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

Bige Kahraman[†] Sida Li[‡] Anthony Limburg[§]

October 31, 2025

Abstract

The rise of self-indexed ETFs – funds tracking indices created by their own issuers (or affiliates) – marks a shift from the traditional model of index provision dominated by external providers. This study evaluates how this practice shapes the landscape of ETF investing. Contrary to predictions of fee competition and increased portfolio differentiation, we find that self-indexed ETFs sponsored by issuers who are also investment advisors feature higher fees, exhibit more portfolio similarity, and deliver no performance gains compared to peers. Further analyses support a self-preferencing interpretation within a search-cost framework, where these issuers promote their own higher-fee products. Self-indexed ETFs from other issuers, however, are associated with greater portfolio differentiation at comparable fees.

[*]We thank the discussants and participants at the SEC Conference on Financial Market Regulation (CFMR), the Midwest Finance Association (MFA) Annual Meeting, the Rising Scholar in Finance Conference (UZH), the FMA Europe Conference, EDHEC Business School, VU Amsterdam University, the Tinbergen Institute Amsterdam, and the University of Oxford Said Business School for their useful comments. Anthony Limburg acknowledges support from the Clarendon Fund, the Templeton Education and Charity Trust (TECT) and the Saïd Business School Foundation. 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

The rise of Exchange-Traded Funds (ETFs) has fundamentally reshaped financial markets investing, increasingly becoming an important investment vehicle for many American households.[1] Traditionally, most ETFs are "public-indexed" which means relying on indices constructed by established industry leaders, such as S&P Dow Jones, 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. About one third of all ETF fees is estimated to be paid to index providers in the form of licensing fees, and 60% of these licensing fees are estimated to be markups.[2] The market dominance of these external providers and their fees have drawn scrutiny from U.S. regulators, which are now questioning whether these firms should be subject to more stringent oversight.[3]

The emergence of "self-indexed" ETFs has recently disrupted this traditional model. Self-indexing refers to the practice in which ETF issuers, or their affiliated entities, construct and manage their own indices, thus avoiding the expenses associated with licensing third-party benchmarks. A crucial factor in the expansion of self-indexed ETFs was

[1]According to ICI reports, U.S.-listed ETF assets have surged from approximately 1.75 trillion in 2013 to over $8.1 trillion by year-end 2023, leading to increased ownership by households. For instance, over the last decade, the number of households directly owning ETFs has increased by over 163%.

[2]See An et al. (2023) for a structural model.

[3]See SEC Release No. IA-6050 (June 15, 2022). The SEC issued this *Request for Comment* to collect public input on whether index providers should be regulated as investment advisers and be subject to fiduciary duty requirements.

the SEC's 2013 decision to relax prior requirements on self-indexing, making it easier for issuers to use their own indices.[4]

Although relatively uncommon just a decade ago, self-indexed ETFs now account for nearly 20% of ETFs within popular investment styles such as broad equity and large cap. The cumulative growth in AUM of self-indexed ETFs has doubled that of externally indexed counterparts during our sample period and several established asset managers (e.g., Goldman Sachs, Fidelity, Wisdom Tree) are now offering self-indexed ETFs. Given these developments, this article focuses on examining how self-indexing practice shapes the landscape of ETF investing. Despite its increasing importance economically and regulatory concerns, the index provision market remains understudied, and this paper aims to contribute to this literature by focusing on self-indexing in the ETF market.

To this end, we consider three hypotheses. First, the removal of licensing costs may empower issuers to offer more competitive fees, an argument consistent with established theories of competitive markets (e.g., Tirole (1988)) and adopted by many industry commentators.[5] This assertion rests on the assumption that competitive pressures will manifest as price competition. However, fund issuers may struggle to compete down prices if they face an inelastic demand for well-established indices (An et al., 2023) or lack sufficient cost-effectiveness to sustain a price war. According to our second hypothesis, in this environment, fund issuers can resort to product differentiation, consistent with basic

[4]See, SEC IM-INFO-2013-09.

[5]See, for instance, in the Financial Times, Barron's and WatersTechnology.

models of monopolistic competition (e.g., Dixit and Stiglitz (1977)). Our third hypothesis posits a self-preferencing mechanism within a search costs framework building on Hortaçsu and Syverson (2004) and Roussanov et al. (2021). Index proliferation heightens search frictions as it becomes more difficult for investors to easily compare and choose among them. This enables issuers to sell nearly homogeneous portfolios at higher fees.[6] These sales are facilitated because some self-indexed ETF issuers are also private wealth managers, which allows them to promote their self-indexed products in their advisory businesses.

To empirically examine these contrasting perspectives, we start by comparing the fees of self-indexed and public-indexed ETFs. Within our sample of passive U.S. equity ETFs, we document that self-indexed ETFs charge significantly *higher* net expense ratios compared to their public-indexed counterparts. Specifically, self-indexed ETFs have fees that are about 10-13 percent higher than those of public-indexed funds.

We then test the notion that self-indexing provides issuers with greater latitude in crafting distinct investment strategies which may fit clients' particular preferences and potentially provide financial value. Using both portfolio holdings similarity as well as return correlation measures, we find that the investment strategies of self-indexed ETFs are actually more *similar* to their peers within the same style, rather than being more distinct. This lack of distinction is also reflected in their performance outcomes: across all

[6] For S&P 500 index funds, Hortaçsu and Syverson (2004) estimate search costs at 11-20bps and Roussanov et al. (2021) estimate at 39bps for actively managed funds.

our tests, we find no significant evidence that self-indexed ETFs outperform their public-indexed counterparts. This conclusion is established through multiple analyses, using different factor models, daily and monthly data, and regression analyses at both the time-series portfolio and the cross-sectional fund levels.

Our findings on fees, portfolio differentiation and performance are different from related studies as our robust testing controls for: (i) unique investment strategies (e.g., Kostovetsky and Warner (2025)), (ii) "closet active management" by passive ETFs (e.g., Akey et al. (2021); Easley et al. (2021); Cheng et al. (2019)), (iii) ETF secondary market liquidity (e.g., Brown et al. (2024); Khomyn et al. (2024)), (iv) thematic ETFs (e.g., Ben-David et al. (2022)), and (v) smart-beta investment styles (e.g., Huang et al. (2023)). Furthermore, our results are obtained after controlling for a host of fund characteristics (such as size and age) as well as time, style and issuer fixed effects.

Our full-sample findings do not support the first two hypotheses that self-indexing is associated with either price competition or portfolio differentiation. We next examine the third hypothesis, which proposes a self-preferencing mechanism among certain issuers: those that both issue self-indexed ETFs and provide wealth management advisory services may preferentially promote their own funds, particularly to clients facing high search costs. This analysis also allows us to identify cross-sectional variation across issuer types. Overall, the findings align with the predictions of this hypothesis.

To assess this idea empirically, we classify issuers into two groups: specialized fund managers (focused solely on fund management) and investment advisors (offering fund management alongside wealth management advisory services). Consistent with the third hypothesis, we first document that the results thus far are predominantly driven by investment advisors. While these advisors generally offer competitive fees across their other ETFs, their self-indexed ETFs are a notable exception, as these charge 20 percent higher compared with their public-index offerings. Moreover, their funds tend to be more similar to their peers. In contrast, among specialized fund managers, there are no significant fee differences between self- versus publicly-indexed ETFs. Also, within this group, self-indexed funds tend to be more differentiated.

The self-preferencing hypothesis argues that investment advisors favor their self-indexed ETFs in the advisory part of their businesses. This would appear as disproportionately high *Self-Ownership* (fraction of ETF shares reported by its issuer in its 13F filings) to the extent that some of these assets are held in discretionary advisory accounts.[7] Consistent with this, we find that self-indexed ETFs average 14% self-ownership, significantly higher than the 2% observed for public-indexed peers. Crucially, our regression analysis reveals 80 percent of this difference is driven by self-indexed ETFs issued by investment advisors.

As a mechanism test, we evaluate whether model portfolio recommendation is a channel through which issuers promote their self-indexed ETFs. Model portfolios are rec-

[7]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).

ommendation portfolios designed by asset management companies, some of which also issue ETFs, create indices and provide advisory services. About 54 percent of advised assets are reportedly allocated in model portfolios, which therefore makes them a useful proxy for otherwise confidential client assets (Brogaard et al., 2021). We find that, on average, ETF issuers which are also model providers tend to recommend their self-indexed ETFs more frequently compared with public-indexed ETFs, and this is exclusively driven by the ones which are also investment advisors.

Last but not least, we conduct an additional mechanism test to examine whether self-indexed ETFs attract investors with higher search costs. The rationale is that index proliferation increases search frictions, allowing issuers to charge higher fees for similar portfolios, particularly to clients facing higher search costs. If these clients can be can be proxied by limited financial sophistication (Roussanov et al., 2021), we would expect flows to self-indexed ETFs to be more sensitive to market sentiment. Consistent with this prediction, we find that flows to self-indexed ETFs rise during periods of elevated sentiment, whereas flows to publicly indexed ETFs show no significant relationship. These results are consistent with a self-preferencing mechanism operating within a search-cost framework.

Related Literature

While there are numerous studies on ETFs, there is not enough research focusing on the index provision market and its potential influence on investor welfare. Although ETF fees are seemingly low on an annual basis, even small differences in fees can have an important impact on long-term investments.[8] Given the increasing popularity of ETFs with many households, it is crucial to examine how the index provision market affects fees and thus investor welfare. Our paper contributes to this area of research by focusing on the emerging phenomenon of self-indexing.

An et al. (2023) is the first study to quantify the impact of competitive dynamics between external index providers and ETF issuers on investor costs. They estimate that 30 percent of ETF fees are licensing costs, with 60 percent being mark-ups. Xiao and Xiong (2024) later developed a general equilibrium model studying the agency issues stemming from self-indexing and they conclude that a concentrated index provision market can be more beneficial to retail investors.

Legal scholars are increasingly interested in the index market due to the significant expansion of passive investment products. While some work recognizes that self-indexed funds come at higher cost, their primary focus remains on legal and regulatory issues (Robertson, 2019). Similarly, Kostovetsky and Warner (2025) acknowledge self-indexed funds but focus mainly on externally indexed funds.

Our study goes beyond simply documenting fees. We offer a detailed economic examination of self-indexed ETFs, analyzing whether these funds provide meaningful portfolio differentiation or performance advantages to their clients. Most importantly, our findings

[8] See, for instance, French (2008), Sharpe (2013) and this SEC Bulletin.

point to a self-preferencing mechanism driven by certain self-indexing issuers which also act as investment advisors. To the best of our knowledge, this analysis is novel both to the academic literature and the existing regulatory debate. Earlier regulatory discussions raised concerns that self-indexing could facilitate malpractices such as NAV manipulation or internal front-running within fund families.[9] The 2013 SEC exemptive orders sought to address these issues through preventive safeguards, including daily portfolio disclosure and enhanced board oversight. By calling attention to a mechanism that arises when the issuer, index provider, and investment advisor are the same entity, this paper provides timely insights for policy discussions on index provision and the evolving regulatory framework.

Our findings also add to the academic literature investigating the incentives and behavior of financial advisors (Mullainathan et al., 2012; Foerster et al., 2017). Prior research documents misconduct and poor client outcomes in the U.S. financial advice industry, in part attributing these issues to principal-agent problems due to brokers' incentives for mutual fund sales (Bergstresser et al., 2009; Guercio and Reuter, 2014; Kahraman, 2021).[10] The incentives analyzed in most of the earlier literature stem from a now largely outdated commission-based compensation model where brokers require no direct payment from clients and instead draw from product-based commissions, some of which are included

[9]Section 4 empirically examines these concerns and finds no supporting evidence.

[10]Egan et al. (2019); Dimmock et al. (2018); Clifford and Gerken (2021); Gurun et al. (2021) study misconduct in the U.S. financial advice industry. Focusing on Canadian financial advisors, Linnainmaa et al. (2021) find advisors' personal investments align with how they advise their clients.

in fund fees (Christoffersen et al., 2013; Edelen et al., 2012).

The market's shift toward Exchange-Traded Funds (ETFs) is largely accompanied by a change to the client-paid fee model, with the goal of better aligning incentives. In this model, advisors are compensated through client-paid fees which are based on clients' total account size and separate from fund expense ratios. Different from most earlier papers, we examine advisor incentives within this client-paid fee environment, which is relatively understudied. Recently, Edelen et al. (2025) studied the client-paid fee model in Australia, where regulators have introduced a novel auto-drop requirement. Focusing on the U.S market, we study the emerging practice of self-indexing and provide evidence that is suggestive of a self-preferencing mechanism among investment advisors, thereby offering insights into agency concerns in the contemporary financial advice landscape.

Last but not least, this study contributes to a growing body of literature documenting dominated products within the ETF market (Brown et al., 2024). For instance, previous work documents significant underperformance and high fees for thematic ETFs (Ben-David et al., 2022), poor performance linked to index providers' data mining in smart-beta funds (Huang et al., 2023), and evidence of closet active management in passive ETFs (Akey et al., 2021; Easley et al., 2021; Cheng et al., 2019), an ability of liquid ETFs to extract rents from liquidity-sensitive investors (Khomyn et al., 2024). As noted previously, we demonstrate the distinctiveness of our results from these studies and highlight the issues arising from self-indexing practices.

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 13 self-indexing ETFs.

Our sample includes all passive unlevered U.S. equity ETFs from 2012 to 2020.[11] To

[11]It is ambiguous whether levered ETFs can be considered "passive". Nevertheless, we find that the underlying indices (before leverage) are all public indices in our data.

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 portfolio 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 issuer or one of its affiliated companies, 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-indexed ETF, on the other hand, is an ETF that licenses its index from an external third-party index provider.[12]

In the early years of the ETF market, issuers seeking to launch self-indexed ETFs were required to obtain exemptive relief from the SEC and were subject to several requirements.[13] A turning point came on July 10, 2013, when the SEC issued guidance (IM-INFO-2013-09) and new exemptive orders that significantly relaxed prior conditions on self-indexed ETFs. In particular, the Commission eased the requirements, relying instead mostly on daily portfolio transparency and governance.[14] These changes brought self-indexed ETFs closer to regulatory parity with ETFs tracking third-party indices, making the model more attractive to issuers and arguably contributing to the growth of self-indexing over the subsequent 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-indexed ETFs.[15]

We use the variable 'primary_benchmark' from ETF Global to identify ETFs that are

[12] Our definition of self-indexing is based on the formal SEC definition, which states: "A "Self-Indexing Fund" is a Fund for which an Affiliated Person, or a Second-Tier Affiliate, of the Trust or a Fund, of the Adviser, of any Sub-Adviser to or promoter of a Fund, or of the Distributor (each, an "Affiliated Index Provider") will serve as the Index Provider" (SEC Release No. 30560)

[13] These requirements mainly stemmed from concerns of conflicts of interest, such as the manipulation of constituent pricing and internal front-running. These requirements generally included: (i) requirements to promote transparency of the index methodology (e.g., making index rules freely available and announcing methodology changes with at least 60 days' notice); (ii) use of an unaffiliated third party to calculate the index; and (iii) creation of rigid firewall arrangements to separate the fund's adviser and the index provider.

[14] The SEC states its objectives in broad terms, as to protect investors, promote informed investment decisions and facilitate innovation in investment products. Some industry commentators interpreted it as an effort to promote competition in the index provision market. See, for instance, SEC Issues New Relief for Self-Indexing ETFs | Morgan Lewis - JDSupra.

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

self- and public-indexing. Our procedure involves multiple steps. Self-indexed ETFs are 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 third-party indices such as "S&P", "FTSE", "Russell", "CRSP", "MSCI", "Morningstar", "Dow Jones", "NYSE", "NASDAQ", "Nasdaq", "StrataQuant", we mark these observations as public-indexed ETFs. If the text in the benchmark does not include these keywords but instead includes the name of the issuer (or its affiliates), we then mark these as self-indexed ETFs. For instance, the JP Morgan US momentum ETF (ticker: JMOM) is marked as a self-indexed 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.[16] 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-indexed

[16]Examples of ETF issuers being affiliated with index providers include FlexShares ETFs tracking Northern Trust indices and Lattice ETFs tracking Hartford indices.

(public-indexed) identifiers accordingly. Through this procedure, we can identify the self-indexed (public-indexed) status of more than 99% of funds for which ETF Global provides a 'primary_benchmark'. This classification method allows us to define self-indexed ETFs in line with the formal definition offered by the SEC (see above). 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').

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). Although there were only a few self-indexed funds a decade ago, by the end of our sample period, this has increased to 96 ETFs in total (Panel A). Moreover, the total cumulative AUM growth of self-indexed ETFs has doubled the growth of public-indexed ETFs during this time (Panel B). Table 1 shows the distribution of self- and public-indexed ETFs across investment styles.[17] Self-indexed

[17] 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.

funds appear to be most prevalent in investment styles which are also popular among public-indexed funds (e.g., broad equity and large cap). Within these investment styles, self-indexed funds constitute about 20% of all funds by the end of our sample period. Throughout our analysis, we control for style fixed effects.

We classify ETF issuers into three groups based on their mix of self- and public-indexed offerings: ones which offer only self-indexed funds, ones which offer only public-indexed funds, and ones which offer a mix of self- and public-indexed funds. Figure 2 shows the share of issuers in each of these groups over time. While most issuers only offered public-indexed ETFs a decade ago, this has been changing over time. The count of issuers which explicitly focus on self-indexed ETFs, and issuers offering both types of funds has been steadily increasing since 2012. By the end of our sample period, 51% of issuers in our sample are pure public-indexed issuers, 28% are pure self-indexed and 20% are issuers which offer a mix of public and self-indexed ETFs.[18]

Internet Appendix Table IA.1 lists the 10 largest ETF issuers within each group.[19] 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-indexed ETFs. For instance, within our sample of U.S

[18]When companies start offering self-indexed ETFs, they generally do this by starting new funds instead of converting the existing ones, hence we oberver very few index-switching events in the data.

[19]For brevity, tables reported in the Internet Appendix are labeled with abbreviation “IA”.

equity passive ETFs, Fidelity offers 6 self- and 11 public-indexed ETFs; JP Morgan offers 5 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-indexed 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-indexed funds include some of the largest players in the market such as BlackRock and Vanguard. 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-indexed 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-indexed 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-indexed ETF. Issuers which already offer self-indexed ETFs are more likely to offer another one.

Panel A of Table 2 shows the descriptive statistics on various fund characteristics for self- versus public-indexed funds. Panel B and C report summary information for our similarity measures, to be defined in the next section. $Net\ Expense\ Ratio\ (\%)$ is the annual net expense ratio 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.

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. $Gross\ Carhart\ Alpha\ (\%)$ is the fund's excess return estimated over a 36-month rolling window using the 4-factor Carhart model. Following the literature, we define fund flow as:

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

As expected, public-indexed ETFs are on average older ($Log(Age)$ is 2 versus 1.37) and consequently, they tend to be larger ($Log(AUM)$ is 19.4 versus 18.2). Self-indexed funds, however, seem to be receiving more flows – *Fund Flow* is 0.031 vs 0.019. ETF performance measures seem comparable. *Gross Return* for public- vs self-indexed ETFs is 1.05 vs 1.02, and *Cahart Alpha* is -0.166 and -0.184. In the following sections, we will formally examine the systematic differences in fees, performances and flows between the two types of ETFs in regression models where we control for various fund characteristics, investment styles, issuer and time fixed effects.

This panel also reports summary statistics on our ownership variable (available at quarterly frequency). $Institutional\ Self\ Ownership$ is the fraction of total ETF shares reported in 13F filings of the ETF issuing company or one of its affiliates. Self-indexed ETFs have 14% self-ownership, while public-indexed ETFs have only 2% on average. Later in the paper, we formally examine also the ownership differences between the two groups of ETFs.

3 Do self-indexed ETFs provide financial benefits to investors?

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. Table 3 reports the results. Throughout all regressions, standard errors are two-way clustered by month-issuer interaction, and we find that self-indexed ETFs have fees that are statistically higher.

In column 1, we first regress $Net\ Expense\ Ratio\ (\%)$ on $SelfIndexer?$, which is an indicator variable that equals one if the fund is identified as a self-indexed ETF, including month-style fixed effects. The coefficient estimate for $SelfIndexer?$, is 0.055 (5.5 basis points) and statistically significant. In the second column, we extend the model by including several control variables $Log(AUM)$, $Log(Age)$, $Log(Holdings)$, $SmartBeta?$ (an indicator variable taken from Morningstar which equals one for smart beta ETFs) as well as issuer fixed effects. The coefficient estimate for $SelfIndexer?$ stands similarly significant at 0.052.

As ETFs fees are in basis points on average, these estimates are economically meaningful: they imply that self-indexed ETFs on average charge 13% more their public-indexed counterparts. Although ETF fees are seemingly low on an annual basis, even small differences in fees can have an important impact on long-term investments.[20] These results

[20] See, for instance, French (2008), Sharpe (2013) and this SEC Bulletin.

are interesting in that they represent contrary evidence to the popular view expressed by numerous industry commentators that self-indexing can be a way for ETF issuers to offer more competitive fees.[21]

We next assess our main findings against related research and demonstrate how they differ from previous work. It is useful to first highlight two points. We incorporate style fixed effects throughout our analysis, to ensure that our findings are not driven by specific style categories, such as thematic ETFs (as discussed in Ben-David et al. (2022)). We also include issuer fixed effects to confirm that these results are not attributable to potential differences in issuer-level factors, such as differences in brand recognition, customer services or other business models that might influence their fees or strategies (Hortaçsu and Syverson, 2004; Kostovetsky and Warner, 2025).

Kostovetsky and Warner (2025) document that certain ETFs employ distinct investment strategies compared to their peers, often accompanied by higher fees. To ensure this does not account for our results, we follow them and control for *Uniqueness* (the 12-month rolling absolute difference in a fund's gross returns versus its style category average). Consistent with the literature, *Uniqueness* is associated with higher fees. Crucially, however, our core finding persists after controlling for this measure (column 3). We will formally examine portfolio differentiation in the next section to evaluate related

[21] For instance, Todd Rosenbluth, director of ETF and Mutual Fund Research at CFRA has recently said "Self-indexing is a reaction to investor preference for low-cost strategies (...) licensing a third-party index can be expensive." For more, also see in the Financial Times, Barron's and WatersTechnology.

hypotheses around it.

In columns 4 and 5, we evaluate the potential role of the "closet active management" of passive ETFs (Akey et al., 2021; Easley et al., 2021; Cheng et al., 2019). It is useful to note that we control for smart-beta in our baseline regressions. Moreover, we address this point by two additional proxies of activeness, following the related literature. In column 4, we include $Activeness$, defined as $(1 - R^2)$ from regressing daily gross returns on the Carhart four-factor model (Amihud and Goyenko, 2013), in our regression. In column 5, we introduce *TurnRatio*, which is ETF's yearly portfolio turnover ratio (Pástor et al., 2017). With both measures, our finding remain stable, indicating self-indexing is empirically distinct from proprietary indexing practices identified by Akey et al. (2021) and more broadly, closet active management.

In column 6, we control for the influence of ETF secondary market liquidity, which is a dimension of ETF competition (Khomyn et al., 2024) and thus a potential confounder for fund fees. As a proxy for share liquidity, we use an ETFs bid-ask spread, *MAvgSpread*, which is defined as the monthly average difference between the ask and bid as reported in CRSP, scaled by the price. Our result is virtually unchanged. Finally, in column 7, we consider all of the additional control variables together alongside month-style and month-issuer fixed effects, and confirm that our results are robust to additional extensions of model specification. To summarize, this section demonstrates that self-indexed ETFs charge significantly higher fees, contradicting the predicted outcome of fee competition with public-indexed funds.

3.2 Differentiated products?

In this section, we build on our earlier discussion and delve deeper into whether self-indexing provides issuers the flexibility to create distinct investment strategies that may provide performance gains to customers. Our analysis reveals that self-indexed ETFs are, in fact, more similar to their peers within the same style, rather than being distinct. They also do not provide significantly better performance. Overall, our results contradict the differentiated products hypothesis.

3.2.1 Differences in holdings and return correlations

We start by introducing a cosine similarity measure between each pair of ETFs within our sample, to examine the differences in holdings (Sias et al., 2016). 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. 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, $Cos_{A,B}$.[22] Fund-level cosine similarity is constructed by averaging the cosine similarity of each fund against all other "peer" funds within the same style category and same month. As an alternative, for each ETF-month, we also calculate average pairwise daily return correlations (against its peer funds within the same style). Panel B and C of Table 2 report summary statistics for fund-level cosine similarity and return correlations, respectively. Return correlations tend to be quite high (around 0.90) while cosine similarity appears relatively lower (around 0.3), arguably due to synthetic replication, value-weighting (with value weighting, a handful of large companies drive a significant portion of index returns) and exponential changes in cosine similarity measure due to its definition.

Our focus is to evaluate the differences in similarity measures between self- and public-indexed ETFs. We start by conducting network analyses to visualize the potential differentiation. Based on $Cos_{A,B}$, we calculate a minimum spanning tree representing the

[22] 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 their measures of cosine similarity are greater than 0.998.

network of all ETFs in our sample. Intuitively, this method calculates a parsimonious network which connects all funds, minimizing the necessary connections, where funds with more similar holdings become bunched together, forming "clusters". In Figure 3, we plot this network using the Fruchterman Reingold layout, which minimizes overlapping connections and spreads the network evenly on a canvas. The network diagrams clearly illustrate that self-indexed ETFs are generally situated at the center of various clusters, which demonstrates a strong similarity with peer funds in the same style. We reproduce the same diagrams using return correlations in IA.2. Our conclusions carry over.

Table 4 presents a formal regression analysis. In columns 1 and 2, the dependent variable is the average monthly cosine similarity of a fund to its peers. We regress this variable on various ETF characteristics including $SelfIndexer?$, $Log(AUM)$, $Log(Age)$, $Log(Holdings)$ and $SmartBeta?$ as well as issuer, style and month fixed effects. Our findings consistently show that, when compared to their public-indexed peers, self-indexed ETFs have more *similar* (i.e. less differentiated) portfolio holdings. In columns 3 and 4, we replace cosine similarity with average monthly return correlations (as before, measured against peers within the same style). Results are similar.[23]

In summary, we find no evidence of portfolio differentiation. If anything, we find self-indexed ETFs are actually more similar to their peer funds within the same style categories, compared with their public-indexed counterparts. Although our analysis shows

[23] Table IA.3 shows the robustness of these findings when we also control for additional explanatory variables (Uniqueness, Activeness, Turnover andSpread) as in Table 3.

no differentiation, investors may still perceive these products as distinct due to marketing and branding. This perception aligns with our third hypothesis, which will be examined in a later section.

3.2.2 Differences in performance

We next move to evaluating the potential differences in fund performances. We conduct multiple analyses, using different factor models, daily and monthly data, and tests at both the time-series portfolio and cross-sectional fund levels. Across all these tests, we never find significant evidence that self-indexed ETFs outperform their public-indexed counterparts.

We start by presenting results of various portfolio tests in which we estimate alphas to the 'spread' portfolio – the portfolio that goes long on self-indexed ETFs and short on public-indexed ETFs. Columns 1 and 2 of Table 5 report the monthly alphas, while columns 3 and 4 report the daily alphas. The daily analysis not only provides a higher statistical power for our tests but also serves as an independent check on the results derived from the monthly alphas. For both monthly and daily alphas, we report results for value- and equal-weighted portfolio returns, and use three different factor models: the CAPM, the Fama-French three-factor and Carhart four-factor model. In total, we estimate alphas under 12 different scenarios. All returns are before fees and we require at least 20 observations in each portfolio of funds.[24] We use Newey-West standard errors al-

[24]Effectively, this moves the start date for this analysis to July 2015, marginally shortening the sample

lowing correlation up to 6 months. We find that all 12 alphas for the spread portfolios are statistically indistinguishable from zero, indicating no significant performance difference between the two types of ETFs.

Next, we move to cross-sectional fund level regression analyses where we control for the influence of fund characteristics on fund performance (Table 6). The dependent variable is the gross risk-adjusted return where returns are adjusted using a Carhart four-factor model with loadings to each factor estimated from 36-months rolling windows. Columns 1 and 2 use monthly, and columns 3 and 4 use daily alphas, significantly increasing the number of observations. 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-indexed ETFs.[25]

In Table 7, we consider the degree to which self- and public-indexed 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. Self-indexed ETFs appear to have somewhat lower volatility, however estimates are eco-

period. Cross-sectional regression tests (presented next) provide an alternative empirical approach and show the robustness of our findings.

[25]Internet Appendix Table IA.3 shows the results when we control for additional explanatory variables as in Table 3. There is no evidence that self-indexed ETFs perform better. If anything, in the analysis with daily alphas, their performance seems weakly negative.

nomically modest (around a 2-4% difference). In a similar vein, self-indexed ETFs seem to have either similar or weakly lower Sharpe Ratios. Taken altogether, our findings highlight that self-indexing ETFs do not deliver superior performance outcomes compared to their public-indexed counterparts.

4 Self-preferencing by issuer-advisors?

Our analyses thus far reveal that, on average, self-indexed funds neither perform better nor provide portfolio differentiation. In this section, we examine the "self-preferencing hypothesis": index proliferation increases search frictions and ETF issuers which are also in the business of providing advisory services promote their own high-fee self-indexed ETFs to arguably high search-cost clients. The results of this section are consistent with this hypothesis.

4.1 Investment advisors and self-indexed ETFs

We start by analyzing whether self-indexed ETFs issued by issuers which are likely to be more prone to self-preferencing charge higher fees. To test this idea, we classify issuers into two groups: 'specialized fund managers' and 'investment advisors' (or 'issuer-advisors'). While the former group explicitly focuses on fund management (e.g., Global X, Defiance ETFs, Van Eck), the latter group offers a range of financial services including fund management as well as wealth management advisory services (e.g., Fidelity, State Street, Goldman Sachs) and is more prone to self-preferencing. There are 105 unique issuers in our 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 8 tests the differences in fees between the two groups of issuers. We take the specification of column 2 of Table 3 and introduce the interaction term $SelfIndexer?$ x $InvAdvisor?$, where $InvAdvisor?$ is an indicator variable equal to one for ETFs issued by issuers classified as 'investment advisors'. The regression includes month-style fixed effects alongside the control variables from Table 3. The results are striking. Investment advisors, on average, charge lower fees, which is consistent with these companies being able to offer more competitive prices for their products. However, their self-indexed ETFs systematically charge higher fees. $SelfIndexer?$ x $InvAdvisor?$ interaction term is significantly positive. The coefficient for $SelfIndexer?$ alone, however, is indistinguishable from zero: among ETFs issued by specialized fund managers, there is no significant difference in fees between self- and public-indexed ETFs.

In columns 2 and 3, respectively, we check whether self-indexed ETFs from investment advisors perform differently or have more differentiated portfolios. We find 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 category. Self-indexed ETFs issued by specialized fund managers, on the other hand, seem to offer more differentiated portfolios.

In column 4, we analyze the differences in investor flows. We first observe that $SelfIndexer?$ is significantly positive, consistent with our previous discussion that self-indexed funds have been gaining traction over the past decade attracting more flows than public-indeed funds. Moreover, the coefficient for $InvAdvisor?$ is also significantly positive (albeit smaller in magnitude), demonstrating that ETFs offered by such issuers on average attract more capital. However, $SelfIndexer? : InvAdvisor?$ is negative. Overall estimates indicate that, on average, these self-indexed ETFs have flows comparable with those of a typical public-indexed ETF with similar characteristics.

In Table 9, we focus on testing how flows to self-indexed ETFs vary with aggregate market sentiment. Under the self-preferencing hypothesis, index proliferation amplifies search frictions, allowing issuers to market similar portfolios at higher fees, arguably to high search-cost clients. These clients can be characterized by limited financial sophistication (Roussanov et al., 2021). If self-indexed ETFs are popular with such clients, we can expect flows of self-indexed funds to be more sensitive to market sentiment. Using the University of Michigan Consumer Sentiment Index as a proxy for aggregate market sentiment (e.g., Giannetti and Kahraman (2017)), we confirm that this is indeed the case. During periods of heightened market sentiment, self-indexed funds experience significantly higher flows, while publicly indexed funds display no comparable sensitivity to sentiment.

4.2 Institutional self-ownership of self-indexed ETFs

Under the self-preferencing hypothesis, investment advisors promote their self-indexed ETFs through the advisory part of their businesses. These assets would appear in the 13F filings of the investment advisor to the extent that some of these assets are purchased via (and held in) discretionary advisory accounts.[26] If this is the case, we would expect to find $Institutional\ Self\text{-}Ownership$ (fraction of ETF's total shares outstanding which are reported by the ETF issuer, or an affiliated company, in 13F filings) to be disproportionately high for self-indexed ETFs issued by investment advisors.

The results reported in Table 10 confirm this prediction. As noted previously, when we look at the summary statistics (Table 2), we observe that self-indexed ETFs have 14% self-ownership, while public-indexed ETFs have only 2% on average. Regression analyses formally establishes the significance of this difference. In column 1, the estimate for the self-index coefficient is 11%, revealing a significant difference in $Institutional\ Self\text{-}Ownership$ on average. In column 2, we include the $SelfIndexer?$ x $InvAdvisor?$ interaction and find that the difference in self-ownership is largely driven by investment advisors. In column 2, the self-indexed dummy has a positively significant but economically small coefficient (around 3.5%). However, the estimate for $SelfIndexer?$ x $InvAdvisor?$ stands at 9.5%, which is both statistically significant and economically important. The investment advi-

[26]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).

sor estimate alone is statistically not different from zero. Overall, our findings provide supporting evidence for the self-preferencing hypothesis.

One might wonder whether self-indexed ETFs are bundled with advisory services and their fees, in part, reflect compensation for such services. This is highly unlikely for two reasons. First, just like public-indexed ETFs, self-indexed ETFs are exchange traded and also widely available for direct purchase by individual investors without an advisor. Second, investment advisors are compensated for selling or recommending ETFs through "client-paid" fees, which are *separate* from ETF expense ratios.[27] Client-paid fees are recurring annual fixed or percentage charges based on the total value of a client's account. This is different from the older commissions model used by broker-sold mutual funds where part of the advisory compensation could be paid via a fund's expense ratio (specifically, 12b-1 fees).

4.3 Model portfolio recommendations and self-indexed ETFs

In this section, we provide a mechanism test by evaluating whether model portfolio recommendation is one channel through which issuers promote their self-indexed ETFs. Model portfolios are recommendation portfolios designed by asset managers and strategists for financial advisors. Some asset management companies play multiple roles such as managing and issuing funds, creating indices as well as providing investment advice

[27] For instance, see 2025 SEC Bulletin.

and model portfolio recommendations. Brogaard et al. (2021) show that asset manager model providers recommend their affiliated ETFs more frequently and therefore enjoy significant flows. Building on these findings, we are interested in examining whether ETF issuers which are model providers are more likely to include their self-indexed ETFs in their recommended model portfolios compared to their public-indexed ETFs. Reportedly, 54% of advised assets are invested in model portfolios.[28] Therefore, 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 out of 785 unique ETFs are at least part of one model portfolio, and 29 out of 105 unique issuers are identified as model portfolio providers. Of these 29 ETF issuers, 20 are investment advisors. This means that a much higher percentage of investment advisors provide model portfolios. Only 12% of the specialized fund managers offer model portfolios (9 out of 71). The same figure is 58% for investment advisors (20 out of 34).

We classify an ETF as affiliated to a model provider if the ETF issuing company 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

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

ETFs in their model portfolios.[29] 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 11. Explanatory variables include $Log(AUM)$, $Log(Age)$, $Net\ Expense\ Ratio\ (\%)$, $AvgNetRet(Yr)\ (\%)$ and a smart-beta flag. Standard errors are clustered by quarter-issuer interaction.

Issuers which are model providers tend to recommend their self-indexed ETFs more often than their public-indexed ETFs. Note that for issuers that are not model providers the dependent variable equals zero for both self- and public-indexed ETFs. This effect is economically large, given that the average count of affiliated model portfolio membership is around 0.6 (the estimates for $SelfIndexer?$ are 0.788 and 1.171 in columns 2-3). In the last column, we introduce the $SelfIndexer?$ x $InvAdvisor?$ interaction to examine the behavior of investment advisors. The estimate for $InvAdvisor?$ is significantly positive – ETFs issued by investment advisors are part of more affiliated model portfolios, consistent with most model providers being classified as investment advisors. Importantly,

[29] 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.

$SelfIndexer?$ x $InvAdvisor?$ is significantly positive. That is, investment advisors tend to recommend their self-indexed ETFs more frequently than their public-indexed ETFs in model portfolios, indicating that model portfolio recommendations are indeed a channel through which issuer-advisors promote their self-indexed ETFs to investors.

4.4 Policy discussion

Unlike previous regulatory discussions, this paper highlights a self-preferencing mechanism among issuer-advisors, drawing attention to a previously overlooked agency concern within the indexing market. Earlier commentary focused on potential malpractices, such as manipulation of constituent pricing to inflate NAV and internal front-running within fund families.[30] While these concerns had been widely discussed, they are not the primary focus of our study, as our analysis finds limited evidence supporting their relevance in our sample.

First of all, the risk of price manipulation is limited given our focus on liquid U.S. equities. Second, we believe that the risk of front-running is arguably more acute for public-indexed ETFs. Their rebalancing and reconstitution dates are pre-announced, making them highly susceptible to large-scale front-running by the broader market given the predictable nature of such trades and the vast aggregate capital tracking these indices (Chinco and Sammon, 2024; Li, 2021). In contrast, self-indexed funds are unlikely to gen-

[30] See, for instance, https://www.etfstream.com/articles/self-indexing-etf-issuers-must-answer-conflict-of-interest-question. For more, see Invesco's SEC comment letter on Rule 6c-11.

erate large price impact in comparison, since their issuers can avoid these issues (e.g., by not pre-announcing or shifting rebalancingdates). However, for completeness, we empirically tested for possible internal front-running and found no supporting evidence (Table IA.4). For each self-indexed ETF, we examine the trading activities of all other ETFs offered by the same issuer around the self-indexed funds' rebalancing dates. We find no trace of strategic trading that would suggest that other funds are front-running self-indexed ETFs (panel A). We also examine whether the presence of self-indexed ETFs in a fund family is associated with superior performance among the issuer's actively managed mutual funds. This would suggest that active funds are capitalizing on informational advantages. However, we find no such relationship (Panel B).

5 Conclusion

This papers examine self-indexing as a potential competitive mechanism and evaluate how this practice shapes the landscape of ETF investing. Contrary to predictions of fee competition and increased portfolio differentiation, self-indexed ETFs sponsored by issuers who are also investment advisors feature higher fees, exhibit more portfolio similarity, and deliver no performance gains compared to peers. Further analyses indicate a self-preferencing interpretation by these issuers. Self-indexed ETFs from other issuers, however, are associated with greater portfolio differentiation at comparable fees.

References

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

Amihud, Yakov and Ruslan Goyenko, "Mutual fund's R 2 as predictor of performance," *The Review of Financial Studies*, 2013, *26* (3), 667–694.

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.

Bergstresser, Daniel, John M. R. Chalmers, and Peter Tufano, "Assessing the Costs and Benefits of Brokers in the Mutual Fund Industry," *The Review of Financial Studies*, 05 2009, *22* (10), 4129–4156.

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," *forthcoming at Critical Finance Review*, 2024.

Cheng, Si, Massimo Massa, and Hong Zhang, "The unexpected activeness of passive investors: a worldwide analysis of ETFs," *The Review of Asset Pricing Studies*, 2019, *9* (2), 296–355.

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

Christoffersen, Susan EK, Richard Evans, and David K Musto, "What do consumers' fund flows maximize? Evidence from their brokers' incentives," *The Journal of Finance*, 2013, *68* (1), 201–235.

Clifford, Christopher P and William C Gerken, "Property rights to client relationships and financial advisor incentives," *The Journal of Finance*, 2021, *76* (5), 2409–2445.

Dimmock, Stephen G, William C Gerken, and Nathaniel P Graham, "Is fraud contagious? Coworker influence on misconduct by financial advisors," *The Journal of Finance*, 2018, *73* (3), 1417–1450.

Dixit, Avinash K and Joseph E Stiglitz, "Monopolistic competition and optimum product diversity," *The American Economic Review*, 1977, *67* (3), 297–308.

Easley, David, David Michayluk, Maureen O'Hara, and Talis J Putnins, "The active world of passive investing," *Review of Finance*, 2021, *25* (5), 1433–1471.

Edelen, Roger M., Kingsley Y.L. Fong, and Jingyi Han, "Regulating inattention in fee-based financial advice," *Journal of Financial Economics*, 2025, *164*, 103985.

_ **, Richard B. Evans, and Gregory B. Kadlec**, "Disclosure and agency conflict: Evidence from mutual fund commission bundling," *Journal of Financial Economics*, 2012, *103* (2), 308–326.

Egan, Mark, Gregor Matvos, and Amit Seru, "The market for financial adviser misconduct," *Journal of Political Economy*, 2019, *127* (1), 233–295.

Foerster, Stephen, Juhani T Linnainmaa, Brian T Melzer, and Alessandro Previtero, "Retail financial advice: does one size fit all?," *The Journal of Finance*, 2017, *72* (4), 1441–1482.

French, Kenneth R, "Presidential address: The cost of active investing," *The Journal of Finance*, 2008, *63* (4), 1537–1573.

Giannetti, Mariassunta and Bige Kahraman, "Open-End Organizational Structures and Limits to Arbitrage," *The Review of Financial Studies*, 07 2017, *31* (2), 773–810.

Guercio, Diane Del and Jonathan Reuter, "Mutual fund performance and the incentive to generate alpha," *The Journal of Finance*, 2014, *69* (4), 1673–1704.

Gurun, Umit G, Noah Stoffman, and Scott E Yonker, "Unlocking clients: The importance of relationships in the financial advisory industry," *Journal of Financial Economics*, 2021, *141* (3), 1218–1243.

Hortaçsu, Ali and Chad Syverson, "Product differentiation, search costs, and competition in the mutual fund industry: A case study of S&P 500 index funds," *The Quarterly journal of economics*, 2004, *119* (2), 403–456.

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

Kahraman, Bige, "Fee complexity and investor mistakes in retail financial markets," *Quarterly Journal of Finance*, 2021, *11* (01), 2150004.

Khomyn, Marta, Talis Putnins, and Marius Zoican, "The value of ETF liquidity," *The Review of Financial Studies*, 2024, *37* (10), 3092–3148.

Kostovetsky, Leonard and Jerold Warner, "Investor heterogeneity and the market for fund benchmarks: Evidence from passive ETFs," *Journal of Banking & Finance*, 2025, *173*, 107412.

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

Linnainmaa, Juhani T, Brian T Melzer, and Alessandro Previtero, "The misguided beliefs of financial advisors," *The Journal of Finance*, 2021, *76* (2), 587–621.

Mullainathan, Sendhil, Markus Noeth, and Antoinette Schoar, "The market for financial advice: An audit study," Technical Report, National Bureau of Economic Research 2012.

Pástor, L'uboš, Robert F Stambaugh, and Lucian A Taylor, "Do funds make more when they trade more?," *The Journal of Finance*, 2017, *72* (4), 1483–1528.

Robertson, Adriana Z, "Passive in name only: Delegated management and index investing," *Yale Journal on Regulation*, 2019, *36*, 795.

Roussanov, Nikolai, Hongxun Ruan, and Yanhao Wei, "Marketing mutual funds," *The Review of Financial Studies*, 2021, *34* (6), 3045–3094.

Sharpe, William F, "The arithmetic of investment expenses," *Financial Analysts Journal*, 2013, *69* (2), 34–41.

Sias, Richard, Harry J Turtle, and Blerina Zykaj, "Hedge fund crowds and mispricing," *Management Science*, 2016, *62* (3), 764–784.

Tirole, Jean, *The theory of industrial organization*, MIT press, 1988.

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

Figures

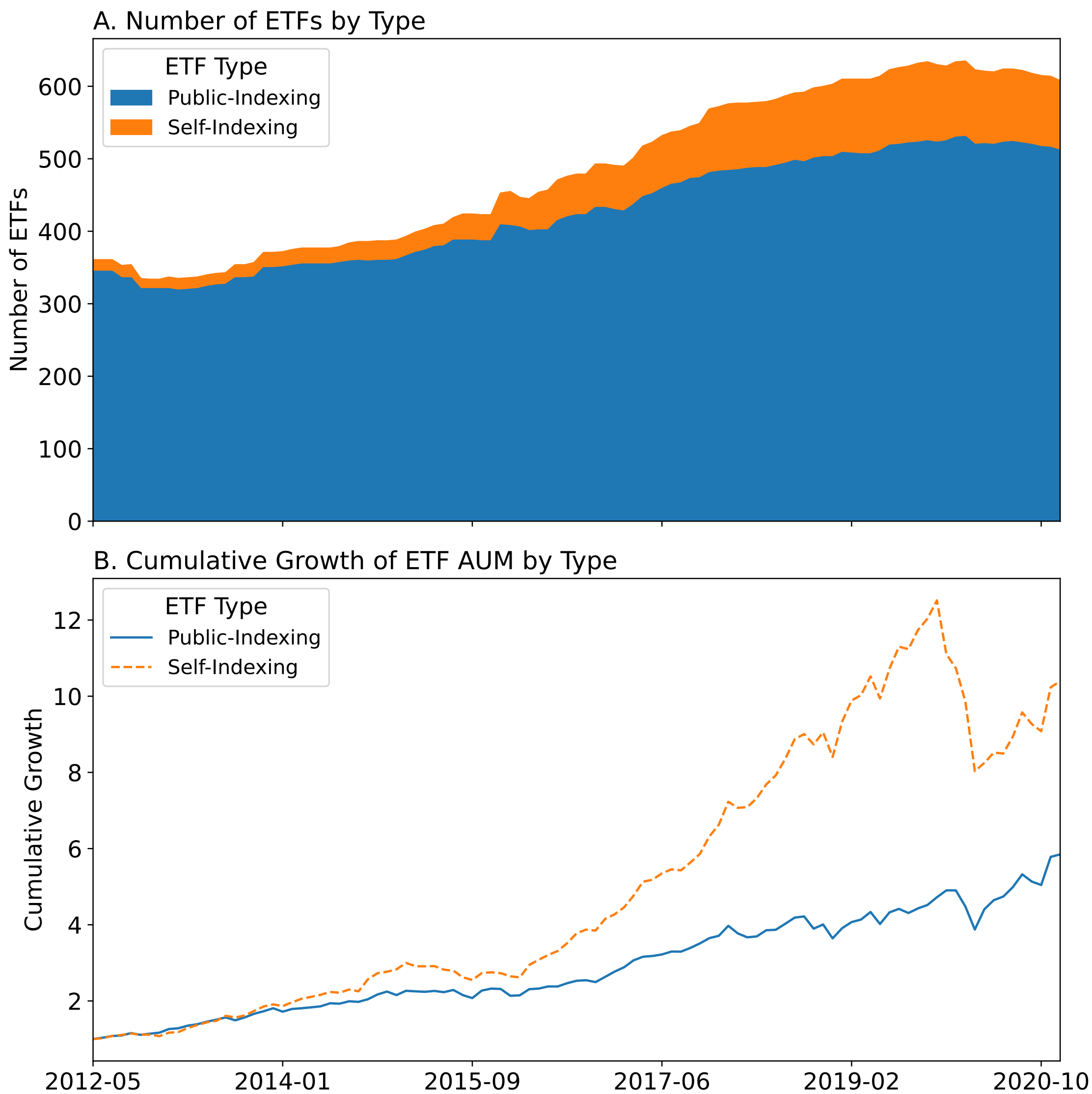


Figure 1. Number of self- and public-indexed ETFs over time.
Panel A presents the monthly number of self-indexed and public-indexed ETFs in our sample from May 2012 until December 2020. Panel B shows the cumulative growth of total ETF assets under management (AUM) for both types of ETFs over the same time frame (AUM is indexed to 1 in May 2012). Self-indexed ETFs are ETFs which track an index created and maintained by the ETF issuer themselves or one of its affiliated companies. Public-indexed ETFs are ETFs which track indices from unaffiliated index providers such as S&P Dow Jones or FTSE Russell.

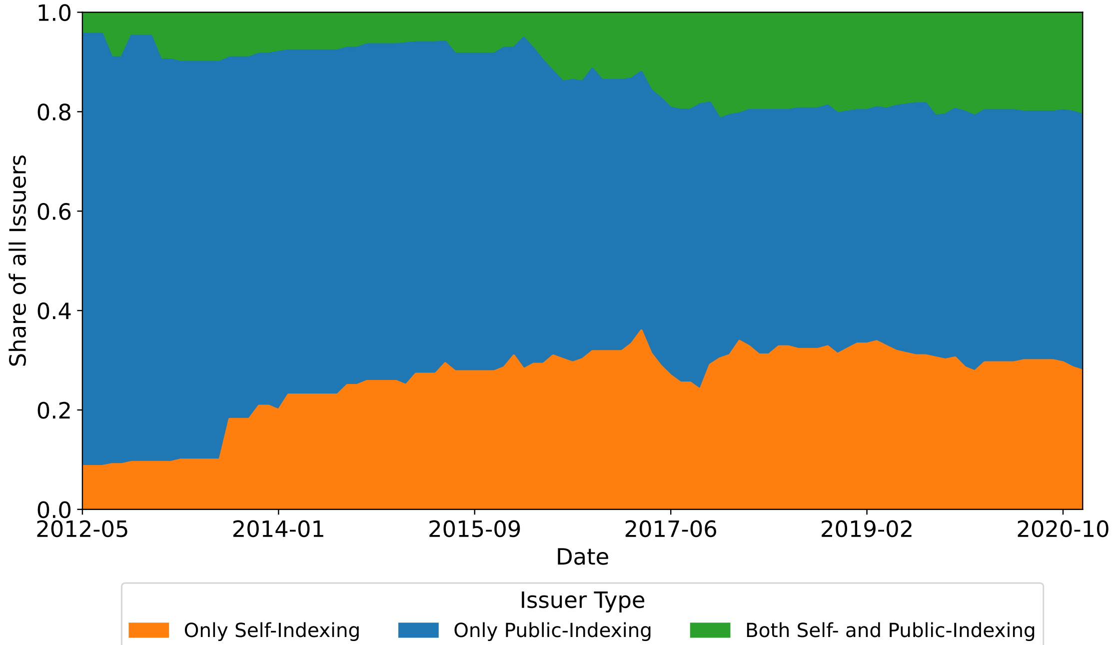


Figure 2. Share of issuer types over time.
This figure presents the monthly share of issuers in our sample from May 2012 until December 2020 which provide only self-index, only public-index, or both types of ETFs.

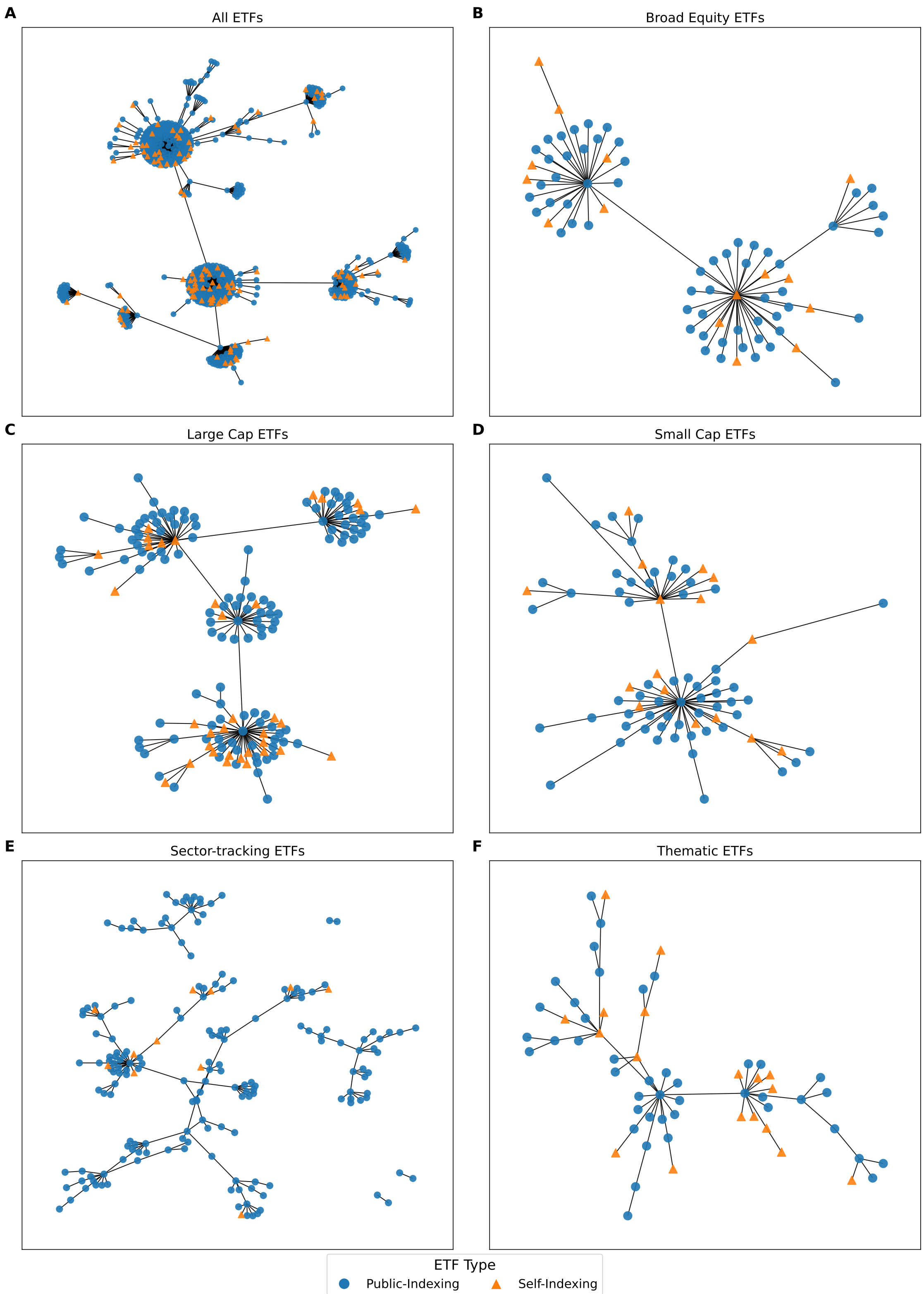


Figure 3. ETF networks based on cosine similarity of stock holdings.
This figure presents minimum spanning tree network plots of ETFs based on their monthly average pair-wise cosine similarity of stock holdings. The networks are plotted using the Fruchterman Reingold Layout. Panel A presents the network 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-Indexed ETFs are ETFs which track indices provided by the ETF issuer or an affiliated company. Public-Indexed 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	77	79.4	20	20.6	97
Buywrite	2	100.0	0	0.0	2
High Dividend Yield	27	69.2	12	30.8	39
Large Cap	165	79.3	43	20.7	208
Long/Short	12	85.7	2	14.3	14
Micro Cap	4	100.0	0	0.0	4
Mid Cap	43	87.8	6	12.2	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	74	79.6	19	20.4	93
Target Outcome	1	100.0	0	0.0	1
Theme	41	70.7	17	29.3	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. $Volatility$ 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. $Inst.SelfOwnership$ 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. Panel C reports the average return correlation with other ETFs in the same style group. We report the statistics for the full sample and the five largest style groups.

	Public-Indexing			Self-Indexing		
Variable	N	Mean	Std. Dev.	N	Mean	Std. Dev.
Panel A: Summary Statistics of Key Variables						
Net Expense Ratio (%)	44261	0.419	0.541	6173	0.428	0.265
Gross Return (%)	44111	1.05	5.44	6091	1.02	5.41
Volatility	44104	0.169	0.13	6086	0.164	0.136
Gross Carhart Alpha (%)	24942	-0.166	2.69	2314	-0.184	2.07
Log(AUM)	43876	19.4	2.42	6081	18.2	1.9
Log(Age)	44256	2	0.748	6173	1.37	0.693
Log(Holdings)	43414	4.88	1.4	6009	5.07	1.46
Fund Flow	43454	0.0198	0.119	5963	0.0319	0.124
Smart Beta?	38888	0.533	0.499	5381	0.775	0.418
Avg. Return Corr to Style-Peers	42797	0.702	0.218	5833	0.737	0.218
Inst. Self Ownership	15036	0.0183	0.0937	2083	0.14	0.283
# Affiliated Models	15050	0.695	4.01	2104	0.874	4
Panel B: Average Cosine Similarity of Holdings Compared to Style-Peers						
Full Sample	551	0.263	0.147	106	0.267	0.146
Large Cap	119	0.318	0.137	34	0.328	0.133
Mid Cap	40	0.267	0.109	5	0.288	0.037
Small Cap	58	0.257	0.112	15	0.231	0.0815
Broad Equity	61	0.249	0.14	11	0.246	0.107
Theme	40	0.201	0.13	16	0.139	0.0968
Panel C: Average Return Correlation with Style-Peers						
Full Sample	590	0.874	0.131	113	0.876	0.199
Large Cap	141	0.911	0.0497	35	0.919	0.0562
Mid Cap	39	0.937	0.024	6	0.876	0.189
Small Cap	62	0.891	0.166	14	0.925	0.039
Broad Equity	62	0.914	0.0489	15	0.902	0.062
Theme	36	0.881	0.0668	15	0.852	0.109

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. $Uniqueness$ 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 and Warner (2025). $Activeness$ is defined as $(1 - R^2)$ from regressing daily gross returns on the Carhart four-factor model. $TurnRatio$ 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)	(5)	(6)	(7)
$SelfIndexer?$	0.055***	0.052***	0.057***	0.056***	0.045***	0.052***	0.035***
	(0.006)	(0.008)	(0.011)	(0.009)	(0.008)	(0.008)	(0.013)
$Log(AUM)$		−0.023***	−0.024***	−0.025***	−0.022***	−0.023***	−0.027***
		(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
$Log(Holdings)$		−0.012***	0.000	0.005**	−0.010***	−0.012***	0.026***
		(0.001)	(0.002)	(0.002)	(0.001)	(0.001)	(0.004)
$Log(Age)$		0.070***	0.075***	0.078***	0.066***	0.070***	0.083***
		(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)
$SmartBeta?$		0.001	0.001	0.007	0.000	0.001	0.004
		(0.010)	(0.012)	(0.010)	(0.010)	(0.010)	(0.012)
$Uniqueness$			0.056***				0.046***
			(0.004)				(0.004)
$Activeness$				0.712***			0.796***
				(0.059)			(0.073)
$TurnRatio$					0.017***		0.038***
					(0.003)		(0.009)
$MAvgSpread$						−0.003	−0.005
						(0.003)	(0.005)
Num. Obs.	50 434	43 214	35 644	42 938	42 096	43 173	34 969
R^2	0.23	0.47	0.50	0.49	0.48	0.47	0.55
FE: Issuer		X	X	X	X	X	
FE: Month-Style	X	X	X	X	X	X	X
FE: Month-Issuer							X

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

Table 4. Explaining ETF differentiation.
In columns (1) - (2) we predict the monthly average portfolio cosine similarity of an ETF to its style-peers and in columns (3) - (4) we predict the average return correlation between an ETF and its style-peers within a month. $SelfIndexer?$ is a dummy variable that equals one if an ETF is self-indexed. $NetExp.Ratio(\%)$ is the ETFs annual expense ratio in percentage terms. $Log(AUM)$ 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.

	Cosine Similarity		Return Correlation	
	(1)	(2)	(3)	(4)
$SelfIndexer?$	0.019***	0.021***	0.015***	0.018***
	(0.002)	(0.002)	(0.004)	(0.004)
$NetExp.Ratio(\%)$	−0.007***	−0.006***	−0.013***	−0.011***
	(0.001)	(0.001)	(0.002)	(0.002)
$Log(AUM)$	0.001***	0.001***	−0.003***	−0.003***
	(0.000)	(0.000)	(0.000)	(0.000)
$Log(Holdings)$	0.028***	0.035***	0.020***	0.023***
	(0.001)	(0.001)	(0.001)	(0.001)
$Log(Age)$	−0.002	−0.001	0.015***	0.012***
	(0.001)	(0.002)	(0.002)	(0.002)
$SmartBeta?$	−0.008***	−0.006***	0.010***	0.010***
	(0.001)	(0.001)	(0.001)	(0.001)
Num. Obs.	29 902	29 902	41 887	41 887
R^2	0.76	0.78	0.83	0.86
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 5. Alphas from portfolio tests with monthly and daily returns.
Self-Indexers (-) Public-Indexers represents alphas from returns of the portfolio of self-indexed ETFs minus the returns of the portfolio of public-indexed ETFs. We report alphas from predictions of gross returns in excess of the risk-free rate. Gross returns are calculated by backing out net expense ratios from net (NAV) returns. 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 (-) Public-Indexers	-0.135	0.061	0.006	0.001
	(0.088)	(0.049)	(0.007)	(0.002)
Panel B: FF3				
Self-Indexers (-) Public-Indexers	-0.011	0.033	0.004	0.000
	(0.069)	(0.045)	(0.006)	(0.002)
Panel C: FF3 + Mom				
Self-Indexers (-) Public-Indexers	-0.012	0.033	0.004	0.001
	(0.069)	(0.045)	(0.006)	(0.002)

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

Table 6. Performance regressions.
The dependent variable is the monthly (daily) gross Carhart four-factor excess return, defined as the gross returns minus predicted returns using factor loadings from 36 month (126 day) 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. Standard errors are two-way clustered by time-issuer interaction and reported in parentheses.

	Gross Carhart Four-Factor Excess Returns (%)			
	Monthly Returns		Daily Returns	
	(1)	(2)	(3)	(4)
$SelfIndexer?$	−0.045	−0.042	−0.003	−0.003
	(0.088)	(0.099)	(0.003)	(0.003)
$Log(AUM)$	−0.012	−0.013	0.000	0.000
	(0.010)	(0.011)	(0.000)	(0.000)
$Log(Holdings)$	−0.020	−0.022	0.000	0.000
	(0.020)	(0.021)	(0.001)	(0.001)
$Log(Age)$	0.002	0.000	−0.001	−0.001
	(0.054)	(0.057)	(0.001)	(0.001)
$SmartBeta?$	−0.088**	−0.096***	−0.003**	−0.003**
	(0.034)	(0.035)	(0.001)	(0.001)
$Volatility$	−0.063	−0.544	−0.010	−0.025
	(0.653)	(0.675)	(0.029)	(0.029)
Num. Obs.	23 929	23 929	843 273	843 273
R^2	0.54	0.60	0.53	0.60
FE: Issuer	X		X	
FE: Month-Style	X	X		
FE: Month-Issuer		X		
FE: Day-Style			X	X
FE: Day-Issuer				X

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

Table 7. 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			
$SelfIndexer?$	-0.005***	-0.007***	-0.008***
	(0.001)	(0.001)	(0.001)
$Log(AUM)$		-0.003***	-0.003***
		(0.000)	(0.000)
$Log(Holdings)$		-0.002***	-0.003***
		(0.000)	(0.000)
$Log(Age)$		0.007***	0.006***
		(0.001)	(0.001)
$SmartBeta?$		-0.006***	-0.006***
		(0.001)	(0.001)
Num. Obs.	50190	43074	43074
R^2	0.90	0.92	0.93
Panel B: Sharpe Ratio			
$SelfIndexer?$	-3.266	-7.239*	-9.197*
	(2.386)	(3.989)	(5.490)
$Log(AUM)$		-0.455	-0.589
		(0.480)	(0.511)
$Log(Holdings)$		2.180	3.031
		(2.040)	(2.724)
$Log(Age)$		0.068	0.420
		(1.136)	(1.126)
$SmartBeta?$		1.166	1.416
		(1.526)	(1.707)
Num. Obs.	50189	43073	43073
R^2	0.14	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 8. 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. In column (4) we predict 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. $InvAdvisor?$ 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. $AvgNetRet(Yr)(\%)$ is the 12-month rolling average net (NAV) ETF return. Standard errors are two-way clustered by month-issuer interaction and reported in parentheses.

	Net Expense Ratio (%)	Gross Return (%)	Cosine Similarity	Fund Flow (t+1)
	(1)	(2)	(3)	(4)
$SelfIndexer?$	0.002	0.136	−0.020***	0.011**
	(0.013)	(0.086)	(0.002)	(0.005)
$InvAdvisor?$	−0.320***	−0.006	0.015***	0.005**
	(0.011)	(0.042)	(0.002)	(0.002)
$Log(AUM)$	−0.024***	0.018*	0.004***	−0.002***
	(0.001)	(0.009)	(0.000)	(0.000)
$Log(Holdings)$	−0.045***	0.001	0.032***	0.001
	(0.003)	(0.014)	(0.001)	(0.001)
$Log(Age)$	0.061***	−0.022	−0.010***	−0.016***
	(0.004)	(0.025)	(0.001)	(0.002)
$SmartBeta?$	0.049***	−0.050**	−0.011***	0.000
	(0.008)	(0.025)	(0.001)	(0.002)
$SelfIndexer?{:}InvAdvisor?$	0.090***	−0.120	0.013***	−0.019***
	(0.016)	(0.093)	(0.003)	(0.006)
$Volatility$		1.141		−0.003
		(0.735)		(0.017)
$AvgNetRet(Yr)(\%)$				0.017***
				(0.001)
Num. Obs.	43 214	43 074	29 902	35 955
R^2	0.21	0.88	0.70	0.17
FE: Month-Style	X	X	X	X

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

Table 9. Sentiment-flow sensitivity.
In all specifications we predict 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. $InvAdvisor?$ 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. $Log(MichSent)$ is the natural logarithm of the University of Michigan Consumer Sentiment Index, $MichSentNorm$ is its normalized value and $MichSentTophalf$ a dummy variable that equals to one when the index is above its median value. Standard errors are two-way clustered by month-issuer interaction and reported in parentheses.

	Fund Flow (t+1)		
	(1)	(2)	(3)
$SelfIndexer?$	−0.335***	−0.010**	−0.015***
	(0.087)	(0.004)	(0.005)
$AvgNetRet(Yr)(\%)$	0.011***	0.011***	0.011***
	(0.001)	(0.001)	(0.001)
$Log(AUM)$	−0.003***	−0.003***	−0.003***
	(0.001)	(0.001)	(0.001)
$Log(Holdings)$	−0.001	−0.001	−0.001
	(0.001)	(0.001)	(0.001)
$Log(Age)$	−0.020***	−0.020***	−0.020***
	(0.002)	(0.002)	(0.002)
$SmartBeta?$	0.001	0.001	0.001
	(0.002)	(0.002)	(0.002)
$Volatility$	0.039***	0.038***	0.037***
	(0.007)	(0.007)	(0.006)
$Log(MichSent)$	0.009		
	(0.012)		
$SelfIndexer?$:$Log(MichSent)$	0.073***		
	(0.019)		
$MichSentNorm$		0.001	
		(0.002)	
$SelfIndexer?$:$MichSentNorm$		0.011***	
		(0.003)	
$MichSentTophalf$			0.002
			(0.003)
$SelfIndexer?$:$MichSentTophalf$			0.014***
			(0.005)
Num. Obs.	35 955	35 955	35 955
R^2	0.04	0.04	0.04
FE: Issuer	X	X	X
FE: Style	X	X	X

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

Table 10. Explaining institutional self ownership.
The dependent variable is Institutional Self Ownership, defined as 13f ownership by the same or related investment management companies. $SelfIndexer?$ 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.

	Institutional Self Ownership	
	(1)	(2)
$SelfIndexer?$	0.108***	0.035**
	(0.017)	(0.015)
$Log(AUM)$	0.008***	0.008***
	(0.001)	(0.001)
$Log(Holdings)$	0.011***	0.009***
	(0.003)	(0.002)
$Log(Age)$	−0.052***	−0.053***
	(0.006)	(0.006)
$SmartBeta?$	0.033***	0.029***
	(0.005)	(0.005)
$AvgNetRet(Yr)(\%)$	−0.001	−0.001
	(0.002)	(0.002)
$InvAdvisor?$		0.000
		(0.004)
$SelfIndexer?$:$InvAdvisor?$		0.095***
		(0.025)
Num. Obs.	12 307	12 307
R^2	0.18	0.19
FE: Quarter-Style	X	X

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

Table 11. 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)
$SelfIndexer?$	−0.083	0.788*	1.171**	−1.190***
	(0.184)	(0.454)	(0.513)	(0.164)
$NetExp.Ratio(\%)$	−0.018	−0.021	−0.029	0.088***
	(0.032)	(0.042)	(0.041)	(0.029)
$Log(AUM)$	0.540***	0.737***	0.769***	0.530***
	(0.063)	(0.091)	(0.094)	(0.062)
$Log(Age)$	−0.788***	−2.036***	−2.156***	−0.853***
	(0.175)	(0.315)	(0.330)	(0.186)
$SmartBeta?$	0.367***	−0.442***	−0.425***	0.336***
	(0.080)	(0.089)	(0.090)	(0.082)
$AvgNetRet(Yr)(\%)$	−0.091	−0.084	−0.043	−0.088
	(0.082)	(0.080)	(0.085)	(0.082)
$InvAdvisor?$				0.339***
				(0.130)
$SelfIndexer?{:}InvAdvisor?$				1.351***
				(0.258)
Num. Obs.	12 385	12 385	12 385	12 385
R^2	0.10	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.
$AvgNetRet(Yr)(\%)$	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'.
$Inst.SelfOwnership$	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.
$Activeness$	$(1 - R^2)$ from regressing daily gross returns on the Carhart four-factor model.
$TurnoverRatio$	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.
$Log(MichSent)$, $MichSentNorm$, $MichSentTophalf$	The natural logarithm, normalized value and dummy for when the the University of Michigan Consumer Sentiment Index is above its median value.
$InvestmentAdvisor?$	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-indexed ETFs operated by an issuer. Self-indexed ETFs are defined using data from ETF Global (see above).

Internet Appendix to

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

(Not for Publication)

Figures

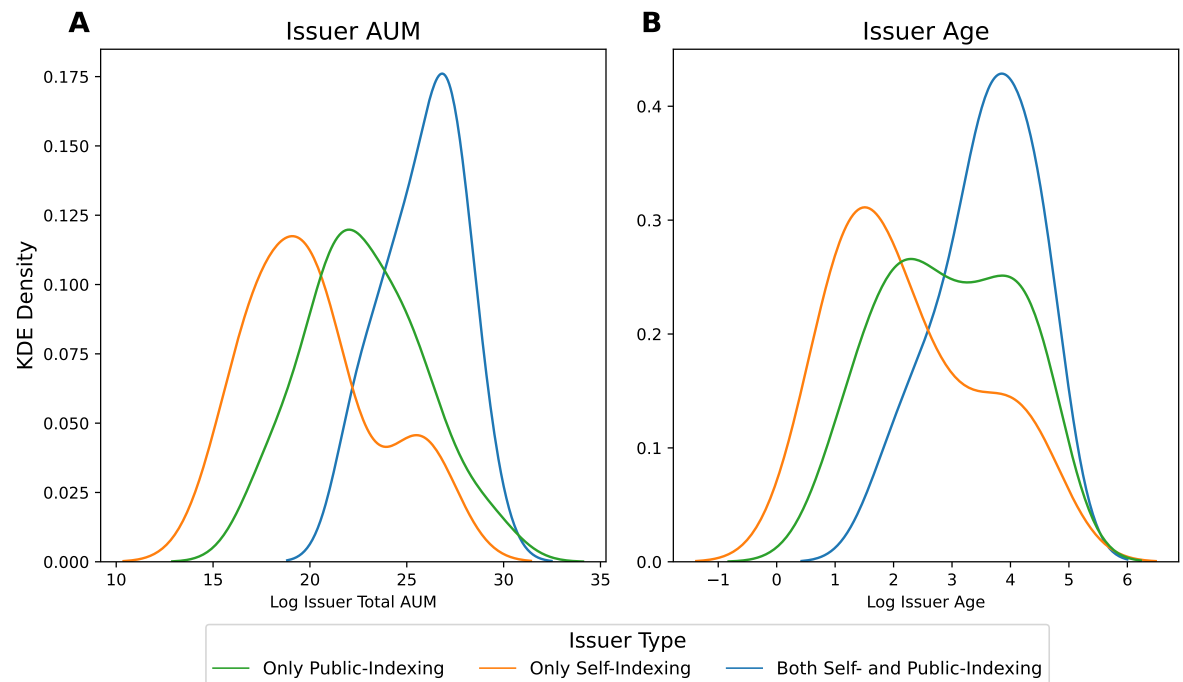


Figure IA.1. Density plots of issuer characteristics by issuer type.
This figure presents KDE density plots for key issuer characteristics, grouped by issuer type, as of December 2020. Panel A plots the densities for the logarithm of total issuer AUM, calculated as the total AUM of all outstanding funds 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.

Figure IA.2. ETF networks based on pairwise ETF return correlation.
This figure presents minimum spanning tree network plots of ETFs based on their monthly average pairwise return correlations. 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 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				
WisdomTree	15588.73	13	1006.86	1
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
Panel C: Only Public-Indexing Issuers				
Blackrock	-	-	961544.19	93
Vanguard	-	-	866575.55	43
First Trust	-	-	55674.09	54
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
Franklin Templeton Investments	-	-	1520.72	3

Table IA.2. Explaining the issuance of self-indexed ETFs at the issuer-month level.
The dependent variable is a dummy which equals one if the issuer increases the amount of outstanding self-indexed 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-indexed 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)	0.042			−1.918		
	(0.051)			(2.457)		
$IssuerLog(Age)$	0.004	0.004	0.004	0.312	0.237	−1.749
	(0.007)	(0.006)	(0.011)	(0.357)	(0.343)	(2.538)
$IssuerLog(AUM)$	−0.002	−0.002	−0.012*	−0.197	−0.157	−1.020
	(0.003)	(0.003)	(0.007)	(0.155)	(0.116)	(1.068)
$IssuerNrETFs$	0.000	0.000	0.001	0.020	0.019	0.047
	(0.000)	(0.000)	(0.001)	(0.014)	(0.014)	(0.042)
$IssuerNrSelfIndexETFs$	0.007***	0.007***	0.010	0.264***	0.266***	0.760
	(0.002)	(0.002)	(0.006)	(0.052)	(0.048)	(0.757)
Num. Obs.	1521	1521	1521	1521	705	286
R^2	0.03	0.05	0.12	0.13	0.17	0.16
FE: Month		X			X	
FE: Issuer			X			X

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

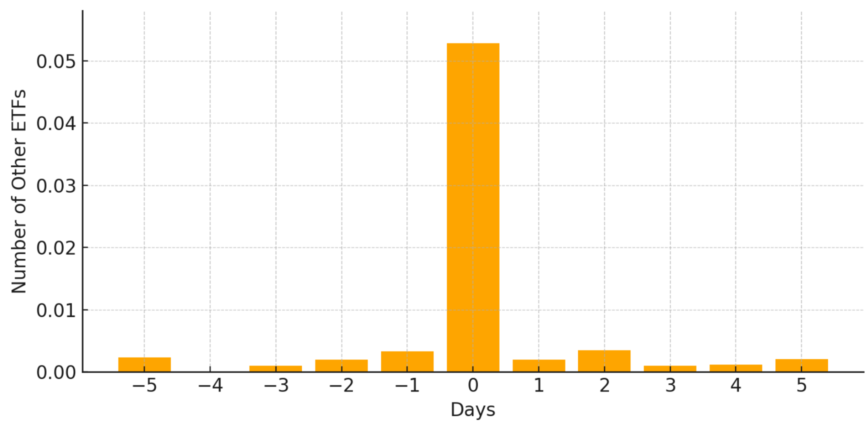
Table IA.3. Robustness Tests.
In column (1) we predict the monthly average portfolio cosine similarity of an ETF to its style-peers and in column (2) we predict the average return correlation between an ETF and its style-peers within a month. In columns (3) - (4) the dependent variable is the monthly (daily) gross Carhart four-factor excess return, defined as the gross returns minus predicted returns using factor loadings from 36 month (126 day) rolling windows $SelfIndexer?$ is a dummy variable that equals one if an ETF is self-indexed. $NetExp.Ratio(\%)$ is the ETFs annual expense ratio in percentage terms. $Log(AUM)$ 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 and Warner (2025). $TurnRatio$ is an ETFs yearly portfolio turnover ratio. $Activeness$ is defined as $(1 - R^2)$ from regressing daily gross returns on the Carhart four-factor model. $MAvgSpread$ is the monthly average ETF secondary market price spread scaled by price. $Volatility$ 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.

	Cosine Similarity	Return Correlation	Monthly Excess Ret	Daily Excess Ret
	(1)	(2)	(3)	(4)
$SelfIndexer?$	0.018***	0.028***	−0.138	−0.010**
	(0.003)	(0.006)	(0.138)	(0.005)
$NetExp.Ratio(\%)$	0.001*	0.001	−0.022	0.001
	(0.001)	(0.002)	(0.076)	(0.003)
$Log(AUM)$	0.001**	−0.004***	0.002	0.000
	(0.000)	(0.000)	(0.014)	(0.000)
$Log(Holdings)$	0.026***	0.013***	−0.048*	−0.001
	(0.001)	(0.001)	(0.028)	(0.001)
$Log(Age)$	−0.009***	0.011***	−0.030	−0.001
	(0.002)	(0.002)	(0.068)	(0.002)
$SmartBeta?$	−0.006***	0.004***	−0.076*	−0.003**
	(0.002)	(0.001)	(0.040)	(0.001)
$Uniqueness$	−0.019***	−0.018***	−0.003	0.000
	(0.001)	(0.001)	(0.050)	(0.002)
$TurnRatio$	−0.001***	0.001	−0.056	−0.001
	(0.000)	(0.001)	(0.041)	(0.001)
$Activeness$	−0.036***	−0.233***	−0.649*	−0.014
	(0.005)	(0.011)	(0.366)	(0.014)
$MAvgSpread$	0.002	0.005*	0.543*	0.004**
	(0.003)	(0.003)	(0.330)	(0.002)
$Volatility$			−0.900	−0.032
			(0.758)	(0.034)
Num. Obs.	22 943	34 181	20 013	693 988
R^2	0.83	0.88	0.62	0.62
FE: Month-Style	X	X	X	
FE: Month-Issuer	X	X	X	
FE: Day-Style				X
FE: Day-Issuer				X

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

Table IA.4. Performance of Active Funds.
Panel A plots the trading behavior of public-indexed ETFs around same-family self-indexed ETF trades. For each self-indexed ETF trade of a stock ticker i on day T, we count the number of same-issuer public-indexed ETFs trading the same ticker in the same direction in [T-5,T+5]. The figure plots the average count. Panel B reports performance regressions for active funds. In Panel B, the dependent variable is monthly net (NAV) return or Net Carhart Alpha using factor loadings from a 36-month rolling window. $IssuerHasSelfIndexer?$ is a dummy variable that equals one if an issuer offers a self-indexed ETF. $Log(AUM)$ is the natural logarithm of a fund's assets under management. $TurnRatio$ is the yearly portfolio turnover ratio. $Log(Age)$ is the natural logarithm of a fund's age (in years). $Volatility$ is the standard deviation of daily net returns within a month. $IssuerLog(AUM)$ is the natural logarithm of the issuer's assets under management. The sample includes all active equity funds in the CRSP MF database from 2012–2020. Standard errors are two-way clustered by month–issuer interaction.

Panel A: Within-Issuer Front-running Test



Panel B: Performance Regressions

	Net Returns (%)			Net Carhart Alpha (%)		
	(1)	(2)	(3)	(4)	(5)	(6)
$IssuerHasSelfIndexer?$	−0.061**	−0.041**	−0.017	−0.072***	−0.029*	−0.035
	(0.027)	(0.017)	(0.021)	(0.023)	(0.016)	(0.023)
$Log(AUM)$	0.016***	0.013***	0.012***	0.010***	0.011***	0.007***
	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)
$TurnRatio$	−0.004*	−0.004**	0.000	−0.004	−0.004	−0.006*
	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)
$Log(Age)$	−0.045***	−0.035***	−0.034***	−0.024**	−0.027***	−0.034***
	(0.006)	(0.005)	(0.005)	(0.009)	(0.007)	(0.008)
$Volatility$	3.420***	2.501***	2.360***	−0.673***	−0.530**	−0.417*
	(0.260)	(0.242)	(0.250)	(0.230)	(0.264)	(0.245)
$IssuerLog(AUM)$	0.014***	0.014***	−0.002	0.003	0.004**	−0.018**
	(0.002)	(0.001)	(0.006)	(0.002)	(0.002)	(0.007)
Num. Obs.	493 044	492 988	492 986	286 151	286 100	286 096
R^2	0.74	0.87	0.87	0.12	0.57	0.58
FE: Month	X			X		
FE: Issuer			X			X
FE: Style	X			X		
FE: Month-Style		X	X		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	-