Index Disruption: The Promise and Pitfalls of Self-Indexed ETFs*

Bige Kahraman† Sida Li‡ Anthony Limburg§

August 29, 2024

Abstract

Self-indexed ETFs, which track indices created and maintained by the ETF issuer themselves, rather than relying on external index providers, are disrupting the index provision market. By avoiding licensing fees to independent index providers, self-indexed ETFs are expected to offer competitive fees for investors. Surprisingly, we find that self-indexed ETFs often charge *higher* fees than their public-index counterparts. This disparity cannot be explained by product differentiation but rather by conflicts of interest. ETF issuers which offer both self-indexed ETFs and wealth management services are incentivised to promote their own high-cost self-indexed products to clients. Increased competition is not always a panacea for end-investors.

JEL Classification: G11, G14, G15, G23

Keywords: exchange-traded funds, ETFs, benchmarks, index providers, self indexing, indices, mutual funds, financial intermediation

*We thank the participants at the FMA Europe Conference and the University of Oxford FAME seminar for their useful comments. Anthony Limburg acknowledges support from the Clarendon Fund and the Templeton Education and Charity Trust (TECT). Sida Li acknowledges support from ACCESS for supercomputing allocations.

†Saïd Business School, University of Oxford. Email: bige.kahraman@sbs.ox.ac.uk.

‡Brandeis International Business School. Email: sidali@brandeis.edu.

§Saïd Business School, University of Oxford. Email: anthony.limburg@sbs.ox.ac.uk.

1 Introduction

Traditionally, exchange-traded funds (ETFs) relied on indices constructed by established industry leaders, such as S&P, FTSE Russell, and MSCI. Wielding significant market power, these prominent index providers often levy substantial fees on ETF issuers which are subsequently passed onto end-investors.[1]

Recently the index provision market has been disrupted by self-indexing – the practice where the ETF issuers create and manage their own proprietary indices, avoiding the need to pay licensing fees to third-party index providers. In 2013, the SEC relaxed rules on self-indexing, making it easier for ETF issuers to use their own indices as benchmarks. This regulatory shift, coupled with potential savings on licensing fees, led to a significant rise in self-indexing ETFs. While self-indexing was uncommon a decade ago, it has gained traction in recent years. For instance, within the popular investment styles of broad equity and large cap, nearly 20% of ETFs are now self-indexing. Given this trend, this paper asks whether self-indexing ETFs ultimately benefit their end-investors. Despite the significance of indices in passive investments, the index provision market remains understudied, and this paper aims to contribute to this literature by focusing on self-indexing in the ETF market.

Proponents of self-indexing celebrate the increased competition in the index provision market, arguing that self-indexing empowers issuers to offer more competitive fees due to the elimination of licensing fees paid to third-party index providers. However, there are also skeptical views. One view points out that it can be difficult to compete down fees due to the inelasticity of demand for the industry leaders' indices (An et al., 2023) and another one highlights the potential conflicts of interest inherent in self-indexing which can ultimately harm investors (Xiao and Xiong, 2024).

We begin our empirical analysis by comparing the fees of self-indexed and public-indexed ETFs. Through various specifications, we find that self-indexing ETFs charge significantly higher (not lower) net expense ratios compared to their public-indexing coun-

[1] An et al. (2023) show that S&P, FTSE Russell, and MSCI collectively control 75% of the market. When we also consider CRSP and NASDAQ, this figure becomes 95%. Moreover, the authors estimate that about one-third of all ETF fees are paid to index providers in the form of licensing fees, and that about 60% of these licensing fees are markups.

terparts. Within our sample of passive U.S. equity funds, self-indexing ETFs have fees between 12% and 14% higher than those of public-indexing funds. Notably, this result persists after controlling for a host of fund characteristics as well as style, issuer, and time fixed effects.

We offer and test two competing hypotheses to elucidate this finding. The first hypothesis, termed the "differentiated products hypothesis," centers on the notion that self-indexing provides issuers greater latitude in crafting investment strategies, and such differentiation translates into financial advantages for investors. The alternative hypothesis, the "conflicts of interest hypothesis", focuses on potential conflicts of interest arising from the dual role of some self-indexing ETF issuers. The idea is that issuers of self-indexing ETFs may also offer wealth management advisory services, and this arm of the business can present an opportunity to promote their own self-indexed products which might be challenging to achieve otherwise. Our results support the conflicts of interest hypothesis.

Our initial focus is on testing the differentiated products hypothesis. To this end, we first evaluate the differences in ETF portfolios using holdings data from ETF Global. Within each investment style, we calculate the average pairwise differences in daily holdings for self- versus public-index ETFs and find that there is no significant difference between the two groups. Furthermore, using a network analysis in which more similar funds are bunched together, we clearly visualize that most self-index ETFs are quite similar to their public-index counterparts. Second, we evaluate the potential differences in fund performances using numerous types of portfolio and panel regression tests, and we do not find any evidence that self-index ETFs perform any better.[2] Third, we consider self-index ETFs with slightly differentiated portfolios and ask whether such self-index ETFs charge higher fees and (or) perform better. Results show that this is not the case. While the public-index ETFs which show more resemblance to their peers charge lower fees (consistent with competitive pressures), self-index ETFs that are more alike charge higher. There is no difference in performance outcomes. In sum, our results do not support the differentiated product hypothesis.

Next, we examine the conflicts of interest hypothesis in three separate tests. We start

[2]While portfolio holdings are similar, fund performance (before fees) can vary due to differences in trading. In Appendix B, we entertain the possibility of strategic trading – for instance, prior to index reconstitutions. We find that self-index ETFs trade no differently than public-index ETFs.

by analyzing whether self-index ETFs issued by issuers which are likely to suffer from conflicts of interest (due to their dual roles as investment advisors) charge higher fees. To test this idea, we classify issuers into two groups: 'specialized fund managers' and 'investment advisors'. While the former group explicitly focuses on fund management (e.g., Global X, Defiance ETFs, Van Eck), the latter one offers various financial services ranging from fund management to wealth management advisory services (e.g., Charles Schwab, the Big 3, Goldman Sachs). This arguably makes them more susceptible to the conflicts of interest concern.

In our first test, we test the differences in fees between self- vs public-index funds offered by these two groups of issuers. Investment advisors, on average, charge significantly lower fees, which is consistent with these companies being larger and hence being able to offer more competitive prices for their products. However, for their self-index ETFs, they systematically charge higher fees. Amongs ETFs issued by specialized fund managers, there is no significant differences in fees between self- and public-index ETFs.[3]

In our second and third tests, we analyze the differences in investor flows and the characteristics of investors holding shares of public- versus self-index ETFs. If it is the case that high-fee self-index ETFs are primarily used to extract rents from conflicted issuers' advisory clients, this should be reflected in fund flows and ETF holdings data. Fund flows are on average expected to be lower for self-index ETFs issued by investment advisors, since those will primarily attract flows from advisory clients of the issuing company. Furthermore, such self-index ETFs should appear to have disproportionately high institutional ownership by the issuing company ('institutional self-ownership') to the extent that these assets are held in discretionary advisory accounts.[4] Our findings are consistent with these predictions.

Controlling for issuer fixed effects, on average, we do not find significant differences in flows between self-index and public-index ETFs. However, when we compare ETFs issued by investment advisors to the others, interesting findings emerge. ETFs issued by

[3]To shed light on the aforementioned result of self-index ETFs which are more alike their peers charging higher fees, we also document that self-index ETFs offered by investment advisors tend to be more similar to their peer groups, albeit without providing any significant performance benefits.

[4]Institutional investment managers in the United States that exercise investment discretion over $100 million in Section 13(f) securities are required to file Form 13F including the assets that they own or the client assets that manage. See Section 13(f)(1)(b) of the Securities Exchange Act, and 15 U.S.C. 78c(a)(35).

investment advisors on average have higher flows, but their self-indexed ETFs, which are on average more expensive compared with their peer groups, attract significantly less investor capital.

Next, for each ETF, we calculate institutional self-ownership, which is the fraction of total shares reported by its issuer in 13F filings and find a significant difference: self-index ETFs have 18% self-ownership, while public-index ETFs have only 2%. For comparison, we also calculate other institutional ownership, which is the fraction of total ETF shares reported by all institutional investors except the ETF issuer. This ownership is generally similar between the two groups of ETFs, with public-index ETFs having slightly more. Motivated by these striking statistics, we carry out formal analyses to find that the difference in institutional self-ownership is predominantly driven by ETFs issued by investment advisors. We repeat the analysis for other-institutional ownership and do not find any effect as such. Taken all together with the results on fees and flows, our findings support the conflicts of interest hypothesis.

Finally, we close with a mechanism test whereby we evaluate whether model portfolio recommendation is a channel through which issuers promote their self-index ETFs. Model portfolios are recommendation portfolios designed by asset managers, some of which play a dual role by simultaneously managing funds. Brogaard et al. (2021) show that asset manager model providers recommend their affiliated ETFs more frequently, and hence enjoy significant flows especially from clients who are not particularly sensitive to fees. Building on these findings, we ask whether ETF issuers which are also model providers are more likely to include their self-index ETFs in their recommended model portfolios, compared to their public index ETFs. We confirm that this is indeed the case.

Related Literature

While there are numerous studies on ETFs, there is not enough research focusing on issues regarding the index provision market despite its relevance. Traditionally, most ETF issuers rely on external index providers and pass on the licensing fees to end-investors. Although ETF fees are seemingly low on an annual basis, even small differences in fees can have an important impact on long-term investments. Given the increasing popularity of ETFs with many households, it is crucial to examine how the index provision market

affects investor fees. Our paper contributes to this area of research by focusing on the emerging phenomenon of self-indexing.

An et al. (2023) study for the first time the competition structure between index providers and ETF issuers, and the consequences of this structure for ETF investors. They show that index providers wield strong market power and charge remarkably high markups to ETFs that are passed on to investors. Specifically, one-third of ETF expense ratios are paid as licensing fees to index providers, and an estimated 60% of these fees are markups charged by the index providers.

In a theoretical model, Xiao and Xiong (2024) focus on the fundamental question of why index providers exist when funds can simply self-index, and what value they deliver to end-investors. Using a general equilibrium model, they argue that self-indexing leads to a conflicts of interest problem between end-investors and ETF issuers, and the presence of index providers address this problem. Although the specific conflicts of interest mechanism that they consider is different to the ones we consider in our paper, in terms of its main message, our analysis provides support for their assertion on self-indexing.

Kostovetsky and Warner (2021) examine how the brand name of an ETF's index affects its popularity with investors. For ETFs using external indices, they show that strong brand names create additional investor flows even after controlling for ETF issuer characteristics. While they recognize and control for self-indexers in their analysis, these ETFs are the primary focus of our research. We provide a comprehensive analysis of self-indexing ETFs, examining their fees, investment strategies, performance and potential conflicts of interest. We characterize the circumstances when such conflicts arise and how they related to fees.

Broadly speaking, our research also adds to the body of knowledge about ETFs which highlights their negative aspects. Focusing on high fees and significant underperformance by thematic ETFs, Ben-David et al. (2022) show that these ETFs cater to investor sentiment. In a similar vein, Huang et al. (2023) show poor performance by smart-beta funds and relate this finding to index providers' data mining practices. Akey et al. (2021) show the existence of closet active management of passive ETFs and how their performance is negatively related to activeness. Brown et al. (2024) document the surprising popularity of ETFs with returns that are highly correlated, with those of cheaper, more liq-

uid competitors. Finally, Khomyn et al. (2020) show, for a given index, ETFs with higher secondary market liquidity extract significant rents from investors who are arguably more sensitive to liquidity than to fees.

2 Data and sample construction

Our main source of data is ETF Global which covers all ETFs listed in the U.S. and Canada with no survivorship biases. ETF Global provides information on a wide range of monthly ETF characteristics (e.g., fund name, issuer, inception date, benchmark index, AUM, leverage ratio, listing exchange, sector exposures, investment region, fund focus, asset class, active management dummy, currency and sector exposure, put and call options volume, short interest, management fee, and total/net expenses) as well as daily (monthly) holdings data for opaque (nonopaque) ETFs.

ETF Global has three key benefits over the CRSP Mutual Fund database. First and most importantly, it provides information on each ETF's benchmark index (variable labeled 'primary_benchmark'). This allows us to identify self-indexers as we will explain below. Second, the database provides daily ETF holdings, based on which we assess the similarity of investment strategies among ETFs. Thirdly, ETF Global updates fund characteristics monthly, enabling more accurate analyses. One downside of ETF Global is that the data are only available from 2012 onwards. However, this does not have an important impact on our study since self-indexing is a recent phenomenon. To illustrate, at the start of our sample in January 2012, we find only 2 self-indexing ETFs.

Our sample includes all passive unlevered U.S. equity ETFs from 2012 to 2020.[5] To identify the active (passive) status of the fund, we use the active management identifier provided by ETF Global. We remove all data points with missing information on fund primary benchmark and net expense ratios. We also eliminate cases where total expenses minus fee waivers are not equal to the ETFs net expenses.

We complement our data with supplementary information from Morningstar Direct, 13F institutional filings from Thomson/Refinitiv, and the CRSP stock and Mutual Fund databases. From Morningstar Direct, we extract a smart-beta identifier and model port-

[5]It is ambiguous whether levered ETFs can be considered "passive". Also, their holdings are generally swap-based, which are not covered by our dataset.

folio recommendations. We match approximately 90% of our sample with Morningstar using tickers and ETF names. We use 13F institutional filings and data on shares outstanding from the CRSP stock database, to calculate the share of institutional ownership in our sample of ETFs. We use the CRSP Mutual Fund database to obtain daily, ETF share-split adjusted, NAV (net asset value) returns as well as to construct characteristics pertinent to ETF issuing investment management companies (e.g. age and total assets under management). Finally, we obtain data on factor returns from Kenneth R. French's website.

A self-indexed ETF is an exchange-traded fund that tracks an index created and maintained by the ETF issuers or one of its affiliated, rather than tracking an index provided by an independent third-party index provider, such as S&P Dow Jones Indices, MSCI, or FTSE Russell. A public-index ETF, on the other hand, is an ETF that licenses an external third-party index provider.

When the ETF market was still in its infancy, ETF sponsors had to obtain exemptive orders from the Securities and Exchange Commission to start offering self-index funds. This changed when new SEC orders, issued on July 10, 2013, permitted ETF sponsors and their underlying indexes to operate more flexibly than before. In particular, the new orders removed most of the requirementswhich the previous system had imposed on self-index ETFs.[6] In this way, the 2013 orders allowed the self-index ETFs to operate in essentially the same manner as ETFs based on unaffiliated, third-party indexes. Leveling the requirements for the two types of ETFs arguably made the self-indexing model more attractive, contributing to the observed trend over the last decade. The 2019 "ETF Rule" by the SEC removed "exemptive order" regulations, instead modernizing the regulation of ETFs by establishing a clear and consistent framework for the majority of ETFs operating today. The 2019 rule did not change the status quo rules and requirements for self-index ETFs.[7]

We use the variable 'primary_benchmark' from ETF Global to identify ETFs that are self- and public-indexing. Our procedure involves multiple steps. Self-index ETFs are

[6]These conditions generally included: (i) requirements to promote transparency of the underlying index methodology (e.g., previous orders required the methodology governing the indexes to be publicly available at no charge, changes to such methodology to be made upon at least 60days' public notice, and index components to be publicly disclosed on a daily basis); (ii) use of an unaffiliated third party to calculate the index; and (iii) creation of certain firewall arrangements to separate the fund's adviser and the index provider.

[7]https://www.sec.gov/news/press-release/2019-190

defined as ETFs which are using a benchmark index that is offered by the same or an affiliated company, offering the ETF. Our first step involves parsing the text reported in 'primary_benchmark', taking the last data point available for a given fund (that is, 2020's cross section snapshot for alive funds and last available one for dead funds). If the text includes one of the key words corresponding to commonly used indices such as "S&P", "FTSE", "Russell", "CRSP", "MSCI", "Morningstar", "Dow Jones", "NYSE", "NASDAQ", "Nasdaq", "WisdomTree", "StrataQuant", we mark these observations as public-index ETFs. If the text in the benchmark does not include these keywords but instead includes the name of the issuer, we then mark these as self-index ETFs. For instance, the JP Morgan US momentum ETF (ticker: JMOM) is marked as a self-index ETF since its benchmark is "JP Morgan US Momentum Factor Index" and the issuer name is "JP Morgan". Second, for the cases that we were unable to identify through text parsing, we manually collect information on funds' benchmark indices using online sources and check whether the index is provided by the same (including its subsidiaries) or an affiliated company offering the ETF.[8] Third, to capture the changes in benchmark index over time for a given fund, we look at changes in the text reported in the 'primary_benchmark' variable and evaluate if it is a genuine change or a change that is simply capturing a name change or change in name abbreviation. We then manually update the self-index (public-index) identifiers accordingly. Through this procedure, we can identify the self-index (public-index) status of more than 99% of funds for which ETF Global provides a 'primary_benchmark'. Our final sample includes 786 unique ETFs.

ETF Global also provides a 'focus' classification for ETFs which broadly classifies the investment style of a fund into one of 26 categories (e.g. 'Large Cap', 'High Dividend Yield', 'Broad Equity'), including sector-specific ETFs (e.g., 'Financials') as well as information on the issuing management company (issuer). For both variables we correct spelling errors (e.g. 'JP Morgan' and 'JPMorgan' and 'High Dividend Yielld' and 'High Dividend Yield').

[8] An example of the ETF issuer being affiliated with index provider is VanEck's ownership of Market Vectors Indexes.

2.1 Descriptive statistics

While self-indexing is a recent phenomenon in the ETF industry, it has been gaining traction over the past decade. Figure 1 shows the number of self- versus public-index ETFs during our sample period. While there were only a few self-index funds a decade ago, by the end of 2020, this has increased to 83 ETFs in total. Table 1 shows the distribution of self- and public-index ETFs across investment styles. Self-index funds appear to be most prevalent in investment styles which are also popular among public-index funds. These include broad equity, large cap, small cap, sector (including all sector-specific funds), and theme.[9] Within these investment styles, self-index funds constitute about 18-26% of all funds by the end of our sample period. Due to their young age, self-index funds are still small in size compared to public-index funds, however our results on fund flows (presented later in the paper) show that their growth rates are on average similar to those of public-index funds, conditional on similar fund characteristics.

We classify ETF issuers into three groups based on their mix of self- and public-index offerings: ones which offer only self-index funds, ones which offer only public-index funds, and ones which offer a mix of self- and public-index funds. Figure 2 shows the share of issuers in each of these groups over time. While most issuers only offered public-index ETFs a decade ago, this has been changing over time. The count of issuers which explicitly focus on self-index ETFs, and issuers offering both types of funds has been steadily increasing since 2012. By the end of our sample period, 53% of issuers in our sample are pure public-index issuers. 28% are pure self-index and 19% are issuers which offer a mix of public and self-index ETFs.

Internet Appendix Table IA.1 lists the 10 largest ETF issuers within each group.[10] Total issuer AUMs reported in this table reflect the total AUMs of ETFs issued by each of these issuers within our sample. Several large investment management companies such as Goldman Sachs, JP Morgan, State Street Global Advisors, and Fidelity stand out as issuers offering a mix of self- and public-index ETFs. For instance, within our sample of U.S equity passive ETFs, Fidelity offers 6 self- and 11 public-index ETFs; JP Morgan offers 5

[9] For expositional purposes, in this table, we group all sector-specific funds as 'sector' funds. this reduces the total number of investment styles to fifteen.

[10] For brevity, tables reported in the Internet Appendix are labeled with abbreviation "IA".

versus 4; Goldman Sachs 3 versus 2; and State Street offers 3 versus 59. Other issuers with a mix offering include Invesco (2 versus 105), Victory Capital Management (7 versus 2) and Pacer Financial with (8 versus 4), and Northern Trust 5 versus 2, and Charles Schwab (1 versus 10).

When we look at the group of issuers which thus far offer only self-index funds, we see that these issuers tend to be smaller. John Hancock, Renaissance and American Century are the largest ones within this group. The issuers which thus far offer only public index funds include some of the largest players in the market such as Blackrock, Vanguard and Wisdom Tree. This group however also includes smaller issuers such as Van Eck, DWS, Principal and Nuveen.

In Table IA.2, we consider the full sample of issuers to examine the systematic relation between issuer characteristics and issuers' propensity to offer self-index ETFs. For each issuer, we extract the total AUM of all outstanding funds (across all asset-classes), as well as the age relative to the earliest offer date of any fund for each issuer from the CRSP Mutual Fund database. When we look at the role of issuer size (total issuer AUM) and issuer age, and we don't find statistically significant effects. Figure IA.1, which shows the distribution of issuer size and issuer age within each issuer group, provides insights into these findings. Self-index ETFs are primarily issued by either large and old or small and young issuers and this drives the coefficient estimates for issuer size and age towards zero in a linear regression model. The only significant effect the regression picks up is the pre-existence of a self-index ETF. Issuers with a self-index ETF are more likely to offer another one.

Panel A of Table 2 shows the descriptive statistics on various fund characteristics for self- versus public-index funds. $Net\ Expense\ Ratio\ (\%)$ is the annual net expenses after fee waivers, as provided by ETF Global. $Log(Age)$ is the natural logarithm of fund age, where age of an ETF is defined as the difference between the month-end date and the inception date, divided by 365. $Log(AUM)$ is the natural logarithm of fund's total assets under management. If information on AUM for an ETF is missing in ETF Global, we extract it from the CRSP Mutual Fund database. We then remove all remaining observations with missing or zero AUM. All variable descriptions are provided in Appendix A. Panel B reports summary information for our holdings similarity measure which we will define

and discuss in the next section.

We examine ETFs' gross (before fees) NAV-based returns ($Gross\ Return\ (\%)$. Since ETFs subtract expense ratios from NAV on a daily basis, we first calculate daily gross returns by adding annual net expense ratio/252 (assuming 252 trading days in a year) to daily NAV net returns. Monthly gross returns are then calculated as the cumulative product of one plus each day's return over a full month, minus one. $Style\text{-}Adjusted\ Return\ (\%)$ is the fund's monthly return minus the average monthly fund return within the same investment style; $Gross\ Carhart\ Alpha\ (\%)$ is the fund's excess return estimated over a 36-month rolling window using the 4-factor Carhart model. Following the literature, we define ETF flow as:

$$FundFlow_{i,t} = \frac{TNA_{i,t} - TNA_{i,t-1} \cdot (1 + R_{i,t})}{TNA_{i,t-1}}. \tag{1}$$

Public-index funds tend to have larger $Log(AUM)$, 19.4 versus 17.7, and they also tend to be older – $Log(Age)$ is 2 versus 1.15 on average. We examine a range of performance metrics such as raw returns, alphas and style-adjusted returns and do not observe significant differences in any of these performance metrics. Moreover, self-index funds on average seem to be charging higher fees than public-index funds. In the next section, we will formally examine the systematic differences in fees and performances between the two types of ETFs in regression models where we control for various fund characteristics, investment styles and time effects.

This panel also reports summary statistics on ownership variables (available at quarterly frequency). $Institutional\ Self\ Ownership$ is the fraction of total ETF shares reported by the 13F institutional investors of the ETF issuing company and $Institutional\ Non\ Self\ Ownership$ is the fraction of ETF shares reported by all 13F institutional investors excluding the ETF issuing company. The most striking difference is in $Institutional\ Self\ Ownership$. Self-index ETFs have 18% self-ownership, while public-index ETFs had only 2% on average. $Institutional\ Non\text{-}Self\ Ownership$ however is largely similar between the two groups, with public-index ETFs having slightly more (45 versus 40%). Later in the paper, we formally examine the ownership differences between the two groups of ETFs.

3 Empirical results

3.1 Differences in annual fees

We start by analyzing the extent to which fees charged by self-indexing ETFs differ from those charged by public-indexing ETFs. We regress net expense ratios on a number of explanatory variables including: $SelfIndexer?$, which is an indicator variable that equals one if the fund is identified as a self-index ETF, $Log(AUM)$, $Log(Age)$, $Log(Holdings)$, $SmartBeta?$ (an indicator variable taken from Morningstar which equals 1 for smart beta ETFs) along with various month, style and issuer fixed effects. Specifically, column 1 includes month-style fixed effects and column 2 adds issuer fixed effects to the specification. Column 3 replaces issuer fixed effects with issuer-time fixed effects. Standard errors are clustered by month and issuer dimensions.

Table 3 reports the results. Throughout all regressions, the coefficient for self-indexing is positive and statistically significant when we include issuer and issuer-time fixed effects (columns 2 and 3, respectively). Specifically, in column 2, the coefficient estimate for the self-index dummy variable is 0.056 and it is statistically significant at the 1% level. In column 3, the estimate is 0.050 and again it is statistically significant at the 1% level. Economically these estimates imply that self-index ETFs on average charge 12% to 14% higher than public-index ETFs. These results are interesting in that they represent contrary evidence to the view that self-indexing is a strategy to become more competitive on fees.

Internet Appendix Table IA.3 reports the results of robustness test for fee regressions. One might wonder whether these results are driven by the possibility that some self-indexing ETFs are specialized ETFs which track specific subsets of assets within a category. As noted previously, our sample includes only passive ETF funds — specialized ETFs are typically classified as active ETFs by our data provider — and we control for smart-beta indicator in the regressions. Nonetheless, to address lingering concerns, we also include a uniqueness measure proposed in Kostovetsky and Warner (2021), which is defined as the 12-month average of a fund's absolute difference in gross returns versus the value-weighted average gross return of all funds within the same style category.

Moreover, we also include the potential influence of turnover ratio (minimum of sales or purchases of securities, divided by 12-month average AUM) and bid-ask spread (difference between the ask and bid as reported in CRSP, scaled by the price) in the regression. We include the turnover ratio to account for a fund's higher operational costs when index reconstitutions are more frequent. An ETF's bid-ask spread serves as a proxy for its liquidity, which was recently shown to be a key dimension of ETF competition (Khomyn et al., 2020). The main conclusion remains intact.[11]

This section demonstrates that self-index ETFs charge higher (not lower) fees compared with public-index counterparts. In the sections to follow, we examine two competing hypotheses which can explain this finding. One hypothesis is based on the idea that self-index ETFs are differentiated products delivering superior performance, and the alternative hypothesis is based on the idea of conflicts of interest stemming from some issuers playing the dual role of being an ETF issuer and wealth management financial advisor.

3.2 Differentiated product hypothesis

In section 3.1, we considered the possibility of self-index ETFs being specialized. In this section, we dive deeper into testing the hypothesis that self-indexing provides issuers greater flexibility in setting the investment strategy and whether such differentiation is associated with financial benefits to customers. Our analysis reveals that there is no significant difference in holdings between self-index and public-index ETFs, and that self-index ETFs do not provide performance benefits to customers. Interestingly, self-index ETFs that are more similar (not differentiated) charge higher fees than the others. Taken altogether, the results do not support the differentiated products hypothesis.

3.2.1 Differences in holdings

We start by introducing a cosine similarity measure between each pair of ETFs within our sample, to examine the differences in holdings. We pull the end-of-month portfolios of all ETFs in the sample from ETF Global and use all equity holdings with a valid ticker.

[11] As an alternative to Uniqueness, in the next section, we develop a holdings-based similarity measure. The main conclusion of the fee regression is robust to controlling for this variable.

For each pair of ETFs within our sample, we identify months where both ETFs have holding data (index for overlapped months is m) and calculate the cosine similarity between the two ETF's portfolios. Specifically, i is the index of stocks held by either ETF, ETF A's holding for stock i is $A_{i,m}$ dollars and ETF B's holding for stock i is $B_{i,m}$, where $A_{i,m}$ and $B_{i,m}$ equal zero if the ETF does not hold stock i. The cosine similarity is calculated as:

$$Cos_{A,B,m} = \frac{\sum_{i=1}^{N} A_{i,m} B_{i,m}}{\sqrt{\sum_{i=1}^{N} A_{i,m}^2} \cdot \sqrt{\sum_{i=1}^{N} B_{i,m}^2}} \tag{2}$$

N is the total number of stocks that either ETF A or B holds. By definition $Cos_{A,B,m} = Cos_{B,A,m}$. We take the numerical average of $Cos_{A,B,m}$ across all months m, and obtain our pairwise cosine similarity measure, $Cos_{A,B}$.[12]

Panel B in Table 2 reports the summary statistics of fund-level cosine similarity for self- vs public-index ETFs and tests the statistical significance of the difference between the two groups. To obtain a fund-level measure, for each fund, we average out all pairwise cosine similarities with all other funds within the same style. We do not find a significant difference between public vs self-index ETFs in holdings similarity (0.263 vs 0.267) in the overall sample. When we focus on popular investment styles (broad equity, large cap, small cap, thematic and sector), we observe the same finding. On average, fund-level cosine similarities are somewhat low (around 0.3), however this is the case for both self- and public-index ETFs.[13] Except for the thematic style where the similarities are overall lower for both groups, we do not find significant differences in holdings similarities between the two groups of ETFs.

We next conduct network analyses to obtain further insights into the variation across funds, specifically, by visualizing the variation in holdings similarity across funds. Based on $Cos_{A,B}$, we calculate a minimum spanning tree representing the network of all ETFs in our sample. Intuitively, this method calculates a parsimonious network which connects all funds, minimizing the necessary connections, where more similar funds become

[12]To confirm our calculations we checked the cosine similarity between well-established S&P 500 ETFs such as SPY, IVV, and VOO and find that the measure is all greater than 0.998.

[13]Relatively low similarity in portfolio holdings is consistent with the practice of synthetic index replication as well as many academic papers documenting passive funds strategically deviating from tracking indices to reduce tracking error (e.g. Brogaard et al. (2021); Dekker et al. (2021); Molk and Robertson (2023)).

bunched together, forming clusters (i.e. 'communities'). In Figure 3, we plot this network using the Fruchterman Reingold layout, which minimizes overlapping connections and spreads the network evenly on a canvas. Panel A shows the network structure in the full sample while Panels B to F focus on the most popular investment styles (listed above).

Figure 3 reveals that, although there is some variation, self-index ETFs are generally centrally located within many clusters, in that their holdings are not significantly different from those of public-index counterparts. Had their holdings been substantially different compared with public-index ETFs, they would be systematically located outside the clusters (perhaps even forming their own). We observe this pattern when analyzing the full sample of ETFs (Panel A) as well as the subsamples based on investment styles (Panels B to F). For instance, in broad equity (Panel B), we observe two main clusters and self-index funds are generally bunched with public-index funds within these clusters. In a similar vein, for large cap and small cap ETFs, the majority of self-index funds seem to be located within the main cluster along with most public-index funds.

To sum up, we do not meaningful differentiation in portfolio holdings between self- and public-index ETFs within the same investment style.

3.2.2 Differences in performance

We next move to evaluating the potential differences in fund performances to see whether the higher expense ratios of self-index ETFs can be justified with better performance outcomes. Despite the similarity in portfolio holdings (as we show in Section 3.2.1), factors such as trading strategies could potentially lead to performance differences.

We first conduct portfolio tests where we estimate and contrast the monthly before-fee alphas of self- and public-indexing ETF portfolios over our sample period. We report results for both value- and equal-weighted portfolio returns, and we use CAPM, Fama-French and Carhart factor models to estimate alphas. In total, we estimate alphas under 6 different scenarios. All returns are before fees and we require at least 20 observations in each portfolio. Standard errors are Newey-West standard errors allowing correlation up to 6 months.

Table 4 reports the results. Panels A, B, and C show CAPM, 3-factor and 4-factor alphas, respectively. Columns 1 and 2 report monthly alphas. When we look at the monthly

alphas of self-index ETF portfolios, we see that alphas are statistically indistinguishable from zero. A similar pattern is observed also for public-index ETF portfolios, except for one scenario where CAPM monthly alpha (with equal-weighting) is weakly negative. Finally, alphas of the spread portfolio – the portfolio that goes long on self-indexers and short on public-indexers – are statistically insignificant throughout all tests. Columns 3 and 4 reproduces the analysis using daily portfolio returns. Given the relatively short sample period of our study, this allows us to increase the statistical power of the tests. The results remain unchanged.

In Internet Appendix Table IA.4, we examine the potential differences between performances in a regression setting where we can control for the potential influence of fund characteristics on fund performance. The dependent variables in Panel A and B, respectively, are gross fund returns and gross risk-adjusted returns where returns are adjusted using Carhart 4-factor model with loadings to each factor estimated from 36-months rolling regressions. Standard errors are clustered by month and issuer. The main variable of interest is $SelfIndexer?$ and other explanatory variables include $Log(AUM)$, $Log(Age)$, $Log(Holdings)$, $SmartBeta?$, $Volatility$ as well as a host of month, style, and issuer fixed effects. The results are consistent with the portfolio tests. There are no significant performance differences between self- and public-index ETFs.

In Table 5, we consider the degree to which self- and public-index funds vary on volatility and Sharpe ratio. In panel A, the dependent variable is $Volatility$, which is defined as the standard deviation of daily net fund returns within a month, and in panel B, it is $Sharpe\ Ratio$, which is the monthly net fund return divided by the fund volatility. Throughout different specifications, there are no significant differences in the Sharpe Ratios. Self-index ETFs appear to have somewhat lower volatility, however estimates are economically modest (around a 2-4% difference with respect to the mean value of public-index ETFs). Taken together with Table 4, our findings highlight that self-indexing ETFs do not deliver superior performance outcomes compared to their public-index counterparts.

Finally, in this section, we examine whether those self-index ETFs which are somewhat differentiated perform better and (or) charge higher fees. Table 6 shows that this is not the case. In columns 1-2 of Table 6, respectively, we introduce a $SelfIndexer?$ x

$CosineSimilarity$ interaction term to the baseline regression on fees (Table 3) and regression on gross ETF returns. $CosineSimilarity$ is the average cosine similarity of holdings of an ETF compared to all other ETFs in the same style category in that month. There is no evidence that differentiated self-index ETFs perform any differently, but interestingly they charge differently. In column 1 where the dependent variable is fees, $CosineSimilarity$ is significantly negative, showing that public-index ETFs that are more similar to their peers charge lower fees, as one would expect. However, the coefficient estimate reverses for self-index ETFs ($SelfIndexer?$ x $CosineSimilarity$ is significantly positive). Self-index ETFs that are more similar (not differentiated) charge higher fees.[14] Overall, the evidence in this section does not support the differentiated products hypothesis.

3.3 Conflicts of interest hypothesis

Our analyses reveal that on average, self-index funds charge significantly higher fees without providing any obvious performance benefits in return. In this section, we examine the hypothesis that conflicts of interest can arise from the fact that some of the issuers of self-index ETFs also offer wealth management advisory services, and this may give them the incentives to favor their self-index ETFs at the expense of their clients. The results of this section support this hypothesis.

3.3.1 Do investment advisors' self-index ETFs charge more?

We start by analyzing whether self-index ETFs issued by issuers which are likely to suffer from conflicts of interest charge higher fees. To test this idea, we classify issuers into two groups: 'specialized fund managers' and 'investment advisors'. While the former group explicitly focuses on fund management (e.g., Global X, Defiance ETFs, Van Eck), the latter group offer a range of financial services including fund management as well as wealth management advisory services (e.g., Charles Schwab, the Big 3, Goldman Sachs) which arguably makes them more susceptible to the conflicts of interest concern. The advisory services part of the business can create the opportunity of marketing expensive self-index ETFs which might be otherwise difficult. There are 105 unique issuers in our

[14] In Section 3.3.1 we shed light on contributing factors behind this result.

sample, with 34 classified as investment advisor and 71 as specialized fund managers. A full list of these issuers is provided in Internet Appendix Table IA.5.

Column 1 of Table 7 tests the differences in fees between the two groups of issuers. We take the specification of column 1 of Table 3 and introduce the interaction term $SelfIndexer?$ x $InvAdvisor?$, where $InvAdvisor?$ is an indicator variable which equals 1 for ETFs issued by issuers classified as 'investment advisors'. The regression includes month-style fixed effects alongside all control variables. The results are striking. Investment advisors, on average, charge lower fees, which is consistent with these companies being larger and hence being able to offer more competitive prices for their products. However, their self-index ETFs systematically charge higher fees. The $SelfIndexer?$ x $InvAdvisor?$ interaction term is significantly positive. $SelfIndexer?$ alone, however, is indistinguishable from zero: amongs ETFs issued by specialized fund managers, there is no significant differences in fees between self- vs public-index ETFs.

In columns 2 and 3, respectively, we check whether self-indexed ETFs from investment advisors perform differently and have more differentiated portfolios. There is no such evidence. There is no significant difference in performance, and if anything, self-indexed ETFs by investment advisors are more similar to other ETFs in the same style. Altogether, the findings in this section support the conflicts of interest hypothesis.

3.3.2 Investor demand for self-index ETFs

In this section, we analyze the differences in investor flows and the characteristics of investors holding shares of public- versus self-index ETFs. If it is the case that expensive self-index ETFs are primarily used to extract rents from conflicted issuers' advisory clients, this should be reflected in fund flows and ETF holdings data. Fund flows can be expected to be lower for self-index ETFs issued by investment advisors since these ETFs will primarily attract flows from advisory clients of the issuing company. Furthermore, these self-index ETFs should appear to have disproportionately high institutional ownership by the issuing company ('institutional self-ownership') as investment advisors report client assets held in discretionary advisory accounts.

We start by examining fund flows. In Table 8, we regress ETF fund flows on explanatory variables including the indicator variable $SelfIndexer?$, $Log(AUM)$, $Log(Age)$, and

$AvgNetRet(Yr)$ (%) which is the average fund return over the past 12 months. Regressions include month-style and issuer fixed effects. Standard errors are clustered by month and issuer dimensions.

As expected, $AvgNetRet(Yr)$ (%) is significantly positive as investors have a tendency to chase past performance. $Log(AUM)$ and $Log(Age)$ are significantly negative, capturing the fact that smaller and younger funds tend to receive more investor capital proportional to their total net assets. These results are consistent with other studies in the literature (e.g., Evans (2010); Pastor et al. (2015)). Our main variable of interest, $SelfIndexer?$, is statistically indistinguishable from zero. Next, we remove issuer fixed effects and include the interaction term, $SelfIndexer?$ x $InvAdvisor?$. ETFs issued by investment advisors on average have higher flows ($InvAdvisor?$ is weakly positive) but their self-indexed ETFs which tend to be more expensive (in Table 7) attract significantly less investor capital. The $SelfIndexer?$ x $InvAdvisor?$ interaction is negative and statistically significant at the 5% level. This result shows that, in the aggregate, investors penalize the higher fee self-index ETFs offered by investment advisor issuers.

We next look at the characteristics of investors holding shares of public- versus self-index ETFs. Due to data limitations, we do not observe all investor holdings, but we can observe ETF holdings reported in 13F filings. As discussed previously in section 2, we construct two ownership variables from 13F filings: $Institutional\ Self\ Ownership$ and $Institutional\ Non\text{-}Self\ Ownership$.

Under the conflicts of interest hypothesis, investment advisors favor their self-index ETFs in the advisory part of their businesses. Because these ETFs are ultimately purchased by they clients, they would appear in the 13F filings of the investment advisor to the extent that these assets are held in discretionary advisory accounts. If this is the case, we would expect to find $Institutional\ Self\text{-}Ownership$ to be mainly high for self-index ETFs issued by investment advisors. Panel A of Table 9 confirms that this is indeed the case. In column 1, the Self-index estimate is 16%, revealing a significant difference in $Institutional\ Self\text{-}Ownership$ on average. In column 2, we include the $SelfIndexer?$ x $InvAdvisor?$ interaction and find that this difference largely stems from investment advisors. In column 2, the self-index dummy is positively significant but economically small (around 4%), however the estimate for $SelfIndexer?$ x $InvAdvisor?$ stands at 17% which

is both statistically significant and economically important. The investment advisor estimate, –0.8%, is also statistically significant, but it is small. Overall, these results confirm our conjecture.

In Panel B, we run the same regression with $Institutional\ Non\text{-}Self\ Ownership$ – this serves as useful placebo test to check whether there are other factors that generally influence institutional ownership in a similar way. Neither the $SelfIndexer?$ nor $SelfIndexer?$ x $InvAdvisor?$ interaction estimates are significant. This provides reassurance for our interpretations of findings.

To sum up, in section 3.3, we document supporting evidence for the conflicts of interest hypothesis. Broad investment advisors which issue self-index ETFs are prone to conflicts of interest since they additionally offer wealth management advisory services to clients. The higher fees levied by self-indexing ETFs stem mainly from this type of ETF issuers. There is no fee differential between public- vs self-index ETFs issued by the remaining issuers. Moreover, self-index ETFs attract less investor flows and have disproportionately high institutional ownership by the issuing company, arguably in part driven by client assets held in discretionary advisory accounts. These findings are consistent with these products primarily being used to extract rents from conflicted issuers' advisory clients.[15]

3.3.3 Model portfolio recommendations and self-index ETFs

In this section, we aim to evaluate the idea that issuers promote their self-index ETFs using model portfolio recommendations. Model portfolios are recommendation portfolios designed by asset managers and strategists for financial advisors, and some asset management companies play a dual role by simultaneously managing funds and providing investment model recommendations. Brogaard et al. (2021) show that asset manager model providers recommend their affiliated ETFs more frequently, and hence enjoy significant flows especially from clients who are not particularly sensitive to fees. Building on these findings, we are interested in examining whether ETF issuers which are model

[15]The main focus of our analysis is to study the potential conflicts of interest arising from the fact that some ETF issuers also offer wealth management advisory services. For self-index ETFs, industry commentators and regulators have long expressed their concerns of conflicts of interest that can potentially arise from other factors. We do not find any of these arguments particularly strong, nevertheless, provide a detailed discussion in Appendix C for the interested readers.

providers are more likely to include their self-index ETFs in their recommended model portfolios compared with their public-index ETFs. Brogaard et al. (2021) reports that more than half (54%) of advised assets are in model portfolios.[16] Hence the data on model portfolios serve as a useful proxy for client portfolios which are otherwise confidential.

For each global and US-focused model portfolio in the Morningstar Direct database, we extract quarterly portfolio holdings between 2012 and 2020. 133 unique model providers report 1125 model portfolios with at least one ETF. Within our sample, 322 unique ETFs out of 785 are at least part of one model portfolio, and 29 out of 105 unique issuers are identified as model portfolio providers. Out of these 29 ETF issuers, 20 are investment advisors. This means that a much higher percentage of investment advisors provide model portfolios. While only 12% of specialized fund managers offer model portfolios (9 out of 71), this figure is 58% for investment advisors (20 out of 34).

We classify an ETF as affiliated to a model provider if the ETF issuer is affiliated with the model provider (as in Brogaard et al. (2021)). For example, an iShares S&P 500 ETF in the "BlackRock 50/50 Long-Horizon Allc" model portfolio counts as an affiliated ETF. Out of 29 model providing ETF issuers within our sample, 17 include their ETFs in their model portfolios.[17] Out of these 17 issuers, 11 are classified as investment advisors. In sum, 32% of investment advisors within our sample (11 out of 34) offer model portfolios which include at least one of their own or affiliated ETFs.

For each ETF quarter, we count the number of affiliated model portfolios an ETF is part of. If we cannot match an ETF to any model portfolio, then the ETF is taken to have zero affiliated model portfolios. This is the main dependent variable for the regressions in Table 10. Explanatory variables include $Log(AUM)$, $Log(Age)$, $Net\ Expense\ Ratio\ (\%)$, $AvgNetRet(Yr)\ (\%)$ and a smart-beta flag. Standard errors are double clustered (by quarter and issuer).

Issuers that are model providers tend to recommend their self-indexed ETFs more often than their public-index ETFs. (Note that for issuers that are not model providers the dependent variable equals zero for both self- and public-index ETFs.) This effect is eco-

[16] https://www.broadridge.com/_assets/pdf/broadridge-fa-model-portfolio-may-2019.pdf

[17] The 17 issuers which include their affiliated ETFs in their model portfolios are Vanguard, First Trust, SSgA, WisdomTree, Goldman Sachs, Northern Trust, John Hancock, BlackRock, Charles Schwab, Fidelity, Franklin Templeton Investments, Inspire Investing, Nuveen, Global X, Invesco, Columbia, and JPMorgan.

nomically large, given that the average count of affiliated model portfolio membership is around 0.6 (the estimates for $SelfIndexer?$ are 0.782 and 1.132 in columns 2-3). In the last column, we introduce $SelfIndexer?$ x $InvAdvisor?$ to examine the behavior of investment advisors. $InvAdvisor?$ is significantly positive – ETFs issued by investment advisors have more affiliation to model portfolios, consistent with most model providers being classified as investment advisors. Moreover, $SelfIndexer?$ x $InvAdvisor?$ is significantly positive, that is, self-index ETFs issued by investment advisors have more affiliation to model portfolios, compared to other self-index ETFs. To sum up, this section provides supporting evidence that model portfolio recommendations appear to be one way that these advisors promote self-index ETFs to investors.

4 Conclusion

Traditionally, passive exchange traded funds relied on indices that are licensed by third party providers such as S&P, MSCI and Russell. The market for index providers is a concentrated market where the five largest providers serve approximately 95 percent of the market. Index providers use their market power to charge relatively high fees to ETF issuers for using and tracking their brand-name indices (An et al., 2023). These costs are then passed on to end investors purchasing the shares of these ETFs. With the 2013 SEC orders allowing ETFs that create their own indices in-house (self-index ETFs) to operate essentially in the same way as ETFs relying on external index providers (public-index ETFs), there has been a gradual shift in the marketplace. An increasing number of ETF issuers are creating proprietary indices in-house to avoid paying fees to third party index providers. In this paper, we ask whether self-indexing ETFs offer more competitive prices and find that self-index funds offer higher, not lower, fees to their customers. To explain this, we suggest two hypotheses, one based on product differentiation and the other one based on conflicts of interest. Our results support the latter.

References

Akey, Pat, Adriana Robertson, and Mikhail Simutin, "Closet active management of passive funds," *Rotman School of Management Working Paper*, 2021, (3874582).

An, Yu, Matteo Benetton, and Yang Song, "Index providers: Whales behind the scenes of ETFs," *Journal of Financial Economics*, 2023, *149* (3), 407–433.

Ben-David, Itzhak, Francesco Franzoni, Byungwook Kim, and Rabih Moussawi, "Competition for Attention in the ETF Space," *The Review of Financial Studies*, 08 2022, *36* (3), 987–1042.

Brogaard, Jonathan, Nataliya Gerasimova, and Ying Liu, "Advising the Advisors: Evidence from ETFs," *Available at SSRN 3972562*, 2021.

Brown, David C, Scott Cederburg, and Mitch Towner, "Dominated ETFs," *Available at SSRN 3694592*, 2024.

Chinco, Alex and Marco Sammon, "The passive ownership share is double what you think it is," *Journal of Financial Economics*, 2024, *157*, 103860.

Dekker, Lennart, Jasmin Gider, and Frank De Jong, "How do Funds Deviate from Benchmarks? Evidence from MSCI's Inclusion of Chinese A-shares," *Available at SSRN 3937986*, 2021.

Evans, Richard B, "Mutual fund incubation," *The Journal of Finance*, 2010, *65* (4), 1581–1611.

Huang, Shiyang, Yang Song, and Hong Xiang, "The Smart Beta Mirage," *Journal of Financial and Quantitative Analysis*, 2023, p. 1–32.

Khomyn, Marta, Tālis J Putniņš, and Marius Zoican, "The value of ETF liquidity," *European Finance Association*, 2020.

Kostovetsky, Leonard and Jerold B Warner, "The market for fund benchmarks: Evidence from ETFs," *Available at SSRN 3804002*, 2021.

Li, Sida, "Should passive investors actively manage their trades?," *Available at SSRN 3967799*, 2021.

Molk, Peter and Adriana Robertson, "Discretionary Investing by 'Passive'S&P 500 Funds," *Yale Journal on Regulation, Forthcoming*, 2023.

Pastor, Lubos, Robert F. Stambaugh, and Lucian A. Taylor, "Scale and skill in active management," *Journal of Financial Economics*, 2015, *116* (1), 23–45.

Xiao, Yizhou and Yan Xiong, "Designing Index Provision," *Available at SSRN 4815898*, 2024.

Figures

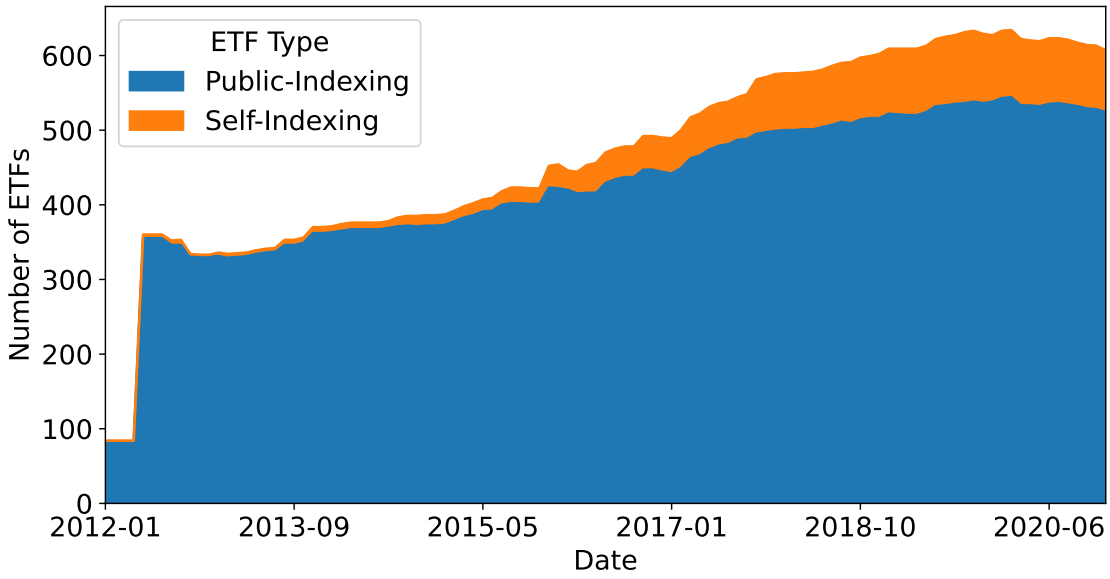


Figure 1. Number of self- and public-index ETFs over time.
This figure presents the monthly number of self-index and public-index ETFs in our sample from January 2012 until December 2020. Self-index ETFs are ETFs which track an index created and maintained by the ETF issuer themselves or one of its affiliated companies. Public-index ETFs are ETFs which track indices from unaffiliated index providers such as S&P Dow Jones or FTSE Russell.

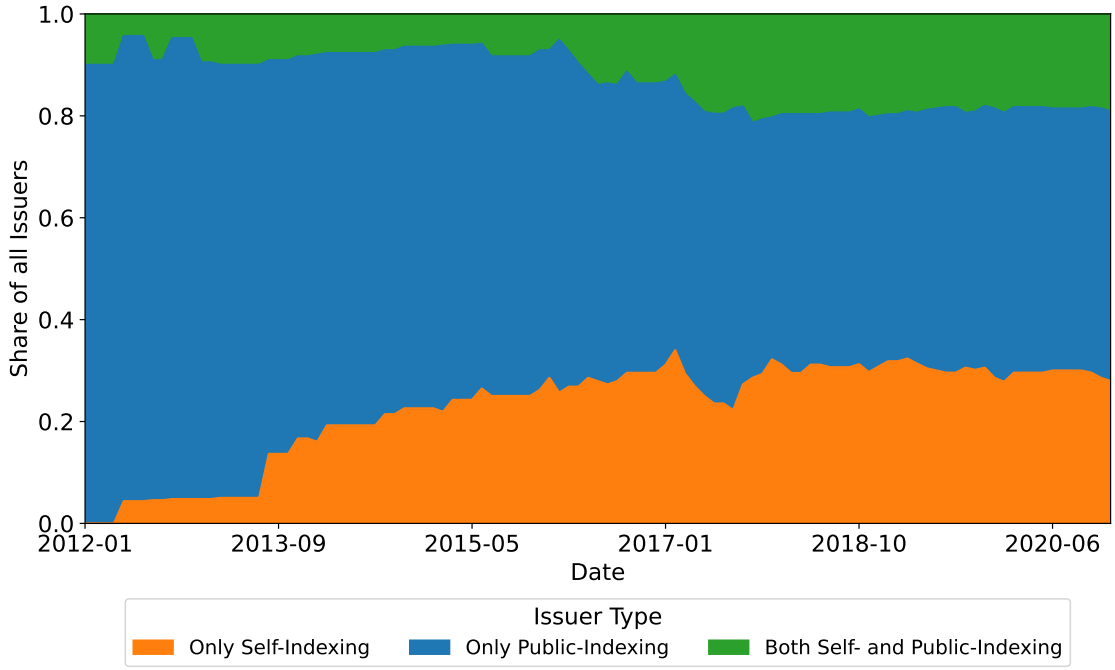


Figure 2. Share of issuer types over time.
This figure presents the monthly share of issuers in our sample from January 2012 until December 2020 which provide only self-index, only public-index, or both types of ETFs.

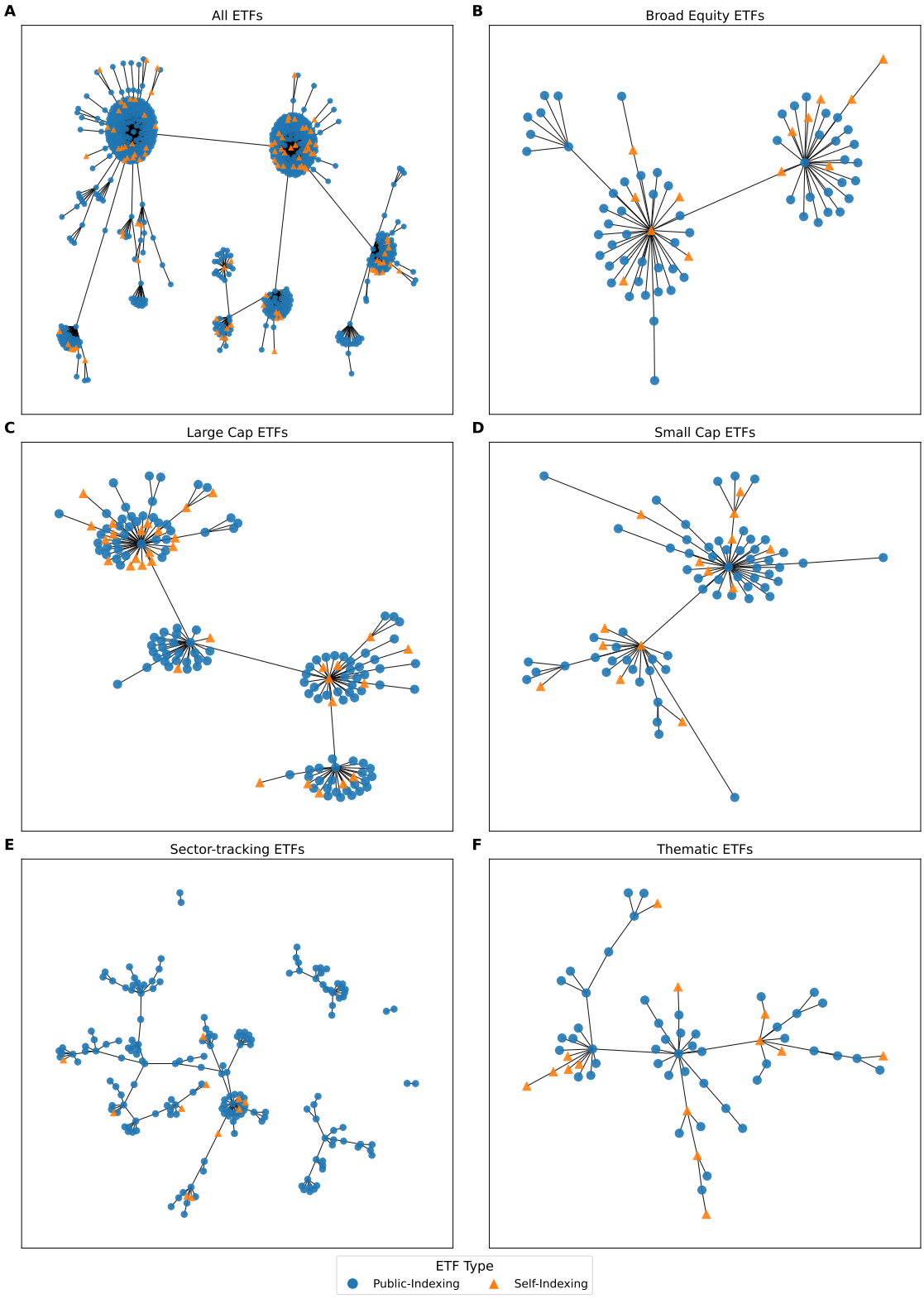


Figure 3. ETF networks based on cosine similarity of stock holdings.
This figure presents minimum spanning tree network plots of ETFs based on their monthly average pairwise cosine similarity of stock holdings. The networks are plotted using the Fruchterman Reingold Layout. Panel A presents the network of all ETFs whereas panels B to F plot networks within ETF subgroups.

Tables

Table 1. Number of ETFs by style and indexing group.
Self-Index ETFs are ETFs which track indices provided by the ETF issuer or an affiliated company. Public-Index ETFs track indices provided by independent third parties. Styles are obtained from the focus grouping within ETF Global. For simplicity, in this table we bunch all sector-related styles into one group.

Style	Public-Indexer (N)	Public-Indexer (%)	Self-Indexer (N)	Self-Indexer (%)	Total (N)
Alpha-Seeking	21	84.0	4	16.0	25
Broad Equity	80	82.5	17	17.5	97
Buywrite	2	100.0	0	0.0	2
High Dividend Yield	31	81.6	7	18.4	38
Large Cap	168	80.8	40	19.2	208
Long/Short	12	85.7	2	14.3	14
Micro Cap	4	100.0	0	0.0	4
Mid Cap	45	91.8	4	8.2	49
Preferred Stock	9	100.0	0	0.0	9
Real Estate	1	50.0	1	50.0	2
Sector	198	92.1	17	7.9	215
Size and Style	1	100.0	0	0.0	1
Small Cap	76	81.7	17	18.3	93
Target Outcome	1	100.0	0	0.0	1
Theme	43	74.1	15	25.9	58

Table 2. Summary statistics
In Panel A we report summary statistics per ETF type. $NetExpenseRatio(\%)$ is the annual net expense ratio. $GrossReturn(\%)$ is the net (NAV) return adjusted for the net expense ratio. $Style\text{-}AdjustedReturn(\%)$ is the $GrossReturn$ minus the style-average gross return in the same time period. $Volatility$ is the standard deviation of net (NAV) returns in a month. $GrossCarhartAlpha(\%)$ is the excess return over a Carhart four-factor model prediction with a 3-year rolling window. $Log(AUM)$ is the natural logarithm of an ETFs assets under management. $Log(Age)$ is the natural logarithm of an ETFs age as measured in days since inception divided by 365. $Log(Holdings)$ is the natural logarithm of an ETFs number of unique portfolio holdings. $FundFlow$ measures the net investment inflow in proportion to the ETFs previous AUM, adjusted for the ETFs returns. $SmartBeta?$ is a dummy variable that is one if an ETF is classified as smart beta. $Avg.ReturnCorrtoStyle\text{-}Peers$ is an ETFs average return correlation compared to other ETFs in the same style category. $Inst.Non\text{-}SelfOwnership$ is 13f ownership minus $Inst.SelfOwnership$, where $Inst.SelfOwnership$ is 13f ownership by the same or related investment management companies. $\#AffiliatedModels$ is the number of model portfolios by related investment management companies which report holding the ETF. Ownership variables and model portfolios are available at quarterly frequency only. Panel B reports the average cosine similarity of ETFs holdings compared to other ETFs in the same style-group, grouped by ETF type. We report the statistics for the full sample and the five largest style groups. For both panels we report the F-statistic for a comparison of means between Self-Index and Public-Index ETFs.

	Public-Indexing			Self-Indexing			
Variable	N	Mean	Std. Dev.	N	Mean	Std. Dev.	Test
Panel A: Summary Statistics of Key Variables							
Net Expense Ratio (%)	45704	0.417	0.533	4730	0.45	0.295	F=17.683***
Gross Return (%)	45546	1.05	5.42	4656	1.03	5.67	F=0.076
Style-Adjusted Return (%)	45546	0.00171	1.98	4656	0.0151	2.55	F=0.182
Volatility	45538	0.168	0.129	4652	0.17	0.143	F=0.427
Gross Carhart Alpha (%)	25815	-0.165	2.65	1441	-0.206	2.47	F=0.334
Log(AUM)	45308	19.4	2.4	4649	17.7	1.73	F=2195.774***
Log(Age)	45699	2	0.743	4730	1.15	0.555	F=5915.198***
Log(Holdings)	44857	4.91	1.39	4566	4.78	1.52	F=37.786***
Fund Flow	44873	0.0195	0.118	4544	0.039	0.136	F=109.714***
Smart Beta?	39965	0.546	0.498	4304	0.719	0.45	F=477.512***
Avg. Return Corr to Style-Peers	44200	0.704	0.216	4430	0.725	0.238	F=37.158***
Inst. Non-Self Ownership	15523	0.454	0.271	1596	0.403	0.345	F=48.621***
Inst. Self Ownership	15523	0.0177	0.0923	1596	0.183	0.311	F=2361.255***
# Affiliated Models	15537	0.735	4.04	1617	0.534	3.67	F=3.688*
Panel B: Average Cosine Similarity of Holdings Compared to Style-Peers							
Full Sample	566	0.263	0.146	91	0.267	0.152	F=0.055
Large Cap	122	0.319	0.136	31	0.325	0.134	F=0.043
Mid Cap	42	0.266	0.106	3	0.313	0.022	F=0.555
Small Cap	60	0.256	0.11	13	0.235	0.0874	F=0.409
Broad Equity	63	0.25	0.139	9	0.237	0.109	F=0.078
Theme	42	0.202	0.127	14	0.126	0.0947	F=4.215**

Table 3. Explaining net expense ratios.
The dependent variable is an ETFs annual Net Expense Ratio in percentage terms. $SelfIndexer?$ is a dummy variable that equals one if an ETF is self-indexed. $Log(AUM)$ is the natural logarithm of an ETFs assets under management. $Log(Holdings)$ is the natural logarithm of an ETFs number of unique portfolio holdings. $Log(Age)$ is the natural logarithm of an ETFs age as measured in days since inception divided by 365. $SmartBeta?$ is a dummy variable that is one if an ETF is classified as smart beta. Standard errors are two-way clustered by month-issuer interaction and reported in parentheses.

	Net Expense Ratio (%)		
	(1)	(2)	(3)
$SelfIndexer?$	0.010	0.056***	0.050***
	(0.007)	(0.008)	(0.009)
$Log(AUM)$	−0.032***	−0.023***	−0.025***
	(0.001)	(0.001)	(0.001)
$Log(Holdings)$	−0.068***	−0.012***	−0.017***
	(0.004)	(0.001)	(0.001)
$Log(Age)$	0.029***	0.070***	0.075***
	(0.006)	(0.004)	(0.004)
$SmartBeta?$	0.065***	0.001	0.001
	(0.007)	(0.010)	(0.010)
Num. Obs.	43 214	43 214	43 214
R^2	0.15	0.47	0.50
FE: Issuer		X	
FE: Month-Style	X	X	X
FE: Month-Issuer			X

* p $<$ 0.1, ** p $<$ 0.05, *** p $<$ 0.01

Table 4. Alphas from portfolio tests with monthly and daily returns.
Self-Indexers represents the portfolio of all self-indexed ETFs in a period. Public-Indexers represents the portfolio of all public-indexed ETFs in a period. Gross returns are calculated by backing out net expense ratios from net (NAV) returns. We report alphas from predictions of gross returns in excess of the risk-free rate. To predict excess gross returns we use the CAPM (Panel A), the Fama-French three-factor model (Panel B) and the Carhart four-factor model (Panel C). Only observations with at least 20 ETFs per portfolio are included. We use Newey-West Standard Errors with six lags and report them in parentheses.

	Monthly Returns (%)		Daily Returns (%)	
	Gross, VW	Gross, EW	Gross, VW	Gross, EW
Panel A: CAPM				
Self-Indexers	-0.060	-0.166	0.001	-0.008
	(0.070)	(0.108)	(0.005)	(0.006)
Public-Indexers	0.014	-0.226*	0.001	-0.008
	(0.040)	(0.135)	(0.003)	(0.007)
Self-Indexers (-) Public-Indexers	-0.074	0.061	-0.004	0.000
	(0.051)	(0.054)	(0.006)	(0.003)
Panel B: FF3				
Self-Indexers	0.011	-0.031	0.001	-0.002
	(0.065)	(0.062)	(0.004)	(0.003)
Public-Indexers	0.048	-0.036	0.001	-0.001
	(0.034)	(0.056)	(0.003)	(0.003)
Self-Indexers (-) Public-Indexers	-0.037	0.005	-0.004	-0.001
	(0.052)	(0.042)	(0.005)	(0.002)
Panel C: FF3 + Mom				
Self-Indexers	0.012	-0.031	0.001	-0.003
	(0.065)	(0.062)	(0.004)	(0.003)
Public-Indexers	0.048	-0.036	0.001	-0.002
	(0.034)	(0.055)	(0.002)	(0.003)
Self-Indexers (-) Public-Indexers	-0.037	0.005	-0.005	-0.001
	(0.053)	(0.042)	(0.005)	(0.002)

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 5. Explaining other ETF characteristics.
In Panel A we predict ETF volatility, defined as the standard deviation of daily net (NAV) returns within a month. In Panel B we predict Sharpe Ratios, defined as the monthly net (NAV) returns divided by the volatility. $SelfIndexer?$ is a dummy variable that equals one if an ETF is self-indexed. $Log(AUM)$ is the natural logarithm of an ETFs assets under management. $Log(Holdings)$ is the natural logarithm of an ETFs number of unique portfolio holdings. $Log(Age)$ is the natural logarithm of an ETFs age as measured in days since inception divided by 365. $SmartBeta?$ is a dummy variable that is one if an ETF is classified as smart beta. Standard errors are two-way clustered by month-issuer interaction and reported in parentheses.

	(1)	(2)	(3)
Panel A: Volatility			
$SelfIndexer?$	-0.008***	-0.006***	-0.007***
	(0.001)	(0.001)	(0.001)
$Log(AUM)$	-0.004***	-0.003***	-0.003***
	(0.000)	(0.000)	(0.000)
$Log(Holdings)$	-0.003***	-0.002***	-0.003***
	(0.000)	(0.000)	(0.000)
$Log(Age)$	0.011***	0.007***	0.006***
	(0.001)	(0.001)	(0.001)
$SmartBeta?$	-0.005***	-0.006***	-0.006***
	(0.001)	(0.001)	(0.001)
Num. Obs.	43074	43074	43074
R^2	0.91	0.92	0.93
Panel B: Sharpe Ratio			
$SelfIndexer?$	-4.834	-7.214*	-9.253
	(3.888)	(4.032)	(5.664)
$Log(AUM)$	-0.393	-0.451	-0.584
	(0.480)	(0.480)	(0.510)
$Log(Holdings)$	1.965	2.174	3.022
	(1.636)	(2.037)	(2.718)
$Log(Age)$	0.451	0.050	0.398
	(0.951)	(1.140)	(1.131)
$SmartBeta?$	2.240	1.156	1.404
	(2.019)	(1.521)	(1.702)
Num. Obs.	43073	43073	43073
R^2	0.13	0.13	0.18
FE: Issuer		X	
FE: Month-Style	X	X	X
FE: Month-Issuer			X

* p $<$ 0.1, ** p $<$ 0.05, *** p $<$ 0.01

Table 6. Exploring cross-sectionality.
In column (1) we predict annual Net Expense Ratios in percentage terms and in column (2) we predict gross returns, defined as net (NAV) returns adjusted for net expense ratios. $SelfIndexer?$ is a dummy variable that equals one if an ETF is self-indexed. $CosineSimilarity$ is the average cosine similarity of holdings of an ETF compared to all other ETFs in the same style category. $Log(AUM)$ is the natural logarithm of an ETFs assets under management. $Log(Holdings)$ is the natural logarithm of an ETFs number of unique portfolio holdings. $Log(Age)$ is the natural logarithm of an ETFs age as measured in days since inception divided by 365. $SmartBeta?$ is a dummy variable that is one if an ETF is classified as smart beta. $Volatility$ is the standard deviation of daily net (NAV) returns within a month. Standard errors are two-way clustered by month-issuer interaction and reported in parentheses.

	Net Expense Ratio (%)	Gross Return (%)
	(1)	(2)
$SelfIndexer?$	0.008	0.013
	(0.010)	(0.165)
$CosineSimilarity$	−0.413***	0.414
	(0.036)	(0.358)
$Log(AUM)$	−0.025***	0.019
	(0.001)	(0.012)
$Log(Holdings)$	−0.016***	−0.009
	(0.002)	(0.024)
$Log(Age)$	0.070***	0.001
	(0.005)	(0.039)
$SmartBeta?$	−0.058***	−0.061*
	(0.012)	(0.034)
$SelfIndexer?$:$CosineSimilarity$	0.264***	−0.092
	(0.031)	(0.678)
$Volatility$		1.432
		(0.928)
Num. Obs.	29 902	29 787
R^2	0.51	0.90
FE: Issuer	X	X
FE: Month-Style	X	X

* p < 0.1, ** p < 0.05, *** p < 0.01

Table 7. Exploring the role of investment advisors.
In column (1) we predict annual Net Expense Ratios in percentage terms. In column (2) we predict Gross Returns, defined as net (NAV) returns adjusted for net expense ratios. In column (3) we predict the average cosine similarity of holdings of an ETF compared to all other ETFs in the same style category. $SelfIndexer?$ is a dummy variable that equals one if an ETF is self-indexed. $InvAdvisor?$ is a dummy variable which is one if the ETF issuer is classified as investment advisor as opposed to a fund specialist. $Log(AUM)$ is the natural logarithm of an ETFs assets under management. $Log(Holdings)$ is the natural logarithm of an ETFs number of unique portfolio holdings. $Log(Age)$ is the natural logarithm of an ETFs age as measured in days since inception divided by 365. $SmartBeta?$ is a dummy variable that is one if an ETF is classified as smart beta. $Volatility$ is the standard deviation of daily net (NAV) returns within a month. Standard errors are two-way clustered by month-issuer interaction and reported in parentheses.

	Net Expense Ratio (%)	Gross Return (%)	Cosine Similarity
	(1)	(2)	(3)
$SelfIndexer?$	0.001	0.138	−0.019***
	(0.013)	(0.086)	(0.002)
$InvAdvisor?$	−0.312***	−0.010	0.012***
	(0.011)	(0.041)	(0.002)
$Log(AUM)$	−0.025***	0.018*	0.004***
	(0.001)	(0.009)	(0.000)
$Log(Holdings)$	−0.043***	0.001	0.032***
	(0.003)	(0.014)	(0.001)
$Log(Age)$	0.059***	−0.018	−0.008***
	(0.004)	(0.025)	(0.001)
$SmartBeta?$	0.054***	−0.051**	−0.013***
	(0.008)	(0.025)	(0.002)
$SelfIndexer?$:$InvAdvisor?$	0.059***	−0.092	0.032***
	(0.015)	(0.097)	(0.003)
$Volatility$		1.150	
		(0.736)	
Num. Obs.	43 214	43 074	29 902
R^2	0.20	0.88	0.70
FE: Month-Style	X	X	X

* p < 0.1, ** p < 0.05, *** p < 0.01

Table 8. Explaining fund flows.
The dependent variable is next-period Fund Flow, defined as the net investment inflow in proportion to the ETFs previous AUM, adjusted for returns. $SelfIndexer?$ is a dummy variable that equals one if an ETF is self-indexed. $Log(AUM)$ is the natural logarithm of an ETFs assets under management. $Log(Holdings)$ is the natural logarithm of an ETFs number of unique portfolio holdings. $Log(Age)$ is the natural logarithm of an ETFs age as measured in days since inception divided by 365. $SmartBeta?$ is a dummy variable that is one if an ETF is classified as smart beta. $NetExp.Ratio(\%)$ is the annual net expense ratio in percentage terms. $AvgNetRet(Yr)(\%)$ is the 12-month rolling average net (NAV) ETF return. $InvAdvisor?$ is a dummy variable which is one if the ETF issuer is classified as investment advisor as opposed to a fund specialist. Standard errors are two-way clustered by month-issuer interaction and reported in parentheses.

	Fund Flows (t+1)		
	(1)	(2)	(3)
$SelfIndexer?$	0.002	−0.002	0.011**
	(0.005)	(0.004)	(0.005)
$Log(AUM)$	−0.001	−0.003***	−0.002***
	(0.001)	(0.001)	(0.000)
$Log(Age)$	−0.030***	−0.018***	−0.017***
	(0.002)	(0.002)	(0.002)
$SmartBeta?$	0.001	0.002	0.000
	(0.002)	(0.002)	(0.002)
$NetExp.Ratio(\%)$		−0.001	−0.001
		(0.001)	(0.001)
$AvgNetRet(Yr)(\%)$		0.017***	0.017***
		(0.001)	(0.001)
$InvAdvisor?$			0.005**
			(0.002)
$SelfIndexer?$:$InvAdvisor?$			−0.019***
			(0.006)
Num. Obs.	43 096	36 145	36 145
R^2	0.16	0.18	0.17
FE: Issuer	X	X	
FE: Month-Style	X	X	X

* p < 0.1, ** p < 0.05, *** p < 0.01

Table 9. Explaining institutional ownership. In Panel A we predict Institutional Self Ownership, defined as 13f ownership by the ETF issuer or related investment management companies. In Panel B we predict Institutional Non-Self Ownership, defined as total 13f ownership minus Institutional Self Ownership. $SelfIndexer?$ is a dummy variable that equals one if an ETF is self-indexed. $Log(AUM)$ is the natural logarithm of an ETFs assets under management. $Log(Holdings)$ is the natural logarithm of an ETFs number of unique portfolio holdings. $Log(Age)$ is the natural logarithm of an ETFs age as measured in days since inception divided by 365. $SmartBeta?$ is a dummy variable that is one if an ETF is classified as smart beta. $AvgNetRet(Yr)(\%)$ is the 12-month rolling average net (NAV) ETF return. $InvAdvisor?$ is a dummy variable which is one if the ETF issuer is classified as investment advisor as opposed to a fund specialist. Standard errors are two-way clustered by quarter-issuer interaction and reported in parentheses.

	(1)	(2)
Panel A: Institutional Self Ownership		
$SelfIndexer?$	0.160***	0.043***
	(0.021)	(0.015)
$Log(AUM)$	0.007***	0.007***
	(0.001)	(0.001)
$Log(Holdings)$	0.013***	0.010***
	(0.003)	(0.002)
$Log(Age)$	-0.041***	-0.039***
	(0.005)	(0.005)
$SmartBeta?$	0.035***	0.028***
	(0.005)	(0.004)
$AvgNetRet(Yr)(\%)$	-0.002	-0.002
	(0.002)	(0.002)
$InvAdvisor?$		-0.008**
		(0.004)
$SelfIndexer?$:$InvAdvisor?$		0.170***
		(0.030)
R^2	0.22	0.25
Panel B: Institutional Non-Self Ownership		
$SelfIndexer?$	0.008	-0.015
	(0.022)	(0.027)
$Log(AUM)$	0.025***	0.025***
	(0.003)	(0.003)
$Log(Holdings)$	-0.004	-0.003
	(0.005)	(0.004)
$Log(Age)$	-0.016	-0.013
	(0.010)	(0.010)
$SmartBeta?$	0.002	0.000
	(0.009)	(0.009)
$AvgNetRet(Yr)(\%)$	-0.006	-0.006
	(0.005)	(0.005)
$InvAdvisor?$		-0.019*
		(0.010)
$SelfIndexer?$:$InvAdvisor?$		0.035
		(0.038)
R^2	0.14	0.14
Num. Obs.	12307	12307
FE: Quarter-Style	X	X

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 10. Explaining affiliated model portfolio membership. The dependent variable is the number of affiliated model portfolios an ETF is a part of. $SelfIndexer?$ is a dummy variable that equals one if an ETF is self-indexed. $NetExp.Ratio(\%)$ is the annual net expense ratio. $Log(AUM)$ is the natural logarithm of an ETFs assets under management. $Log(Age)$ is the natural logarithm of an ETFs age as measured in days since inception divided by 365. $SmartBeta?$ is a dummy variable that is one if an ETF is classified as smart beta. $AvgNetRet(Yr)(\%)$ is the 12-month rolling average net (NAV) ETF return. $InvAdvisor?$ is a dummy variable which is one if the ETF issuer is classified as investment advisor as opposed to a fund specialist. Standard errors are two-way clustered by quarter-issuer interaction and reported in parentheses.

	# Affiliated Model Portfolios			
	(1)	(2)	(3)	(4)
$SelfIndexer?$	-0.915^{***}	0.782^{*}	1.132^{**}	-1.268^{***}
	(0.148)	(0.455)	(0.514)	(0.165)
$NetExp.Ratio(\%)$	-0.015	-0.021	-0.029	0.106^{***}
	(0.031)	(0.042)	(0.041)	(0.031)
$Log(AUM)$	0.544^{***}	0.737^{***}	0.768^{***}	0.532^{***}
	(0.064)	(0.091)	(0.094)	(0.063)
$Log(Age)$	-0.916^{***}	-2.035^{***}	-2.154^{***}	-0.983^{***}
	(0.175)	(0.314)	(0.329)	(0.186)
$SmartBeta?$	0.393^{***}	-0.441^{***}	-0.425^{***}	0.403^{***}
	(0.079)	(0.089)	(0.090)	(0.080)
$AvgNetRet(Yr)(\%)$	-0.080	-0.085	-0.046	-0.078
	(0.082)	(0.080)	(0.085)	(0.081)
$InvAdvisor?$				0.496^{***}
				(0.129)
$SelfIndexer?$:$InvAdvisor?$				0.408^{*}
				(0.230)
Num. Obs.	12 385	12 385	12 385	12 385
R^2	0.11	0.16	0.19	0.11
FE: Issuer		X		
FE: Quarter-Style	X	X	X	X
FE: Quarter-Issuer			X	

* p $<$ 0.1, ** p $<$ 0.05, *** p $<$ 0.01

Appendix A: Variable Definitions

Table A.1. Variable Definitions

Variable	Definition
$SelfIndexer?$	Dummy variable that equals one if an ETF is self-indexed. Self-indexed ETFs are ETFs which track an index created and maintained by the ETF issuer themselves. ETF indices obtained from 'primary_benchmark' column in ETF Global.
$NetExp.Ratio(\%)$	Yearly net expense ratio after fee waivers from ETF Global.
$Log(AUM)$	Natural logarithm of an ETF's assets under management from ETF Global. If this information is missing in ETF Global, we use data from the CRSP MF database.
$Log(Age)$	Natural logarithm of fund age, where age of an ETF is defined as the difference between the month-end date and the inception date, divided by 365. The inception date is from ETF Global.
$Log(Holdings)$	Natural logarithm of the unique number of held assets by an ETF, obtained from ETF Global.
$FundFlow$	Increase in total net assets of an ETF not attributable to asset performance, as share of previous period's total net assets (see equation 1).
$Volatility$	Standard deviation of daily net (NAV) ETF returns within a calendar month. Daily NAV returns are obtained from the CRSP daily stock file.
$GrossReturn(\%)$	We add $NetExp.Ratio(\%)/252$ to daily net (NAV) returns. Monthly returns are then calculated as the cumulative product of one plus each day's return over a full month, minus one. Daily NAVs are from the CRSP daily stock file.
$Style\text{-}AdjustedReturn(\%)$	$GrossReturn(\%)$ minus the style-average $GrossReturn(\%)$ in the same month. Style categories are obtained from the 'focus' columns in ETF Global.
$AvgNetRet(Yr)(\%)$	12-month rolling average net (NAV) ETF return, calculated using NAVs from the CRSP MF database.
$GrossCarhartAlpha(\%)$	ETF's excess $GrossReturn(\%)$ estimated over a 36-month rolling window using the 4-factor Carhart model.
$SharpeRatio$	Monthly net (NAV) ETF return divided by the $Volatility$, using NAVs from the CRSP MF database.
$Avg.RetCorrtoStyle\text{-}Peers$	Average of all pairwise daily net (NAV) return correlations of a given ETF with all other style-peers in each month. Daily NAV returns are obtained from the CRSP daily stock file and style refers to the column 'focus' in ETF Global.
$CosineSimilarity$	Monthly average cosine similarity of ETF holdings compared to all other ETFs in the same style. Data is obtained from ETF Global where style refers to column 'focus'.
$Inst.Non\text{-}SelfOwnership$	Fraction of ETF shares held by all 13F institutional investors excluding the ETF issuing company. Quarterly holdings are from Thomson/Refinitiv's S34 file.
$Inst.SelfOwnership$	Fraction of ETF shares held only by the ETF issuing company. Quarterly holdings are from Thomson/Refinitiv's S34 file.
$SmartBeta?$	Dummy variable that is one if an ETF is classified as smart beta. The smart beta classification is taken from Morningstar Direct.
$Uniqueness$	12-month average of an ETFs absolute difference in $GrossReturn(\%)$ versus the value-weighted average $GrossReturn(\%)$ of all ETFs within the same style category. Style categories are obtained from the 'focus' columns in ETF Global.
$TurnoverRatio$	Minimum of aggregate sales or purchases of securities by the ETF, divided by the 12-month average TNA. Obtained from the CRSP MF database.
$MavgSpread$	Monthly average difference between the daily ask and bid, scaled by the price. All variables obtained from the CRSP daily stock file.
$InvestmentAdvisor?$	Dummy variable which is one if the ETF issuer is manually classified as investment advisor as opposed to a fund specialist.
$IssuerLog(AUM)$	Natural logarithm of all assets under management of an issuer across all asset classes in the CRSP MF database.
$IssuerLog(Age)$	Natural logarithm of the number of days divided by 365 since the first observation of an issuer in the CRSP MF database.
$IssuerNrETFs$	Number of passive equity ETFs operated by an issuer in the CRSP MF database.
$IssuerNrSelfIndexETFs$	Number of passive self-index ETFs operated by an issuer. Self-index ETFs are defined using data from ETF Global (see above).

Appendix B: Index Rebalancing Trades

We investigate whether self-indexing provides greater flexibility in trade execution and enables fund companies to save on trading costs. Rebalancing trades are crowded and hence costly for public-index ETFs tracking flagship indices (Chinco and Sammon, 2024; Li, 2021) since most ETFs trade on the rebalancing date of the index reconstitution, and since such events are pre-announced, they are susceptible to front running by arbitrageurs.

On the one hand, self-index ETFs can control the rebalance date of the index reconstitutions and reduce the costs associated with trading due to index rebalancing. On the other hand, they may simply follow public-index ETFs if they are largely copying their portfolio strategies.

We proxy for ETF rebalancing trades due to index reconstitution events by focusing on trades that are either brand new purchases or sell-offs.[18] To diagnose whether the ETFs are gradually executing these trades over multiple days, we construct $TradeSpeed_{i,j,T}$.

$$TradeSpeed_{i,j,T} = \begin{cases} \frac{TradeSize_{i,j,T}}{\frac{1}{5}\sum_{i=1}^{5} HoldingSize_{i,j,T+i}} & \text{if } TradeDirection_{i,j,T} = 1 \\ \frac{TradeSize_{i,j,T}}{\frac{1}{5}\sum_{i=1}^{5} HoldingSize_{i,j,T-i}} & \text{if } TradeDirection_{i,j,T} = -1, \end{cases} \tag{3}$$

where i is the index for ETF, j is the index for the underlying stock, and T is the trade day. $TradeSize$ is defined as the size of a rebalancing trade in bps of the stock's market capitalisation. Thus, if a buy (sell) trade happens gradually, the 5-day average holding size after (before) the trade would be greater than the trade size of only one day. On the other hand, if the trade is concentrated within day T, the average holding size were expected to be the same with $TradeSize$, and we will have $\mathbb{E}[TradeSpeed] = 1$. We present the key rebalancing trade variables in Table B.1. The difference in mean $TradeSize$ (2.63 versus 8.65) reflects the fact that public-index ETFs are larger than self-index ETFs. $TradeSpeed$ equals ≈ 1 for self- and public-index ETFs, indicating that both ETF types are concentrating their trades on day T – no gradual trading is observed.

[18] Stock price fluctuations and fund flows can cause the ETF holdings change day-to-day, but in these cases the ETF does not conduct trades of new purchases or selloffs. First, due to float market-cap weighting, when a stock's price fluctuates, its weight in the index will fluctuate proportionally, and there is no need for the fund tracking the index to trade. Second, fund flows proportionally inflate or deflate the portfolio size while not changing the composition of the portfolio, and associated share creations and redemptions are generally conducted through the authorized participants of the ETFs.

Table B.1. Exploring ETF rebalancing behaviour. This table reports key rebalancing trading metrics for ETFs in our sample. We report statistics separately for the full sample of ETFs (Panel A), self-index ETFs only (Panel B) and public-index ETFs only (Panel C). $TradeSize$ is defined as the size of a rebalancing trade in bps of the stock's market capitalisation, $TradeDirection$ is a variable which equals to 1 if the ETF opens a new position and -1 if the ETF closes the position. $TradeSpeed$ is defined as in equation 3.

	Mean	1%	25%	50%	75%	99%	St. Dev.	N
Panel A: All ETFs								
TradeSize	7.4	0	0.17	1.32	6.17	80.14	15.52	126096
TradeDirection	0.25	-1	-1	1	1	1	0.97	126096
Trade Speed	1	0.5	0.98	1	1.03	1.28	0.14	126096
Panel B: Self-Index ETFs								
TradeSize	2.63	0	0.08	0.5	2.33	35.85	6.97	20329
TradeDirection	0.21	-1	-1	1	1	1	0.98	20329
TradeSpeed	0.99	0.42	0.97	1	1.02	1.4	0.17	20329
Panel C: Public-Index ETFs								
TradeSize	8.65	0	0.22	1.72	7.93	83.04	16.8	101190
TradeDirection	0.25	-1	-1	1	1	1	0.97	101190
TradeSpeed	1.01	0.53	0.98	1	1.03	1.28	0.13	101190

In Figure B.1, we examine the timing of self-indexers' rebalancing trades due to index reconstitution events. For each self-indexer's rebalancing trade on day T, we count the number of public-indexing ETFs trading the same ticker in the same direction in $[T-5, T+5]$.[19] Self-indexers neither front-run nor lag the trading of public-indexers. When a self-indexer is trading a stock, on average there are slightly more than 0.4 public-indexing ETFs trading the same stock in the same direction. Almost no other ETFs are trading the same asset either before or after a self-indexer rebalances.

[19] Extending the time window does not alter the interpretations of our findings.

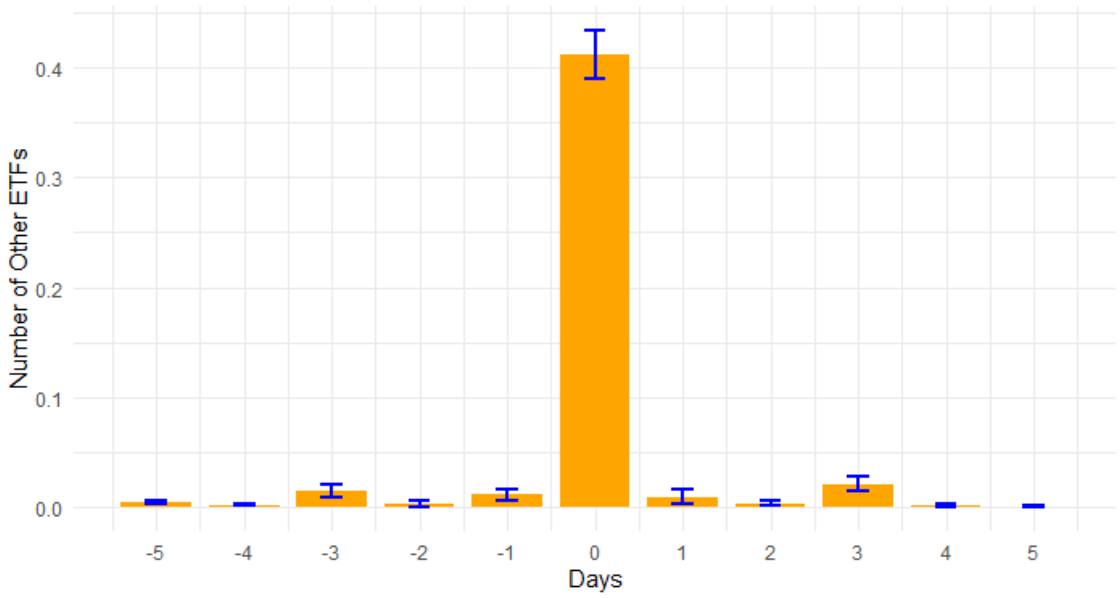


Figure B.1. Rebalancing timing of self-index ETFs.
This figure presents number of public-index ETFs which trade the same stock around the day that a self-index ETF trades ($T = 0$). The error bars represent 95% confidence intervals.

Appendix C: Other forms of conflicts of interest

Industry commentators and regulators have long raised concerns about other ways in which conflicts of interest can arise with self-indexing.[20] One commonly expressed concern is that self-index ETFs may manipulate the constituent pricing to inflate their NAVs. We did not focus on this issue, because while this might be a concern for ETFs investing in illiquid securities, it should not be an important concern for our sample ETFs whose holdings are largely liquid U.S. equities.

Another concern is that in-house production of an index can potentially incentivize funds offered by the same management company front run the company's own ETF tracking the index, due to information leaks within the company regarding the rebalancing dates. We find this argument unfounded. First, it is unlikely to realize an important gain from front-running self-index ETFs. Due to their size, the price impact of their trades is arguably small. Our findings on the trading behavior of self-index ETFs supports this notion. In Appendix B we find no evidence of front running. Second, even when self-index ETFs may grow larger, the price impact of their trades can still remain modest as these funds have the flexibility to control their rebalance dates unlike public-index ETFs. Rebalancing trades are crowded and costly for public-index ETFs (Chinco and Sammon, 2024; Li, 2021) and rebalancing dates of public-index reconstitutions are pre-announced, which makes them more susceptible to front running by arbitrageurs. Nevertheless, in Table C.1, we test whether active funds issued by fund families which also offer self-index ETFs have superior performance. If there were significant degree of front running exploited by active mutual funds, this would help them improve their performance. We find that this is not the case. The coefficient for $IssuerHasSelfIndexer?$, which is a dummy variable for whether an active fund's fund family also offers self-index ETFs in the same month, is statistically not different from zero in all specifications.[21]

[20] https://www.etfstream.com/articles/self-indexing-etf-issuers-must-answer-conflict-of-interest-question, https://www.morningstar.co.uk/uk/news/97786/self-indexing-conflict-in-the-etf-industry.aspx

[21] Since mutual funds holdings data are available at quarterly frequency, it is difficult to design a more detailed test around the rebalancing dates.

Table C.1. Performance regressions for active funds.
In Panel A the dependent variable is monthly net (NAV) return and in Panel B the dependent variable is the Net Carhart Alpha using factor loadings from 36 month rolling windows. $IssuerHasSelfIndexer?$ is a dummy variable that equals one if an issuer offers a self-index ETF. $Log(AUM)$ is the natural logarithm of an funds assets under management. $TurnRatio$ is a funds yearly portfolio turnover ratio. $Log(Age)$ is the natural logarithm of an funds age as measured in days since inception divided by 365. $Volatility$ is the standard deviation of daily net (NAV) returns within a month. $IssuerLog(AUM)$ is the natural logarithm of an issuers assets under management. The sample includes all active equity funds in the CRSP MF database from 2012-2020. Standard errors two-way clustered by month-issuer interaction.

	(1)	(2)	(3)
Panel A: Net Returns (%)			
$IssuerHasSelfIndexer?$	-0.032	-0.024	0.001
	(0.023)	(0.017)	(0.021)
$Log(AUM)$	0.016***	0.014***	0.012***
	(0.002)	(0.001)	(0.001)
$TurnRatio$	-0.004*	-0.004**	0.000
	(0.002)	(0.002)	(0.002)
$Log(Age)$	-0.044***	-0.034***	-0.032***
	(0.007)	(0.005)	(0.005)
$Volatility$	3.646***	2.710***	2.559***
	(0.292)	(0.261)	(0.274)
$IssuerLog(AUM)$	0.015***	0.014***	-0.006
	(0.002)	(0.001)	(0.008)
Num. Obs.	482047	482047	482047
R^2	0.73	0.87	0.87
Panel B: Net Carhart Alpha (%)			
$IssuerHasSelfIndexer?$	-0.023	-0.008	0.001
	(0.019)	(0.016)	(0.024)
$Log(AUM)$	0.011***	0.010***	0.006***
	(0.002)	(0.001)	(0.002)
$TurnRatio$	-0.003	-0.003	-0.009***
	(0.003)	(0.003)	(0.004)
$Log(Age)$	-0.020**	-0.024***	-0.033***
	(0.010)	(0.007)	(0.007)
$Volatility$	-0.765***	-0.532**	-0.457*
	(0.240)	(0.270)	(0.253)
$IssuerLog(AUM)$	0.004*	0.005***	-0.042***
	(0.002)	(0.002)	(0.012)
Num. Obs.	277258	277258	277258
R^2	0.12	0.57	0.57
FE: Month	X		
FE: Issuer			X
FE: Style	X		
FE: Month-Style		X	X

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Internet Appendix to

Index Disruption: The Promise and Pitfalls of Self-Indexed ETFs

(Not for Publication)

Figures

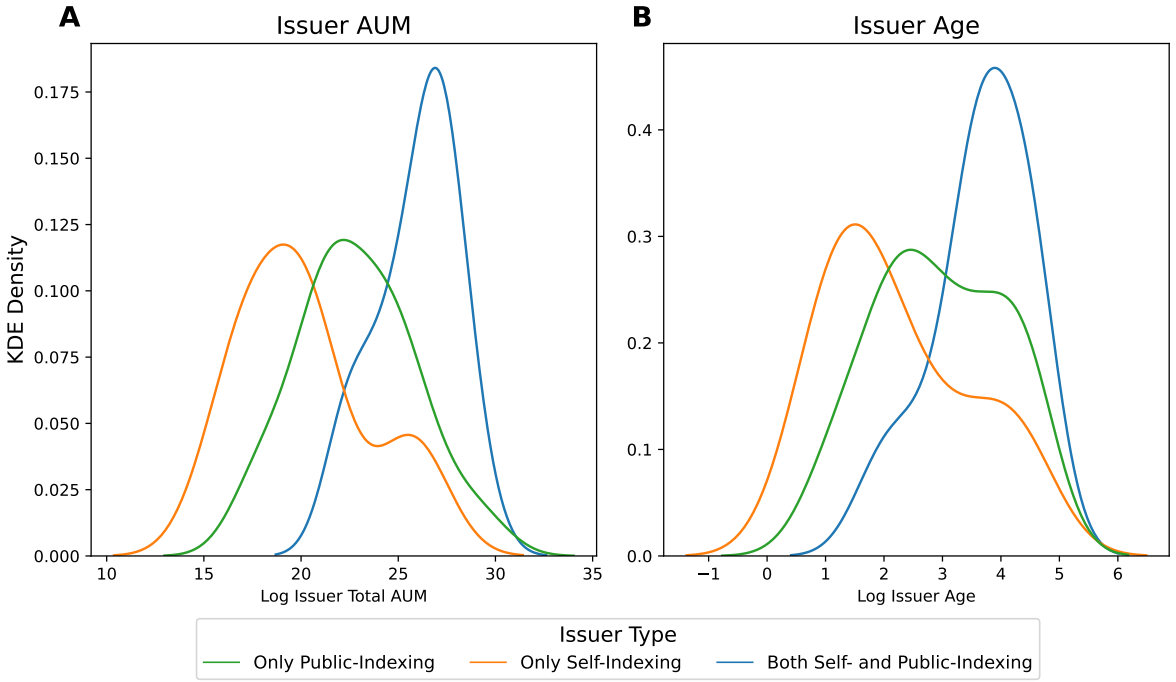


Figure IA.1. Density plots of issuer characteristics by issuer type.
This figure presents KDE density plots for key issuer characteristics, grouped by issuer type, as of December 2020. Panel A plots the densities for the logarithm of total issuer AUM, calculated as the total AUM of all outstanding funds by an issuer (across all asset-classes) from the CRSP MF database. Panel B plots the densities for the logarithm of issuer age, defined as the difference in days between December 31st 2020 and the first observation of the issuer in the CRSP MF database, divided by 365.

Tables

Table IA.1. Largest issuers of ETFs by AUM as of December 2020.
AUM and the number of unique ETFs are from our sample of non-active equity ETFs in ETF Global. Mixing issuers are issuers which offer both self- and public-index ETFs.

	Self-Indexing		Public-Indexing	
Issuer	AUM ($Mn.)	Unique ETFs	AUM ($Mn.)	Unique ETFs
Panel A: Only Self-Indexing Issuers				
John Hancock	854.76	11	-	-
Renaissance	734.30	1	-	-
American Century	384.50	2	-	-
Motley Fool Asset Management	367.76	1	-	-
Tortoise	360.65	1	-	-
Hartford	269.96	2	-	-
Inspire Investing	237.59	2	-	-
Janus Henderson	236.10	2	-	-
Alpha Architect	218.68	2	-	-
Distillate Capital Partners	201.09	1	-	-
Panel B: Mixing Issuers				
Goldman Sachs	11400.19	3	621.38	2
Pacer Financial	2722.28	8	871.15	4
Victory Capital Management	2365.95	7	448.53	2
Northern Trust	2080.11	5	1452.45	2
Charles Schwab	1320.58	1	107147.66	10
Fidelity	980.20	6	17625.00	11
SSgA	976.39	3	633741.90	59
JPMorgan	737.89	5	1280.29	4
ProShares	278.79	2	9032.34	13
Invesco	185.71	2	228106.61	105
Panel C: Only Public-Indexing Issuers				
Blackrock	-	-	961544.19	93
Vanguard	-	-	866575.55	43
First Trust	-	-	55674.09	54
WisdomTree	-	-	16595.59	14
Global X	-	-	6132.46	15
Alps	-	-	5936.66	6
Van Eck	-	-	4300.81	4
DWS	-	-	3444.60	4
Principal	-	-	2858.50	5
Nuveen	-	-	1904.51	6

Table IA.2. Explaining the issuance of self-index ETFs at the issuer-month level.
The dependent variable is a dummy which equals one if the issuer increases the amount of outstanding self-index ETFs. In columns (1) - (3) we use an OLS estimation and in columns (4) - (6) we use probit estimation. $IssuerLog(Age)$ is the natural logarithm of the number of days divided by 365 since the first observation of an issuer in the CRSP MF database. $IssuerLog(AUM)$ is the natural logarithm of all assets under management of an issuer across all asset classes in the CRSP MF database. $IssuerNrETFs$ is the number of non-active equity ETFs operated by an issuer and $IssuerNrSelfIndexETFs$ is the number of self-index ETFs operated by an issuer in our sample from ETF global. Standard errors are two-way clustered by month-issuer interaction and reported in parentheses.

	OLS			Logit		
	(1)	(2)	(3)	(4)	(5)	(6)
(Intercept)	0.040			−2.173		
	(0.049)			(2.719)		
$IssuerLog(Age)$	0.004	0.003	0.007	0.122	−0.063	−0.165
	(0.006)	(0.006)	(0.010)	(0.492)	(0.470)	(3.246)
$IssuerLog(AUM)$	−0.002	−0.002	−0.013*	−0.166	−0.101	−1.194
	(0.003)	(0.003)	(0.007)	(0.176)	(0.121)	(1.170)
$IssuerNrETFs$	0.000	0.000	0.001	0.022	0.022	0.071
	(0.000)	(0.000)	(0.001)	(0.015)	(0.015)	(0.051)
$IssuerNrSelfIndexETFs$	0.007***	0.008***	0.006	0.325***	0.368***	0.474
	(0.003)	(0.003)	(0.005)	(0.077)	(0.095)	(0.727)
Num. Obs.	1521	1521	1521	1521	629	250
R^2	0.02	0.04	0.13	0.11	0.15	0.16
FE: Month		X			X	
FE: Issuer			X			X

* p < 0.1, ** p < 0.05, *** p < 0.01

Table IA.3. Explaining net expense ratios with extended controls.
The dependent variable is an ETFs annual Net Expense Ratio. $SelfIndexer?$ is a dummy variable that equals one if an ETF is self-indexed. $Log(AUM)$ is the natural logarithm of an ETFs assets under management. $Log(Holdings)$ is the natural logarithm of an ETFs number of unique portfolio holdings. $Log(Age)$ is the natural logarithm of an ETFs age as measured in days since inception divided by 365. $SmartBeta?$ is a dummy variable that is one if an ETF is classified as smart beta. $Uniqueness$ is the 12-month rolling average absolute difference of ETF gross returns versus the value-weighted average ETF gross return in the same style-category, as in Kostovetsky & Warner (2022). $TurnRatio$ is an ETFs yearly portfolio turnover ratio. $MAvgSpread$ is the monthly average ETF secondary market price spread scaled by price. Standard errors are two-way clustered by month-issuer interaction and reported in parentheses.

	Net Expense Ratio (%)			
	(1)	(2)	(3)	(4)
$SelfIndexer?$	0.059***	0.050***	0.044***	0.031***
	(0.012)	(0.012)	(0.012)	(0.011)
$Log(AUM)$	−0.024***	−0.025***	−0.022***	−0.023***
	(0.002)	(0.002)	(0.002)	(0.002)
$Log(Holdings)$	0.000	0.001	0.001	0.003
	(0.002)	(0.002)	(0.002)	(0.002)
$Log(Age)$	0.075***	0.079***	0.068***	0.074***
	(0.004)	(0.004)	(0.004)	(0.004)
$SmartBeta?$	0.001	0.004	−0.004	−0.004
	(0.012)	(0.012)	(0.013)	(0.013)
$Uniqueness$	0.056***	0.060***	0.055***	0.059***
	(0.004)	(0.005)	(0.005)	(0.005)
$TurnRatio$			0.032***	0.041***
			(0.007)	(0.008)
$MAvgSpread$			−0.008	−0.008
			(0.006)	(0.006)
Num. Obs.	35 644	35 644	35 000	35 000
R^2	0.50	0.53	0.51	0.53
FE: Issuer	X		X	
FE: Month-Style	X	X	X	X
FE: Month-Issuer		X		X

* p < 0.1, ** p < 0.05, *** p < 0.01

Table IA.4. Monthly performance regressions.
In Panel A we predict the Gross Return of ETFs, calculated by backing out net expense ratios from net (NAV) returns. In Panel B we predict Gross Carhart four-factor excess returns, defined as the gross returns minus predicted returns using factor loadings from a 36 month rolling windows. $SelfIndexer?$ is a dummy variable that equals one if an ETF is self-indexed. $Log(AUM)$ is the natural logarithm of an ETFs assets under management. $Log(Holdings)$ is the natural logarithm of an ETFs number of unique portfolio holdings. $Log(Age)$ is the natural logarithm of an ETFs age as measured in days since inception divided by 365. $SmartBeta?$ is a dummy variable that is one if an ETF is classified as smart beta. $Volatility$ is the standard deviation of daily net (NAV) returns within a month. $Uniqueness$ is the 12-month rolling average absolute difference of ETF gross returns versus the value-weighted average ETF gross return in the same style-category, as in Kostovetsky & Warner (2022). Standard errors are two-way clustered by month-issuer interaction and reported in parentheses.

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Gross Returns (%)						
$SelfIndexer?$	0.077*	-0.027	-0.021	0.136**	-0.024	0.021
	(0.045)	(0.055)	(0.060)	(0.061)	(0.099)	(0.093)
$Log(AUM)$	0.018*	0.012	0.007	0.020**	0.009	0.004
	(0.009)	(0.009)	(0.008)	(0.010)	(0.009)	(0.009)
$Log(Holdings)$	-0.001	0.009	0.010	0.003	0.010	0.014
	(0.014)	(0.015)	(0.019)	(0.016)	(0.017)	(0.021)
$Log(Age)$	-0.018	0.005	0.005	-0.024	0.022	0.024
	(0.026)	(0.027)	(0.027)	(0.029)	(0.031)	(0.032)
$SmartBeta?$	-0.054**	-0.050*	-0.055*	-0.034	-0.014	-0.015
	(0.025)	(0.029)	(0.029)	(0.028)	(0.034)	(0.034)
$Volatility$	1.161	1.180	0.706	0.602	0.678	0.376
	(0.735)	(0.731)	(0.788)	(0.691)	(0.702)	(0.757)
$Uniqueness$				0.043	0.044	0.047
				(0.048)	(0.050)	(0.052)
Num. Obs.	43074	43074	43074	35641	35641	35641
R^2	0.88	0.88	0.90	0.87	0.87	0.90
Panel B: Gross Carhart Four-Factor Excess Returns (%)						
$SelfIndexer?$	0.042	-0.045	-0.042	0.060	-0.142	-0.153
	(0.079)	(0.088)	(0.099)	(0.100)	(0.112)	(0.137)
$Log(AUM)$	-0.005	-0.012	-0.013	-0.009	-0.017	-0.018
	(0.009)	(0.010)	(0.011)	(0.013)	(0.013)	(0.014)
$Log(Holdings)$	-0.012	-0.020	-0.022	-0.032	-0.037	-0.031
	(0.019)	(0.020)	(0.021)	(0.022)	(0.023)	(0.024)
$Log(Age)$	0.011	0.002	0.000	-0.003	-0.035	-0.027
	(0.043)	(0.054)	(0.057)	(0.049)	(0.062)	(0.068)
$SmartBeta?$	-0.085**	-0.088**	-0.096***	-0.089**	-0.092**	-0.105**
	(0.034)	(0.034)	(0.035)	(0.039)	(0.041)	(0.043)
$Volatility$	-0.348	-0.063	-0.544	-0.387	-0.071	-0.718
	(0.643)	(0.653)	(0.675)	(0.742)	(0.760)	(0.755)
$Uniqueness$				-0.018	-0.029	-0.017
				(0.046)	(0.051)	(0.050)
Num. Obs.	23929	23929	23929	20162	20162	20162
R^2	0.54	0.54	0.60	0.55	0.55	0.62
FE: Issuer		X			X	
FE: Month-Style	X	X	X	X	X	X
FE: Month-Issuer			X			X

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table IA.5. Sample of issuers classified as investment advisors and fund specialists.
We classify issuers as 'Investment Advisors' if our manual search shows they offer a range of financial services including fund management and (wealth) advisory services. 'Fund Specialists' are issuers which focus explicitly on fund management.

Investment Advisors	Fund Specialists	
Advisors Asset Management	AGFiQ	LocalShares
Blackrock	Alpha Architect	M-CAM International
Calamos	AlphaClone	Metaurus Advisors
Cboe Vest Financial	Alps	Motley Fool Asset Management
Charles Schwab	American Century	Nuveen
Deutsche Bank	Amplify	O'Shares
Diamond Hill	Aptus Capital Advisors	Oppenheimer
DWS	Beyond Investing	Pacer Financial
Fidelity	BioShares	Pax World
Goldman Sachs	BNY Mellon Investment Management	ProShares
Guggenheim	Cambria	QuantX Funds
Hartford	Change Finance	Reality Shares
Invesco	Columbia	Recon Capital
Janus Henderson	Compass	Renaissance
John Hancock	Cushing	RevenueShares
JPMorgan	Defiance ETFs	Salt Financial
Legg Mason	Direxion	SerenityShares
Nationwide	Distillate Capital Partners	SL Advisors
Northern Trust	Elkhorn	SP Funds
PIMCO	EntrepreneurShares	Sprott
Point Bridge Capital	ETF Managers Group	Syntax
Principal	ETF Securities	Timothy Plan
Redwood	Exchange Traded Concepts	TriLine Index Solutions
Royal Bank of Scotland	Exponential ETFs	TWM Funds
Russell	Falah Capital	US Commodity Funds
Scottrade	FFCM	USAA
SoFi	First Trust	ValueShares
SSgA	Franklin Templeton Investments	Van Eck
Tortoise	Global Beta Advisors	VelocityShares
Vanguard	Global X	Vident Financial
Victory Capital Management	GraniteShares	Virtus
Wahed Invest	Horizons	WBI
WisdomTree	Hoya Capital	-
-	Impact Shares	-
-	IndexIQ	-
-	Innovation Shares	-
-	Innovator Capital Management	-
-	Inspire Investing	-
-	KraneShares	-
-	Lattice	-