The Smart Beta Mirage

Shiyang Huang
The University of Hong Kong HKU Business School
huangsy@hku.hk

Yang Song
University of Washington Foster School of Business
songy18@uw.edu

Hong Xiang
The Hong Kong Polytechnic University Faculty of Business School of Accounting and Finance
hong.xiang@polyu.edu.hk (corresponding author)

Abstract

We document and explain the sharp performance deterioration of smart beta indexes after the corresponding exchange-traded funds (ETFs) are launched for investment. While smart beta is purported to deliver excess returns through factor exposures, the market-adjusted return of smart beta indexes drops from about 3% "on paper" before ETF listings to about -0.50% to -1% after ETF listings. This performance decline cannot be explained by variation in factor premia, strategic timing, or diminishing returns to scale. Instead, we find strong evidence of data mining in the construction of smart beta indexes, which helps ETFs attract flows, as investors respond positively to backtests.

I. Introduction

Exchange-traded funds (ETFs) have grown remarkably during the past two decades. In the ETF marketplace, the fastest-growing and possibly hottest segment is "smart beta" ETFs, which ETF sponsors often refer to as one of the most important financial innovations supported by asset pricing research. 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/growth and momentum). Since their first appearance in 2000, the assets under management (AUM) of U.S. equity smart beta ETFs increased to about $700 billion as of

2018, accounting for more than 20% of the aggregate ETF market (Morningstar (2019)). BlackRock 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 their so-called rules-based formulaic index construction.[1] Although smart beta indexes are claimed to be built upon formulaic rules, the degrees of freedom over smart beta index construction are actually quite large in practice. With large-scale data availability and strong computing power, it is not difficult to develop a smart beta index with outstanding 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 (Financial Times (2017), Morningstar (2017), and ETF Stream (2019)).[2] Presumably, the better the backtested performance of smart beta indexes, the easier it is to sell ETFs to investors.[3] This tendency 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 strong backtests, raises several important questions: Does the claimed "smartness" of smart beta exist outside of backtests? If not, to what extent is the construction of smart beta indexes susceptible to data mining? 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. Moreover, although ETF sponsors typically market their smart beta ETFs as "smart" financial innovations endorsed by academic research,[4] it is not clear whether the proliferation of smart beta ETFs adds value for investors.

To address these questions, we compile a sample of U.S. domestic equity smart beta ETFs listed between 2000 and 2018 and manually collect their underlying smart beta indexes and the index returns before and after ETF listing. Our sample is representative, covering about 80% of the total AUM of the U.S. 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 article. Probably the most disappointing finding to smart beta investors is that the claimed "smart" performance of

---

[1] For example, the prospectus for the JP Morgan U.S. Momentum Factor ETF states, "The J.P. Morgan U.S. Momentum Factor Index uses a rules-based risk allocation and factor selection process developed by J.P. Morgan Asset Management."

[2] Morningstar (2017) notes that smart beta marketing relies almost exclusively on backtested data, possibly due to the lack of sufficient real-life track records. According to a senior market participant cited 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."

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

[4] For example, Investco (2018) cites more than 10 academic papers to endorse its smart beta ETFs.

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 capital asset pricing model (CAPM) alpha of 2.77% ($t = 4.59$) per year over an average 13-year period before ETF listings.[5] Once smart beta ETFs are launched to investors, we observe a substantial decline in 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- and post-listing periods (referred to as the performance decline hereafter) is 3.22% on an annualized basis ($t = 4.63$).[6]

We also find that the post-ETF-listing performance decline is prevalent across smart beta indexes of different factor themes. For example, the value/growth smart beta indexes, on average, have a significant annualized performance drop of 1.76%, while the smart beta indexes of other factor themes have even sharper average performance declines (ranging from 2.70% to 4.91%). We also conduct a large set of robustness checks to further confirm our findings.

Note also that we mainly use smart beta index returns rather than ETF returns to measure the post-ETF-listing performance. Using smart beta index returns has two advantages: First, ETF returns do not exist before ETF listings, and focusing on index returns makes the pre- and post-listing performance directly comparable. Second, because we use smart beta index returns to gauge post-listing performance, the transaction/implementation costs in replicating indexes are not driving the performance decline.[7] When we use ETF returns over the post-ETF-listing period, we observe an even slightly sharper performance decline.

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

We first show that our findings are *not* explained by the general time trend and variation in factor premia. Under this explanation, factor profitability declines over time; therefore, the post-ETF-listing return decline of smart beta would be a natural outcome of this general trend. To examine this explanation, we control for *factor-wide* return variation when measuring smart beta performance. Specifically, we categorize smart beta indexes by the Morningstar-classified factor themes and then

---

[5]We follow Fama and French (2015) and use as market returns the value-weighted returns of all U.S. firms listed in the CRSP database. This choice is reasonable because investors can invest in the aggregate market index through the Vanguard Total Stock Market ETF (VTI) at a very low cost (i.e., 3 BPS per year).

[6]The results are very similar if we simply measure performance based on the difference between smart beta returns and aggregate market returns. Moreover, if we use returns of the SPDR S&P 500 ETF (SPY) as the benchmark to estimate the CAPM alphas, the performance decline is even sharper, averaging 3.94% ($t = 5.23$) on an annualized basis (from 3.07% before listing to -0.87% after listing).

[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.

benchmark each index relative to the earliest-constructed index in its factor theme category.[8] Alternatively, we use the academic asset pricing factors, like the Fama and French (1993) 3 factors and Carhart (1997) momentum factor, as the factor theme benchmark. Strikingly, smart beta indexes even outperform their factor theme benchmarks on paper. That is, in the backtest before ETF listing, a typical smart beta index produces significantly positive alphas *even after controlling for its targeted factor returns.* The magnitude of the post-listing performance decline is quantitatively similar to the performance decline measured by CAPM alpha.[9]

Our research focus is not on the reliability of asset pricing factors (anomalies) or whether smart beta efficiently delivers factor premia. In fact, because we also control for the factor theme benchmark (either the earliest index or the academic factor) when evaluating smart beta performance, our findings are mostly *distinct* from whether factor premia exist or persist and are not driven by factor-wide return variation.

The "strategic timing" hypothesis cannot explain our findings either. Under this hypothesis, ETF sponsors might strategically launch products when the tracked factors have recently outperformed the aggregate market. By adjusting for factor theme benchmark returns, we effectively control the time trend in factor profitability, including the potential mean-reversion of factor returns. Yet, we still observe a significant performance decline after ETF listings.

Another possibility is that investment flows into smart beta ETFs could make factor-based strategies overcrowded, 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 proxy for the influence of scale effects on smart beta returns using various measures, including ETF AUM, the number of stock holdings (Pástor, Stambaugh, and Taylor (2020)), and the average portfolio liquidity based on the measure of Amihud (2002). Neither across all ETFs nor within factor theme categories do we find evidence that larger ETFs, ETFs with more concentrated positions, or ETFs that hold more illiquid stocks experience more significant post-listing performance decline. Thus, it is unlikely that the scale effects drive our results.

The fourth explanation is that the decline in performance after ETF listing is driven by the deterioration of factor performance after academic publication on the factor (e.g., McLean and Pontiff (2016)). To rule out this channel, we restrict our analysis of smart beta performance to the period after the corresponding academic publication date. We still find a significant post-listing performance decline pattern, which suggests that our results are not driven by the publication effects.

Because these seemingly plausible explanations fail to account for the observed patterns, we explore the data mining hypothesis. We find supporting evidence that the construction of smart beta indexes involves data mining so that

---

[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.

[9]Moreover, 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 12 well-studied factors declines by about 30 BPS per year.

FIGURE 1

An Indicative Case of Data Mining

Figure 1 plots the cumulative net asset value (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 gray) was constructed in Dec. 1997. The "enhanced" version of the first index (in red) was constructed in Dec. 2014, and the ETF tracking this index was listed shortly thereafter. Note that the returns of the "enhanced" index before the ETF listing are "on paper" and are not investable.
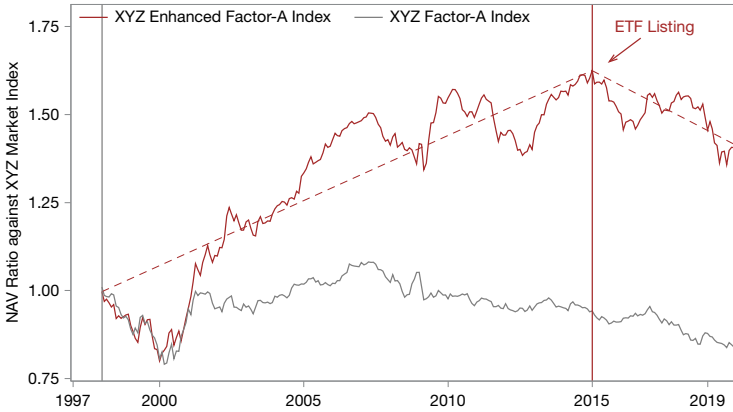
backtested performance is not sustained after ETFs go live. To illustrate data mining in action, Figure 1 plots cumulative returns of two smart beta indexes of the same factor theme constructed by one of the world's major index providers: the XYZ Value Index and XYZ Enhanced Value Index.[10] The XYZ Value Index was released in 1997. After years of no excess returns relative to the market, the index provider constructed the XYZ Enhanced Value Index in Dec. 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 on-paper superior performance of the enhanced index disappeared and trailed the aggregate market.

Motivated by this real-world example, we argue that data mining largely drives the performance decline of smart beta indexes. To support this argument, we provide five pieces of evidence as follows.

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. Multi-factor indexes can freely combine multiple factors so that the degree of freedom in building multi-factor indexes is higher by nature than for single-factor indexes. 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

_____

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

conjecture, we find that before ETF listings, multi-factor indexes have the best backtests and the sharpest post-listing performance decline. In contrast, value and growth indexes have the smallest performance decline after ETF listings.

Second, we explore the discretion ETF sponsors have during index construction. As strong 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.[11] 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 following two nonexclusive 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 6 months of the ETF listing date.[12] 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 influence in index construction. In line with our conjecture, the performance decline for high-ETF-discretion smart beta indexes is much larger than for 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 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.

Fourth, we explore the heterogeneity in the constraints of index construction. In particular, a subset of the smart beta indexes in our sample was derived from well-known existing indexes.[13] We argue that these smart beta indexes are limited to selecting stocks within a pre-specified stock pool (e.g., S&P 500 index constituents), allowing less room for data mining.[14] Consistent with this argument, we find that smart beta indexes derived from existing indexes experience a milder performance decline after ETF listings than other indexes.

Fifth, we compare the performance decline between a group of smart beta indexes that are tracked by multiple ETFs versus those tracked by only one ETF. We conjecture that the discretion of data mining in index construction can be largely reduced when the same index is tracked by multiple ETFs. Indeed, we find that these smart beta indexes experience a much smaller performance decline compared with those tracked by a single ETF.

Thus far, our evidence is consistent with data mining in indexation leading to a large performance decline after the launch of smart beta ETFs. Although our

---

[11]When ETF sponsors collaborate with third-party index providers to construct smart beta indexes, the sponsors can be extensively involved in index construction, and index providers are likely to cater to their demands (Financial Times (2015), Weinberg (2018)).

[12]About 60% of 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).

[13]For instance, the S&P 500 momentum index is a smart beta index derived from the S&P 500 index and by overweighting S&P 500 index constituents that exhibit persistence in their relative performance.

[14]Compared with some mutual funds that can freely select stocks from the entire market, smart beta indexes (particularly those indexes derived from existing indexes) could be more constrained in stock selection since they are designed to track asset pricing factors.

findings suggest that data mining could be detrimental to smart beta investors, we find that professional financial services and information providers fail to highlight the potential risk of data mining in smart beta products. Hence, 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 1-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 mining in index construction, because this behavior gets rewarded through flows.[15]

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 for investors in terms of returns. However, one might argue that the proliferation of smart beta ETFs could still be valuable if these products help investors get higher or cheaper exposure to asset pricing factors. Unfortunately, our further analyses show that this is not the case; instead, 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, we find that the majority of smart beta ETFs seem to add little value for investors both in terms of excess returns beyond the broad market and in terms of desired factor exposures. These results raise concerns about the prolifer-ation of smart beta ETFs. 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 and index builders have set up intricate and arbitrary rules in smart beta indexation, they may believe that backtests best represent future performance. At the very least, our findings suggest that investors need to be aware that backtesting is not, on average, an effective mechanism to identify the future outperformance of smart beta ETFs.

II.   Literature Review

Our article contributes to the literature on 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

---

[15]In Table A.11 in the Supplementary Material, we further show that flows into smart beta ETFs significantly negatively predict future ETF performance, reinforcing our conclusion that ETF investors are likely to be unsophisticated.

Zhang (2009), and Wermers (2012)), misreport their portfolios (e.g., Bollen and Pool (2009), (2012), Cici, Gibson, and Merrick (2011), Chen, Cohen, and Gurun (2021), and Agarwal, Barber, Cheng, Hameed, and Yasuda (2022)), take an excess risk to boost performance (e.g., Carhart, Kaniel, Musto, and Reed (2002), Kaniel and Parham (2017)), and obfuscate fee structures and disclosures (e.g., Barber, Odean, and Zheng (2005), Edelen, Evans, and Kadlec (2012), and DeHaan, Song, Xie, and Zhu (2021)). 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.[16] While our study focuses on smart beta ETFs, our findings also have important implications for other segments of the ETF market, such as sector ETFs and thematic ETFs,[17] that also involve discretionary index construction.

In another important contribution to this literature, Evans (2010) shows that mutual fund firms use an incubation strategy when launching new mutual funds that harms mutual fund investors.[18] Although both strategies 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 investment firm deploys 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, typically less than 3 years (Evans (2010)). However, the average length of on-paper backtested index returns is about 13 years in our sample.

Our article is also related to the growing literature on ETFs, particularly to the debate 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), Agarwal, Hanouna, Moussawi, and Stahel (2018), Ben-David, Franzoni, and Moussawi (2018), Da and Shive (2018), and Pan and Zeng (2019)) and that ETF investors often make suboptimal decisions (e.g., Ben-David et al. (2021), Brown, Cederburg, and Towner (2021a)). On the other hand, some argue that ETFs improve market efficiency (e.g., Box, Davis, Evans, and Lynch (2021), Glosten, Nallareddy, and Zou (2021), and Huang, O'Hara, and Zhong (2021)). 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.

---

[16]Suhonen, Lennkh, and Perez (2017) analyze rules-based strategies across five asset classes offered by 15 investment banks. Similar to our findings on smart beta ETFs, they observe significant performance drop-offs from backtests when these strategies go live. Several practitioners (Li and West (2017), Pattabiraman (2020)) have also observed that smart beta indexes experience return declines after ETFs are listed, suggesting the possibility of data mining. These prior studies neither examine the underlying mechanism nor explore the influences on investment flows.

[17]Ben-David, Franzoni, Kim, and Moussawi (2021) provide evidence that thematic ETFs are listed to cater to investor sentiment.

[18]Mutual fund incubation is a strategy that mutual fund firms use to test multiple new funds, with only outperforming funds being open to the public.

III.   Background and Data on Smart Beta ETFs

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

A.   Smart Beta and the Role of Data Mining

In the booming ETF market, smart beta ETFs have experienced the largest growth, accounting for more than 20% of the overall ETF marketplace as of 2018 (Morningstar (2019)).[19] In the simplest terms, smart beta ETFs follow an indexation approach, but they do so differently from traditional broad cap-weighted market indexes. Smart beta ETFs emphasize alternative index construction rules and claim to provide investors excess returns through exposure 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.

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), Weinberg (2018)).[20]

Although smart beta sponsors claim to follow rules-based and formulaic indexation procedures, the discretion they have in constructing smart beta indexes is very large 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, or 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, Section B of the Supplementary Material presents the "rules" for determining the number of constituents of a value smart beta index with the ETF offered by a large ETF sponsor.[21] These index rules are quite intricate and hard to justify by economic or intuitive rationale. Given this large discretion, it is not difficult to build a "smart" value index that produces strong backtests out of thousands or even millions of trials. Value is arguably the most-studied factor; thus, the degree of freedom for other smart beta factor themes can only be higher, with multi-factor smart beta indexes being free to choose among multiple factors.

---

[19]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 have the most smart beta ETF products.

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

[21]The index name and the ETF sponsor name are hidden on purpose. As of June 2020, this ETF has an AUM of more than 50 billion U.S. dollars.

Furthermore, backtests play a key role in launching and promoting smart beta products, as strong backtested returns help attract investment flows (Financial Times (2017), Morningstar (2017), and ETF Stream (2019)).[22] 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 strong backtests. Moreover, index providers are paid according to the amount managed against their benchmarks, known as index-licensing fees (Financial Times (2019), An, Benetton, and Song (2021)). In short, there are plenty of incentives and space for data mining in smart beta index construction.

B.   Data Description

We next describe how we construct the sample of smart beta ETFs and 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.[23] For each ETF, we obtain the ETF name, the ETF listing date, and the factor theme from Morningstar Direct. Specifically, "factor theme" refers to the broad-type factor category of smart beta ETFs classified by Morningstar.[24] In the second step, we use the CRSP Mutual Fund database to cross-check these smart beta ETFs based on the ticker and names. After this step, we obtain 379 smart beta ETFs.

In the third step, for each smart beta ETF, we manually identify the underlying index and collect the index performance start date, index release date, and return history of the index.[25] 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). For index names, we collect index information and the entire history of index returns from Morningstar Direct or from the websites of the index providers. We next want to compare the smart beta index performance before and after ETF listings. Therefore, in the fourth step, we require the underlying index of a smart beta ETF to have nonmissing returns 12 months before and 12 months

---

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

[23]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."

[24]Morningstar classifies smart beta ETFs into nine broad-type factor theme categories: i) Multi-factor, ii) Value, iii) Growth, iv) Risk/Volatility, v) Dividend, vi) Momentum, vii) Quality, viii) Fundamentals, and ix) Others. Our sample contains 215 smart beta ETFs under categories (i)–(viii) and an additional 23 under the "Others" category. Smart beta ETFs in the "Others" category can be further classified into several subcategories, including "equal-weighted," "buyback," "beta," "reversal," and so forth. Because each of these subcategories only contains a few numbers of ETFs, we follow Morningstar classification by grouping them into an "Others" category.

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

FIGURE 2

Aggregate AUM and Number of Smart Beta ETFs by Factor Theme

In Figure 2, the left axis shows the number of smart beta ETFs listed each year under different smart beta factor themes based on our sample, and the right axis shows the total AUM across all ETFs in our sample.
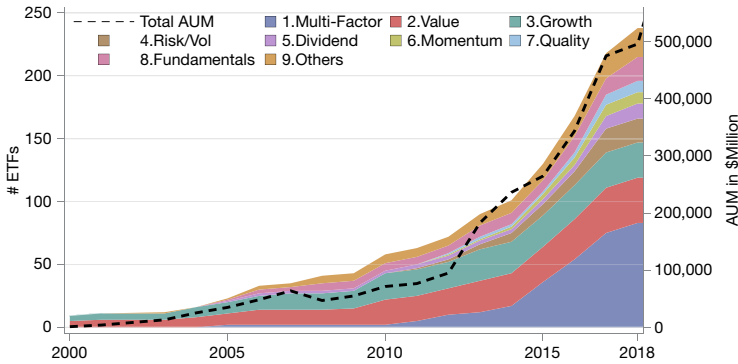
after the ETF listing.[26] Our final sample has 238 U.S. domestic equity smart beta ETFs listed between 2000 and 2018, tracking 223 unique smart beta indexes.[27]

Based on the AUM at the end of 2019, the 238 smart beta ETFs in our final sample 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 high percentage suggests that our sample is representative of the U.S. equity smart beta ETF market, with 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. The figure clearly shows 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).

Table 1 reports summary statistics of our sample. About 16% of smart beta ETFs were launched before 2010, 24% were launched between 2010 and 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. 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.

Note that each smart beta index has "on-paper" returns and "live" returns, as illustrated in Figure 3. From the index performance start date to the ETF listing date, smart beta returns are not accessible to investors and thus are denoted as "on paper"

---

[26]We discard 141 ETFs in this sample selection procedure. Among them, 88 ETFs are discarded because we cannot find their historical index return data, and 53 ETFs are discarded due to missing index return data in the 12 months before and after ETF listing dates. We compare the total net assets (TNAs) of the 141 discarded ETFs and the 238 ETFs in our final sample and find that the average TNA of the discarded ETFs is much smaller than that of ETFs in the final sample. Because our sample selection procedure filters out smaller ETFs that are more likely to engage in data mining, our analysis based on the 238 ETFs is likely to be a lower-bound estimate for the influence of data mining on index performance.

[27]Ten indexes are tracked by multiple smart beta ETFs. When evaluating smart beta index performance, we only count these indexes once.
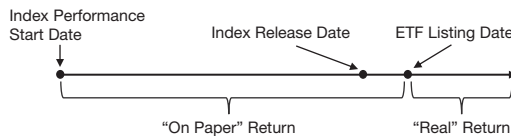
TABLE 1

Summary Statistics

Table 1 summarizes the 238 smart beta ETFs offered by 41 ETF sponsors and their underlying smart beta indexes. Panel A reports the distribution of annualized expense ratio, correlations between the monthly ETF return and the underlying index return, and the 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.

	Mean	Std. Dev.	25th	50th	75th
Panel A. Smart Beta ETF					
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					
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

FIGURE 3

An Illustration of "On Paper" and "Real" Returns

Figure 3 illustrates the timeline for "on paper" and "real" returns.

or backtested returns. The "live" or "real" returns are those after the ETF listing date and are accessible to investors. Panel B of Table 1 shows that the average smart beta index has 228 months of returns available, which includes 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 the Supplementary Material, we also use the index release date as the cutoff, and we get even sharper results.[28]

IV.   "Smart" Performance Only Exists in Backtests

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

---

[28]As Table 1 reports, more than half of the smart beta ETFs in our sample were listed within 5 months of the index release date.

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 ETF sponsors often claim that the primary goal for their ETFs 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.[29] Specifically, we use the value-weighted return of all CRSP U.S. firms as the market return, following Fama and French (2015). In fact, investors can easily invest in broad market indexes through, for example, the Vanguard Total Stock Market ETF (VTI) at a very low cost (i.e., 3 BPS per year).[30]

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 in CAPM alphas between the periods before and after ETF listings is 3.22% per year and is highly statistically significant. In untabulated exercises, we also measure performance by the difference between smart beta returns and aggregate market returns, and we find very similar results.

Note that in our main analyses, we use smart beta index returns rather than the corresponding ETF returns over the post-ETF-listing period. Using smart beta index returns has two advantages. First, ETF returns do not exist before ETF listings; thus, focusing on index returns can make the pre- and post-listing performance directly comparable. Second, because we use smart beta index returns to measure the post-ETF-listing performance, the performance decline cannot be attributed to transaction/implementation 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.

As an alternative measure 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. We note that smart beta ETFs charge higher fees than SPY. When

---

[29]We manually examine the official webpage, fact sheet, or summary prospectus of each smart beta ETF to identify what types of historical performance (index returns or Sharpe ratio) are presented. We find that while all 238 ETFs in our sample report the historical returns of their own underlying smart beta indexes, only 33 ETFs additionally report the Sharpe ratio of the underlying smart beta index. We therefore focus on the index returns/alphas instead of Sharpe ratios in the main analysis. In untabulated results, we examine the smart beta indexes' pre- and post-listing Sharpe ratios and information ratios (relative to the market index). The results show post-listing declines in both Sharpe ratios and information ratios.

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

TABLE 2

Smart Beta Performance Before and After ETF Listing

Table 2 reports the performance of smart beta indexes/ETFs before and after ETF listing. The sample period ends in Dec. 2019. 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 1-year, 2-year, and 3-year windows before and after the ETF listing date. Standard errors are clustered by factor theme categorized by Morningstar. t-statistics are reported in parentheses. *, **, and *** denote the 10%, 5%, and 1% significance levels, respectively.

Panel A. Pre- and Post-Listing Index Performance

	Before	After	Diff.
	1	2	3
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

	Before	After	After	Diff.	Diff.
		(Gross)	(Net)	2–1	3–1
	1	2	3	4	5
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)

we take these fee differences into account, smart beta ETFs *underperform* SPY even more, by 1.43% ($t = 3.69$) per year on average. Thus, smart beta ETFs deliver strictly lower returns relative to SPY after listing.

We also perform a panel regression analysis to investigate the post-ETF-listing performance decline pattern. Specifically, in a sample consisting of ETF-by-month-level observations, we regress monthly index CAPM alphas on a dummy variable that indicates the post-ETF-listing period. We control for ETF fixed effects and cluster the standard errors by time and by ETF. Table 4 shows that the CAPM alpha of smart beta indexes is significantly lower in the post-listing period. For instance, column 2 in Panel A reports that the index CAPM alpha is 0.27% per month (3.24% on an annualized basis) lower in the post-listing period than in the pre-listing period (t-stat. $= -3.09$). The magnitude of the post-listing performance decline shown in the panel regressions is similar to that reported in Tables 2 and 3.

We conduct several robustness tests. First, 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.1 in the Supplementary Material). Second, we show that the results in Table 2 are robust when we require ETFs to have a longer history

TABLE 3

Smart Beta Performance Before and After ETF Listing: Benchmark Against SPY

Table 3 performs a similar analysis as in Table 2 but uses the SPDR S&P 500 ETF (SPY) returns as the benchmark to estimate CAPM alpha. The sample period ends in Dec. 2019. 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 categorized by Morningstar. *t*-statistics are reported in parentheses. *, **, and *** denote the 10%, 5%, and 1% significance levels, respectively.

Panel A. Pre- and Post-Listing Index Performance

	Before	After	Diff.
	1	2	3
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

	Before	After	After	Diff.	Diff.
		(Gross)	(Net)	2–1	3–1
	1	2	3	4	5
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)

(at least 3 years) of nonmissing index returns before ETF listings (see Table A.2 in the Supplementary Material). Last, when we loosen our sample restriction and require the sample smart beta ETFs to have nonmissing index returns in the 3-month window before and after ETF listing, the results are robust (see Table A.3 in the Supplementary Material).

In sum, the results in this section and in Section A of the Supplementary Material show that the purported 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 that of broad market indexes.

B.   Discussion of Potential Explanations

In this section, we explore five plausible explanations for the performance deterioration of smart beta indexes: i) the time trend in factor premia, ii) strategic timing of ETF listings, iii) diminishing returns to scale caused by ETF flows, iv) a publication effect, and v) data mining in index construction. In this section, we argue that none of the first four explanations can materially explain the sharp

TABLE 4

Regression Analysis of Post-Listing Performance Decline

In Table 4, we analyze the post-listing performance decline through panel regressions. The sample consists of ETF-by-month-level observations, and the sample period ends in Dec. 2019. In Panel A, both pre- and post-listing performance are 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 and 2, the dependent variable is monthly CAPM alpha (in percent) relative to the CRSP value-weighted market index. In columns 3 and 4, the dependent variable is monthly CAPM alpha (in percent) relative to SPY ETF returns. The key independent variable (Post Listing) is a dummy variable that indicates whether an observation is in the post-listing period. ETF fixed effects are included in columns 2 and 4. Standard errors are double clustered by time and by ETF. t-statistics are reported in parentheses. *, **, and *** denote the 10%, 5%, and 1% significance levels, respectively.

	Dependent Variable			
	CAPM_ALPHA_CRSP		CAPM_ALPHA_SPY_ETF	
	1	2	3	4
Panel A. Pre- and Post-Listing Index Performance				
POST_LISTING	−0.29***	−0.27***	−0.33***	−0.32***
	(−4.11)	(−3.09)	(−3.69)	(−2.83)
ETF FE	No	Yes	No	Yes
No. of obs.	52,117	52,117	50,807	50,807
Adj. R^2	0.004	0.007	0.005	0.007
Panel B. Post-Listing ETF Performance				
POST_LISTING	−0.27***	−0.27***	−0.32***	−0.35***
	(−3.44)	(−3.00)	(−3.35)	(−3.04)
ETF FE	No	Yes	No	Yes
No. of obs.	52,117	52,117	50,807	50,807
Adj. R^2	0.002	0.002	0.003	0.004

decline in index performance after ETF listings. We provide supporting evidence of data mining in Section V.

1. Time Trend in Factor Premia

One possible explanation for the performance drop-off 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 and 3.

To examine this explanation, 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.[31] Specifically, we measure smart beta performance using both excess returns and alphas relative to the factor theme benchmark. In a similar exercise, we also use the academic asset pricing factors as the factor theme benchmark to estimate alphas.[32] Table 5 reports the results.

---

[31]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.

[32]In this exercise, we also control for market excess returns. The academic factors we use for the value/growth, momentum, quality, and risk/volatility indexes are HML, UMD, QMJ, and VOL, respectively. Here, QMJ is the quality factor from the AQR website, and VOL is the total volatility factor constructed following Ang, Hodrick, Xing, and Zhang (2006).

TABLE 5

Controlling for Factor Return Variation

Table 5 reports the performance of smart beta indexes relative to their factor theme benchmarks. Specifically, we group smart beta indexes based on the factor theme following Morningstar's classification, and we use the first-constructed index in each category or the corresponding academic asset pricing factor 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. In Panels A and B, we report index excess returns and alphas relative to the benchmark indexes, respectively. In Panel C, we report index alphas relative to the corresponding academic factors (also controlling for the aggregate market factor). Standard errors are clustered by factor theme categorized by Morningstar. t-statistics are reported in parentheses. *, **, and *** denote the 10%, 5%, and 1% significance levels, respectively.

	Before	After	Diff.
Panel A. Return in Excess of Factor Theme Benchmark			
All years before and after listing	2.74%***	−0.93%*	−3.67%***
	(6.26)	(−1.78)	(−4.17)
(−1 Year, +1 Year) around listing date	1.79%***	0.17%	−1.62%***
	(2.80)	(0.43)	(−3.42)
(−2 Year, +2 Year) around listing date	0.99%***	−0.44%	−1.42%***
	(2.75)	(−0.81)	(−3.96)
(−3 Year, +3 Year) around listing date	1.34%***	−0.34%	−1.68%***
	(3.91)	(−0.61)	(−4.08)
Panel B. Alpha Relative to Factor Theme Benchmark			
All years before and after listing	2.44%***	−1.33%**	−3.77%***
	(5.51)	(−2.32)	(−3.98)
(−1 Year, +1 Year) around listing date	1.32%**	−0.18%	−1.50%***
	(2.04)	(−0.40)	(−3.34)
(−2 Year, +2 Year) around listing date	0.57%	−0.75%	−1.32%***
	(1.28)	(−1.25)	(−3.37)
(−3 Year, +3 Year) around listing date	0.96%**	−0.62%	−1.58%***
	(2.43)	(−0.93)	(−3.43)
Panel C. Alpha Relative to the Academic Factor			
All years before and after listing	2.14%***	−0.66%***	−2.81%***
	(4.06)	(−2.94)	(−3.98)
(−1 Year, +1 Year) around listing date	0.73%**	0.09%	−0.63%
	(2.09)	(0.26)	(−1.40)
(−2 Year, +2 Year) around listing date	0.72%**	−0.15%	−0.86%*
	(2.16)	(−0.38)	(−1.68)
(−3 Year, +3 Year) around listing date	0.98%***	−0.17%	−1.15%**
	(3.06)	(−0.42)	(−2.23)

After controlling for factor return variations, 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 (i.e., the earliest-constructed index) in each factor theme category, the excess return of smart beta indexes is, on average, 2.74% per year before ETF listings, dropping to −0.93% per year post ETF listings. The performance decline is 3.67% per year and is highly statistically significant. When using alphas relative to the benchmark index or academic factors as the performance measure, we observe only slightly smaller performance deterioration (see Panels B and C of Table 5). We also perform a regression analysis for Table 5 and find consistent results (see Table A.4 in the Supplementary Material).

Because we control for the factor theme benchmark (either the earliest index or the academic factor), these results imply that before ETF listing, an average smart beta index generates significantly positive alphas even *after controlling for its targeted factor returns.* That is, regardless of the average return of a factor, a typical

smart beta index of the exact factor theme significantly outperforms the factor in backtests. In this sense, our findings are *distinct* from whether factor premia exist/persist and whether smart beta ETFs have meaningful factor exposures.[33]

Quantitatively, we also note that 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 12 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.

2.   Strategic Timing of ETF Listings

Another plausible explanation for the index deterioration we document is that smart beta ETF sponsors strategically time their tracked factors when launching the 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 would therefore not be as good as the pre-listing performance. To test this possibility, we control for factor-wide fluctuations (including the potential factor return mean-reversion) in Table 5 and still find similar results. Thus, this "factor timing" hypothesis fails to explain the post-ETF-listing performance decline of smart beta indexes.

One 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 is also unlikely to explain our results. As shown in Tables 2–5, 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.

3.   Diminishing Returns to Scale

One might also argue that investment flows into smart beta ETFs could make certain factor-based strategies overcrowded, and consequently, impair the performance of smart beta indexes due to decreasing returns to scale. Note that we also use index returns rather than ETF returns in the post-ETF-listing period. Our findings are thus not driven by transaction costs in implementing indexes, which could be a source of decreasing returns to scale. We conduct several tests to rule out this explanation.

In the first test, we split smart beta ETFs into two groups based on their average post-listing AUM. We then analyze the pre- and post-listing index performance for these two groups as in Table 2. To alleviate concerns that different factor-based

---

[33]Studies that examine the reliability of asset pricing factors/anomalies include, for example, Harvey, Liu, and Zhu (2016), McLean and Pontiff (2016), Novy-Marx (2016), Yan and Zheng (2017), and Hou, Xue, and Zhang (2020). A contemporary paper by Johansson, Sabbatucci, and Tamoni (2022) finds that investors cannot get factor exposures effectively through smart beta ETFs.

TABLE 6

Analyzing the Influence of Decreasing Returns to Scale

Table 6 analyzes the effect of decreasing returns to scale on post-listing performance decline. In Panels A and C, we divide all smart beta indexes into two groups by their ETFs' average post-listing AUM and average number of stock holdings, respectively. In Panels B and D, within each factor theme category, we divide smart beta indexes into two groups by their ETFs' average post-listing AUM and average number of stock holdings, respectively. In all 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 categorized by Morningstar. *t*-statistics are reported in parentheses. *, **, and *** denote the 10%, 5%, and 1% significance levels, respectively.

	Before	After	Diff.	Diff.-in-Diff.
	1	2	3	4
Panel A. AUM Across All ETFs				
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. AUM Within Factor Theme Category				
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)	
Panel C. # Holdings Across All ETFs				
Below median	3.23%***	0.21%	−3.02%***	−0.30%
	(9.90)	(0.62)	(−6.90)	(−0.44)
Above median	2.32%***	−0.99%***	−3.31%***	
	(6.03)	(−3.33)	(−6.40)	
Panel D. # Holdings Within Factor Theme Category				
Below median	3.32%***	0.38%	−2.93%***	−0.45%
	(9.90)	(1.17)	(−6.54)	(−0.66)
Above median	2.27%***	−1.11%***	−3.38%***	
	(6.10)	(−3.63)	(−6.72)	

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 6 reports the results. As one can see, whether we rank ETFs by their AUM across the whole sample (Panel A) or within the factor theme category (Panel B), the indexes of smaller-sized ETFs experience even sharper performance declines after ETF listings than larger ETFs, which is inconsistent with the argument of diminishing returns to scale.

In additional tests, we also use the number of stock holdings (Pástor et al. (2020)) (Panels C and D of Table 6) or the portfolio-weighted-average Amihud measure (Amihud (2002)) (Table A.5 in the Supplementary Material) to proxy for the extent to which smart beta ETFs are subject to the scale effect. Whether across all ETFs or ETFs within a given factor theme category, the tables show that the number of stock holdings and the average portfolio liquidity have no significant relationship with the magnitude of the performance decline. Thus, we interpret these findings as rejecting the explanation of decreasing returns to scale.

4.   Publication Effect

Another alternative explanation for the decline in smart beta performance is the publication effect. Specifically, prior studies document that factor performance

TABLE 7

Ruling Out the Publication Effect

Table 7 reports the smart beta index performance before and after the ETF listing date. To rule out the publication effect, we only use the index performance in the sample period when the paper on the corresponding academic factor has been published. To measure the index performance, we use CAPM alpha relative to the CRSP value-weighted market index in row 1, alpha relative to the SPY ETF returns in row 2, returns in excess of the factor theme benchmark index in row 3, alpha relative to the factor theme benchmark in row 4, and alpha relative to the corresponding academic factor in row 5. All alphas and returns in this table are annualized. Standard errors are clustered by factor theme categorized by Morningstar. t-statistics are reported in parentheses. *, **, and *** denote the 10%, 5%, and 1% significance levels, respectively.

	Before	After	Diff.
1. CAPM Alpha	2.56%***	−0.46%*	−3.02%***
	(4.63)	(−1.85)	(−4.56)
2. Alpha relative to SPY	2.85%***	−0.88%***	−3.74%***
	(4.94)	(−3.41)	(−5.22)
3. Return in excess of factor theme benchmark	2.52%***	−0.93%*	−3.45%***
	(5.91)	(−1.77)	(−4.11)
4. Alpha relative to factor theme benchmark	2.21%***	−1.33%**	−3.54%***
	(5.30)	(−2.30)	(−3.92)
5. Alpha relative to academic factor	1.98%***	−0.68%***	−2.66%***
	(3.89)	(−3.04)	(−3.89)

is significantly lower after academic publication on the factor (e.g., McLean and Pontiff (2016)). Because the smart beta ETF listing dates are usually later than the academic publication dates of the corresponding factors, the difference between pre- and post-listing performance potentially could be driven by the publication effect. In this subsection, we show that the publication effect is not driving our findings.

For this analysis, we assign academic publication dates to 215 smart beta ETFs in our sample.[34] To rule out the publication effect in our analysis, we only use the smart beta index return data from the post-publication period (i.e., after the academic paper on the corresponding factor has been published) to compare pre- and post-listing performance. Table 7 shows that the post-listing performance decline still displays an economically and statistically significant pattern. As robustness checks, we perform a regression analysis to rule out the publication effect (see Table A.6 in the Supplementary Material), and we lag the publication date by 3 years and reperform the analysis in Table 7 (see Table A.7 in the Supplementary Material). All the results are robust and suggest that the post-listing performance is not driven by the publication effect.

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

---

[34]In addition to broad-type factor themes, Morningstar also provides finer factor theme categories that allow us to identify the academic publication date for multi-factor smart beta indexes. For example, if a multi-factor smart beta index belongs to both "value" and "momentum" factor themes under the finer categories, we use the average publication date for the value and momentum factors as the publication date of this index. Since smart beta indexes under the "fundamentals" and "others" categories cannot be directly assigned to an academic factor, we manually check these smart beta indexes and assign academic factors.

V.   Data Mining and ETF Flows

In this section, we provide supporting evidence that data mining 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. Together, our findings raise concerns about the large wave of smart beta offerings.

A.   Evidence of Data Mining in Index Construction

While smart beta indexes are purported to be built under rules-based and formulaic indexation procedures, the discretion ETF sponsors have in constructing indexes is large in practice. This discretion comes from the multitude of ways factors can be defined and the flexibility in selecting stocks and choosing their weights in the index. Meanwhile, ETF sponsors often advocate backtested results (e.g., Morningstar (2017)). Presumably, strong performance in backtests can help attract investment flows, even though the outperformance only exists on paper. Therefore, ETF sponsors have incentives to overexploit data in the backtests. To demonstrate that data mining is the key driving force for the smart beta performance decline in Tables 2–5, we conduct five 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 for single-factor smart beta indexes. Intuitively, multi-factor indexes can freely choose among multiple factors, and the flexibility in combining factors leads to larger discretion in index construction. Moreover, because the category of multi-factor smart beta ETFs has the most ETF offerings, multi-factor smart beta ETFs are likely to face the most competitive product market. Therefore, their sponsors will likely desire superior 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.

We indeed find strong supporting evidence. As shown in Panel A of Table 8, before ETF listings, multi-factor smart beta indexes have the highest average CAPM alpha at 4.11% per year, and value and growth smart beta indexes have the lowest average CAPM alpha at 1.16% per year. After ETF listings, multi-factor smart beta indexes have the largest performance drop, 4.91% per year; in comparison, value and growth smart beta indexes have the smallest performance decline, 1.76% per year on average.

To provide further support, 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 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.

TABLE 8

Evidence of Data Mining in Index Construction

Table 8 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 and Growth, and Others) based on factor themes. 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 6 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 Panel D, we split the smart beta ETFs into two groups based on whether the underlying smart beta index is derived from an existing well-known index (e.g., the S&P 500 index). In Panel E, we classify ETFs into two groups based on whether other ETFs track the same smart beta index as the given ETF. 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 categorized by Morningstar. *t*-statistics are reported in parentheses. *, **, and *** denote the 10%, 5%, and 1% significance levels, respectively.

	#ETF	Before	After	Diff.	Diff.-in-Diff.
	1	2	3	4	5
Panel A. Multi-Factor Versus Single-Factor Smart Beta					
Factor Theme					
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 and Growth	64	1.16%**	−0.60%**	−1.76%***	
		(2.33)	(−2.30)	(−2.63)	
Panel B. High-ETF-Discretion Versus Low-ETF-Discretion					
High Discretion					
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 Versus Others					
B/V/S					
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)	
Panel D. Indexes Derived from Existing Indexes Versus Others					
Existing Index					
No	154	3.26%***	−0.55%*	−3.81%***	−1.68%**
		(6.41)	(−1.80)	(−6.28)	(−2.16)
Yes	84	1.88%***	−0.25%	−2.13%**	
		(2.77)	(−0.82)	(−2.56)	
Panel E. ETFs Sharing Index Versus Others					
Sharing Index					
No	213	3.04%***	−0.46%	−3.50%***	−2.72%**
		(6.08)	(−1.54)	(−5.46)	(−1.98)
Yes	25	0.48%	−0.30%	−0.79%	
		(0.33)	(−0.57)	(−0.39)	

To test this argument, we classify smart beta indexes into "high-ETF-discretion" and "low-ETF-discretion" groups. To be classified as high discretion, a smart beta index must satisfy at least one of two nonexclusive 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 6 months of the ETF listing date. The second condition implies that the index is specifically tailored for the ETF, indicating the ETF sponsor's control over the index construction. Although this classification is crude, 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 8, 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, whereas low-ETF-discretion indexes have much smaller performance deterioration, averaging 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. Because 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 test this supposition, 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 of these two groups of ETF sponsors separately.

As shown in Panel C of Table 8, 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.[35]

In the fourth test, we explore the heterogeneity in constraints around index construction. Specifically, some smart beta indexes are derived from other existing and well-known indexes; thus, they are restricted to select stocks within a pre-specified stock pool. For instance, the S&P 500 Momentum index is a smart beta index derived from the S&P 500 index that overweights the S&P 500 index constituents that exhibit persistence in their relative performance. We posit that data mining is more constrained for these smart beta indexes than for those not derived from an existing index. To test this conjecture, we split the smart beta indexes into two groups based on whether they are derived from an existing and well-known index, such as the S&P 500 index. We then examine the performance of the two groups before and after ETF listings.

As shown in Panel D of Table 8, the smart beta indexes derived from existing indexes experience a smaller performance decline than other indexes. The difference between the post-listing and pre-listing index CAPM alpha is -2.13% per year for indexes derived from existing indexes, compared to -3.81% for other indexes.

In the fifth test, we examine the cases in which multiple ETFs track the same smart beta index. We conjecture that when a smart beta index is tracked by multiple ETFs, there is less discretion for data mining in index construction. To test this conjecture, we identify 25 smart beta ETFs that share a common underlying index

---

[35]Note also 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%).

with at least one other smart beta ETF. We classify these 25 ETFs into the "shared-index" group and the rest of the smart beta ETFs into the other group.

As shown in Panel E of Table 8, the difference between the post-listing and pre-listing index CAPM alpha is -0.79% per year on average for smart beta ETFs in the shared-index group, compared to -3.50% for the other smart beta ETFs on average. This finding suggests that the discretion in data mining is largely removed when a smart beta index is tracked by multiple ETFs.

In an additional analysis, we examine whether the discretion in data mining is larger when the volatility of the underlying index is larger. We classify smart beta ETFs into higher and lower volatility groups based on the pre-ETF-listing return volatility of the underlying smart beta index. Then, we compare the degree of post-listing performance decline of these two groups. Table A.8 in the Supplementary Material shows that the difference between the post-listing and pre-listing index CAPM alpha is -1.95% per year for indexes with lower volatility, compared to -4.48% per year for indexes with higher volatility. This finding confirms that higher volatility of index returns is associated with greater discretion in data mining.

In summary, we explore the extent to which smart beta indexes are susceptible to data mining and find strong evidence that data mining in indexation accounts for the post-listing performance deterioration of smart beta indexes. Although these findings suggest that data mining could be detrimental to smart beta investors, we find that professional financial services and information providers fail to highlight the potential risk of data mining in smart beta products.[36] 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 provide evidence that the backtested performance of smart beta indexes 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 examine whether investors respond to backtested smart beta index returns, we estimate the following regression:

$$(1) \qquad \mathrm{FLOW}_i^{\mathrm{post},k} = \pi_t + \tau_f + \lambda \cdot \alpha_i^{\mathrm{pre},s} + \varepsilon_i.$$

Here, $\mathrm{FLOW}_i^{\mathrm{post},k}$ 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, respectively; and $\alpha_i^{\mathrm{pre},s}$ 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

---

[36]For example, Morningstar Direct provides a separate webpage for each smart beta index that provides an overview of the index. However, we do not find any caution against data mining on these webpages. In addition, we used "strategic beta" (or "smart beta") as the keyword to search for relevant articles/videos on Morningstar websites. We found a total of 410 (196) articles/videos, only six of which briefly mention potential data-mining issues in the smart beta products.

TABLE 9

Pre-Listing Index Performance and Post-Listing ETF Flows

Table 9 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 of ETF flows over the k-month period ($k = 6$ in columns 1–4 or $k = 12$ in columns 5–8) after ETF listing. The independent variables are the annualized benchmark-adjusted returns of the underlying index over the 1-year, 2-year, 3-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 5. ETF-listing-year fixed effects and factor theme fixed effects are included. Standard errors are double clustered by factor theme categorized by Morningstar and by ETF listing year. t-statistics are reported in parentheses. *, **, and *** denote the 10%, 5%, and 1% significance levels, respectively.

	$k = 6$				$k = 12$			
	1	2	3	4	5	6	7	8
Panel A. Index Performance Measured by CAPM Alpha								
ONE_YEAR_BEFORE_LISTING	1.13				0.69			
	(1.28)				(1.25)			
TWO_YEARS_BEFORE_LISTING		1.51*				0.80*		
		(1.80)				(1.75)		
THREE_YEARS_BEFORE_LISTING			3.02***				1.60**	
			(2.59)				(2.36)	
ENTIRE_PERIOD_BEFORE_LISTING				3.86***				1.69**
				(2.58)				(2.46)
Listing year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Factor theme FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
No. of 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								
ONE_YEAR_BEFORE_LISTING	0.71				0.74			
	(1.02)				(1.65)			
TWO_YEARS_BEFORE_LISTING		0.53				0.62		
		(0.76)				(1.31)		
THREE_YEARS_BEFORE_LISTING			3.74***				2.34***	
			(3.50)				(2.98)	
ENTIRE_PERIOD_BEFORE_LISTING				3.30***				1.63***
				(3.43)				(3.12)
Listing year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Factor theme FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
No. of 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

also consider different values of k and s. In this regression, λ is the estimated flow-to-backtest sensitivity and captures how investors respond to the on-paper returns of smart beta indexes.[37]

Panel A of Table 9 reports the results. The data show that investors respond strongly to backtested index returns, although such returns are not sustained after ETF listings. For example, a 1-standard-deviation increase in the 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 9). These results are

---

[37]In Table A.9 in the Supplementary Material, we consider another backtested performance measure. Specifically, we examine how post-listing ETF flows respond to the pre-listing index returns in excess of "similar ETFs," defined as listed ETFs under the same factor theme category of a given ETF.

consistent with the claim of Vanguard (2012) that backtests contribute to the viability of new ETFs.

An interesting pattern from Table 9 is that the flow sensitivity to backtested performance in the first 6 months after ETF listing is larger than that in the next 6 months after ETF listing. A potential explanation for this pattern is that, after the launch of the ETF, investors can observe the actual performance of the ETFs, and they tend to rely on actual performance rather than the backtested performance to choose ETFs. In Table A.10 in the Supplementary Material, we indeed find that the flow sensitivity to backtested performance becomes insignificant in the second 6 months after ETF listing. In addition, the flow sensitivity to the actual performance of ETFs is positive and significant.

The results here suggest that investors chase hypothetical backtested returns of smart beta ETFs. To further support the argument that an average investor in smart beta ETFs is not sophisticated, we analyze the predictability of ETF flows on future ETF performance in Table A.11 in the Supplementary Material. Either through portfolio sorting or 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 1-year flows significantly underperforms the bottom-flow quintile by about 30 BPS over the next month. Moreover, we show that the results are similar when controlling for past-1-year returns and ETF size. This "dumb money" effect suggests that those who invest in smart beta ETFs are likely to be unsophisticated.[38]

In sum, the analysis of flow-to-backtest sensitivity in this section sends up red flags about data mining in constructing smart beta indexes, as investors respond strongly to the 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 listings. These results suggest 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 exposure to asset pricing factors. We find the opposite in the data. That is, despite investors paying 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 illustrate this point, 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/volatility) and the independent variables are the Fama and French (1993) 3 factors, Carhart (1997) momentum factor, the QMJ factor (Asness, Frazzini, and Pedersen (2019)), and the total

---

[38]Bhattacharya, Loos, Meyer, and Hackethal (2017) and Brown, Davies, and Ringgenberg (2021b) also show that ETF flows are largely uninformed.

volatility factor (Ang et al. (2006)).[39] For comparison, we also estimate similar regressions for the first smart beta index in each of the five-factor theme categories.[40]

Graph A of Figure 4 plots the exposures of smart beta indexes on their designated factors, and Table 10 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 et al. (2022), who argue that investors cannot effectively harvest factor exposures through smart beta ETFs.

Graph B of Figure 4 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 Graphs 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 for investors either in terms of excess returns beyond the aggregate market or in terms of factor exposures.

VI.   Evidence from Other Developed Markets

For external validity, 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 U.S. market in AUM.[41] Like in the United States, we find that smart beta indexes experience a sharp performance decline in these countries, suggesting that data mining in constructing smart beta indexes is quite prevalent.

We conduct our analysis as follows: First, following the same procedure as in Section III, we collect 77 non-U.S. equity smart beta ETFs whose index returns are available before ETF listings.[42] Then, we analyze the performance of these

---

[39]Here, we do not include the smart beta themes of "dividend" or "fundamentals" because there are no clear and widely recognized asset pricing factors of "dividend" and "fundamentals." Also, we do not consider the theme of "size" but only include the SMB factor as a control. Size is not one of the broad factor themes classified by Morningstar, and size is often combined with other factor themes when constructing a smart beta index. Nonetheless, we identified eight smart beta ETFs that can be categorized into the "size" theme. We compare the first smart beta index against the later-constructed indexes under the size theme and find no significant differences in their SMB exposures and expense ratios.

[40]For robustness, we also examine the post-ETF-listing index returns and find similar results, presented in Table A.12 in the Supplementary Material.

[41]For example, the AUM of European smart beta ETFs grew from 0 in 2005 to $57.4 billion in 2018.

[42]We take the following steps to obtain the list of smart-beta ETFs that invest in equities outside the United States. First, we use the Smart Beta and Global Investment Fund Sector (GIFS) data fields to

FIGURE 4

Factor Exposures and ETF Expense Ratios

Graph A of Figure 4 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 later-constructed smart beta indexes; and the gray bar shows the difference in factor exposure between the earliest index and later-constructed indexes. For growth indexes, we report the factor exposure on the "LMH" factor, which is the negative of the Fama–French HML factor. The 95% confidence bands are shown around the bars. Graph 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 later-constructed indexes.
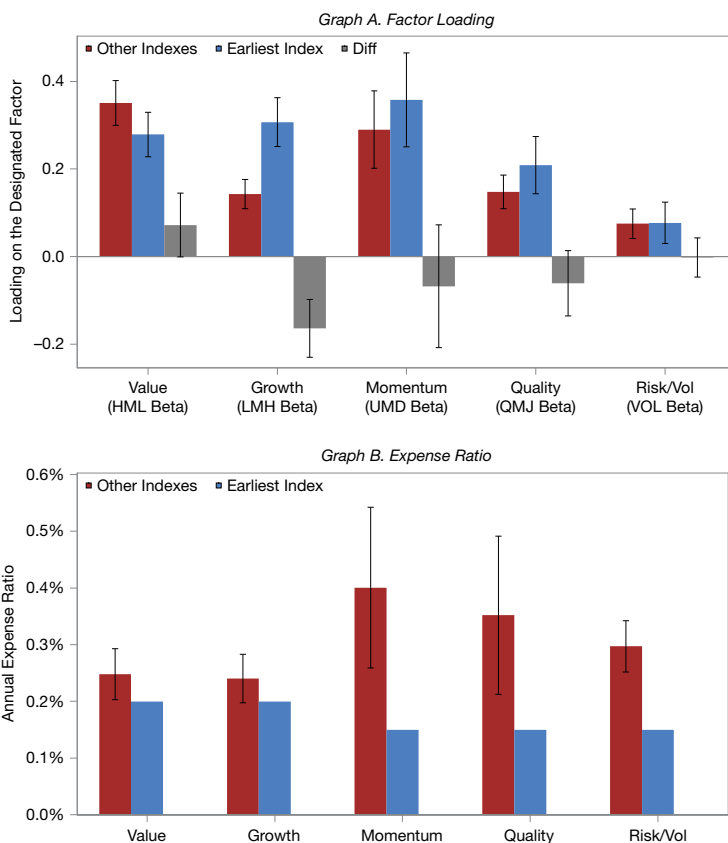
smart beta indexes relative to their corresponding aggregate market indexes before and after ETF listings as in Table 2. Here, the European smart beta indexes are benchmarked to the MSCI Europe Index. Similarly, the Canadian, Australian, and UK smart beta indexes are benchmarked to the MSCI Canada Index, the MSCI Australia Index, and the MSCI United Kingdom Index, respectively. Table 11 reports the results.

---

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 nonmissing index returns in the 6 months before and after the ETF listing date. After this procedure, our sample comprises 37 European smart beta ETFs, 29 Canadian ETFs, 6 Australian ETFs, and 5 UK ETFs.

TABLE 10

Factor Exposure and ETF Expense Ratio

Panel A of Table 10 reports the loadings of smart beta indexes on their designated factors. The sample period starts in Jan. 2001 and ends in Dec. 2019. 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 the following set of factors: 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, and UMD are from Kenneth French's data library. QMJ is from the AQR website. VOL is a total volatility factor, which represents the returns of a value-weighted portfolio that 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 a Newey–West correction of 12 lags. 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. *t*-statistics associated with the differences are reported in parentheses. *, **, and *** denote the 10%, 5%, and 1% significance levels, respectively.

Factor Theme Category	First Index	Other Indexes	Diff.
Panel A. Index Loading on the Designated Factor			
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			
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)

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. The results are consistent if we conduct a regression analysis on the performance decline (see Table A.13 in the Supplementary Material). These findings suggest that the use of data mining in constructing smart beta indexes is also common in other developed markets.

VII.   Conclusion

In recent 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 they often use academic research to endorse their smart beta products. While smart beta

TABLE 11

Smart Beta Performance Deterioration: International Evidence

Table 11 reports the sample comprises 77 smart beta ETFs that invest in the European, Canadian, Australian, or UK equity market and were listed by Dec. 2018. The sample period ends in Dec. 2019. 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 categorized by Morningstar. *t*-statistics are reported in parentheses. *, **, and *** denote the 10%, 5%, and 1% significance levels, respectively.

	Before	After	Diff.
	1	2	3
Panel A. Annualized Return in Excess of Market Index			
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			
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)

ETFs are purported to follow predetermined and formulaic rules in indexation, the discretion sponsors have in devising the rules is very large in practice. Potentially, the claimed "smartness" of smart beta ETFs is just a mirage arising from data mining. If indeed considerable data mining is used in index construction, it is unclear whether the proliferation of smart beta products adds value for investors. 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 mining in index construction, meaning that strong performance only exists in backtests and has no predictive power for "real" performance. Although these findings suggest data mining could be detrimental to smart beta investors, professional financial services and information providers often fail to highlight the potential risk of data mining in smart beta products. Our results highlight the risk of data mining in the proliferation of ETF offerings because investors respond strongly to backtests.

Supplementary Material

Supplementary Material for this article is available at https://doi.org/10.1017/S0022109023000674.

References

Agarwal, V.; B. M. Barber; S. Cheng; A. Hameed; and A. Yasuda. "Private Company Valuations by Mutual Funds." *Review of Finance*, 27 (2022), 693–738.

Agarwal, V.; P. Hanouna; R. Moussawi; and C. Stahel. "Do ETFs Increase the Commonality in Liquidity of Underlying Stocks?" Working Paper, Georgia State University (2018).

Amihud, Y. "Illiquidity and Stock Returns: Cross-Section and Time-Series Effects." *Journal of Financial Markets*, 5 (2002), 31–56.

An, Y.; M. Benetton; and Y. Song. "Index Providers: Whales Behind the Scenes of ETFs." Available at SSRN 3855836 (2021).

Ang, A.; R. J. Hodrick; Y. Xing; and X. Zhang. "The Cross-Section of Volatility and Expected Returns." *Journal of Finance*, 61 (2006), 259–299.

Asness, C. S.; A. Frazzini; and L. H. Pedersen. "Quality Minus Junk." *Review of Accounting Studies*, 24 (2019), 34–112.

Barber, B. M.; T. Odean; and L. Zheng. "Out of Sight, Out of Mind: The Effects of Expenses on Mutual Fund Flows." *Journal of Business*, 78 (2005), 2095–2120.

Ben-David, I.; F. Franzoni; B. Kim; and R. Moussawi. "Competition for Attention in the ETF Space." National Bureau of Economic Research (2021).

Ben-David, I.; F. Franzoni; and R. Moussawi. "Do ETFs Increase Volatility?" *Journal of Finance*, 73 (2018), 2471–2535.

Bhattacharya, U.; B. Loos; S. Meyer; and A. Hackethal. "Abusing ETFs." *Review of Finance*, 21 (2017), 1217–1250.

Bloomberg. "BlackRock Projects Smart Beta ETF Assets Will Reach $1 Trillion by 2020, and $2.4 Trillion by 2025" (2016). 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, N. P., and V. K. Pool. "Do Hedge Fund Managers Misreport Returns? Evidence from the Pooled Distribution." *Journal of Finance*, 64 (2009), 2257–2288.

Bollen, N. P., and V. K. Pool. "Suspicious Patterns in Hedge Fund Returns and the Risk of Fraud." *Review of Financial Studies*, 25 (2012), 2673–2702.

Box, T.; R. Davis; R. Evans; and A. Lynch. "Intraday Arbitrage Between ETFs and Their Underlying Portfolios." *Journal of Financial Economics*, 141 (2021), 1078–1095.

Brown, D. C.; S. Cederburg; and M. Towner. "Dominated ETFs." Working Paper, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3694592 (2021a).

Brown, D. C.; S. W. Davies; and M. C. Ringgenberg. "ETF Arbitrage, Non-Fundamental Demand, and Return Predictability." *Review of Finance*, 25 (2021b), 937–972.

Brown, K. C.; W. V. Harlow; and H. Zhang. "Staying the Course: The Role of Investment Style Consistency in the Performance of Mutual Funds." Working Paper, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1364737 (2009).

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

Carhart, M. M.; R. Kaniel; D. K. Musto; and A. V. Reed. "Leaning for the Tape: Evidence of Gaming Behavior in Equity Mutual Funds." *Journal of Finance*, 57 (2002), 661–693.

Chan, L. K.; H.-L. Chen; and J. Lakonishok. "On Mutual Fund Investment Styles." *Review of Financial Studies*, 15 (2002), 1407–1437.

Chen, H.; L. Cohen; and U. G. Gurun. "Don't Take Their Word for It: The Misclassification of Bond Mutual Funds." *Journal of Finance*, 76 (2021), 1699–1730.

Chordia, T.; A. Subrahmanyam; and Q. Tong. "Have Capital Market Anomalies Attenuated in the Recent Era of High Liquidity and Trading Activity?" *Journal of Accounting and Economics*, 58 (2014), 41–58.

Cici, G.; S. Gibson; and J. J. Merrick Jr. "Missing the Marks? Dispersion in Corporate Bond Valuations Across Mutual Funds." *Journal of Financial Economics*, 101 (2011), 206–226.

Da, Z., and S. Shive. "Exchange-Traded Funds and Asset Return Correlations." *European Financial Management*, 24 (2018), 136–168.

Dannhauser, C. D., and J. Pontiff. "FLOW." Working Paper, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3428702 (2020).

DeHaan, E.; Y. Song; C. Xie; and C. Zhu. "Obfuscation in Mutual Funds." *Journal of Accounting and Economics*, 72 (2021), 101429.

Edelen, R. M.; R. B. Evans; and G. B. Kadlec. "Disclosure and Agency Conflict: Evidence from Mutual Fund Commission Bundling." *Journal of Financial Economics*, 103 (2012), 308–326.

ETF Stream. "ETF Insight: The Dangers of Smart Beta Backtesting" (2019). Available at https://www.etfstream.com/features/etf-insight-the-dangers-of-smart-beta-backtesting/.

Evans, R. B. "Mutual Fund Incubation." *Journal of Finance*, 65 (2010), 1581–1611.

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

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

Financial Times. "Smart Beta's New Kids on the Index Block" (2015). Available at https://www.ft.com/content/1fb9d574-43c6-11e5-b3b2-1672f710807b.

Financial Times. "Lack of Smart Beta Benchmarks Fuels Concerns" (2017). Available at https://www.ft.com/content/d4f03f5e-a9f0-11e7-ab66-21cc87a2edde.

Financial Times. "Index Companies to Feel the Chill of Fund Managers' Price War" (2019). Available at https://www.ft.com/content/e886b2d2-e852-3071-85c1-c9a57113d8a5.

Glosten, L.; S. Nallareddy; and Y. Zou. "ETF Activity and Informational Efficiency of Underlying Securities." *Management Science*, 67 (2021), 22–47.

Harvey, C. R.; Y. Liu; and H. Zhu. "… and the Cross-Section of Expected Returns." *Review of Financial Studies*, 29 (2016), 5–68.

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

Huang, S.; M. O'Hara; and Z. Zhong. "Innovation and Informed Trading: Evidence from Industry ETFs." *Review of Financial Studies*, 34 (2021), 1280–1316.

Invesco. "Understanding Smart Beta" (2018). Available at https://www.invesco.com/us-rest/contentdetail?contentId=a8569b81e1c0e510VgnVCM100000c2f1bf0aRCRD.

Israeli, D.; C. M. Lee; and S. A. Sridharan. "Is There a Dark Side to Exchange-Traded Funds? An Information Perspective." *Review of Accounting Studies*, 22 (2017), 1048–1083.

Johansson, A.; R. Sabbatucci; and A. Tamoni. "Tradable Risk Factors for Institutional and Retail Investors." Working Paper, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3594064 (2022).

Kaniel, R., and R. Parham. "WSJ Category Kings–The Impact of Media Attention on Consumer and Mutual Fund Investment Decisions." *Journal of Financial Economics*, 123 (2017), 337–356.

Li, F., and J. West. "Live from Newport Beach. It's Smart Beta!" Research Affiliates, available at https://www.advisorperspectives.com/commentaries/2017/08/24/live-from-newport-beach-its-smart-beta (2017).

McLean, R. D., and J. Pontiff. "Does Academic Research Destroy Stock Return Predictability?" *Journal of Finance*, 71 (2016), 5–32.

Morningstar. "Be Wary of Strategic-Beta ETFs with Rosy Back-Tested Results" (2017). Available at https://www.morningstar.com/articles/805594/be-wary-of-strategic-beta-etfs-with-rosy-back-tested-results.

Morningstar. "A Global Guide to Strategic-Beta Exchange-Traded Products." Technical Report (2019).

Novy-Marx, R. "Testing Strategies Based on Multiple Signals." Working Paper, University of Rochester (2016).

Pan, K., and Y. Zeng. "ETF Arbitrage Under Liquidity Mismatch." Working Paper, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2895478 (2019).

Pástor, L.; R. F. Stambaugh; and L. A. Taylor. "Fund Tradeoffs." *Journal of Financial Economics*, 138 (2020), 614–634.

Pattabiraman, M. "Data Mining in Index Construction: Why Investors Need to be Cautious" (2020). Available at https://freefincal.com/data-mining-index-construction/.

Weinberg, A. "ETF Providers Going in New Directions with Indexes." Pensions & Investments, available at https://www.pionline.com/article/20181001/PRINT/181009995/etf-providers-going-in-new-directions-with-indexes (2018).

Suhonen, A.; M. Lennkh; and F. Perez. "Quantifying Backtest Overfitting in Alternative Beta Strategies." *Journal of Portfolio Management*, 43 (2017), 90–104.

Vanguard. "Joined at the Hip: ETF and Index Development" (2012). Available at https://www.vanguardcanada.ca/documents/joined-at-the-hip.pdf.

Wermers, R. "Matter of Style: The Causes and Consequences of Style Drift in Institutional Portfolios." Working Paper, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2024259 (2012).

Yan, X., and L. Zheng. "Fundamental Analysis and the Cross-Section of Stock Returns: A Data-Mining Approach." *Review of Financial Studies*, 30 (2017), 1382–1423.