of the smart beta indexes used by the smaller ETF sponsors drops from 3.31% per year before ETF listings to −0.56% per year after ETF listings, while the indexes of the top-three ETF sponsors experience a much milder decline.[26]

In summary, through exploring the extent to which smart beta indexes are susceptible to data overexploitation, we find strong evidence that data mining in indexation accounts for the post-listing performance deterioration of smart beta indexes. Nevertheless, even though the pre-ETF-listing performance of smart beta indexes cannot be sustained, in the next section, we show that investors still respond strongly to backtested returns.

B. Investors Respond Positively to Backtested Returns

In this section, we show that the backtested performance of smart beta indeed has a strong positive influence on investment flows. This analysis sheds light on investor welfare and strengthens our argument that data mining is likely the cause of the performance decline after ETF listings, as this behavior is rewarded through investment flows.

To show that investors respond to backtested smart beta index returns, we estimate the following regression:

$$\text{Flow}_i^{\text{post},k} = \pi_t + \tau_f + \lambda \cdot \alpha_i^{\text{pre},s} + \epsilon_i. \tag{1}$$

Here, $\text{Flow}_i^{\text{post},k}$ is the average monthly flow for smart beta ETF i over the k-month period after the ETF listing, π_t and τ_f are the listing-year and smart-beta-theme fixed effects, and $\alpha_i^{\text{pre},s}$ is the monthly CAPM alpha of the underlying smart beta index over s months before the ETF listing date. For robustness, we also replace CAPM alpha with the factor-theme-adjusted return, and we also consider different values of k and s. Clearly, in this regression,

[26]It is also worth noting that only 28.4% of the smart beta indexes that the top-three ETF sponsors use are "high-ETF-discretion" indexes, while the smaller ETF sponsors mostly use "low-ETF-discretion" smart beta indexes (73.2%).

λ is the estimated flow-to-backtest sensitivity and captures how investors respond to "on-paper" returns of smart beta indexes.

Panel A of Table 7 reports the results. As one can see, investors respond strongly to backtested index returns, although such returns are not sustained after ETF listings. For example, a one-standard-deviation increase in pre-listing CAPM alpha is associated with an increase in monthly ETF flows of 6% over the first post-listing year. Considering that the median flow is 11% per month over this period, the effect of backtested performance in attracting investment flows is economically significant. Moreover, we obtain similar results when replacing CAPM alpha with the factor-theme-adjusted return (see Panel B of Table 7). These results are consistent with the claim of Vanguard (2012) that backtests contribute to the viability of new ETFs.

[Insert Table 7 here]

To add more color to investor behavior, we further analyze the relationship between ETF flows and future ETF performance in Table A.5 of Appendix A. Either through portfolio sorting or through panel regressions, we find that investment flows significantly and negatively predict the future performance of smart beta ETFs. For example, the top quintile of smart beta ETFs (AUM-weighted) by past one-year flows significantly underperform the bottom-flow quintile by about 30 bps over the next month. Moreover, we show that the results are similar when controlling for past one-year return and ETF size. This "dumb money" effect suggests that investors of smart beta ETFs are likely to be unsophisticated.[27]

In sum, the analysis of flow-to-backtest sensitivity in this section cautions data mining in constructing smart beta indexes as investors respond strongly to the stellar "on paper" performance.

C. Smart Beta ETFs and Factor Exposures

We show in Section IV that smart beta indexes/ETFs underperform the aggregate market and the first-offered index of the same factor-theme (the factor-theme benchmark) after ETF

[27]Bhattacharya, Loos, Meyer, and Hackethal (2017) and Brown, Davies, and Ringgenberg (2019) also show that ETF flows are largely uninformed.

listings, suggesting that the large offerings of smart beta ETFs add little value to investors in terms of returns.

However, one might still argue that the proliferation of smart beta ETFs is beneficial as investors could potentially get higher or cheaper exposures to asset pricing factors. We find the opposite in the data. That is, despite investors needing to pay significantly higher fees to access later-constructed indexes, these smart beta indexes mostly fail to deliver higher average exposure to their designated asset pricing factors than the first-offered index of the same factor theme.

To see this, we estimate factor exposures of smart beta indexes through multivariate regressions, where the dependent variable is the average index return for each of the five major factor themes (value, growth, momentum, quality, and risk/vol) and the independent variables are the Fama-French-Carhart four factors, the QMJ factor (Asness, Frazzini, and Pedersen, 2019), and the total volatility factor (Ang, Hodrick, Xing, and Zhang, 2006).[28] For comparison, we also estimate similar regressions for the first smart beta index in each of the five factor-theme categories.[29]

[Insert Table 8 here]

Panel A of Figure 4 plots the exposures of smart beta indexes on their designated factors, and Table 8 reports the detailed results. While the later-constructed value indexes deliver marginally higher average exposure to the value factor, smart beta indexes in the four other factor-theme categories provide investors lower average exposure to the designated factor than the first-constructed smart beta index of the same factor theme. For example, momentum smart beta indexes have an average beta to the momentum factor of 0.290, which is lower than that of the first momentum smart beta index, the MSCI USA Momentum Index, by 0.068. Quality smart beta indexes have an average beta to the quality factor of 0.148, while the first quality index, the MSCI USA Quality Index, has a quality beta of 0.209. These findings are also consistent with Johansson, Sabbatucci, and Tamoni (2020),

[28]Here, we do not include the smart beta theme of "dividend" because there is no clear and widely-recognized asset pricing factor of "dividend."

[29]For robustness, we also focus on the post-ETF-listing index returns and find similar results in Table A.6.

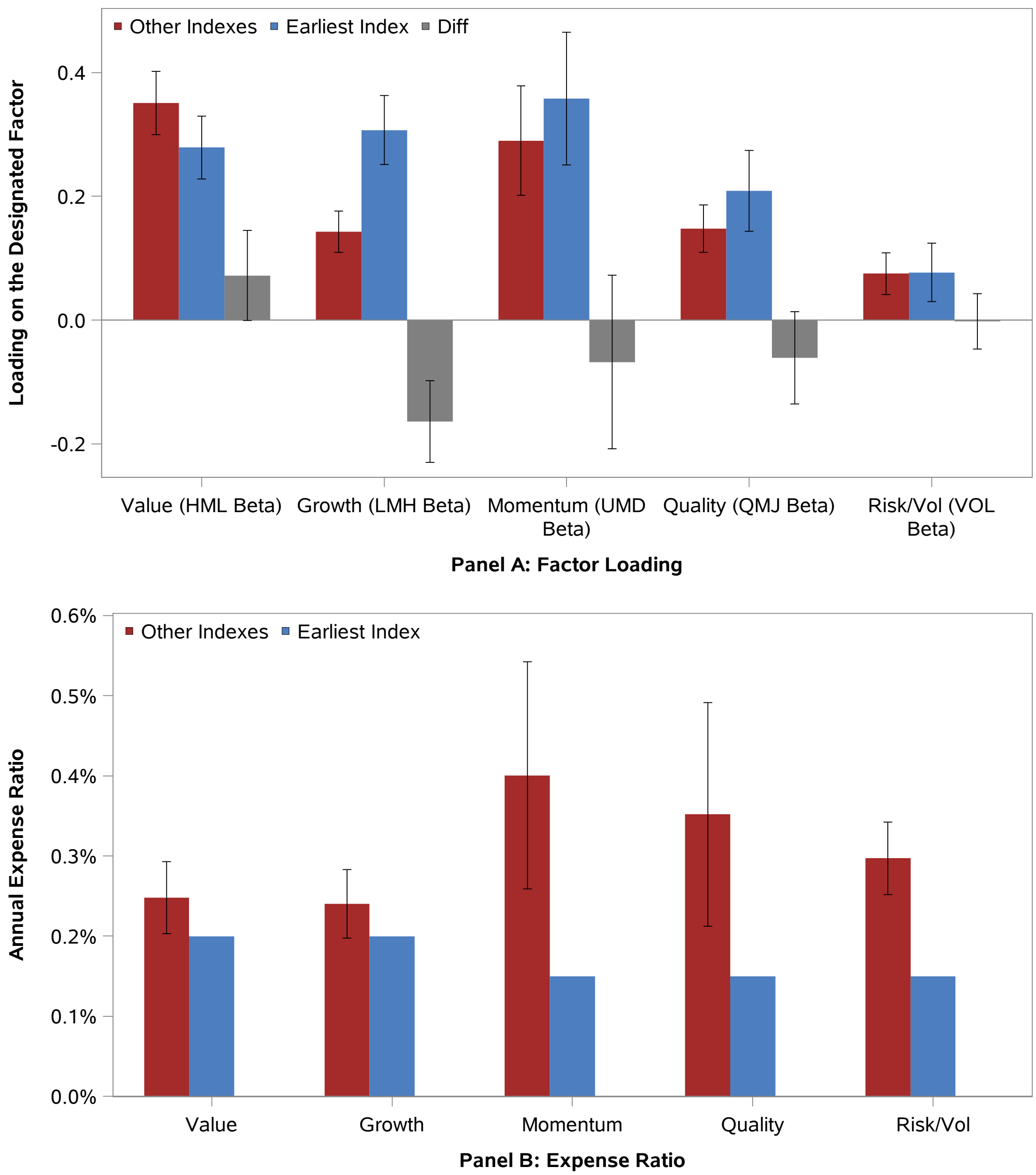

Figure 4: **Factor exposures and ETF expense ratio.** Panel A plots the exposures of smart beta indexes on their designated factors. Within each factor-theme, the blue bar indicates the factor exposure of the earliest index, the red bar indicates the average factor exposure of latter-constructed smart beta indexes, and the gray bar shows the difference in factor exposure between the earliest index and latter-constructed indexes. For growth indexes, we report the factor exposure on "LMH" factor, which is the negative of Fama-French HML factor. The 95% confidence bands are shown around the bars. Panel B plots annualized expense ratios of smart beta ETFs that track the indexes. The blue bar shows the expense ratio of the ETF that tracks the earliest index, and the red bar shows the average expense ratio of ETFs that track latter-constructed indexes.

who argue that investors cannot effectively harvest factor risk premia through smart beta ETFs.

Panel B of Figure 4 further shows management fees charged by the smart beta ETFs tracking the indexes. To invest in the later-constructed smart beta indexes, investors need to pay significantly higher management fees. For example, investors would need to pay an additional 25 bps and 20 bps per year on average to access the later-built momentum and quality indexes, respectively. Taking Panels A and B of Figure 4 together, we conclude that the proliferation of smart beta indexes fails to deliver higher or cheaper factor exposure to investors.

In summary, our results suggest that the large wave of smart beta ETF offerings does not add value to investors both in terms of excess returns beyond the aggregate market and in terms of factor exposures.

VI. Evidence from Other Developed Markets

To provide more evidence, we extend our study to other developed markets, including Europe, the UK, Canada, and Australia. According to Morningstar (2019), smart beta ETFs are also growing rapidly in these markets, although they are not as large as the US market.[30] Like in the US, we also find that smart beta indexes experience a sharp performance decline in these countries, suggesting that data overexploitation in constructing smart beta indexes is quite prevalent.

We conduct our analysis as follows. First, following the same procedure in Section III, we collect 77 non-US equity smart beta ETFs whose index returns are available before ETF listings.[31] Then, we analyze the performance of these smart beta indexes relative to their corresponding aggregate market indexes before and after ETF listings as in Table 2. Here, the European smart beta indexes are benchmarked to the MSCI Europe Index. Similarly, the

[30]For example, the AUM of European smart beta ETFs grew from zero in 2005 to $57.4 billion in 2018.

[31]We take the following steps to obtain the list of smart beta ETFs that invest in equities outside the US. Firsty, we use the *Smart Beta* and *Global Investment Fund Sector (GIFS)* data fields to extract the list of smart beta ETFs from Morningstar Direct. Second, we collect information on the underlying indexes, and we require an ETF to have non-missing index returns in the six months before and after the ETF listing date. After this procedure, our sample includes 37 European smart beta ETFs, 29 Canadian ETFs, 6 Australian ETFs, and 5 UK ETFs.

Canadian, Australian, and UK smart beta indexes are benchmarked to the MSCI Canada Index, the MSCI Australia Index, and the MSCI United Kingdom Index, respectively. Table 9 reports the results.

[Insert Table 9 here]

Before ETF listings, these 77 smart beta indexes significantly outperform their corresponding market indexes by 3.02% per year on average. However, once the ETFs are listed, these smart beta indexes trail the market indexes by 0.15% per year, yielding a return decline of 3.17% per year ($t = 5.80$). The results are similar if we use CAPM alpha to gauge the pre- and post-listing performance. These findings suggest that data mining in constructing smart beta indexes is also common in other developed markets.

VII. Conclusion

In the past decades, smart beta ETFs have experienced rapid growth, outpacing other segments of the booming ETF market. ETF sponsors profess that smart beta indexes deliver excess returns through exposures to investment factors and often use academic research to endorse their smart beta products. While smart beta ETFs claim to follow pre-determined and formulaic rules in indexation, the discretion in setting up the rules is enormous in practice. Potentially, the claimed "smartness" of smart beta ETFs is just a mirage arising from data overexploitation. If there indeed exists considerable data mining in index construction, it is unclear whether the proliferation of smart beta products adds value to investors or not. The answer to this question is important and urgent, given the sheer size and the rapid growth of smart beta ETFs.

In this study, we compile a comprehensive sample of equity smart beta ETFs and systematically examine the performance of smart beta indexes before and after ETF listings. Our study shows that smart beta indexes can only outperform the aggregate market "on paper" *before* ETF listings. Smart beta indexes quickly trail the broad cap-weighted market index after the corresponding ETFs are listed. We further explore potential explanations for the performance deterioration of smart beta indexes. We find that it can neither be attributed to ETF sponsors' strategic timing in ETF listings nor factor return fluctuations. Instead, we find strong support for data overexploitation in index construction, meaning that stellar performance only exists in backtests and has no predictive power for "real" performance. Our results caution against the risk of data mining in the proliferation of ETF offerings because investors respond strongly to the stellar performance in backtests.

REFERENCES

Agarwal, Vikas, Brad M. Barber, Si Cheng, Allaudeen Hameed, and Ayako Yasuda, 2019, Private company valuations by mutual funds, Working Paper, Georgia State University.

Agarwal, Vikas, Paul Hanouna, Rabih Moussawi, and Christof Stahel, 2018, Do ETFs increase the commonality in liquidity of underlying stocks?, Working Paper, Georgia State University.

Ang, Andrew, 2014, *Asset management: A systematic approach to factor investing* (Oxford University Press).

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

Arnott, Robert, Noah Beck, Vitali Kalesnik, and John West, 2016, How can "smart beta" go horribly wrong?, Available at: `https://www.researchaffiliates.com/en_us/publications/articles/442_how_can_smart_beta_go_horribly_wrong.html`.

Arnott, Robert D., Campbell R. Harvey, Vitali Kalesnik, and Juhani T. Linnainmaa, 2019, Alice's adventures in factorland: Three blunders that plague factor investing, Working paper, Duke University.

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

Bams, Dennis, Roger Otten, and Ehsan Ramezanifar, 2017, Investment style misclassification and mutual fund performance, Working Paper, Available at SSRN 2648375.

Barber, Brad M., Terrance Odean, and Lu Zheng, 2005, Out of sight, out of mind: The effects of expenses on mutual fund flows, *Journal of Business* 78, 2095–2120.

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

Ben-David, Itzhak, Jiacui Li, Andrea Rossi, and Yang Song, 2019, What do mutual fund investors really care about?, Working Paper, The Ohio State University.

Ben-David, Itzhak, Jiacui Li, Andrea Rossi, and Yang Song, 2020, Style investing, positive feedback loops, and asset pricing factors, Working paper, Univeristy of Washington.

Bhattacharya, Utpal, Benjamin Loos, Steffen Meyer, and Andreas Hackethal, 2017, Abusing ETFs, *Review of Finance* 21, 1217–1250.

Bloomberg, 2016, Blackrock projects smart beta ETF assets will reach $1 trillion by 2020, and $2.4 trillion by 2025, Available at `https://www.bloomberg.com/press-releases/2016-05-12/blackrock-projects-smart-beta-etf-assets-will-reach-1-trillion-globally-by-2020-and-2-4-trillion-by-2025`.

Bollen, Nicolas P.B., and Veronika K. Pool, 2009, Do hedge fund managers misreport returns? Evidence from the pooled distribution, *Journal of Finance* 64, 2257–2288.

Bollen, Nicolas P.B., and Veronika K. Pool, 2012, Suspicious patterns in hedge fund returns and the risk of fraud, *Review of Financial Studies* 25, 2673–2702.

Box, Travis, Ryan Davis, Richard B. Evans, and Andrew A. Lynch, 2020, Intraday arbitrage between ETFs and their underlying portfolios, *Journal of Financial Economics* Forthcoming.

Brown, David C., Scott Cederburg, and Mitch Towner, 2020, (Sub)optimal asset allocation to ETFs, Working paper.

Brown, David C., Shaun Davies, and Matthew Ringgenberg, 2019, ETF arbitrage, non-fundamental demand, and return predictability, Working Paper, Available at SSRN 2872414.

Brown, Keith C., W. Van Harlow, and Hanjiang Zhang, 2009, Staying the course: The role of investment style consistency in the performance of mutual funds, Working Paper, Available at SSRN 1364737.

Budiono, Diana, and Martin Martens, 2009, Mutual fund style timing skills and alpha, Working Paper, Available at SSRN 1341740.

Cao, Jie, Jason C. Hsu, Zhanbing Xiao, and Xintong Zhan, 2019, Smart beta, 'smarter' flows, Working Paper, Available at SSRN 2962479.

Carhart, Mark M., Ron Kaniel, David K. Musto, and Adam V. Reed, 2002, Leaning for the tape: Evidence of gaming behavior in equity mutual funds, *Journal of Finance* 57, 661–693.

Chan, Louis K.C., Hsiu-Lang Chen, and Josef Lakonishok, 2002, On mutual fund investment styles, *Review of Financial Studies* 15, 1407–1437.

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

Chen, Joseph, Harrison Hong, Wenxi Jiang, and Jeffrey D. Kubik, 2013, Outsourcing mutual fund management: Firm boundaries, incentives, and performance, *Journal of Finance* 68, 523–558.

Chordia, Tarun, Avanidhar Subrahmanyam, and Qing Tong, 2014, Have capital market anomalies attenuated in the recent era of high liquidity and trading activity?, *Journal of Accounting and Economics* 58, 41–58.

Cici, Gjergji, Scott Gibson, and John J. Merrick Jr, 2011, Missing the marks? Dispersion in corporate bond valuations across mutual funds, *Journal of Financial Economics* 101, 206–226.

Da, Zhi, and Sophie Shive, 2018, Exchange-traded funds and asset return correlations, *European Financial Management* 24, 136–168.

Dannhauser, Caitlin D., and Jeffrey Pontiff, 2020, FLOW, Working paper, Boston College.

Easley, David, David Michayluk, Maureen O'Hara, and Tālis J Putniņš, 2020, The active world of passive investing, Working Paper, Available at SSRN 322084.

Edelen, Roger M., Richard B. Evans, and Gregory B. Kadlec, 2012, Disclosure and agency conflict: Evidence from mutual fund commission bundling, *Journal of Financial Economics* 103, 308–326.

Egan, Mark L., Alexander MacKay, and Hanbin Yang, 2020, Recovering investor expectations from demand for index funds, Working Paper, National Bureau of Economic Research.

Engle, Robert F., and Debojyoti Sarkar, 2006, Premiums-discounts and exchange-traded funds, *The Journal of Derivatives* 13, 27–45.

ETF Stream, 2019, ETF insight: The dangers of smart beta backtesting, Available at: `https://www.etfstream.com/features/etf-insight-the-dangers-of-smart-beta-backtesting/`.

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

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

Financial Times, 2015, Smart beta's new kids on the index block, Available at `https://www.ft.com/content/1fb9d574-43c6-11e5-b3b2-1672f710807b`.

Financial Times, 2017, Lack of smart beta benchmarks fuels concerns, Available at `https://www.ft.com/content/d4f03f5e-a9f0-11e7-ab66-21cc87a2edde`.

Financial Times, 2019, Index companies to feel the chill of fund managers' price war, Available at `https://www.ft.com/content/e886b2d2-e852-3071-85c1-c9a57113d8a5`.

Glosten, Lawrence, Suresh Nallareddy, and Yuan Zou, 2020, ETF activity and informational efficiency of underlying securities, *Management Science,* Forthcoming.

Hartzmark, Samuel M., and Abigail B. Sussman, 2019, Do investors value sustainability? a natural experiment examining ranking and fund flows, *Journal of Finance* 74, 2789–2837.

Harvey, Campbell R., 2017, Presidential address: The scientific outlook in financial economics, *Journal of Finance* 72, 1399–1440.

Harvey, Campbell R., and Yan Liu, 2019, A census of the factor zoo, Working paper. Available at SSRN 3341728.

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

Hou, Kewei, Chen Xue, and Lu Zhang, 2020, Replicating anomalies, *Review of Financial Studies* Forthcoming.

Huang, Shiyang, Maureen O'Hara, and Zhuo Zhong, 2021, Innovation and informed trading: Evidence from industry ETFs, *Review of Financial Studies* Forthcoming.

Huang, Shiyang, Yang Song, and Hong Xiang, 2020, Noise trading and asset pricing factors, Working Paper, Available at SSRN 3359356.

Investco, 2018, Understanding smart beta, Available at https://www.invesco.com/us-rest/contentdetail?contentId=a8569b81e1c0e510VgnVCM100000c2f1bf0aRCRD.

Investment Company Institute, 2019, *Investment company fact book (2019)* (Investment Company Institute).

Israeli, Doron, Charles M.C. Lee, and Suhas A. Sridharan, 2017, Is there a dark side to exchange-traded funds? An information perspective, *Review of Accounting Studies* 22, 1048–1083.

Jiang, Wenxi, and Mindy Z. Xiaolan, 2017, Growing beyond performance, Working Paper, Available at SSRN 3002922.

Johansson, Andreas, Riccardo Sabbatucci, and Andrea Tamoni, 2020, Smart beta made smart, Working Paper, Available at SSRN 3594064.

Kaniel, Ron, and Robert Parham, 2017, WSJ category kings - the impact of media attention on consumer and mutual fund investment decisions, *Journal of Financial Economics* 123, 337–356.

Lettau, Martin, and Ananth Madhavan, 2018, Exchange-traded funds 101 for economists, *Journal of Economic Perspectives* 32, 135–54.

Li, Feifei, and John West, 2017, Live from newport beach. it's smart beta!, Available at: https://www.advisorperspectives.com/commentaries/2017/08/24/live-from-newport-beach-its-smart-beta.pdf.

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

Morningstar, 2017, Be wary of strategic-beta ETFs with rosy back-tested results, Available at https://www.morningstar.com/articles/805594/be-wary-of-strategic-beta-etfs-with-rosy-back-tested-results.

Morningstar, 2019, A global guide to strategic-beta exchange-traded products, Technical Report.

Novy-Marx, Robert, 2016, Testing strategies based on multiple signals, Working Paper, University of Rochester.

Pan, Kevin, and Yao Zeng, 2019, ETF arbitrage under liquidity mismatch, Working Paper, Available at SSRN 2895478.

Pattabiraman, Murari, 2020, Data mining in index construction: Why investors need to be cautious, Available at: https://freefincal.com/data-mining-index-construction/.

Pensions & Investments, 2018, ETF providers going in new directions with indexes, Available at https://www.pionline.com/article/20181001/PRINT/181009995/etf-providers-going-in-new-directions-with-indexes.

Rajan, Uday, Amit Seru, and Vikrant Vig, 2010, Statistical default models and incentives, *American Economic Review* 100, 506–10.

Rajan, Uday, Amit Seru, and Vikrant Vig, 2015, The failure of models that predict failure: Distance, incentives, and defaults, *Journal of Financial Economics* 115, 237–260.

Sensoy, Berk A., 2009, Performance evaluation and self-designated benchmark indexes in the mutual fund industry, *Journal of Financial Economics* 92, 25–39.

Song, Yang, 2020, The mismatch between mutual fund scale and skill, *Journal of Finance* Forthcoming.

Suhonen, Antti, Matthias Lennkh, and Fabrice Perez, 2017, Quantifying backtest overfitting in alternative beta strategies, *Journal of Portfolio Management* 43, 90–104.

Swinkels, Laurens, and Liam Tjong-A-Tjoe, 2007, Can mutual funds time investment styles?, *Journal of Asset Management* 8, 123–132.

Vanguard, 2012, Joined at the hip: ETF and index development, Available at https://www.vanguardcanada.ca/documents/joined-at-the-hip.pdf.

Wermers, Russ, 2012, Matter of style: The causes and consequences of style drift in institutional portfolios, Working Paper, Available at SSRN 2024259.

Yan, Xuemin, and Lingling Zheng, 2017, Fundamental analysis and the cross-section of stock returns: A data-mining approach, *Review of Financial Studies* 30, 1382–1423.

Table 1 **Summary statistics.** This table summarizes the 238 smart beta ETFs and their underlying smart beta indexes. Panel A reports the distribution of ETFs by listing year and the number of ETF sponsors in each period. Panel A also reports the distribution of annualized expense ratio, correlations between the monthly ETF return and the underlying index return, and number of ETFs offered per ETF sponsor. Panel B reports the summary statistics of the underlying smart beta indexes. *Months of Index Return* is the number of months with returns available (including returns in backtests). *Index Performance Start to ETF Listing* is the number of months between the index performance start date and the listing date of the corresponding ETF. *After ETF Listing* is the number of months with index returns available after ETF listings. *Index Release to ETF Listing* is the number of months between the index release date and listing date of the corresponding ETF.

Panel A: Smart Beta ETF					
ETF Listing Year	#ETFs	PCT	#ETF Sponsors		
2000-2009	38	16%	5		
2010-2014	58	24%	15		
2015-2018	142	60%	38		
Other Characteristics	Mean	SD	25th	50th	75th
Expense Ratio	0.33%	0.16%	0.20%	0.29%	0.40%
Return Correlation	0.99	0.08	1.00	1.00	1.00
#ETFs offered per sponsor	5.80	8.27	2.00	2.00	6.00
Panel B: Smart Beta Index					
	Mean	SD	25th	50th	75th
Months of Index Return	228	88	175	227	272
Index Performance Start to ETF Listing	157	81	88	161	200
After ETF Listing	71	57	29	48	93
Index Release to ETF Listing	36	62	1	5	32

Table 2 **Smart beta performance before and after ETF listing.** This table reports the performance of smart beta indexes/ETFs before and after ETF listing. In Panel A, columns (1) and (2) report across all smart beta indexes, the average annualized CAPM alpha over the pre-ETF-listing period and the post-ETF-listing period, respectively. Column (3) shows the difference in index performance before and after ETF listing. In Panel B, we replace smart beta index returns with smart beta ETF returns over the post-listing period. Specifically, we report the post-listing ETF CAPM alpha before fees and after fees in columns (2) and (3), respectively. Columns (4) and (5) report the post-listing performance decline. We also report smart beta performance around the one-year/two-year/three-year windows before and after the ETF listing date. Standard errors are clustered by factor theme as categorized by Morningstar and t-statistics are reported in parentheses. *, **, and *** denote the 10%, 5%, and 1% significance level, respectively. Our sample ends in December 2019.

Panel A: Pre-listing and Post-listing Index Performance

	(1) Before	(2) After	(3) Diff
All years before and after listing	2.77%***	−0.44%*	−3.22%***
	(4.59)	(−1.74)	(−4.63)
(−1 Year, +1 Year) around listing	1.21%***	0.69%	−0.52%
	(4.04)	(1.32)	(−1.04)
(−2 Year, +2 Year) around listing	1.18%***	0.32%	−0.86%
	(2.66)	(0.87)	(−1.25)
(−3 Year, +3 Year) around listing	1.51%***	0.15%	−1.36%**
	(3.95)	(0.43)	(−2.49)

Panel B: Pre-listing Index and Post-listing ETF Performance

	(1) Before	(2) After *Gross*	(3) After *Net*	(4) Diff *(2)−(1)*	(5) Diff *(3)−(1)*
All years before and after listing	2.77%***	−0.59%*	−0.87%**	−3.37%***	−3.65%***
	(4.59)	(−1.69)	(−2.40)	(−3.96)	(−4.20)
(−1 Year, +1 Year) around listing	1.21%***	0.48%	0.15%	−0.73%	−1.06%*
	(4.04)	(0.80)	(0.25)	(−1.19)	(−1.69)
(−2 Year, +2 Year) around listing	1.18%***	0.14%	−0.18%	−1.04%	−1.36%*
	(2.66)	(0.32)	(−0.38)	(−1.39)	(−1.78)
(−3 Year, +3 Year) around listing	1.51%***	0.46%	0.15%	−1.06%	−1.36%
	(3.95)	(0.64)	(0.21)	(−1.26)	(−1.60)

Table 3 **Smart beta performance before and after ETF listing: using SPY returns as benchmark.** This table performs a similar analysis as in Table 2 but uses SPDR S&P 500 ETF (SPY) returns as the benchmark to estimate CAPM alpha. Panel A reports the annualized alpha of smart beta indexes relative to SPY gross returns. In Panel B, we replace smart beta index returns with smart beta ETF returns over the post-listing period. Specifically, ETF CAPM alphas are estimated using gross and net returns of smart beta ETFs and SPY in columns (2) and (3), respectively. Columns (4) and (5) report the post-listing performance decline. Standard errors are clustered by factor theme as categorized by Morningstar and t-statistics are reported in parentheses. *, **, and *** denote the 10%, 5%, and 1% significance level, respectively. Our sample ends in December 2019.

Panel A: Pre-listing and Post-listing Index Performance

	(1) Before	(2) After	(3) Diff
All years before and after listing	3.07%***	−0.87%***	−3.94%***
	(4.87)	(−3.29)	(−5.23)
(−1 Year, +1 Year) around listing	1.64%***	0.60%	−1.04%**
	(3.69)	(1.26)	(−2.07)
(−2 Year, +2 Year) around listing	1.20%***	0.11%	−1.09%*
	(2.66)	(0.33)	(−1.74)
(−3 Year, +3 Year) around listing	1.60%***	−0.08%	−1.68%***
	(4.32)	(−0.21)	(−3.39)

Panel B: Pre-listing Index and Post-listing ETF Performance

	(1) Before	(2) After *Gross*	(3) After *Net*	(4) Diff (2)−(1)	(5) Diff (3)−(1)
All years before and after listing	3.07%***	−1.24%***	−1.43%***	−4.31%***	−4.50%***
	(4.87)	(−3.33)	(−3.69)	(−4.68)	(−4.79)
(−1 Year, +1 Year) around listing	1.64%***	0.17%	−0.07%	−1.47%***	−1.71%***
	(3.69)	(0.32)	(−0.13)	(−2.79)	(−3.25)
(−2 Year, +2 Year) around listing	1.20%***	−0.29%	−0.52%	−1.49%**	−1.71%**
	(2.66)	(−0.69)	(−1.17)	(−2.27)	(−2.56)
(−3 Year, +3 Year) around listing	1.60%***	0.02%	−0.19%	−1.59%**	−1.80%**
	(4.32)	(0.03)	(−0.26)	(−2.01)	(−2.24)

Table 4 **Ruling out strategic timing.** This table reports the performance of smart beta indexes relative to their factor-theme benchmarks. Specifically, we group smart beta indexes based on factor theme following Morningstar's classification, and we use the first-constructed index in each category as the benchmark for indexes in that category. Similar to Table 2, for each smart beta index, we calculate the excess returns and alphas relative to its factor-theme benchmark over different time windows before and after ETF listing. Then we report the average annualized index returns and alphas across all smart beta ETFs in Panels A and B, respectively. Standard errors are clustered by factor theme as categorized by Morningstar and the associated t-statistics are reported in parentheses. *, **, and *** denote the 10%, 5%, and 1% significance level, respectively.

Panel A: Return in Excess of Factor-Theme Benchmark			
	Before	After	Diff
All years before and after listing	2.85%***	−0.42%	−3.27%***
	(5.66)	(−1.29)	(−5.04)
(−1 Year, +1 Year) around listing	2.16%***	0.42%	−1.74%***
	(4.51)	(1.30)	(−3.95)
(−2 Year, +2 Year) around listing	1.45%***	−0.03%	−1.48%***
	(4.68)	(−0.07)	(−3.98)
(−3 Year, +3 Year) around listing	1.79%***	0.11%	−1.68%***
	(4.85)	(0.28)	(−4.09)
Panel B: Alpha Relative to Factor-Theme Benchmark			
	Before	After	Diff
All years before and after listing	2.80%***	−0.33%	−3.13%***
	(4.20)	(−0.89)	(−5.40)
(−1 Year, +1 Year) around listing	2.18%***	0.56%	−1.61%***
	(4.88)	(1.62)	(−3.74)
(−2 Year, +2 Year) around listing	1.56%***	0.18%	−1.38%***
	(3.06)	(0.36)	(−3.41)
(−3 Year, +3 Year) around listing	1.92%***	0.31%	−1.61%***
	(3.35)	(0.84)	(−3.37)

Table 5 **ETF size and post-listing performance decline.** This table reports the relationship between smart beta index performance and ETF size. In Panel A, we divide all smart beta indexes into two groups by their ETFs' average post-listing AUM. In Panel B, within each factor-theme category, we classify smart beta indexes into two groups by their ETFs' average post-listing AUM. In both panels, columns (1) and (2) show the average annualized CAPM alpha before and after ETF listing across all indexes. Column (3) shows the average after-minus-before-listing difference in index alphas. Column (4) shows the difference in the average after-minus-before-listing index alphas between the two groups. Standard errors are clustered by factor theme as categorized by Morningstar and the associated t-statistics are reported in parentheses. *, **, and *** denote the 10%, 5%, and 1% significance level, respectively.

Panel A: Based on ETF AUM				
	(1) Before	(2) After	(3) Diff	(4) Diff-in-Diff
below median	3.31%***	−0.85%***	−4.16%***	−1.88%***
	(8.08)	(−3.21)	(−9.02)	(−2.72)
above median	2.24%***	−0.04%	−2.28%**	
	(2.74)	(−0.14)	(−2.49)	
Panel B: Based on Within-Category ETF AUM				
	(1) Before	(2) After	(3) Diff	(4) Diff-in-Diff
below median	3.17%***	−0.99%***	−4.16%***	−1.82%**
	(8.29)	(−5.19)	(−9.68)	(−2.38)
above median	2.40%***	0.06%	−2.34%**	
	(2.74)	(0.16)	(−2.30)	

Table 6 **Evidence of data mining in index construction.** This table explores the degree to which smart beta indexes are susceptible to data mining. In Panel A, we classify smart beta indexes into three groups (Multi-factor, Value&Growth, and Others) based on factor theme. In Panel B, we classify a smart beta index into the high-ETF-discretion group if (*i*) the index is constructed "in-house" by the ETF sponsor or (*ii*) the index name contains the ETF sponsor's name or the index's release date is within six months of the ETF listing date. In Panel C, we classify smart beta indexes into two groups depending on whether their corresponding ETFs are managed by BlackRock iShares, Vanguard, or State Street, the top-three ETF sponsors by assets. In each panel, columns (2) and (3) show the average annualized CAPM alphas before and after ETF listing across all indexes in a given group. Column (4) shows the average after-minus-before-listing difference in CAPM alphas. Column (5) shows the difference in the average after-minus-before-listing alphas between groups. Standard errors are clustered by factor theme as categorized by Morningstar and the associated t-statistics are reported in parentheses. *, **, and *** denote the 10%, 5%, and 1% significance level, respectively.

Panel A: Multi-Factor v.s. Single-Factor Smart Beta

Factor-theme	(1) #ETF	(2) Before	(3) After	(4) Diff	(5) Diff-in-Diff
Multi-Factor	83	4.11%***	−0.79%**	−4.91%***	−3.15%***
		(11.34)	(−2.03)	(−9.33)	(−4.02)
Others	91	2.68%***	−0.02%	−2.70%***	
		(7.31)	(−0.05)	(−5.79)	
Value & Growth	64	1.16%**	−0.60%**	−1.76%***	
		(2.33)	(−2.30)	(−2.63)	

Panel B: High-ETF-Discretion v.s. Low-ETF-Discretion

High-discretion	(1) #ETF	(2) Before	(3) After	(4) Diff	(5) Diff-in-Diff
Yes	141	3.64%***	−0.62%*	−4.26%***	−2.57%***
		(11.24)	(−1.91)	(−11.41)	(−4.43)
No	97	1.51%**	−0.18%	−1.69%**	
		(2.43)	(−0.79)	(−2.43)	

Panel C: BlackRock/Vanguard/State Street v.s. Others

B/V/S	(1) #ETF	(2) Before	(3) After	(4) Diff	(5) Diff-in-Diff
No	164	3.31%***	−0.56%*	−3.87%***	−2.11%***
		(8.32)	(−1.76)	(−7.24)	(−3.65)
Yes	74	1.58%**	−0.19%	−1.76%***	
		(2.21)	(−0.94)	(−2.63)	

Table 7 **Pre-listing index performance and post-listing ETF flows.** This table estimates the relationship between the pre-listing backtested performance and the post-listing investment flows. In both panels, the dependent variable is the average monthly percentage ETF flows over the k-month period ($k = 6$ or 12) after ETF listing. The independent variables are the annualized benchmark-adjusted returns of the underlying index over the one-year, two-year, three-year, and entire time window before ETF listing. In Panel A, the independent variables are based on CAPM alphas. In Panel B, the independent variables are based on index returns in excess of the factor-theme benchmark as in Table 4. In all regressions, ETF-listing-year fixed effects and factor-theme fixed effects are included. *, **, and *** denote the 10%, 5%, and 1% significance level, respectively.

Panel A: Index Performance Measured by CAPM Alpha								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		$k = 6$				$k = 12$		
One year before listing	1.13				0.69			
	(1.33)				(1.30)			
Two years before listing		1.51*				0.80*		
		(1.88)				(1.79)		
Three years before listing			3.02***				1.60**	
			(2.66)				(2.41)	
Entire period before listing				3.86***				1.69***
				(2.84)				(2.68)
Listing-year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Factor-theme FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
No. Obs.	209	209	209	209	209	209	209	209
Adj. R^2	0.19	0.19	0.22	0.23	0.11	0.10	0.12	0.12

Panel B: Index Performance Measured by Returns in Excess of Factor-Theme Benchmark								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		$k = 6$				$k = 12$		
One year before listing	0.91				0.87*			
	(1.19)				(1.74)			
Two years before listing		0.66				0.67		
		(1.08)				(1.50)		
Three years before listing			4.17***				2.54***	
			(3.71)				(3.09)	
Entire period before listing				4.23***				2.04***
				(3.76)				(3.34)
Listing-year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Factor-theme FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
No. Obs.	209	209	209	209	209	209	209	209
Adj. R^2	0.18	0.18	0.22	0.23	0.11	0.10	0.14	0.12

Table 8 **Factor exposure and ETF expense ratio.** Panel A reports the loadings of smart beta indexes on their designated factors. For each factor theme in a given month, we form an equal-weighted portfolio of smart beta indexes and calculate the portfolio returns. We then estimate the portfolio's loadings on a set of factors, including market (MKTRF), size (SMB), value (HML), momentum (UMD), quality (QMJ), and volatility (VOL). For growth indexes, we replace the HML factor with an LMH factor, which is the negative of the HML factor. MKTRF, SMB, HML, UMD are from French's data library. QMJ is from the AQR website. VOL is a total volatility factor that represents the returns of a value-weighted portfolio, which longs (shorts) stocks in the bottom (top) quintile of total return volatility (see Ang et al. (2006)). We also estimate the factor loadings for the earliest smart beta index in each factor theme. For each factor theme, we report the loadings on the designated factor. t-statistics in parentheses are computed based on standard errors with Newey-West correction of 12 lags. *, **, and *** denote the 10%, 5%, and 1% significance level, respectively. The sample period starts from January 2001 and ends in December 2019. Panel B shows the annualized expense ratio of smart beta ETFs that track the indexes. For each factor-theme category, we compare the expense ratio of the ETF that tracks the earliest index and the average expense ratio of the ETFs that track later-constructed indexes.

Panel A: Index Loading on the Designated Factor			
Factor-Theme Category	First Index	Other Indexes	Diff
Value (HML Factor)	0.279***	0.351***	0.072*
	(11.00)	(13.76)	(1.95)
Growth (LMH Factor)	0.307***	0.143***	−0.164***
	(11.02)	(8.58)	(−4.98)
Momentum (MOM Factor)	0.358***	0.290***	−0.068
	(6.69)	(6.56)	(−0.97)
Quality (QMJ Factor)	0.209***	0.148***	−0.061
	(6.40)	(7.69)	(−1.64)
Risk/Vol (VOL Factor)	0.077***	0.075***	−0.002
	(3.27)	(4.43)	(−0.09)
Panel B: Expense Ratio of ETFs that Track the Indexes			
Factor-Theme Category	First Index	Other Indexes	Diff
Value	0.20%	0.25%	0.05%**
			(1.96)
Growth	0.20%	0.24%	0.04%***
			(3.85)
Momentum	0.15%	0.40%	0.25%***
			(3.42)
Quality	0.15%	0.35%	0.20%***
			(6.83)
Risk/Vol	0.15%	0.30%	0.15%**
			(2.19)

Table 9 **Smart beta performance deterioration: international evidence.** The sample consists of 77 smart beta ETFs which invest in the European, Canadian, Australian, or the UK equity market and are listed by December 2018. Following Table 2, we calculate the market-adjusted performance of these smart beta indexes before and after the corresponding smart beta ETFs are listed. For each smart beta ETF, we use the corresponding regional market index from MSCI as the benchmark. In Panel A, we report the index performance using annualized index returns in excess of the corresponding regional market index returns. In Panel B, we calculate the annualized index CAPM alpha relative to the corresponding regional market index. Standard errors are clustered by factor theme as categorized by Morningstar and the associated t-statistics are reported in parentheses. *, **, and *** denote the 10%, 5%, and 1% significance level, respectively. Our sample ends in December 2019.

Panel A: Annualized Return in Excess of Market Index			
	Before (1)	After (2)	Diff (3)
All years before and after listing	3.02%***	−0.15%	−3.17%***
	(15.07)	(−0.38)	(−5.80)
(−1 Year, +1 Year) around listing	0.74%	0.14%	−0.60%
	(0.77)	(0.15)	(−0.67)
(−2 Year, +2 Year) around listing	1.70%***	0.16%	−1.54%***
	(3.02)	(0.26)	(−3.13)
(−3 Year, +3 Year) around listing	1.75%***	0.25%	−1.50%***
	(4.26)	(0.53)	(−2.81)
Panel B: Annualized CAPM Alpha			
	Before (1)	After (2)	Diff (3)
All years before and after listing	3.43%***	0.26%	−3.16%***
	(12.26)	(0.51)	(−5.95)
(−1 Year, +1 Year) around listing	1.22%	0.89%	−0.33%
	(1.24)	(1.06)	(−0.45)
(−2 Year, +2 Year) around listing	2.09%***	0.68%	−1.42%**
	(2.75)	(1.02)	(−2.36)
(−3 Year, +3 Year) around listing	2.04%***	0.74%	−1.31%***
	(3.52)	(1.26)	(−2.65)

Appendix A. Additional Results

This section reports additional results and robustness checks. For the results in Section IV, we conduct four robustness tests: (*i*) we show that the post-listing performance decline pattern is also robust under panel regressions (see Table A.1); (*ii*) when we compare index performance before and after the index release date rather than the ETF listing date, the results are even stronger (see Table A.2); (*iii*) we conduct sub-sample analysis and show that the patterns in Table 2 are robust and are stronger among more recent listed smart beta ETFs (see Table A.3); (*iv*) we show that the result is robust when we require ETFs to have a longer history of index returns (at least three years) before ETF listing date (see Table A.4).

Furthermore, we analyze the relationship between ETF flows and future ETF performance. Either through portfolio exercise or through panel regressions, we find that investment flows significantly and negatively predict the future performance of smart beta ETFs (see Table A.5). Lastly, we conduct a robustness check for Table 8 by using both pre-listing and post-listing index returns, and the results are similar (see Table A.6).

Table A.1 **Robustness check of Table 2: panel regressions.** In this table, we analyze the post-listing performance decline through panel regressions. In Panel A, both pre-listing and post-listing performance is measured by index returns/alphas. In Panel B, pre-listing performance is measured by index returns/alphas, and post-listing performance is measured by ETF gross returns/alphas. In columns (1)-(2), the dependent variable is monthly CAPM alpha (in percent). In columns (3)-(4), the dependent variable is monthly returns in excess of the factor-theme benchmark index (in percent). The key independent variable in all columns is a dummy variable, which indicates whether an index-month/ETF-month observation is in the post-listing period. Standard errors are clustered by time and by factor theme. *, **, and *** denote the 10%, 5%, and 1% significance level, respectively.

Panel A: Pre- and Post-listing Index Performance				
	(1)	(2)	(3)	(4)
DepVar:	CAPM Alpha		Factor-Theme Adj Ret	
Post Listing	-0.29^{***}	-0.27^{***}	-0.25^{***}	-0.26^{***}
	(-4.11)	(-3.09)	(-3.30)	(-2.76)
ETF FE	No	Yes	No	Yes
No. Obs.	52,117	52,117	52,117	52,117
Adj. R^2	0.004	0.007	0.003	0.002
Panel B: Post-listing ETF Performance				
	(1)	(2)	(3)	(4)
DepVar:	CAPM Alpha		Factor-Theme Adj Ret	
Post Listing	-0.27^{***}	-0.27^{***}	-0.20^{**}	-0.23^{**}
	(-3.44)	(-3.00)	(-2.35)	(-2.31)
ETF FE	No	Yes	No	Yes
No. Obs.	52,117	52,117	52,117	52,117
Adj. R^2	0.002	0.002	0.001	0.001

Table A.2 **Smart beta index performance before and after index release.** This table reports index performance before and after the index release date rather than the ETF listing date. The analysis follows Table 2. The sample period ends in December 2019. Standard errors are clustered by factor-theme as categorized by Morningstar and the associated t-statistics are reported in parentheses. *, **, and *** denote the 10%, 5%, and 1% significance level, respectively.

Annualized index CAPM alpha			
	Before	After	Diff
All years before and after index release	3.49%***	−0.41%	−3.90%***
	(9.58)	(−1.28)	(−8.65)
(−1 Year, +1 Year) around index release	1.79%***	0.29%	−1.50%***
	(6.60)	(0.55)	(−2.69)
(−2 Year, +2 Year) around index release	2.26%***	−0.22%	−2.48%***
	(6.97)	(−0.44)	(−4.65)
(−3 Year, +3 Year) around index release	2.34%***	−0.16%	−2.50%***
	(9.76)	(−0.36)	(−5.78)

Table A.3 **Subsample analysis of Table 2: ETFs listed before and after 2015.** In this table, we re-perform the analysis in Table 2 in the subsamples of smart beta indexes whose corresponding ETFs are listed before 2015 and after 2015, respectively. Annualized CAPM alphas are reported in both panels. Standard errors are clustered by factor theme as categorized by Morningstar and the associated t-statistics are reported in parentheses. *, **, and *** denote the 10%, 5%, and 1% significance level, respectively.

Panel A: ETFs Listed before 2015				
	# ETF	Before	After	Diff
All years before and after	96	2.47%***	−0.18%	−2.65%***
		(3.22)	(−0.64)	(−4.23)
(−1 Year, +1 Year) around listing	96	2.53%***	2.02%*	−0.52%
		(6.21)	(1.69)	(−0.51)
(−2 Year, +2 Year) around listing	96	1.97%**	1.34%*	−0.64%
		(1.99)	(1.84)	(−0.39)
(−3 Year, +3 Year) around listing	96	2.14%**	1.03%***	−1.11%
		(2.51)	(2.66)	(−1.05)
Panel B: ETFs Listed after 2015				
	# ETF	Before	After	Diff
All years before and after	142	2.98%***	−0.62%**	−3.60%***
		(5.99)	(-1.97)	(-5.10)
(−1 Year, +1 Year) around listing	142	0.32%	−0.20%	−0.52%
		(0.74)	(−0.97)	(−1.14)
(−2 Year, +2 Year) around listing	142	0.65%*	−0.36%	−1.01%***
		(1.95)	(−1.30)	(−2.89)
(−3 Year, +3 Year) around listing	142	1.09%***	−0.44%	−1.52%***
		(4.14)	(−1.35)	(−4.27)

Table A.4 **Robustness check of Table 2: requiring a longer return history before ETF listing.** In this table, we re-perform the analysis in Table 2 with the subsample of smart beta indexes that have non-missing returns over at least three years before ETF listing. Standard errors are clustered by factor-theme as categorized by Morningstar and the associated t-statistics are reported in parentheses. *, **, and *** denote the 10%, 5%, and 1% significance level, respectively.

Annualized Index CAPM Alpha				
	#ETF	Before	After	Diff
All years before and after listing	223	2.91%***	−0.43%	−3.34%***
		(4.94)	(−1.63)	(−4.88)
(−1 Year, +1 Year) around listing	223	1.27%***	0.71%	−0.56%
		(4.76)	(1.30)	(−1.05)
(−2 Year, +2 Year) around listing	223	1.21%***	0.33%	−0.88%
		(2.74)	(0.88)	(−1.28)
(−3 Year, +3 Year) around listing	223	1.57%***	0.16%	−1.41%***
		(4.28)	(0.42)	(−2.64)

Table A.5 **ETF flows and future performance.** This table analyzes the relationship between ETF flows and future ETF performance. Panel A reports results from portfolio-sorting exercises. At each month-end, we sort smart beta ETFs into quintiles based on their average flows over the past 6 or 12 months. We hold the ETF portfolios in the next month and compute the AUM-weighted returns. We report average monthly excess returns, CAPM alpha, and FFC4 alpha (in percent) during the holding period of January 2007 to December 2019. Standard errors are calculated using the Newey-West correction of 12 lags. Panel B reports the results from panel regressions. The dependent variable is the ETF return in month t+1. The key independent variable is the average monthly percentage flow over the past 6 or 12 months. We include the past 6-/12-month ETF returns and the ETF AUM at the end of month t as control variables. ETF and time fixed effects are included. Standard errors are clustered by time. *, **, and *** denote the 10%, 5%, and 1% significance level, respectively.

Panel A: ETF Portfolios Sorted by Past Flow			
	Sort by Past 6-Month Flow		
Portfolio	Excess	CAPM	FFC4
1 (Low Flow)	0.68*	−0.10	−0.05
	(1.65)	(−1.24)	(−0.69)
5 (High Flow)	0.46	−0.31***	−0.26***
	(1.10)	(−3.70)	(−2.97)
5−1	−0.23**	−0.21*	−0.21**
	(−2.44)	(−1.89)	(−2.16)
	Sort by Past 12-Month Flow		
Portfolio	Excess	CAPM	FFC4
1 (Low Flow)	0.81*	0.04	0.08
	(1.94)	(0.54)	(1.02)
5 (High Flow)	0.54	−0.23**	−0.20*
	(1.28)	(−2.22)	(−1.90)
5−1	−0.27***	−0.27**	−0.29**
	(−2.69)	(−2.15)	(−2.40)

Panel B: Panel Regressions				
	(1)	(2)	(3)	(4)
	k=6		*k=12*	
Past k-Month Flow	−0.02***	−0.02***	−0.01**	−0.01**
	(−2.62)	(−2.66)	(−2.29)	(−2.46)
Past k-Month Ret		−0.09**		−0.14
		(−2.25)		(−1.45)
AUM		0.00		0.00
		(1.09)		(1.15)
ETF and Time FEs	Yes	Yes	Yes	Yes
No. Obs.	13,863	13,863	13,863	13,863
Adj. R^2	0.586	0.586	0.586	0.586

Table A.6 **Robustness check for Table 8.** For each factor theme in a given month, we form an equal-weighted portfolio of all smart beta indexes whose corresponding ETFs have been listed. We then estimate the portfolio's loading on a set of factors, including market (MKTRF), size (SMB), value (HML), momentum (UMD), quality (QMJ), and volatility (VOL). For indexes in each factor-theme category, we report their loadings on the designated factor. For growth indexes, we report their loadings on LMH, which is the negative of the HML factor. The sample ends in December 2019. *, **, and *** denote the 10%, 5%, and 1% significance level, respectively.

Factor-Theme Category	First Index	Other Indexes	Diff
Value (HML Factor)	0.279***	0.402***	0.123***
	(11.00)	(12.04)	(2.82)
Growth (LMH Factor)	0.307***	0.152***	−0.155***
	(11.02)	(8.90)	(−4.65)
Momentum (MOM Factor)	0.287***	0.243***	−0.045
	(5.86)	(8.93)	(−0.80)
Quality (QMJ Factor)	0.182***	0.098**	−0.085
	(4.36)	(2.09)	(−1.36)
Risk/Vol (VOL Factor)	0.162***	0.163***	0.001
	(3.53)	(3.42)	(0.01)

Appendix B. An Illustration of Index Construction

This section presents the screenshots of the rules in determining the number of constituents of a value index whose ETF is offered by one of the major sponsors. As of June of 2020, this ETF managed more than $50 billion of assets. These screenshots are from the methodology document of the index. We choose not to disclose the name of the index and the identity of the ETF sponsor.

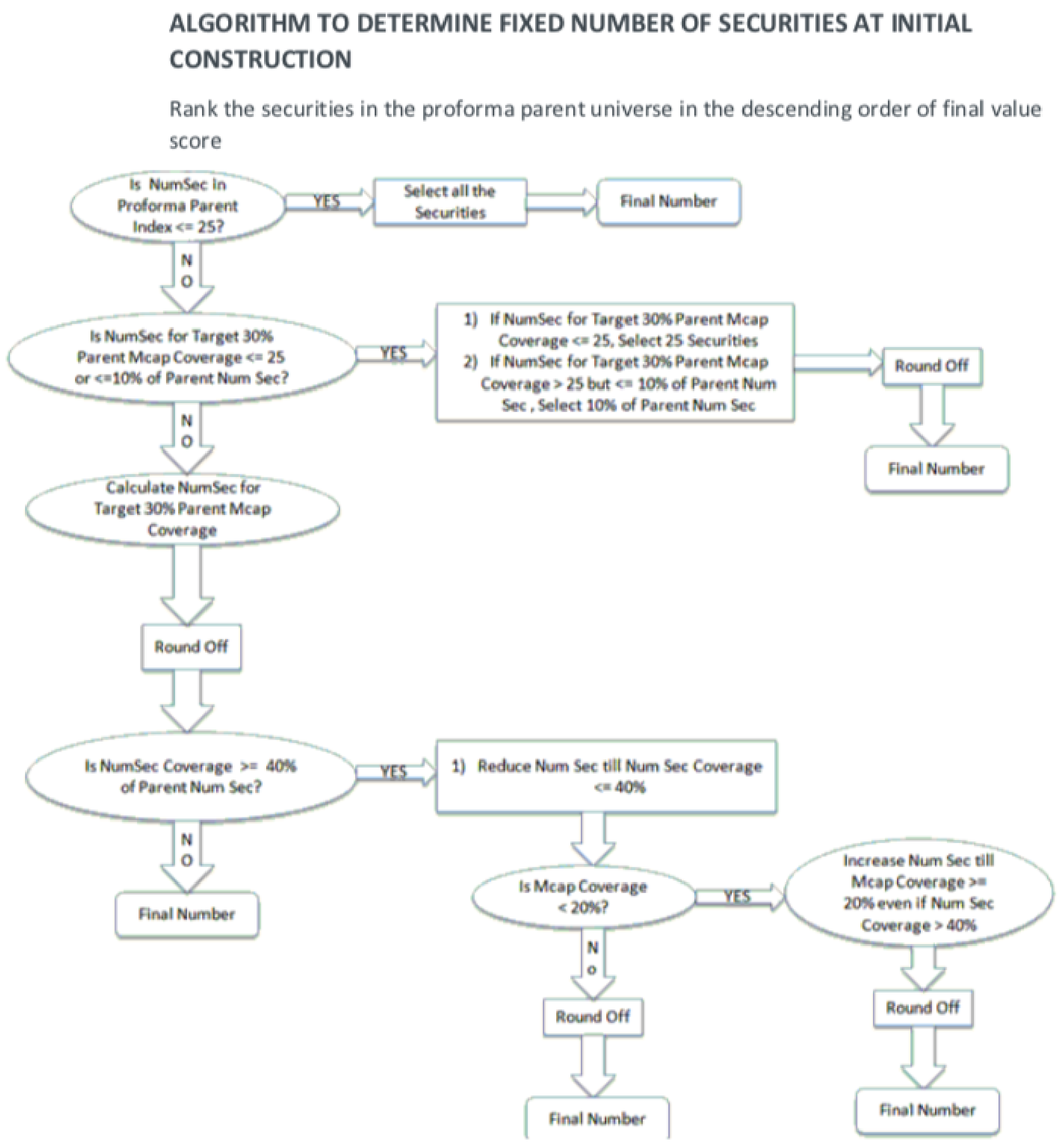

Figure B.1: **Screenshot from the methodology document of a value index.**

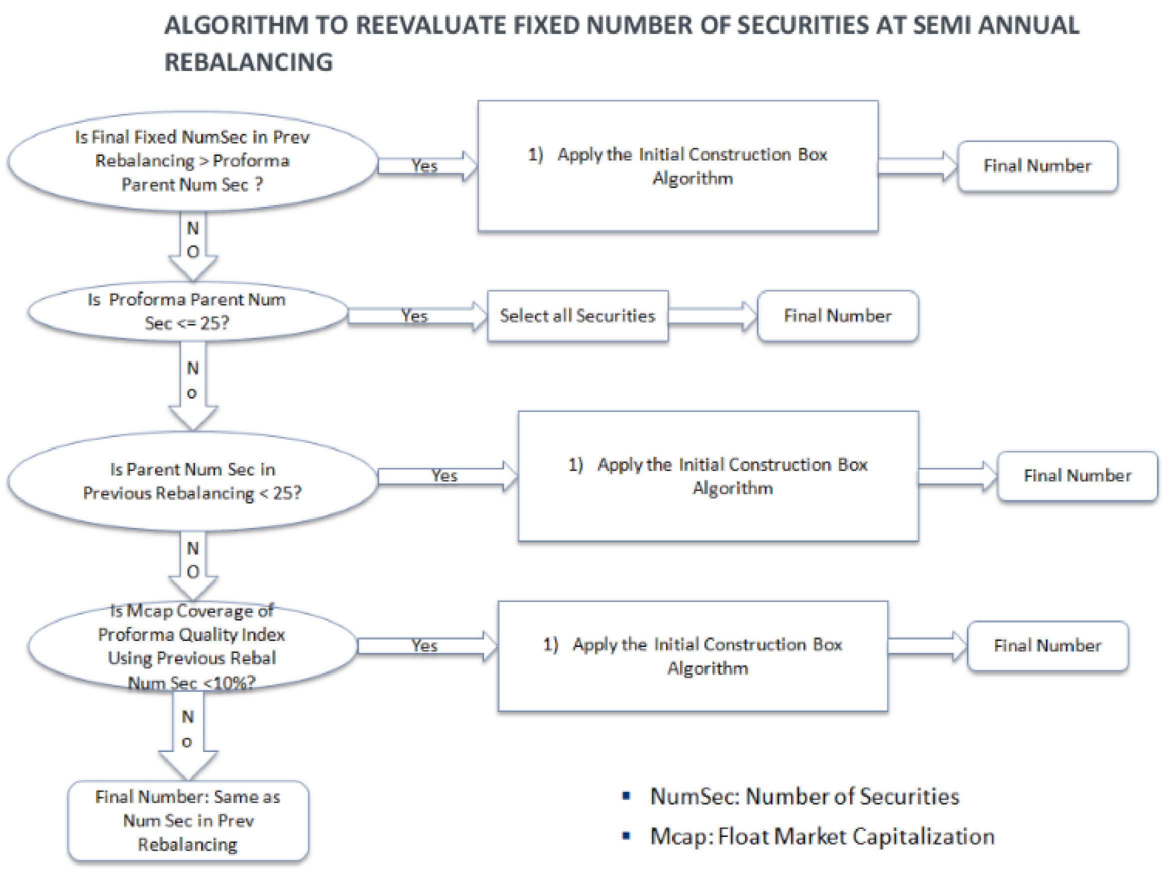

Figure B.2: **Screenshot from the methodology document of a value index.**