The Passive Paradox: Why High-Indexed Stocks Outperform Despite Lower Expected Returns *

Pouya Behmaram†

February 23, 2026

Abstract

Equity markets are now largely dominated by passive investments, this study introduces the *Indexing Inclusion Ratio* (IXI) as a measure of passive ownership to assess its increasing impact on U.S. equity markets. The findings reveal that high-indexed stocks highly outperform their low-indexed counterparts, primarily due to the influx of passive capital flows rather than fundamental value. By analyzing the expected return both ex-ante through implied costs of capital and ex-post by adjusting for passive flows and earnings anomalies, I show that high-indexed stocks exhibit lower expected returns, indicating a potential correction when the trend toward passive investing reaches equilibrium. In addition, I show that the recent underperformance of value and small-cap stocks is intrinsically related to the secular passive shift.

Keywords: Asset pricing, institutional investors, index effect, passive investing, mutual funds, ETFs
JEL Codes: G11, G12, G23

*I am grateful to my supervisors Laurent Barras and David Schumacher for their invaluable guidance and support. I acknowledge helpful comments from Daniel Andrei, Vincent Grégoire, Thomas Rivera, Gregory Weitzner, Begum Ipek Yavuz, and seminar participants at McGill University. All remaining errors are my own. I am thankful for the research support from the Social Sciences and Humanities Research Council of Canada.

†University of Quebec at Montreal, École des sciences de la gestion (behmaram.pouya@uqam.ca)

1 Introduction

Over the past decade, the U.S. equity markets have undergone a transformative shift toward passive investment strategies, with passive equity funds now accounting for 50% of all equity fund investments as of 2019[1]. Although the merits of passive investing, such as lower costs and historically better returns, have been well established, there remains a gap in the literature examining the nuanced repercussions of this seismic shift. This paper aims to fill this void by addressing pivotal questions about the long-term implications of this transition for market equilibrium prices and stock performance, especially for stocks heavily represented in popular indices. Specifically, I show that the rising trend in passive investing could be artificially boosting the prices of highly indexed stocks, subsequently reducing their future expected returns. Given the growing appeal of passive investment products, this trend poses a risk to the long-term returns of retirement funds and pensions. In addition, I explore the potential implications of this shift on the recent underperformance of traditional valuation factors, including size and value.

What can the historical performance of high-indexed stocks reveal about their future prospects? Informed by a model from Kashyap et al. (2021), this paper examines two core hypotheses concerning high-indexed stocks: first, that their prices rise with increasing passive ownership; and second, that they are likely to yield lower expected returns. However, these expectations are confounded by two opposing channels. Firstly, the continued shift toward passive elevates the demand for high-indexed stocks, which, due to their less elastic demand, reap disproportionate price benefits over their fundamentals, thus leading to lower expected returns that incorporate both demand and risk premiums. Secondly, increased passive ownership lowers the price elasticity of the stock, thereby incentivizing firms to undertake riskier growth projects, as highlighted by Buss and Sundaresan (2020), which in turn increases profits and future expected returns. To empirically validate these hypotheses in the presence of opposing channels, I use methods described by Pástor et al. (2022), adapted

[1] Source: Morningstar

to the dynamics of passive and active investing. My results indicate that these competing channels result in negative expected returns overall; however, ongoing inflows into passive investment postpone the realization of these lower expected returns.

My study focuses on the U.S. stock market. I construct a precise measure of passive ownership for each stock that I call *Indexing Inclusion ratio* (IXI) which is not only limited to index mutual funds and ETFs but also includes the so-called closet indexers which are a group of active funds that closely follow their respective benchmark weights. Unlike similar measures such as the Benchmarking Intensity (BMI) proposed by **?**, which relies on fund return and index weight data and assumes that all benchmarked funds strictly adhere to their indices, IXI takes a more nuanced approach. It utilizes actual fund-holding data to gauge passive ownership and accommodates the evolving "active share" of each fund, acknowledging that not all funds fully mimic their benchmark weights. This is particularly important for capturing the direct demand funds exert on stocks through their holdings, offering a more accurate and textured understanding of how active is increasingly becoming more passive over time. Moreover, IXI offers broader coverage than BMI, drawing from 41,200 global funds across 460 indices compared to BMI which only focuses on U.S. equity funds across 38 indices. This is especially relevant given the rise of foreign fund ownership in U.S. stocks [2].

My data sample spans from 2000 to December 2021. Throughout this period, the cap-weighted portfolio of high-indexed stocks consistently outperformed low-indexed stocks, achieving an average annual return of 4%. I label this return difference as HILI (high-indexed minus low-indexed). In particular, HILI boasts an annualized Sharpe ratio of 0.59, surpassing the stock market's 0.34 ratio over the same period. The recent strong performance of high-indexed stocks may seem to suggest promising future returns for passive investments. However, my findings indicate a different story. The notable outperformance of HILI is largely driven by the unexpected passive flows, rather than what the fundamental risk-return profile of these stocks would suggest. My examination of passive capital flows

[2] Source: Foreign Portfolio Holdings of U.S. Securities as of June 30, 2022, Federal Reserve Bank of New York

reveals a significant increase in passive ownership and a steady shift of capital from active to passive investments over the last ten years. In particular, the IXI measure that monitors this trend has nearly doubled. I have observed that sudden fluctuations in IXI, especially when new indices are added, correlate positively with HILI. As a result, as passive ownership grows, high-indexed stocks continually surpass their low-indexed counterparts.

I introduce a key metric called the "indexing premium," which quantifies the divergence in expected returns between stocks with high and low degrees of passive ownership. I estimate these expected returns using a dual-method approach: first, by utilizing ex-ante data through the Implied Cost of Capital (ICC) measure; and second, by employing ex-post data that filters out earnings-related news and market flow shocks. My findings show a predominantly negative indexing premium for most of the sample period. This suggests that there are two contrasting mechanisms at work. On the one hand, increased passive ownership can fundamentally enhance a company's value by inducing higher investments in riskier growth projects, thereby potentially elevating its future returns. However, a surge in inelastic passive demand can often inflate stock prices without corresponding fundamental improvements, thereby depressing expected future returns. My results indicate that the latter effect outweighs the former, leading to an overall negative impact on expected returns. To refine my understanding, I specifically use the second method of measuring the expected return by deriving an ex post estimate of the expected return of HILI. When I control for both the influence of passive investment inflows and unexpected earnings shocks, the counterfactual performance of HILI turns mostly negative. This implies that stocks with high levels of passive ownership could underperform in the long term, particularly if the increasing passive investment flows were to wane.

My primary findings that tie passive ownership shocks to stock returns hinge on the time series of sorted portfolios of stocks based on their degree of passive ownership. I further bolster this analysis by implementing panel regressions on individual stocks, which yield several insights. Firstly, a stock's return displays a positive cross-sectional relationship with its IXI score. However, this relation diminishes when I include a stock-level measure of change in passive flow. This indicates that shocks to a stock's passive flow contribute to the exceptional performance of highly indexed stocks throughout the period in question. Time-series data mirrors these outcomes, showcasing that even with lower expected returns, high-indexed stocks outperform low-indexed stocks due to new index additions and large capital shifts from active to passive investments.

This analytical interpretation of the superior return of high-indexed stocks mirrors the insights of the model from Pástor et al. (2022). In times when both passive ownership and passive investment flows amplify unexpectedly, HILI registers positive returns, driven by the increasing investor preference for passive instruments and, in turn, high-indexed stocks. Outperformance caused by the strengthening of investor passive demand is followed by a lower expected performance of HILI going forward as the shift from active to passive loses speed and finally reaches an equilibrium where the surviving active opportunities increase. That is, a shift in HILI's expected future performance relates inversely to its realized performance.

The pronounced outperformance of high-indexed stocks in recent times provides clarity on the subpar returns of value and small-cap stocks during the 2010s. This decade was particularly challenging for the HML and SMB factors of Fama and French (1993). I explore the modified framework of Pástor et al. (2022), which suggests that asset prices are driven by a two-factor model consisting of the market portfolio and the HILI factor. The HILI factor captures the returns from a strategy that favors high-indexed stocks and disfavors low-indexed ones. The allocation of individual stocks in this strategy is based on their respective IXI scores. My findings indicate that the two-factor model can explain part of the decline in SMB and HML during the specified timeframe. Specifically, from November 2012 to December 2020, HML's monthly CAPM alpha was -60 bps, a value that is significant. However, when evaluated under the two-factor model, HML's alpha stands at an insignificant -30 bps. In parallel observation, the inclusion of the HILI factor reduces the SMB alpha by more than 90%. It is also important to note the inverse correlation between the HILI factor and both HML and SMB. This is mainly because value and small-cap stocks are less frequently included in prominent indices.

Numerous studies, including those of French (2008) and Ben-David et al. (2017), have highlighted the benefits of passive investment strategies, and experts such as Sharpe (1991) have highlighted their potential advantages over active approaches. Bond and Garcia (2022) further discussed the growing trend towards indexing as participation costs decrease. However, there seems to be less exploration of the long-term implications of this move toward passive investing, particularly for high-indexed stocks. My research aims to fill this gap, suggesting that the recent performance of high-indexed stocks might not necessarily predict their future.

Passive funds, particularly ETFs, are now the largest group of institutional investors. Studies such as Gompers and Metrick (2001) have shown that institutional demand for certain stock characteristics (e.g., high liquidity, large-cap) impacts stock prices and returns. Furthermore, Edelen et al. (2016) revealed that the positive influence of institutional demand on future returns could be transient. Building on this, my study distinguishes between realized and expected returns, using the best-performing return proxies such as the implied cost of capital, as evidenced by Lee et al. (2021). Notably, past studies have not tackled the sheer magnitude of the modern passive sector, which, when combined with its rigid demand, could magnify its price influence on stocks.

Recently, Buss and Sundaresan (2020) argued that passive ownership can boost informational efficiency. This fosters riskier growth moves by companies with high passive ownership, sparking active investors to gather more data, which in turn heightens price information and in response lowers the cost of capital for the firm. My findings provide an empirical test for this theoretical model showing that despite the fact that this mechanism fundamentally boosts the value of the firm, the overall effect on the expected returns remains negative. Furthermore, I argue that their current advantages, influenced by passive flows, might diminish as the shift from active to passive peters out.

Historically, the literature on the index effect has provided insights into the relationship between institutional demand and asset prices. The seminal works by Harris and Gurel (1986) and Shleifer (1986) presented arguments about the nature of demand curves for stock prices. Price pressure and imperfect substitute hypotheses offered contrast perspectives, suggesting variations in stock demand (Scholes (1972), Chang et al. (2015)). Although Chang et al. (2015) used a regression discontinuity technique to examine the effect of index inclusion on the stock price, the overarching influence of passive flows was somewhat overlooked. Furthermore, as passive investing becomes more dominant, the short-term demand pressure from new index inclusions becomes less important compared to the consistent growth in the size of popular index providers. **?** moved the discourse forward with the concept of *benchmarking intensity* (BMI). However, their focus remains narrowly tailored to the small threshold of Russell 1000/2000 index reconstitution, specifically finding that an increase in a stock's BMI can be linked to underperformance over a 1-5 year period. This work relies on historical returns within a 1-5 year time frame as a simple proxy for expected returns, which may not accurately capture the influence of the persistent and growing passive flows. In contrast, my study argues that stocks with higher passive ownership have shown consistent outperformance. Furthermore, when the scope is expanded to include stocks in all major indices like the S&P 500, the underperformance suggested by **?** is compensated. This occurs because of long-term secular growth in passive investing, which effectively postpones the realization of lower expected returns until an equilibrium in passive investment is reached sometime into the future.

The inverse relationship between realized returns and shifts in expected returns is a recognized concept in the stock return literature. Although this inverse relationship is clear as discussed by Fama and French (2002) and Cochrane (2008), estimating stock expected returns presents inherent complexities, whereas the ex-ante and ex-post measures of expected returns that are used in this study allow for a more parsimonious and transparent manifestation of this relationship despite the existence of two opposing channels affecting the expected returns.

Finally, my research contributes to the discussion of factor investing and expected returns. Although studies such as Asness et al. (2018), Harvey and Liu (2016), and Hou et al. (2020) document the underperformance of previously known investing factors such as size and value over the past decade, I shed light on a potential cause. In particular, as growth and large-cap stocks dominate passive products, the dip in performance of size and value factors appears to align with the growing prominence of passive investing. My analysis reveals that indexing could be responsible for almost all the observed underperformance in small-cap stocks and contributes to approximately 50% of the decline in value stocks over the past decade.

The paper is organized as follows. In Section 2, I draw some testable hypotheses about the realized and expected returns of the indexed stocks using a simple adapted model. Next, in 3, I introduce the concept of the indexing inclusion measure and provide a detailed overview of the data at hand. Proceeding to Section 5, I explore the realized returns of indexed stocks, using portfolios sorted by the IXI measure. In Section 6, the emphasis moves to assessing expected returns. In this section, I utilize two different approaches: one based on ex-ante data and the other on ex-post data, demonstrating how the expected return becomes negative when considering the effects of IXI and passive flow. In section 7 I construct a stock-level metric for passive flow, offering insights that reinforce the observations made at the portfolio level and shedding light on how the passive shift might underpin the underperformance seen in value and size factors. The paper concludes with Section 8, where I draw together and discuss the overarching themes and findings.

2 Model of Asset Prices with Active and Passive Investors

To show how asset prices and expected returns behave in the presence of a substantial group of indexed investors, I use Kashyap et al. (2021) simple model of delegated asset management with heterogeneous benchmarks that I modify to also include passive managers. In this model, instead of mostly focusing on the heterogeneous fund managers whose performance is compared to an index, I focus more on the passive and closet indexer investors. The primary function of the model is to provide some testable hypotheses about the relationship between indexing inclusion and realized and expected stock returns.

2.1 Model

Consider a typical setting. There are two time periods, $t = 0, 1$. The available investment options include a risk-free bond with an interest rate of zero, which is normalized to zero, and N risky assets indexed by $n = 1, \ldots, N$ that entitle the holder to cash flows $D(n)$ realized in period 1. The vector representation for the period zero risky asset prices is denoted by $S = (S(1), \ldots, S(N))'$ and similarly the vector notation for period one cash flows are denoted by $D = (D(1), \ldots, D(N))'$. In period one, the asset prices are equal to cash flows D. The cash flows D are jointly normally distributed, $D \sim N(\mu, \Sigma)$, where $\mu = (\mu(1), \ldots, \mu(N))'$, $\Sigma_{nn} = \operatorname{Var}(D(n)) = \sigma_n^2$, and $\Sigma_{nm} = \operatorname{Cov}(D(n), D(m)) = \rho_{mn}\sigma_n\sigma_m$. I assume that the matrix Σ is invertible. The risky assets have a fixed total number of shares in supply denoted by $\bar{x} \equiv (\bar{x}(1), \ldots, \bar{x}(N))'$.

All benchmark indices are indexed by $i = 1, \ldots, I$ where each benchmark index i consists of a portfolio of N stocks with shares denoted by $\omega_i = (\omega_i(1), \ldots, \omega_i(N))'$, depending on each index constituents the ω_i can take a value of zero.

There are three agents in this economy. Direct investors are similar to retail investors and manage their own money, which takes up λ_D fraction of the population. Active fund managers are hired by fund investors to manage portfolios of stocks and are evaluated against a benchmark index. The faction of the population of all active managers that are evaluated against a specific benchmark i is indicated by λ_i^A. Similarly, passive fund managers are hired by fund investors to exactly follow their respective index and their population fraction for each index i is denoted by λ_i^P. All agents have a constant absolute risk aversion (CARA) utility function over terminal wealth W, $U(W) = -e^{-\gamma W}$, where $\gamma > 0$ is the coefficient of absolute risk aversion. The direct investor's terminal wealth is $W = W_0 + x'_D(D - S)$, where W_0 is the initial wealth of the investor, and the vector x_D denotes the number of shares held by the direct investor. The compensation of active fund managers for the funds that are benchmarked to index i is denoted by w_i and depends on three components: one is a linear payment based on the absolute performance of the fund, the second component depends on the performance of the fund relative to benchmark i, and the third part is independent of performance.

$$w_i = aR_i + b\left(R_i - B_i\right) + c \tag{1}$$

where $R_i = x'_i(D - S)$ is the fund's portfolio performance and $B_i = \omega'_i(D - S)$ is the performance of index i. The measures a and b determine the compensation sensitivities to absolute and relative performance. The fund manager chooses a portfolio of x_i shares to maximize his utility $U\left(w_i\right)$. A passive fund manager is obligated to hold the same quantity of shares as their designated index. Consequently, they have a notable high value of b, where $b \to \infty$. This strongly incentivizes them to exclusively maintain the benchmark index portfolio, with substantial penalties for any deviations from it. Following a standard mean-variance portfolio optimization, the portfolio demand for direct investors is:

$$x_D = \frac{\mu - S}{\gamma}\Sigma^{-1} \tag{2}$$

Active fund managers have different demand due to their compensation contract, which is given by:

$$x_i^A = \frac{1}{\gamma(a+b)}(\mu - S)\Sigma^{-1} + \frac{b}{a+b}\omega_i. \tag{3}$$

The active fund manager divides their risky asset allocations between two distinct portfolios: the mean-variance portfolio (represented by the first term in equation 3) and the reference portfolio (denoted by the second term). The reference portfolio is established as the fund manager aims to avoid lagging behind the benchmark index. In alignment with the preferred habitat theory, the fund manager shows a greater preference for stocks within their benchmark. It is important to note that the demand for the reference portfolio, symbolized as ω_i remains inelastic. This demand remains independent of the asset's inherent risks and is solely determined by the compensation scheme. In the case of passive fund managers where $a = 0, b = 0$, $x_i^P = \omega_i$.

Using the market clearing condition for the risky assets, $\lambda_D x_D + \sum_{i=1}^{I} x_i(\lambda_i^P + \lambda_i^A) = \bar{x}$, the equlibrium asset price is:

$$S = \mu - \gamma A \Sigma \left(\bar{x} - \frac{b}{a+b} \sum_{i=1}^{I} \lambda_i^A \omega_i + \lambda_i^P \omega_i \right) \tag{4}$$

where $A = \left[\lambda_D + \frac{\sum_i \lambda_i^A}{a+b} \right]^{-1}$ is the adjustment to risk aversion imposed by market clearing.

Equation 4 sheds light on the driving factors behind the index effect in the model. The manifestation of the index effect can be seen through the term $\frac{b}{a+b} \sum_{i=1}^{I} \lambda_i^A \omega_i + \lambda_i^P \omega_i$, which captures the aggregate inelastic demand from fund managers. When a stock is added to an index or its weight within the index increases, its price also increases. An additional insight is that when an increasing number of fund managers, represented by their assets under management (AUM), follow a particular index i, it intensifies the price pressure resulting from indexing, which in turn magnifies the impact of index inclusion. The more indices a stock is part of, and the greater its weight in those indices, the more it attracts attention from fund managers, leading to an upswing in its price.

Next, the model also provides some predictions about the expected stock returns or the cost of equity, which is the main focus of this paper. The expected return for the stock (n), denoted in terms of a per-share return as $\Delta S(n) = \bar{\mu}(n) - S(n)$, is as follows:

$$E\left[\Delta S(n)\right] = \gamma A \sigma_n^2 \left(\bar{x} - \frac{b}{a+b} \sum_{i=1}^{I} \lambda_i^A \omega_i + \lambda_i^P \omega_i \right) \tag{5}$$

Equation 5 suggests that the aforementioned price pressure is enduring and persists for the duration that a stock stays within the index of fund managers. All else being equal, stocks with a higher indexing inclusion ratio, which in this setup is denoted by $\sum_{i=1}^{I} \lambda_i^A \omega_i + \lambda_i^P \omega_i$, will have lower expected returns. Moreover, should the indexing score of a stock rise (perhaps due to its addition to an index), the price should also increase while the expected return decreases.

3 Dateset

The data for mutual fund and ETF stock holdings are sourced from FactSet, which offers a more comprehensive data set compared to the quarterly Thomson Reuters mutual fund holdings Database, which is mostly limited to the U.S. funds.

To estimate the *Indexing Inclusion ratio* (IXI), benchmark index data for each fund is sourced from Morningstar, which provides extensive benchmark coverage for a range of mutual funds and ETFs. This data is then integrated with the fund-holding data from FactSet.

To measure the indexing inclusion ratio, the weights of each stock within each index is measured by pooling together the holdings of all the index ETFs and open-ended mutual funds that share the same index benchmark id. The IXI score is calculated using all active and passive funds that have a corresponding identified benchmark from Morningstar. The passive IXI is calculated only using the funds that are identified as the index.

The FactSet security level holding data is aggregated at the company level using Permco as I require one unique IXI measure for each company. Firm characteristics and price data are extracted from a merge of CRSP-Compustat data, where only the Permno with the highest market cap is kept as the primary for each Permco.

In the interest of data accuracy and reliability, only the securities with share codes 10, 11, 12, and 18, trading on NYSE, AMEX, and NASDAQ, are retained, with corresponding exchange codes equal to 1,2 and 3. The earning announcement dates and news are obtained from the I/B/E/S dataset from WRDS. Finally, the Fama-French factors are obtained from the Kenneth French Data Library.

4 Indexing Inclusion (IXI) as a Measure of Stock Passive Ownership

The aggregated holdings of indexed passive investors for each stock divided by its market cap is the simplest way to proxy for the passive ownership of a stock.However, this approach understates passive ownership because it excludes the substantial capital managed by funds that track indices without explicitly declaring themselves as passive. A comprehensive measure must therefore account for both declared index funds and the so-called closet indexers.

Existing fund-level measures such as the Benchmarking Intensity (BMI) of Pavlova and Sikorskaya (2023) and index tracking error are incentive-based metrics. These measures evaluate whether fund managers follow their stated prospectus benchmarks, a question central to fund performance evaluation and detecting benchmark misfit risk in portfolio construction. However, incentive-based measures face a fundamental limitation when applied to stock-level asset pricing research: they assume that any fund benchmarked to an index holds exactly the index weights at all times. This assumption is particularly problematic for BMI, which attributes the entire assets under management of a benchmarked active fund to the passive ownership of index constituents, regardless of how the fund actually allocates its capital. In practice, actively managed funds even those with stated benchmarks, routinely deviate from benchmark weights in pursuit of alpha, rendering the assumption of perfect index compliance empirically untenable.

I introduce the *Indexing Inclusion Ratio* (IXI), a holdings-based measure of realized passive capital allocation at the stock level. Unlike BMI, which infers passive ownership from fund mandates, IXI directly examines actual portfolio holdings and captures only the passive *portion* of each fund's assets. For actively managed funds benchmarked to an index, IXI adjusts their contribution based on their Active Share which is the degree to which their holdings deviate from benchmark weights. A fund with 40% Active Share contributes only 60% of its assets to the passive measure, reflecting that 40% of its portfolio represents active bets rather than mechanical index tracking.

All benchmarks are indexed by $i = 1, \dots, I$ while the individual funds are indexed by $j = 1, \dots, J$.

$$IXI_t(n) = \frac{\sum_{i=1}^{I} \tilde{A}_{i,t} w_{i,t}(n)}{ME_t(n)} \tag{6}$$

where

$$\tilde{A}_{i,t} = \sum_{j=1}^{J} A_{j,t} \left(1 - \frac{1}{2} \sum_{n=1}^{N} |w_{j,t}(n) - w_{i,t}(n)|\right) \tag{7}$$

Where $w_{i,t}(n)$ refers to the weight of the asset n at time t within index i. $A_{j,t}$ is the total assets under management (AUM) of fund j at time t that is benchmarked to index i and the $ME_t(n)$ denotes the overall market capitalization of stock n at time t. The term in parentheses equals one minus the Active Share of fund j relative to its benchmark, calculated following Cremers and Petajisto (2009). This formulation weights each fund's contribution to passive ownership by the fraction of its portfolio that actually tracks the benchmark. A pure index fund with zero Active Share contributes its full AUM to the measure, while an active fund with 40% Active Share contributes only 60% of its assets, reflecting that the remaining 40% represents active positions that deviate from benchmark weights.

The key methodological distinction between IXI and BMI lies in the treatment of benchmarked active funds. BMI assumes that any fund declaring a benchmark index holds exactly the index weights at all times, effectively treating the full AUM of benchmarked funds as passive capital. This assumption leads to substantial overestimation of realized passive ownership, as actively managed funds routinely deviate from their benchmarks in pursuit of alpha. IXI addresses this limitation by using actual fund holdings to measure the degree of index tracking, capturing only the portion of assets that genuinely follows passive strategies. This distinction is particularly important in the earlier part of my sample when the prevalence of closet indexing was lower and active funds exhibited greater deviation from their benchmarks.

The distinction between incentive-based and realized measurement also relates to the broader literature documenting hidden passive capital. Chinco and Sammon (2024) demonstrate that actual passive ownership is approximately double self-declared index fund holdings when accounting for internally managed index portfolios, closet indexers, and quasi-indexing strategies across all institutional investors. While my measure focuses specifically on mutual funds and ETFs rather than the full institutional investor universe, the underlying intuition is similar: substantial passive capital is hidden within vehicles that do not explicitly declare themselves as passive. IXI captures this hidden passive capital within the fund sector by examining what funds actually hold rather than what they claim to track, identifying closet indexers whose portfolios closely mirror benchmarks despite active classifications and corresponding fee structures.[3]

The distinction between incentive-based and realized measurement has direct implications for interpreting empirical results in asset pricing research. Incentive-based measures answer the question: "How much capital is *supposed* to track indices?" Realized measures answer: "How much capital *actually* tracks indices?" For research questions about how passive capital influences stock prices, and where the relevant economic mechanism works through investors' actual portfolio holdings rather than through contractual mandates, realized measures offer the suitable empirical basis.

The construction methodology yields three distinct measures that decompose passive ownership along economically meaningful dimensions. The main measure, IXI, captures Active Share adjusted passive ownership by incorporating both pure index funds and the passive component of closet indexers. A stock's IXI reflects the fraction of its shares held by

[3]See Cremers et al. (2016) for comprehensive evidence on the prevalence of closet indexing globally, estimating that approximately 20% of mutual fund assets are closet indexed.

Table 1: Fund Universe Summary Statistics by Fund Type

Category	Statistic	Full Sample	2000–2006	2007–2012	2013–2019	2020–2021	2022–2023
All Funds	N Funds	22,936	5,823	9,839	16,437	19,504	22,352
	Total AUM ($B)	11628.8	3342.3	6550.0	16146.6	25263.0	26422.0
	Mean AUM ($M)	713.7	624.9	692.0	1028.6	1306.4	1192.3
	Median AUM ($M)	98.5	99.7	102.9	119.6	131.2	107.4
	Unique Benchmarks	6,837	1,566	2,728	4,774	5,723	6,701
Passive ETF	N Funds	5,359	411	1,289	3,354	4,268	5,201
	Total AUM ($B)	3139.7	359.1	1112.2	3990.6	8786.6	10330.0
	Mean AUM ($M)	835.5	1012.4	900.0	1239.6	2071.3	1992.1
	Median AUM ($M)	62.1	70.2	71.3	71.2	93.5	73.0
	Unique Benchmarks	3,336	248	789	2,081	2,651	3,278
Passive OEF	N Funds	2,363	495	916	1,606	1,641	1,943
	Total AUM ($B)	1025.0	211.5	450.5	1314.3	2550.3	3057.5
	Mean AUM ($M)	819.0	477.3	538.2	935.1	1576.2	1642.8
	Median AUM ($M)	88.4	80.2	78.5	112.8	168.0	126.0
	Unique Benchmarks	1,110	230	434	756	814	970
Active Benchmarked	N Funds	15,214	4,917	7,634	11,477	13,595	15,208
	Total AUM ($B)	7464.1	2771.7	4987.3	10841.6	13926.1	13034.5
	Mean AUM ($M)	654.4	607.4	675.3	979.9	1033.6	861.2
	Median AUM ($M)	115.2	104.8	112.9	136.3	140.3	115.9
	Unique Benchmarks	3,425	1,279	1,880	2,615	3,049	3,423

This table reports summary statistics for the fund universe used to construct IXI, broken down by fund type and time period. Passive ETF and Passive OEF are exchange-traded funds and open-end mutual funds, respectively, classified as index funds based on Morningstar style classification. Active Benchmarked includes all non-index funds with a declared prospectus benchmark. N Funds is the number of unique funds observed during the period. Total AUM is the average annual assets under management in billions USD. Mean and Median AUM are cross-sectional statistics of fund-level average AUM in millions USD. All funds have benchmark assignments from Morningstar. AUM calculated from FactSet holdings data.

investors whose portfolios, in aggregate, track major market indices, regardless of whether those investors explicitly declare themselves as passive. This is the primary measure used throughout the analysis.

The second variant, IXI_{pass}, measures passive ownership arising solely from declared index funds. This measure is comparable to traditional passive ownership estimates based on fund classifications and serves as a lower bound on realized passive ownership, capturing only the capital that is explicitly and transparently committed to index tracking strategies. The third variant, $\text{IXI}_{\text{non-adj}}$, represents total benchmarked capital without Active Share adjustment, effectively assuming that all funds benchmarked to an index fully track it regardless of their actual holdings. This measure is methodologically comparable to BMI in that both rely on the assumption of full benchmark compliance among benchmarked funds. $\text{IXI}_{\text{non-adj}}$ therefore provides an upper bound on passive ownership, and the comparison between IXI and $\text{IXI}_{\text{non-adj}}$ quantifies the degree to which incentive-based measures may overstate realized passive capital allocation when active funds deviate from their stated benchmarks.

By construction, the ordering $IXI_{\text{pass}} \leq IXI \leq IXI_{\text{non-adj}}$ holds for each stock-month observation. The difference between IXI and IXI_{pass}, which I term the Closet Gap, captures the passive capital embedded within funds that charge active fees while holding portfolios that closely mirror benchmark weights. This hidden passive capital would be overlooked by conventional measures relying solely on fund classifications.

The final IXI sample covers 6,767 unique CRSP common stocks from January 2000 through December 2023. The measure incorporates holdings from over 22,000 distinct funds benchmarked to nearly 7,000 unique indices including their variations, providing comprehensive coverage of the passive investing landscape. Table 1 provides an overview of the fund universe by fund type and time period. As the sample progresses, the dominance of pure index funds in terms of assets under management becomes increasingly pronounced, with passive ETFs growing from approximately 359 billion dollars in the early part of the sample to over 10.3 trillion dollars by the end of the sample period in 2023. The breadth of index coverage is particularly important given the proliferation of specialized indices, including sector funds, factor funds, and thematic ETFs, that now constitute a meaningful fraction of passive assets but would be missed by measures focusing only on broad market benchmarks.

Table 2: Properties of Index Inclusion Intensity (IXI)

	Full Sample	2000–2006	2007–2012	2013–2019	2020–2021	2022–2023
Panel A: Descriptive Statistics						
Avg IXI, %	8.1	2.8	6.8	11.0	14.2	14.0
Median IXI, %	4.7	1.8	6.2	9.6	12.4	10.8
SD of IXI, %	9.2	2.9	6.1	9.7	12.1	12.9
Min IXI, %	0.0	0.0	0.0	0.0	0.0	0.0
Max IXI, %	97.9	97.9	97.9	97.9	96.5	95.5
Avg IXI Passive, %	6.3	1.8	4.7	8.6	11.9	12.0
SD of IXI Passive, %	7.6	1.9	4.2	7.9	10.2	11.1
Avg IXI Unadjusted, %	18.3	9.0	20.0	23.1	24.9	22.8
SD of IXI Unadjusted, %	18.3	8.9	18.5	19.8	21.0	20.8
Avg no. of benchmarks	31.5	11.8	21.5	40.3	59.4	65.9
Panel B: Average Contribution of Major Benchmark Families (%)						
S&P 500	42.4	67.5	49.2	40.7	38.5	38.3
Russell 2000	9.8	10.7	14.0	11.4	7.9	7.9
Russell 1000	2.3	2.3	2.5	2.6	2.2	2.0
Russell 3000	2.2	1.1	2.1	2.5	2.1	2.2
CRSP US Total Market	11.7	4.9	9.3	12.7	12.6	12.5
Nasdaq 100	6.4	13.9	9.9	5.6	5.6	6.3
Panel C: Contribution to IXI by Fund Type (%)						
Pure Passive	78.3	73.1	73.8	81.8	86.3	88.3
Closet Indexers	21.7	26.9	26.2	18.2	13.7	11.7

This table reports properties of Index Inclusion Intensity (IXI). IXI measures the realized passive capital allocation at the stock level, weighted by funds' Active Share relative to their declared benchmarks. IXI Passive captures ownership from self-declared index funds only. IXI Unadjusted measures total benchmarked capital without Active Share adjustment. Panel A shows descriptive statistics with IXI values expressed in percentage points. Panel B reports the value-weighted average contribution of major U.S. benchmark families to a stock's IXI, where weights are each stock's total IXI dollar contribution (market capitalization × IXI). Contribution is the ratio of IXI coming from funds benchmarked to each index family to the stock's total IXI. Panel C decomposes IXI into contributions from Pure Passive funds (self-declared index funds) and Closet Indexers (benchmarked funds with Active Share adjustment); these sum to 100% by construction. Sample: CRSP common stocks, 2000–2023.

Table 2 reports properties of the IXI measure. Panel A presents descriptive statistics across the full sample and five sub-periods that capture distinct phases of the passive investing revolution.

Average IXI increased from 2.8% during 2000–2006 to 14% during 2022–2023, a five-fold increase reflecting the dramatic shift toward passive strategies over the sample period. The dispersion of IXI, measured by standard deviation, grew proportionally from 2.9% to approximately 12%, indicating that the rise in passive ownership has not been uniform across stocks but rather has created substantial cross-sectional variation that can be exploited for identification in empirical tests.

The contrast between the two IXI variants is especially revealing. The average $\text{IXI}_{\text{non-adj}}$, a BMI-comparable metric that assumes complete index adherence, is 18.3% across the entire sample which is more than twice the 8.1% average recorded for IXI. This large difference demonstrates that incentive-based measures substantially overstate realized passive ownership by ignoring the active share of benchmarked funds. In the early sample period (2000–2006), the ratio of $\text{IXI}_{\text{non-adj}}$ to IXI exceeds three, reflecting the prevalence of truly active management during this period. By 2022–2023, this ratio has compressed to approximately 1.6, consistent with the widely documented trend toward closet indexing as active managers face competitive pressure from low-cost passive alternatives.

Panel B reports the contribution of major U.S. benchmark families to aggregate IXI. Since there are many variations of the same benchmark which can differ across different currencies, total return vs. net return etc., I calculate each benchmark family's contribution as the ratio of IXI coming from funds benchmarked to that family relative to the stock's total IXI, then compute value-weighted averages where weights are each stock's passive dollar amount (market capitalization multiplied by IXI). This value-weighting approach ensures that the reported contributions reflect where passive capital is actually concentrated rather than giving equal weight to small stocks with idiosyncratic benchmark exposure.

The S&P 500 family dominates passive ownership, contributing 42.4% of total IXI on a value-weighted basis over the full sample. This concentration has declined over time from 67.5% during 2000–2006 to 38.4% during 2020–2023 as total market indices (CRSP US Total Market: 11.7%) and factor-based strategies have gained market share. The Russell 2000 contributes 9.8% of passive ownership, reflecting its role as the dominant small-cap benchmark, while the NASDAQ-100 accounts for 6.4%, concentrated among technology stocks.

Panel C decomposes IXI into contributions from pure passive funds and closet indexers.

Pure passive funds account for 78.3% of total IXI over the full sample, with this share increasing from 73.1% in 2000–2006 to 88.3% in 2022–2023. The rising pure passive share reflects both the growth of explicit index funds and ETFs and the declining prevalence of closet indexing.

Figure 1 presents the time-series evolution of average IXI, IXI_{pass}, and $IXI_{non\text{-}adj}$. Several notable patterns emerge. First, all three measures exhibit strong growth, with a clear acceleration after the 2008 financial crisis, which spurred both retail and institutional investors to allocate more capital to passive vehicles. Second, the gap between $IXI_{non\text{-}adj}$ and IXI—which reflects how much incentive-based measures are overstated, has decreased over time, though it remains significant. Third, the gap between IXI and IXI_{pass}, referred to as the Closet Gap, has leveled off and in recent years has begun to shrink, in line with a decline in closet indexing.

5 Realized Return of High-indexed Stocks

Over the past two decades, stocks with a high IXI score have demonstrated substantial outperformance relative to their low IXI counterparts. The comparative performance of these monthly sorted and value-weighted portfolios, which span from January 2000 to January 2021, is illustrated in Figure 2. The solid blue line in this figure represents high IXI stocks, showcasing the cumulative value-weighted return on the portfolio of stocks that exhibit IXI scores in the top third. In contrast, the dashed blue line delineates the corresponding return for stocks that fall within the bottom third of IXI scores, thus characterizing low IXI stocks. The divergence between these two portfolios became particularly pronounced after 2012, a period that also marked significant shifts from active to passive investing. In particular, the past decade has witnessed the outperformance of highly indexed stocks over low-indexed ones. This disparity has resulted in a cumulative return difference of approximately 325% over our 20-year sample period.

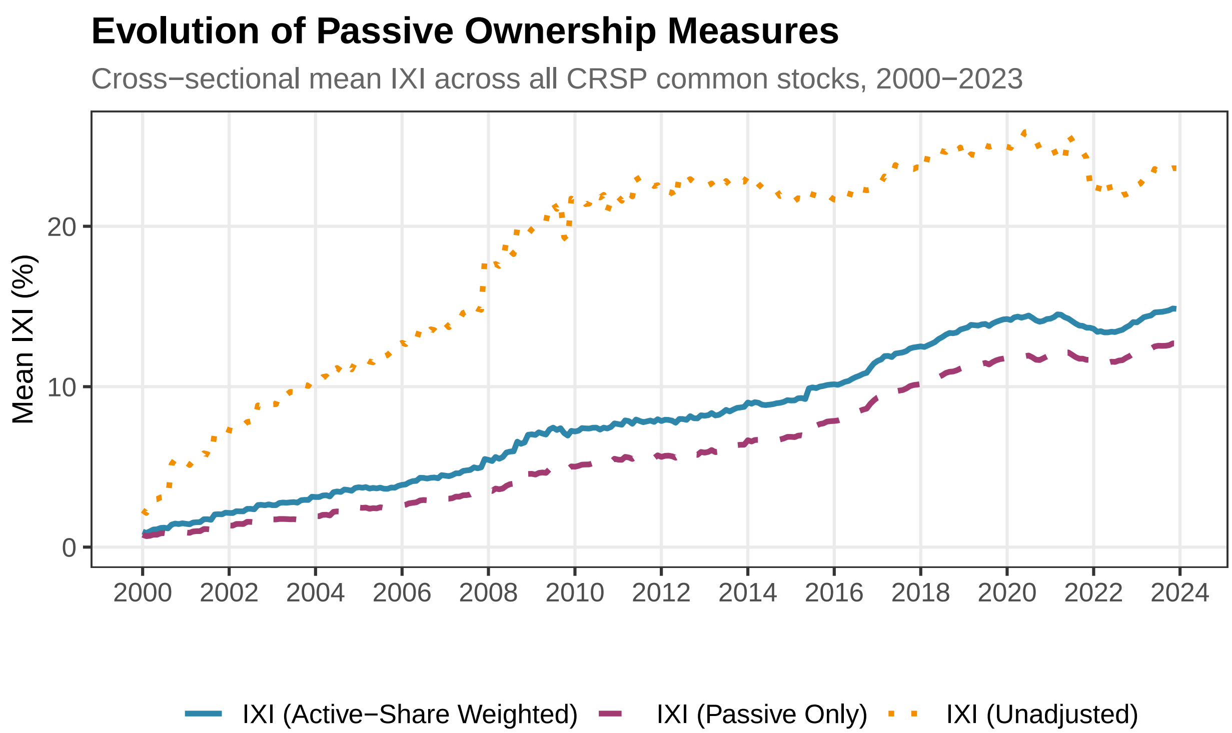

Figure 1: Evolution of IXI Measures (2000-2023)

This figure plots the monthly cross-sectional mean of IXI across all stocks from 2000 to 2023. The Solid Line (IXI) represents the primary measure, defined as the aggregate dollar value of passive capital (from index funds and the passive component of active funds. The Dashed Line (IXI Passive) represents ownership held strictly by declared index mutual funds and ETFs. The Dotted Line (IXI Unadjusted) represents the total ownership of all funds benchmarked to indices without adjusting for Active Share (equivalent to the BMI measure approach).

The robust performance of HILI (high IXI minus low IXI portfolio) cannot be accounted for by exposure to well-documented return factors in asset pricing. The first column of Table 3 confirms that the monthly performance disparity between high and low IXI stocks remains statistically significant. Subsequent columns in Table 3 present the outcomes of regressing HILI on various factors, encompassing those included in the five-factor models proposed by Fama and French (2015), the momentum factor proposed by Carhart (1997), and the green minus brown factor which captures the degree of ESG compliance of stocks by Pástor et al. (2022). In all scenarios, HILI's alpha (the regression constant) is both economically and statistically meaningful. The smallest alpha for HILI, as outlined in Table 3, is reported in column 6, which accounts for the Fama-French five factors. The portfolio's exposures

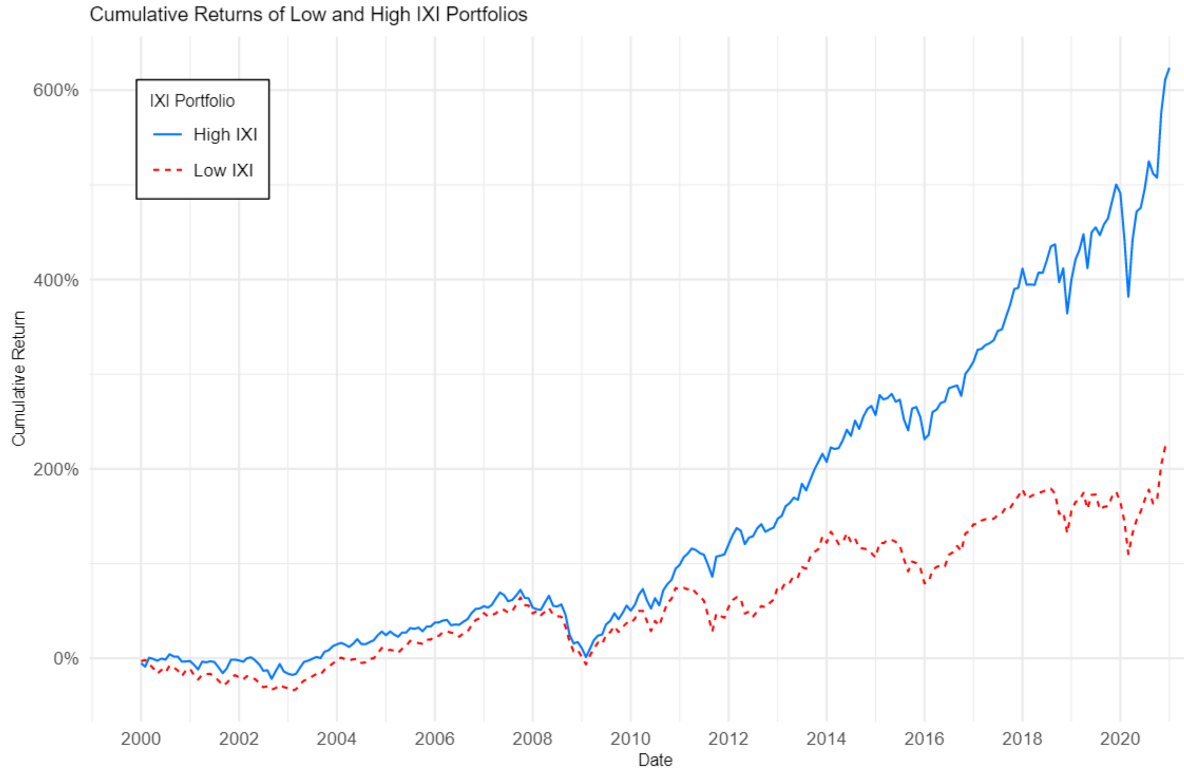

Figure 2: The cummulative returns of High IXI vs. Low IXI portfolios

to SMB, HML, and RMW suggest that HILI leans towards larger, growth-oriented stocks, and firms with higher profitability. After adjusting for these exposures, the alpha of HILI continues to be statistically significant at a 5% confidence level.

To assess the robustness of the findings, Table A.2 in the Appendix replicates the analysis conducted in Table 3, with a notable adjustment: stocks within the top 2.5 percentile of market capitalizations are excluded each month. This methodological refinement ensures that the results are not disproportionately influenced by a small subset of megacap stocks.

6 Indexing Premium

In this section, I investigate the expected return dynamics of stocks categorized into high and low Indexing Inclusion (IXI) portfolios. The concept of an "indexing premium" is introduced, capturing the predicted return discrepancy between high IXI and low IXI stocks. Specifically,

Table 3: High IXI minus Low IXI sorted portfolios performance

	Dependent variable:							
	HILI							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	0.388^{***}	0.370^{***}	0.424^{***}	0.429^{***}	0.344^{**}	0.271^{**}	0.420^{***}	0.450^{**}
	(0.136)	(0.138)	(0.128)	(0.124)	(0.134)	(0.124)	(0.124)	(0.197)
MKT-Rf		0.034	0.086^{**}	0.086^{**}	0.113^{***}	0.145^{***}	0.096^{***}	0.085^{**}
		(0.037)	(0.035)	(0.034)	(0.029)	(0.034)	(0.031)	(0.042)
SMB			-0.259^{***}	-0.257^{***}	-0.197^{***}	-0.210^{***}	-0.261^{***}	-0.243^{***}
			(0.055)	(0.057)	(0.069)	(0.067)	(0.057)	(0.083)
HML				-0.055	-0.110^{**}	-0.201^{***}	-0.048	-0.103^{*}
				(0.045)	(0.043)	(0.058)	(0.048)	(0.059)
RMW					0.152^{*}	0.166^{***}		
					(0.079)	(0.063)		
CMA						0.243^{**}		
						(0.115)		
MOM							0.021	0.075
							(0.031)	(0.051)
GMB								0.057
								(0.085)
Observations	253	253	253	253	253	253	253	144
R^2	0.000	0.006	0.149	0.156	0.181	0.214	0.159	0.204

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

The performance of the high IXI minus low IXI (HILI) portfolio, sorted on a monthly basis, is assessed using data spanning from 2000 to 2021. The outcome variable in this context is the HILI - the difference in returns between the high and low IXI portfolios. The excess market return is represented by Mkt-RF. SMB and HML correspond to size-sorted and value-sorted factors as proposed by Fama and French (1993). Moreover, profitability and investment factors are denoted by RMW and CMA as per Fama and French (2015). MOM is the momentum factor of Carhart (1997). GMB is the green minus brown factor of Pástor et al. (2022). Notably, all returns are expressed in percentages and at a monthly frequency, with robust standard errors enclosed in parentheses.

the equity indexing premium is defined as the expected return on the High-Low indexing ratio (HILI) spread. As expected stock returns are not directly observable, it is necessary to estimate the equity indexing premium. To this end, I adopt two methods based on both ex-ante data using the implied cost of capital (ICC) method and the ex-post data by purging the earning-related news and flow shocks.

6.1 Measuring the Implied Cost of Capital (ICC)

In this section, I outline the method that I use to calculate the ICC for stocks. I adopt the methodology established by Hou et al. (2012), hereafter referred to as "HVZ." This methodology is built on the residual income valuation model proposed by Gebhardt et al. (2001). Unlike the IBES earnings forecasts often used in such calculations, the HVZ methodology employs earnings forecasts derived from regression analyses. My implementation stays as close as possible to that of HVZ, as per the execution by Lee et al. (2021) and Pástor et al. (2022).

The implied cost of capital, denoted as r_e below, is the internal rate of return that aligns the present value of future dividends with the currently observed stock market capitalization. This can be expressed with the following formula:

$$ME_{i,t} = B_{i,t} + \sum_{\tau=1}^{\infty} \frac{\mathbb{E}_t\left[NI_{i,t+\tau}\right] - r_e \times \mathbb{E}_t\left[B_{i,t+\tau-1}\right]}{(1+r_e)^{\tau}} \tag{8}$$

In this equation, $ME_{i,t}$ is the current stock market cap, $NI_{i,t+\tau}$ is the forecast of earnings or more specifically net income before extraordinary items in year $t+\tau$, and $B_{i,t}$ is the book value of equity. The ICC r_e is specific to firm i and time t, but I omit these subscripts for the sake of simplicity in notation, and similarly the i subscript is dropped for $ME_{i,t}$. Implementing a 12-year forecasting horizon with a terminal perpetuity the equation can be simplified to the following:

$$ME_t = B_t + \sum_{\tau=1}^{11} \frac{E_t\left[\left(ROE_{t+\tau} - r_e\right) B_{t+\tau-1}\right]}{\left(1+r_e\right)^{\tau}} + \frac{E_t\left[\left(ROE_{t+12} - r_e\right) B_{t+11}\right]}{r_e\left(1+r_e\right)^{11}} \tag{9}$$

Where ROE_t stands for the return on equity that is NI_t divided by B_t. I estimate the forecasted earnings NI for the first 3 years using the HVZ method. Where for each month, and for each forecasting horizon τ of 1, 2 and 3 years, the following pooled cross-sectional regression is estimated:

$$NI_{i,t+\tau} = \alpha_0 + \alpha_1 A_{i,t} + \alpha_2 D_{i,t} + \alpha_3 DD_{i,t} + \alpha_4 E_{i,t} + \alpha_5 NegE_{i,t} + \alpha_6 AC_{i,t} + \varepsilon_{i,t+\tau} \tag{10}$$

Where $A_{i,t}$ is the total assets, $D_{i,t}$ is the dividend payment, $DD_{i,t}$ is a dummy variable that equals 1 for dividend payers and 0 otherwise, $NegE_{i,t}$ is a dummy variable that equals 1 for firms with negative earnings and 0 otherwise, and $AC_{i,t}$ is accruals. All explanatory variables are measured in dollars as of year t and are winsorized at 1% and 99% levels each year. The pooled regressions are estimated using the previous ten years of data and the forecasted $NI_{i,t+\tau}$ is obtained by multiplying the independent variables for each firm i and year t with the coefficients $(\alpha_1, ..., \alpha_6)$.

When predicting ROE over forecast horizons ranging from $\tau = 4$ to 12 years, I use a linear interpolation method. This interpolates from the forecasted ROE three years ahead, gradually converging to the median ROE within the industry. I operate under the assumption that by year $t + 12$, the ROE will have reverted to its industry's historical median value. Adhering to the methodology outlined by Lee et al. (2021), I calculate the industry's median ROE annually, using data from the preceding decade. In alignment with this classification, I organize firms into 48 Fama-French industry categories. As per the HVZ guidelines, when determining the median industry ROEs, I exclude any loss-making firms.

The future book value per share $B_{i,t+\tau}$ is calculated based on the accounting clean surplus relation:

$$B_{i,t+\tau} = B_{i,t+\tau-1} + NI_{i,t+\tau} - D_{i,t+\tau} \tag{11}$$

where $D_{i,t+\tau}$ is the total dividends in year $t + \tau$. Additional provisions are incorporated to limit instances of the forecasted negative book value of equity, as the current framework permits dividends to assume negative values, implying the firm is issuing new equity. Finally, the ICC estimates are winsorized at the 1st and 99th percentiles.

6.2 Indexing Premium: ICC

The ex-ante method equates each stock's expected return to its Implied Cost of Capital (ICC). The ICC represents the discount rate that aligns the stock's current market price with the present value of its projected future cash flows, using contemporaneous data. The expected HILI return is therefore formulated based on the ICCs of the individual stocks.

Employing an ex-ante approach, I calculate the ICC for each stock-month. The ICC is inferred through market price data and predicted cash flows, in accordance with the traditional discounted cash flow formula. Following the steps of Hou et al. (2012), which is rooted in the seminal framework of Gebhardt et al. (2001) but employs regression-based forecasts in place of analyst earnings forecasts, the ICC is estimated. In line with Pástor et al. (2022), from various ICC methods surveyed by Lee et al. (2021), I select the approach that was found to yield the most precise cross-sectional expected return estimates.

Figure 3 illustrates a time series of ICCs for the high IXI and low IXI portfolios that make up the long and short legs of the HILI. The ICC of each portfolio is calculated as the value-weighted average of the ICCs of its component stocks. Throughout most of the sample period from 2000 to 2021, the high IXI portfolio's ICC consistently falls short of the low IXI portfolio's ICC, denoting a persistently negative equity indexing premium and suggesting a lower expected return for highly indexed stocks. Figure 4 plots the ICC spread between high IXI and low IXI portfolios which average around -0.3% per year during the sample period with the highest peaks during the 2007-2008 financial crisis and the lowest point during the 2020 covid crash.

Assuming that the ICC is a credible proxy for the expected stock return, the consistent negative indexing premium endorses the proposition that the strong performance of HILI during the sample period, with the main exception of the 2007-2008 financial crisis, was largely unexpected.

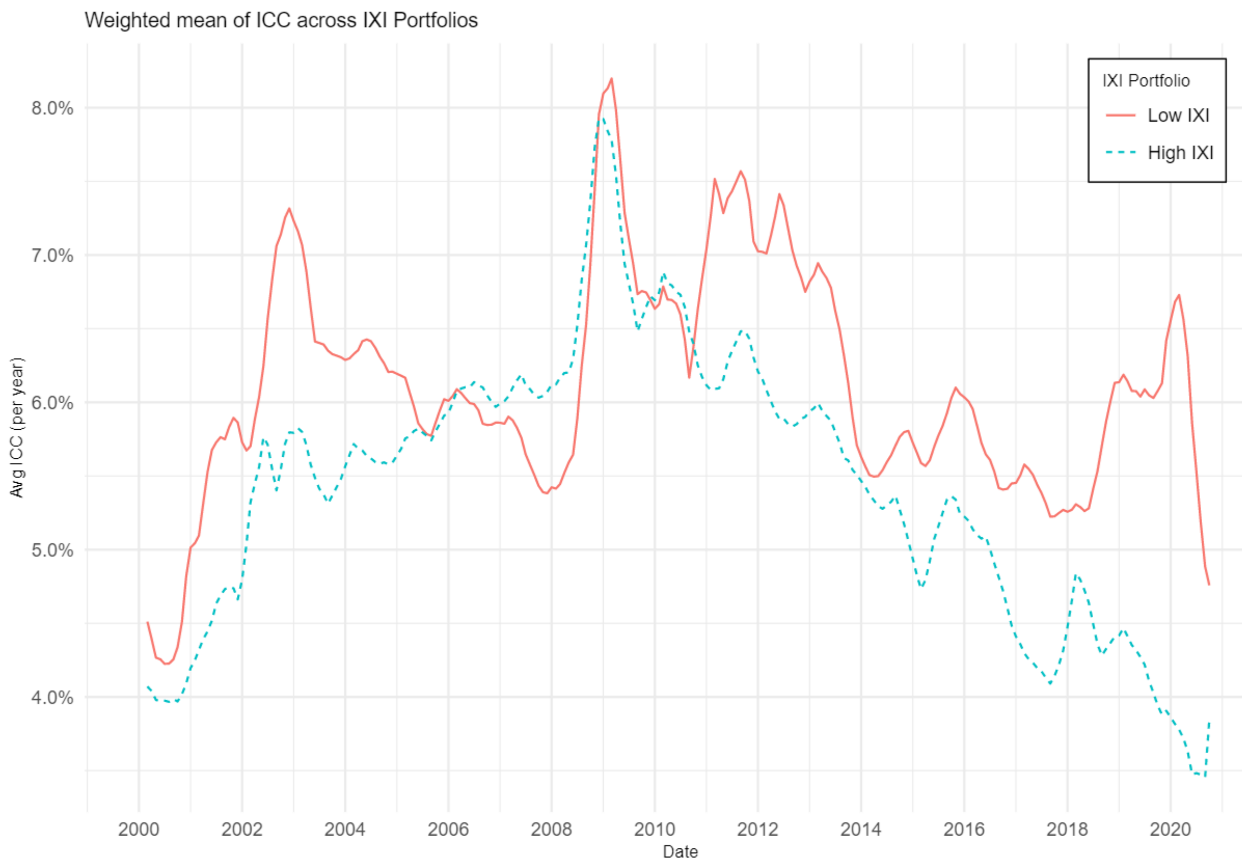

Figure 3: The weighted average of ICC across IXI portfolios

Figure 5 presents the ICC metrics for portfolios sorted by monthly IXI across different categories of stocks. The top two panels delineate the distribution within the size dimension, where each month stocks are segregated into terciles according to market capitalization. The top left panel represents the upper-tercile or large-cap stocks, and the top right panel illustrates the lower-tercile or small-cap stocks. Within each size category, portfolios are further arranged based on the IXI score to assess whether the ICC disparity between high

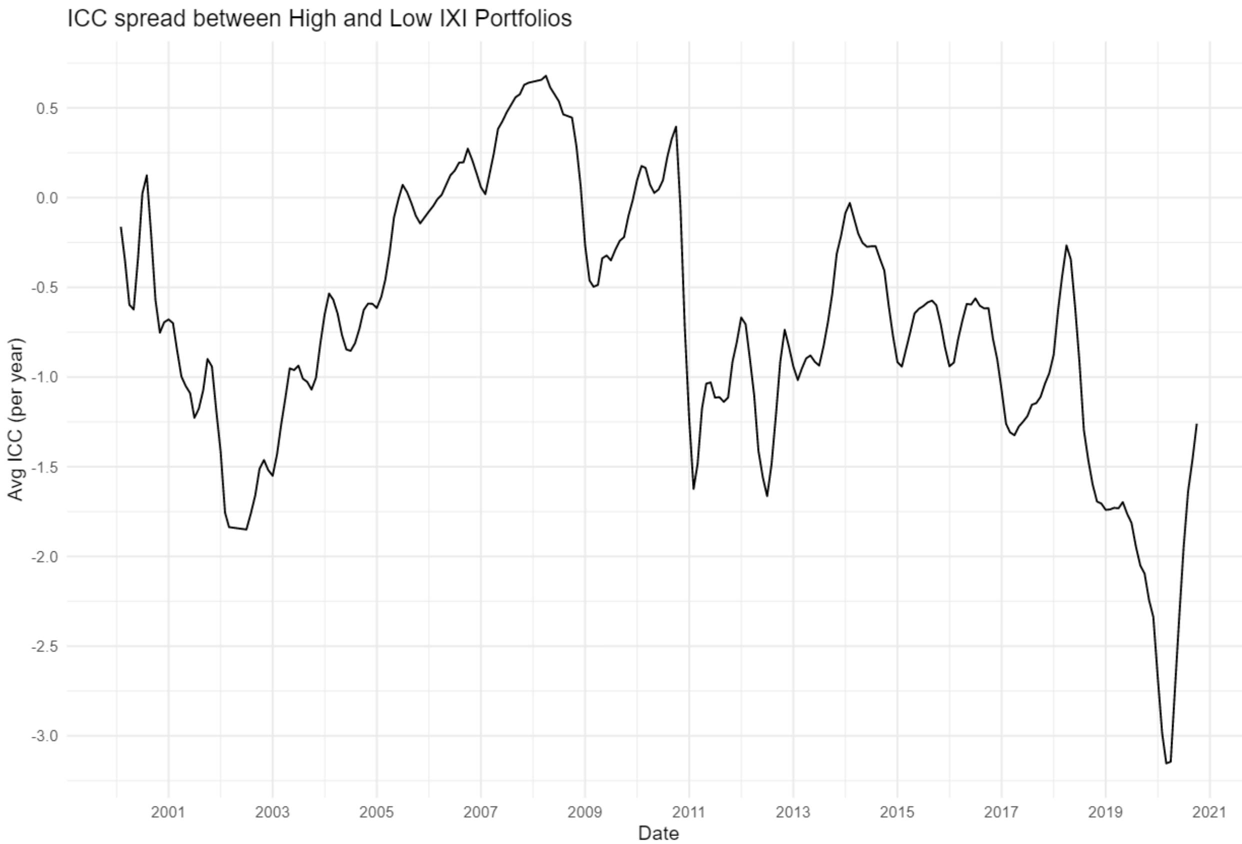

Figure 4: The weighted average of ICC Spread: Indexing Premium

and low IXI groups remains consistent across diverse stock types.

In the context of large-cap stocks, the ICC differential between high and low IXI portfolios is marginal, and the low IXI ICC even dropped below the high IXI from 2013 to 2021. In contrast, for small-cap stocks, the ICC of the low IXI portfolio substantially exceeds the high IXI portfolio beginning in 2011. The expected return of small-cap stocks as proxied by the ICC is predictably higher, averaging around 8.3%, as compared to large-cap stocks which maintain an average ICC of approximately 5.5%.

The lower two panels in Figure 5 illustrate the average ICC of IXI sorted portfolios across three tertiles of the book-to-market (B/M) ratio. The lower left panel represents the bottom tertile of B/M or the growth stocks. Here, the ICC for both high IXI and low IXI portfolios is similar, though the high IXI portfolio's ICC slightly exceeds the low IXI's, particularly towards the end of the sample period. For value stocks, depicted in the lower right panel, the low IXI portfolio's ICC diverges upward from the high IXI portfolio beginning in 2011. In particular, the average ICC for growth stocks remains lower at approximately 4.4%, compared to that of the value stocks, which is nearly double, around 8.7%.

Overall, it can be deduced from the results in Figure 5 that the majority of the negative indexing premium is driven by the value and small-cap stock which also happens to have lower IXI score on average.

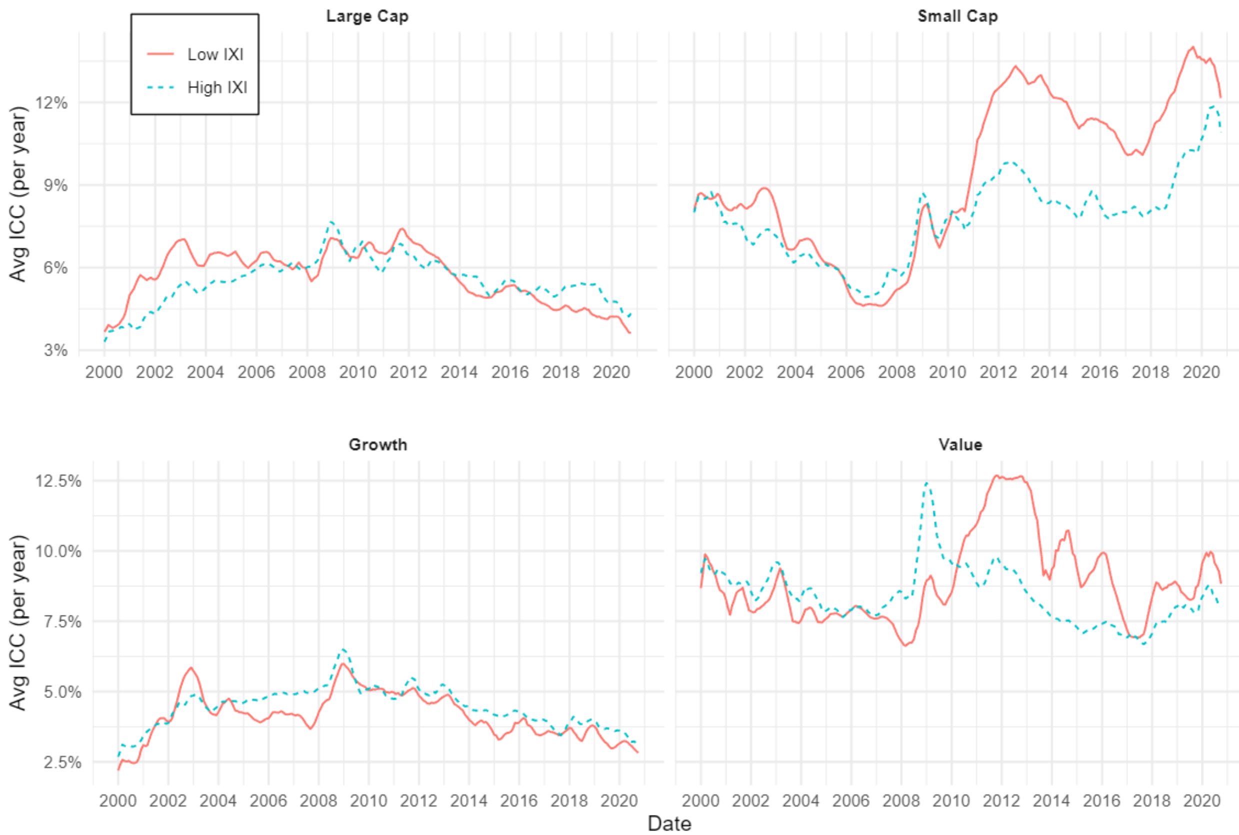

Figure 5: The weighted average of ICC across IXI sub-portfolios of large vs. small and growth vs. value

Further insights into the indexing premium are derived from a panel regression presented in Table 4. This table presents the results of regressing a stock's ICC on its lagged IXI score, while accounting for both month and firm fixed effects. The standard errors are robust by double clustering on firm-month. The significantly negative slope estimate ($t = -17.22$) is notable, and further substantiates the existence of a negative indexing premium.

Table 4: Panel regression of ICC on lagged IXI using both month and firm fixed effects and with robust double clustered SE

	Dependent variable:
	ICC
Lag IXI	-0.133^{***}
	(0.007)
Observations	726,536
Within R^2	0.03
Adjusted R^2	0.55
Month FE	Y
Firm FE	Y
Note:	*p<0.1; **p<0.05; ***p<0.01

6.3 Measuring Indexing Premium from Historical Realizations

One of the fundamental challenges in asset pricing is discerning an asset's unconditional expected return, denoted as $\mu = E\{r_t\}$, from ex-post data. Following Pástor et al. (2022), a straightforward approach would utilize the asset's sample average return, represented as $\bar{r}$, as an estimator for μ. Another, perhaps more nuanced, method incorporates supplementary information from a contemporaneous variable, x_t. This variable, by design, correlates with the return given $E\{x_t\} = 0$. In our setting, this might be instantiated as an earnings release shock.

Given the regression format

$$r_t = a + bx_t + \epsilon_t \tag{12}$$

it is evident that a can be interpreted as μ due to the ex-ante zero mean of x_t. From this, our estimation of μ can be derived from the sample estimate of a. Simplifying further, the OLS intercept is formulated as

$$\hat{a} = \bar{r} - \hat{b}\bar{x} \tag{13}$$

where $\hat{b}$ denotes the OLS slope coefficient and $\bar{x}$ represents the sample average of x_t.

To gain insight from this estimator, consider a scenario where x_t has a positive sign, implying $b > 0$. If the realizations of x_t systematically exceed their expectations, rendering $\bar{x} > 0$, $\bar{r}$ would overstate μ to an extent of $b\bar{x}$. This overestimation is effectively corrected by the estimator $\hat{a}$, which adjusts $\bar{r}$ by the term $\hat{b}\bar{x}$. In contrast, for $\bar{x} < 0$, $\bar{r}$ is likely to underestimate μ, and $\hat{a}$ essentially compensates for this discrepancy. In general, when $\bar{x} \neq 0$, the regression intercept improves the potential distortion in $\bar{r}$. This logic remains robust even when x_t extends to a vector of variables, since their sample means diverge from zero.

6.4 Measuring Earning News Shocks and Passive Flow

To utilize the approach relying on $\hat{a}$, it is crucial to define x_t. Essentially, x_t represents a vector of shocks from two main sources. First, stock-specific information directly influences earnings expectations and returns, hence the incorporation of earnings news into x_t. Secondly, the surge in demand for passive investment products affects the differential in returns between high-indexed and low-indexed stocks. This shift towards passive investment, especially evident in the US, has not only grown rapidly but also indicates passive strategies taking a larger share of the market.

When a significant portion of assets are held in passive investments, its impact on expected stock returns becomes more pronounced, as predicted by the PST model. Subsequently, I will outline the methods to quantify these shock sources. In specifying x_t, I introduce two metrics of earnings news derived from the CRSP and I/B/E/S datasets. The initial metric is influenced by the notion that significant earnings news emerges on the days of firms' earnings-related announcements, as highlighted by Beyer et al. (2010) and subsequently by Pástor et al. (2022). I determine stock returns in excess of the market within a three-day trading window centered on these announcement dates, aggregating these excess returns for distinct events in a specific stock-quarter.

The secondary metric is oriented towards long-term earnings news, which can manifest sporadically between quarterly announcements. This metric uses analysts' predictions of a firm's long-term earnings growth. For a firm i in quarter t, it's computed as the difference between the earliest median forecast for the quarter $t+1$ and the concluding median forecast for the quarter $t-1$. Performing predictions from quarters $t-1$ and $t+1$ ensures the encapsulation of all earnings revelations during quarter t. Even though this metric might include minor details from quarters $t-1$ or $t+1$, such inclusions are inconsequential, since they are unlikely to elucidate returns in quarter t. This metric is winsorized at the 1% threshold.

Considering that HILI identifies a discrepancy in portfolio returns, it is crucial to convert firm-specific earnings metrics into the respective portfolio-level measures for x_t. This is achieved by applying HILI's portoflio constructions, where monthly value-weighted averages of the firm-specific metrics within HILI's high-indexed and low-indexed categories are calculated, and their difference is assessed. It is important to recognize the inherent difficulties in isolating the portion of returns affected by earnings news. The mentioned metrics, although strong, may not encompass all pertinent earnings information. The primary metric may miss flash news outside the three-day announcement window. A notable limitation of the secondary metric is the gap between analysts' and investors' forecasts. Additionally, since analysts' long-term forecasts cover only three to five years, this metric might overlook changes in earnings forecasts beyond a five-year period.

Considering the substantial and continuously growing asset base of passive investing, it is reasonable to suggest that passive investing has significantly influenced HILI's observed returns. The impact of inelastic passive demand could be significant, particularly given research showing that stock prices can react noticeably to even small changes in demand, as highlighted by Koijen and Yogo (2019) and Gabaix and Koijen (2020). I present two variables to explore the impact of passive investing. The first variable uses flows into passive funds as a measure of changes in investor demand for indexed assets. Utilizing data from Morningstar, I collect information on the quarterly total inflows into U.S. passive funds.

These flows, referred to as "passive flows," are adjusted by the average total market capitalization of CRSP stocks for the respective quarter. Interestingly, passive flows experienced a notable increase from 2013 to 2021, with a marked acceleration beginning in 2017. The second investment metric uses the lagged total assets (AUM) of passive funds to gauge the prevailing investor sentiment towards passive investments. This selection is guided by PST's theoretical assertion that expected green-minus-brown returns are inversely proportional to the average intensity of ESG preferences, and the size of the ESG sector is directly linked to such preferences. I calculate the passive fund AUM using Morningstar data and adjust the passive AUM relative to the total market valuation of CRSP stocks.

6.5 Indexing Premium: Earning News and Passive Flow

Columns 1 and 2 of Table 5 display the results of regressions of HILI returns on the two earnings variables and the two flow and asset size variables. Earlier, it was observed that HILI leans toward large growth stocks. To ensure that this inclination towards size and growth is not influencing my findings, I examined columns 3 and 4 of Table 5, which are labeled "HILI Alpha" These columns represent regressions where the dependent variable is redefined as the HILI return adjusted for the three factors presented by Fama and French (1993). I derived this return by adding the intercept to the residual of the time series regression of HILI against these factors. In particular, the slope coefficients in these columns closely match those in the first two columns of the table. A potential concern with regressing returns on contemporaneous flows is reverse causation. It is possible that, rather than flows (or shifts in investors' passive demands) influencing returns, the flows might be pursuing recent returns within the same timeframe. The potential endogeneity issue was addressed by employing the previous quarter's value as an instrument for the passive flow of the same quarter, subsequently estimating the regression through two-stage least squares. This method relies on the reasonable exclusion restriction that flows cannot follow future returns. My analysis reveals significant first-stage t-statistics, confirming that the relevance condition is satisfied and there are no concerns regarding weak instruments.

The coefficients related to passive flows and assets in Table 5 all adhere to predictions. Specifically, passive flows have a positive association with the HILI spread, while passive size has a negative association. For the HILI spread, all coefficients are statistically significant with the exception of Δ Earnings Forecasts. Additionally, the passive size is not significant when the dependent variable is the HILI alpha.

In the absence of passive flow shocks or other shocks to high-indexed vs. low-indexed stock earnings, how would the cumulative retrun of HILI portfolio look? Figure 6 juxtaposes HILI's actual returns with a hypothetical scenario where these shocks are absent. Using the regression results from column 2 of Table 4, I determine the theoretical monthly HILI return by summing the regression intercept with the estimated residual. In Figure 6, the dashed trajectory illustrates the cumulative counterfactual return, while the solid line presents the cumulative realized return.

The comparison in Figure 6 reveals a compelling insight that throughout most of the time series, if we disregard the passive flow and earnings shocks, HILI's trajectory is largely in the negative with a slightly declining trend. This mirrors the negative intercepts found in columns 1 and 2 of Table 4. Importantly, starting from 2003, HILI's counterfactual returns consistently lag behind its realized returns.

7 A Stock Level Measure of Passive Flow

The empirical analysis I have conducted so far revolves around the time series of returns from high-indexed versus low-indexed portfolios. Nevertheless, to ensure that my results are not exclusively tied to portfolio returns, I explore panel regressions focused on individual stocks. A key component for this analysis is the stock-specific passive flow metric. This metric is crucial to assess the potential impact of passive flows on stock returns, considering their respective IXI score. Specifically, passive $\Delta\text{Flow}_{i,t}$ denotes the total change in capital

Table 5: Passive Flows and Size

	Dependent variable:			
	HILI Return		HILI Alpha	
	(1)	(2)	(3)	(4)
Constant	−0.152	−0.155	0.007	0.005
	(0.268)	(0.202)	(0.213)	(0.175)
Earnings Release Returns	0.372**	0.435***	0.437**	0.514***
	(0.176)	(0.148)	(0.179)	(0.155)
Δ Earnings Forecasts	0.016	0.200	0.083	0.196
	(0.180)	(0.126)	(0.130)	(0.138)
Passive Flow		1.103**		0.875**
		(0.456)		(0.406)
Passive Size		−0.814***		−0.448
		(0.293)		(0.285)
Observations	252	252	252	252
Note:				*p<0.1; **p<0.05; ***p<0.01

The performance of the high IXI minus low IXI (HILI) portfolio, sorted on a monthly basis, is assessed using data spanning from 2000 to 2021. The outcome variable in columns 1-2 is the HILI - the difference in returns between the high and low IXI portfolios. Earnings Release Return is the individual stock's sum of excess returns during the three-trading day window around the earnings release date. Δ Earnings Forecasts is the change in IBES analysts' median long-term forecast of earnings growth rate. The passive flow is the quarterly dollar flow into passive funds scaled by the average total CRSP market size. The passive flow is instrumented by its previously quarterly value. The Passive Size is the total AUM of all passive US funds which is scaled by the total CRSP market size. Columns 3-4 are the alpha from a time-series regression of HILI on Fama-French three-factors model. In particular, all regressors are standardized, and all regressions include robust standard errors enclosed in parentheses.

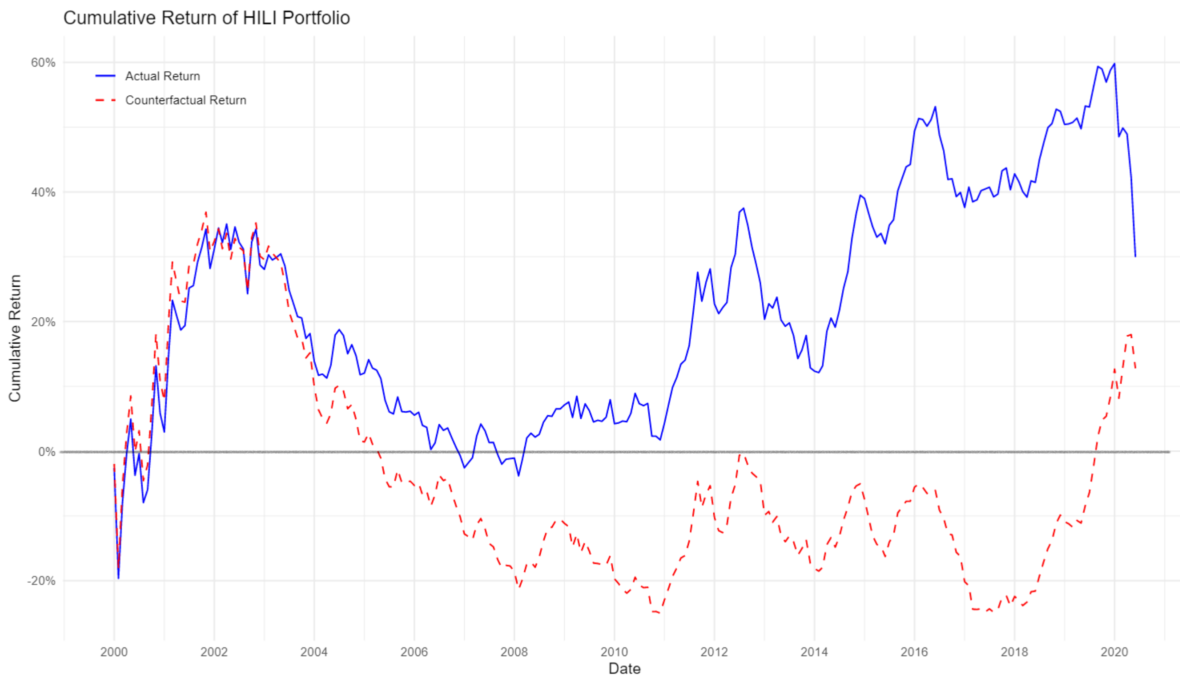

Figure 6: The cumulative return of the counterfactual HILI portfolio. The solid blue line represents the realized cumulative return on HILI portfolio. The dashed red line shows the counterfactual monthly returns of HILI which is estimated from column 1 of table 5. The counterfactual returns are measured by subtracting the fitted regression value minus the intercept from the realized returns.

allocated to stock i at time t by both purely passive and closet indexer mutual funds, including ETFs.

But first, I need a stock-level measure of passive flow, in order to estimate whether the passive flows at the individual level can affect the stock returns given their IXI score. The passive $\Delta\text{Flow}_{i,t}$ is the total change in the dollar amount of capital flowing to stock i at time t of all passive and closet-indexer mutual funds and ETFs. I compute this measure using the number of shares held for each stock i within each passive fund j using the Factset fund holding database and then compute the following:

$$
\begin{aligned}
\Delta\text{Shares}_{i,j,t} &= \text{Shares}_{i,j,t} - \text{Shares}_{i,j,t-1} \\
\text{Flow}_{i,j,t} &= \Delta\text{Shares}_{i,j,t} \times \text{Price}_{i,t} \\
\text{NetFlow}_{i,t} &= \sum_j \text{Flow}_{i,j,t} \\
\widetilde{\text{Flow}}_{i,t} &= \frac{\text{NetFlow}_{i,t}}{\text{AvgMktCap}_t} \\
\Delta\text{Flow}_{i,t} &= \widetilde{\text{Flow}}_{i,t} - \widetilde{\text{Flow}}_{i,t-1}
\end{aligned} \tag{14}
$$

Where $\text{Shares}_{i,j,t}$ represents the number of shares of stock i held by fund j at time t, and AvgMktCap_t is the average CRSP market capitalization for month t.

Table 6 displays the panel regressions in which individual stock returns in month t are analyzed against various predictors. These regressions include time-fixed effects, emphasizing the cross-sectional variation in returns. Column 1 starts with a single predictor, the stock's indexing inclusion score, denoted as $IXI_{i,t-1}$. The following columns add predictors for changes in stock flow and earnings measures. The earnings metrics are firm-specific versions of those used in Table 5. Column 4 introduces additional individual stock controls, such as the log of lagged market equity, the log of the lag book-to-market ratio, and stock returns from months $t-12$ to $t-2$. The final column includes the interaction between the lag IXI and the change in flow, while maintaining the same control variables.

The scenario in which only IXI serves as the regressor shows its notable positive correlation with the returns (observed in column 1). This aligns with the trend where high IXI stocks outperform, as mirrored in HILI. Even after introducing the change in flow and earnings variables in the regression (columns 3 and 4), the coefficient of IXI retains its positive significance. Hence, in line with the HILI findings, the association between indexing inclusion and returns remains distinctly positive, even after factoring in the impacts of passive flow and earnings shocks.

When considering the coefficients tied to the stock passive flow metrics, it can be interpreted that high-indexed stocks tend to outperform during periods of rising passive flow, as evidenced by the positive interaction term in column 5 which is not highly significant but still retains a t-score of 1.35. This observation aligns with the results derived from the HILI-centric regressions in Table 5.

Table 6: Individual Stock Returns and IXI

	Dependent variable:				
	$Ret_{i,t}$				
	(1)	(2)	(3)	(4)	(5)
$IXI_{i,t-1}$	0.048***	0.048***	0.033**	0.036**	0.036**
	(0.016)	(0.016)	(0.015)	(0.015)	(0.015)
$\Delta Flow_{i,t}$		0.039**	0.039***	0.114***	0.133***
		(0.015)	(0.015)	(0.030)	(0.033)
$IXI_{i,t-1} \times \Delta Flow_{i,t}$					0.034
					(0.025)
$Earnings\ Release\ Ret_{i,t}$			2.664***	2.184***	2.184***
			(0.015)	(0.015)	(0.015)
$\Delta Earnings\ Forecasts_{i,t}$			0.225***	0.125***	0.125***
			(0.015)	(0.015)	(0.015)
Controls	No	No	No	Yes	Yes
Observations	520,324	520,320	519,464	460,556	460,556
R^2	0.00002	0.00003	0.058	0.119	0.119

Note: *p<0.1; **p<0.05; ***p<0.01

The performance of individual stock monthly returns, is assessed using data spanning from 2000 to 2021. The $IXI_{i,t-1}$ is the stock's lagged Indexing Inclusion score. $\Delta Flow_{i,t}$ measures the individual change in stock passive flow. $Earnings\ Release\ Ret_{i,t}$ is the individual stock's sum of excess returns during the three-trading day window around the earnings release date. $\Delta Earnings\ Forecasts_{i,t}$ is the change in IBES analysts' median long-term forecast of earnings growth rate. Column (4) regression includes controls for log of lagged stock market cap, log of lagged book-to-market ratio, and the stock's return during the month $t-12$ to $t-2$. Notably, all regressors are standardized and all the regressions are clustered by month with robust standard errors enclosed in parentheses.

7.1 Size and Value

In recent years, there has been a noticeable underperformance of value and small-cap stocks as shown by figures A1 and A2. The small-cap premium identified in the early work of Fama and French (1993) has been elusive in recent decades, as noted by Asness et al. (2018). One potential explanation for this observed underperformance might be the increasing dominance of passive investors' portfolios by large-cap, predominantly growth stocks. As passive investing strategies, especially those tracking major indices, inherently tilt toward the largest market capitalizations and trending growth stocks, this could exert significant influence on stock prices and consequently impact the relative performance of value and small-cap stocks.

Figure 8 shows the percentage of total investors' wealth that is dedicated to the smallest cap stocks in the first quintile. The figure presents the diminishing contribution of the smallest cap stocks in the overall investors' portfolio over time.

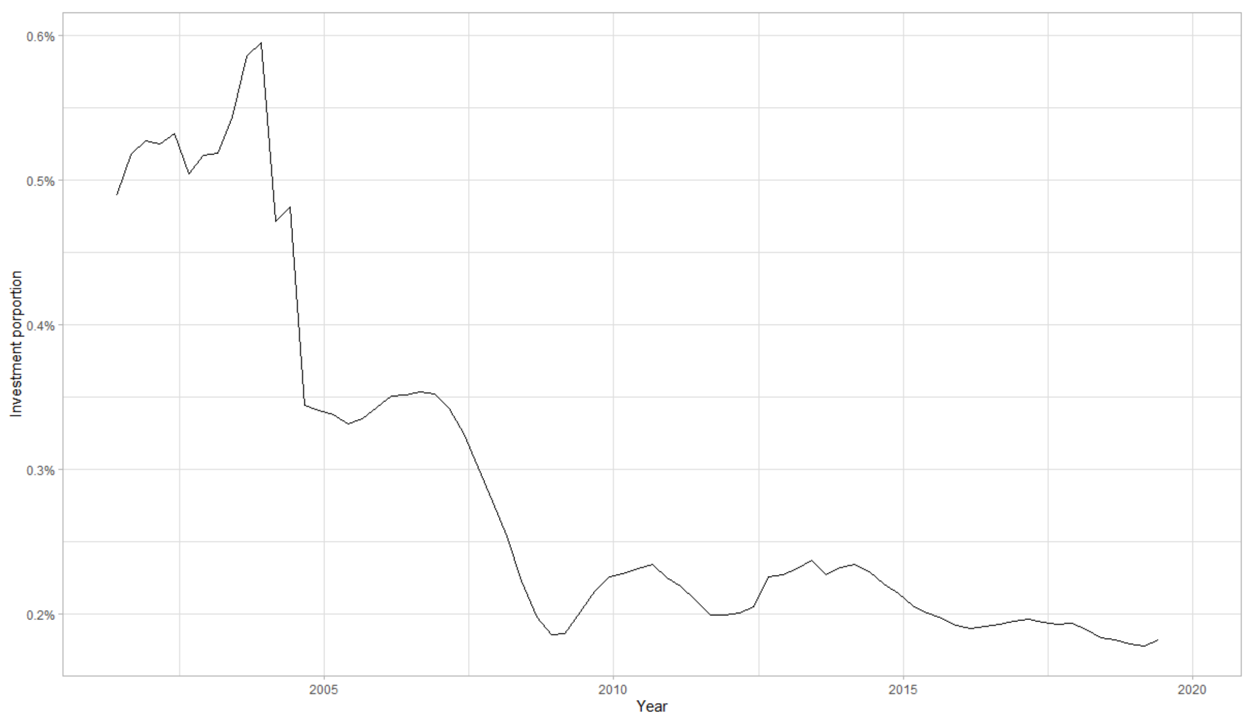

Figure 7: The proportion of average investors' AUM dedicated to smallest cap stocks over time. The y-axis shows what percentage of the total investor holdings is in the smallest cap stocks. The stocks are divided into 5 quintiles at each quarter.

Figure 8 offers a clearer understanding of passive and active ownership of the smallest capitalization stocks. The figure displays the passive and active investors' assets under management dedicated to the smallest cap stocks as a percentage of the overall market, taking into account the general growth of the investment market and the inflow and outflow of funds to these stocks. The figure shows a decreasing trend in active ownership of the smallest cap stocks, with passive ownership making up only a small percentage of the overall ownership of these stocks. Similarly, figure 9 illustrates the passive and active ownership of the largest capitalization stocks. In contrast to the smallest cap stocks, the overall ownership of the largest cap stocks as a percentage of the total market has remained stable over time. Passive ownership is rapidly increasing, while active ownership is decreasing.

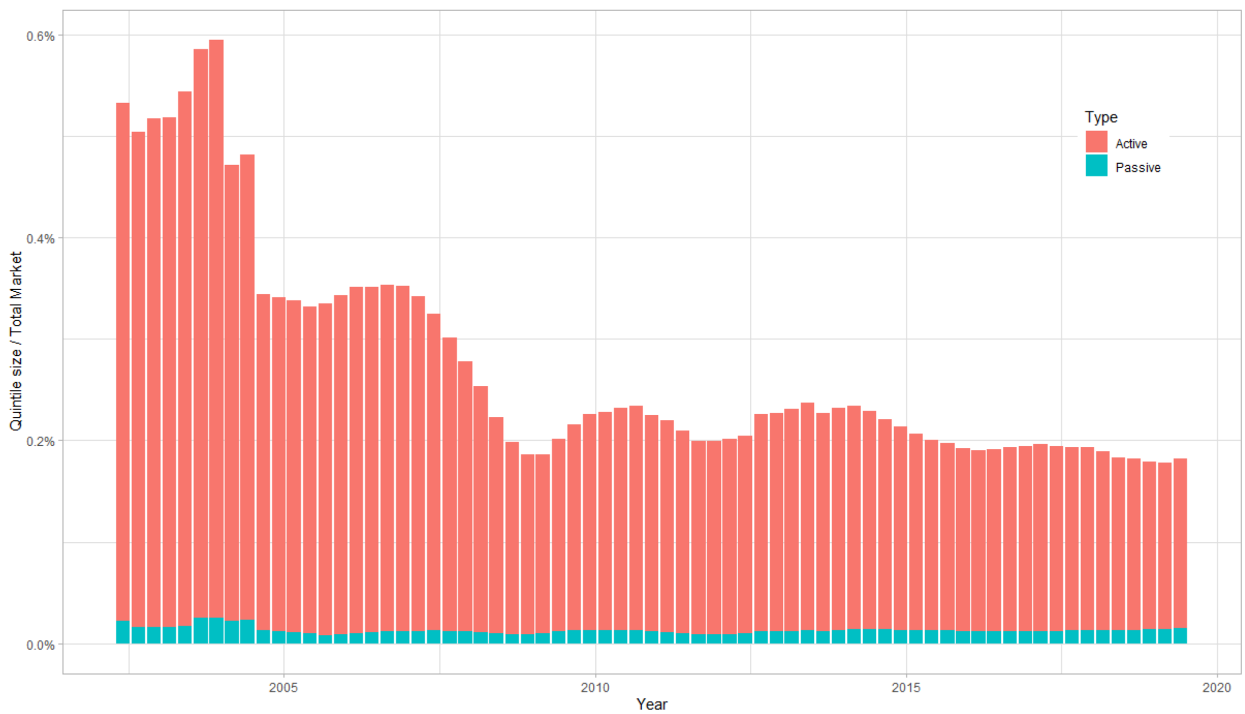

Figure 8: This graph shows the investor composition of the smallest cap stocks as a proportion of the total combined investors' AUM. The y-axis is the total aum that is invested in the smallest size quintile stocks as a percentage of the total aum invested in the whole market at each quarter

The strong correlations of HILI with value and size, as shown in table 3, raise an interesting question: How much has the superior performance of high-indexed stocks compared to low-indexed ones influenced the significant underperformance of value and size sectors over the last ten years?

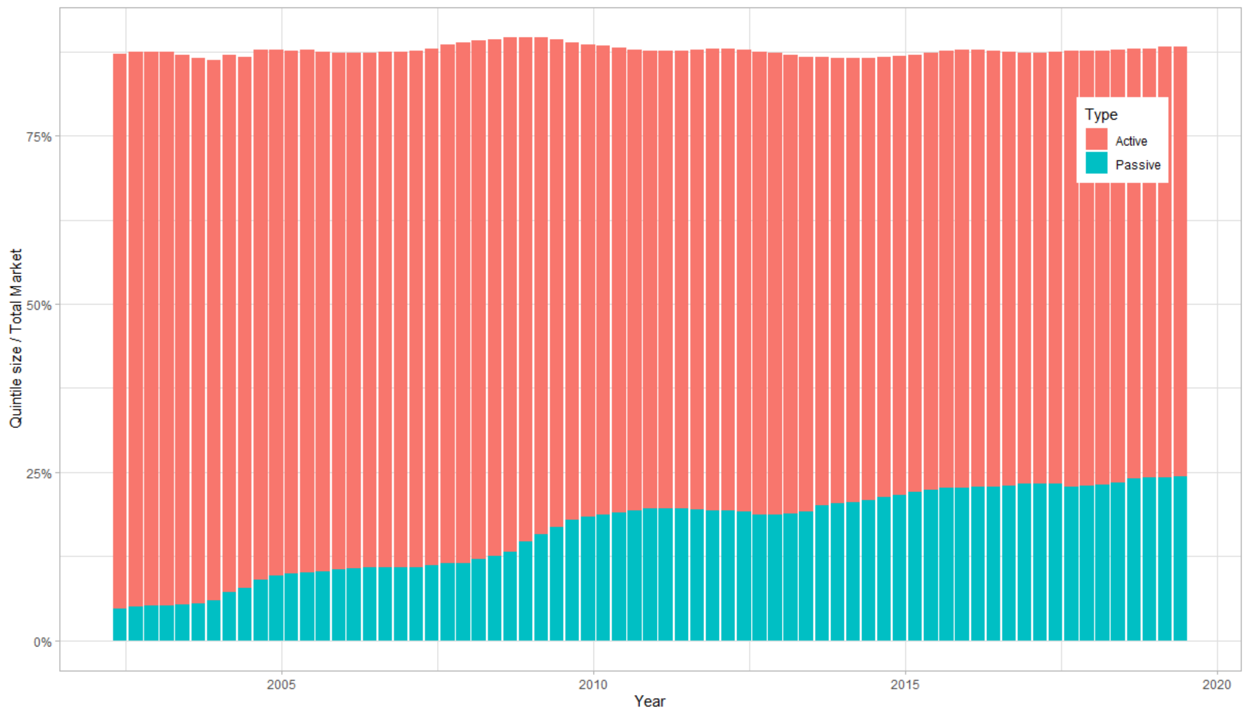

Figure 9: This graph shows the investor composition of the largest cap stocks as a proportion of the total combined investors' AUM. The y-axis is the total aum that is invested in the smallest size quintile stocks as a percentage of the total aum invested in the whole market at each quarter

To address this question, I use the adapted PST's two-factor model, where the key factors are the market portfolio and the HILI factor. As indicated in columns 2 and 4 of Table 7, the HILI coefficients are negative and highly significant. The alphas of HML and SMB in this two-factor model are less pronounced compared to when only the market is considered. For example, HML's initially significant alpha of approximately -61 bps reduces to an insignificant -31 bps, while SMB's significant alpha drops to nearly zero. Importantly, the underperformance significance of HML and SMB vanishes with the inclusion of HILI.

These findings indicate that approximately 50% of HML's negative alpha and almost all of SMB's underperformance can be linked to the strong performance of the IXI factor. Since value stocks generally have low-indexed traits and large stocks typically exhibit high-indexed features (mainly due to the dominance of large-cap and growth stocks in major indices), this analysis highlights the difficulties encountered by value and small-cap strategies, particularly as the move towards passive investing has accelerated over the past ten years.

Table 7: Explaining the role of the Indexing Inclusion Factor on Value and Size. From 2011 to 2021, I analyzed HML (columns 1 and 2) and SMB (columns 3 and 4) returns against the excess market return and HILI return. Returns are shown as percentages per month, and robust standard errors are in parentheses.

	Dependent variable:			
	HML		SMB	
	(1)	(2)	(3)	(4)
Constant	-0.608^{**}	-0.312	-0.389^{*}	-0.011
	(0.298)	(0.270)	(0.219)	(0.185)
MKT-RF	0.134	0.093	0.271^{***}	0.218^{***}
	(0.096)	(0.082)	(0.056)	(0.039)
HILI		-0.409^{***}		-0.522^{***}
		(0.118)		(0.095)
Observations	120	120	120	120
R^2	0.043	0.160	0.194	0.403
Note:				*p<0.1; **p<0.05; ***p<0.01

8 Conclusion

Over the past two decades, the momentum has shifted towards passive investment strategies. Although these strategies offer low costs and more predictable returns, their rising prominence has had significant effects, especially on high-indexed stocks in the US. Despite the recent strong performance of these stocks, there are signs pointing to a less promising future. This upturn is largely attributed to unexpected increases in passive ownership and the broader trend of moving from active to passive investment strategies.

The core of this study is the Indexing Inclusion ratio (IXI), a new measure of passive ownership. The IXI provides a detailed view that captures a broader range of passive investments, including closet indexers. In line with the adapted PST model, the data suggests that as the surge in passive strategies slows down and the market moves towards equilibrium, the expected returns for high-indexed stocks may diminish.

Another key finding is the concept of the *indexing premium*, which underscores the difference in expected returns between high and low-indexed stocks. A consistent negative indexing premium throughout the study period suggests that the strong performance of high-indexed stocks may have caught many off-guard.

This study also clarifies the ambiguous performance patterns of value and small-cap stocks. The rise of passive investment and the dominance of growth and large-cap stocks in passive portfolios can provide insights into their recent underperformance.

In summary, this research sheds light on the intricate relationship between passive investing trends, high-indexed stocks, and market dynamics. It serves as a reminder to investors about the importance of critical evaluation, emphasizing that past performance isn't always a reliable predictor of future results. The landscape of passive investing is evolving, and a measured approach is essential as we move forward.

References

C. Asness, A. Frazzini, R. Israel, T. J. Moskowitz, and L. H. Pedersen. Size matters, if you control your junk. *Journal of Financial Economics*, 129(3):479–509, Sept. 2018. ISSN 0304-405X. doi: 10.1016/j.jfineco.2018.05.006. URL `http://www.sciencedirect.com/science/article/pii/S0304405X18301326`.

I. Ben-David, F. Franzoni, and R. Moussawi. Exchange-Traded Funds. *Annual Review of Financial Economics*, 9(1):169–189, 2017. doi: 10.1146/annurev-financial-110716-032538. URL `https://doi.org/10.1146/annurev-financial-110716-032538`. _eprint: https://doi.org/10.1146/annurev-financial-110716-032538.

A. Beyer, D. A. Cohen, T. Z. Lys, and B. R. Walther. The financial reporting environment: Review of the recent literature. *Journal of accounting and economics*, 50(2-3):296–343, 2010.

P. Bond and D. Garcia. The equilibrium consequences of indexing. *The Review of Financial Studies*, 35(7):3175–3230, 2022.

A. Buss and S. Sundaresan. More risk, more information: How passive ownership can improve informational efficiency. 2020.

M. M. Carhart. On persistence in mutual fund performance. *The Journal of finance*, 52(1): 57–82, 1997.

Y.-C. Chang, H. Hong, and I. Liskovich. Regression Discontinuity and the Price Effects of Stock Market Indexing. *The Review of Financial Studies*, 28(1):212–246, Jan. 2015. ISSN 0893-9454. doi: 10.1093/rfs/hhu041. URL `https://doi.org/10.1093/rfs/hhu041`.

A. Chinco and M. Sammon. The passive ownership share is double what you think it is. *Journal of Financial Economics*, 157:103860, 2024.

J. H. Cochrane. The dog that did not bark: A defense of return predictability. *The Review of Financial Studies*, 21(4):1533–1575, 2008.

K. M. Cremers and A. Petajisto. How active is your fund manager? a new measure that predicts performance. *The review of financial studies*, 22(9):3329–3365, 2009.

M. Cremers, M. A. Ferreira, P. Matos, and L. Starks. Indexing and active fund management: International evidence. *Journal of Financial Economics*, 120(3):539–560, 2016.

R. M. Edelen, O. S. Ince, and G. B. Kadlec. Institutional investors and stock return anomalies. *Journal of Financial Economics*, 119(3):472–488, Mar. 2016. ISSN 0304-405X. doi: 10.1016/j.jfineco.2016.01.002. URL `http://www.sciencedirect.com/science/article/pii/S0304405X16000039`.

E. F. Fama and K. R. French. Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, 33(1):3–56, 1993.

E. F. Fama and K. R. French. The equity premium. *The Journal of Finance*, 57(2):637–659, 2002.

E. F. Fama and K. R. French. A five-factor asset pricing model. *Journal of financial economics*, 116(1):1–22, 2015.

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

X. Gabaix and R. S. J. Koijen. In Search of the Origins of Financial Fluctuations: The Inelastic Markets Hypothesis. SSRN Scholarly Paper ID 3686935, Social Science Research Network, Rochester, NY, Nov. 2020. URL `https://papers.ssrn.com/abstract=3686935`.

W. R. Gebhardt, C. M. Lee, and B. Swaminathan. Toward an implied cost of capital. *Journal of accounting research*, 39(1):135–176, 2001.

P. A. Gompers and A. Metrick. Institutional Investors and Equity Prices. *The Quarterly Journal of Economics*, 116(1):229–259, Feb. 2001. ISSN 0033-5533. doi: 10.1162/003355301556392. URL `https://academic.oup.com/qje/article/116/1/229/1938986`. Publisher: Oxford Academic.

L. Harris and E. Gurel. Price and Volume Effects Associated with Changes in the S&P 500 List: New Evidence for the Existence of Price Pressures. *The Journal of Finance*, 41(4):815–829, 1986. ISSN 1540-6261. doi: https://doi.org/10.1111/j.1540-6261.1986.tb04550.x. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1986.tb04550.x`. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.1986.tb04550.x.

C. R. Harvey and Y. Liu. Does Scale Impact Skill? *SSRN Electronic Journal*, 2016. ISSN 1556-5068. doi: 10.2139/ssrn.2872385. URL `http://www.ssrn.com/abstract=2872385`.

K. Hou, M. A. Van Dijk, and Y. Zhang. The implied cost of capital: A new approach. *Journal of Accounting and Economics*, 53(3):504–526, 2012.

K. Hou, C. Xue, and L. Zhang. Replicating anomalies. *The Review of financial studies*, 33 (5):2019–2133, 2020.

A. K. Kashyap, N. Kovrijnykh, J. Li, and A. Pavlova. The benchmark inclusion subsidy. *Journal of Financial Economics*, 142(2):756–774, 2021.

R. S. J. Koijen and M. Yogo. A Demand System Approach to Asset Pricing. *Journal of Political Economy*, 127(4):1475–1515, 2019. ISSN 0022-3808. doi: 10.1086/701683. URL `https://www.journals.uchicago.edu/doi/full/10.1086/701683`. Publisher: The University of Chicago Press.

C. M. Lee, E. C. So, and C. C. Wang. Evaluating firm-level expected-return proxies: implications for estimating treatment effects. *The Review of Financial Studies*, 34(4):1907–1951, 2021.

L. Pástor, R. F. Stambaugh, and L. A. Taylor. Dissecting green returns. *Journal of Financial Economics*, 146(2):403–424, 2022.

A. Pavlova and T. Sikorskaya. Benchmarking intensity. *The Review of Financial Studies*, 36 (3):859–903, 2023.

M. S. Scholes. The Market for Securities: Substitution Versus Price Pressure and the Effects of Information on Share Prices. *The Journal of Business*, 45(2):179–211, 1972. ISSN 0021-9398. URL `https://www.jstor.org/stable/2352030`. Publisher: University of Chicago Press.

W. F. Sharpe. The arithmetic of active management. *Financial Analysts Journal*, 47(1):7–9, 1991.

A. Shleifer. Do Demand Curves for Stocks Slope Down? *The Journal of Finance*, 41(3):579–590, 1986. ISSN 1540-6261. doi: https://doi.org/10.1111/j.1540-6261.1986.tb04518.x. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1986.tb04518.x`. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.1986.tb04518.x.

Appendices

Appendix A

Table A.1: Summary statistics somparison between IXI and BMI measures

	Time Period			
	Full Sample	**2000-2006**	**2007-2012**	**2013-2018**
Statistics IXI (Active Share Adjusted)				
IXI Mean %	9	3.2	9.6	13.5
IXI SD %	8.8	3.2	7.1	9.5
Min IXI	0	0	0	0
Max IXI	99	66	96	99
Avg no. of Indices	22.2	14.3	20.2	22
Statistics IXI (Not Adjusted)				
IXI Mean %	15.2	6.3	16.5	21.9
IXI SD %	13.5	5.6	11.8	15.1
Min IXI	0	0	0	0
Max IXI	99	98	97	99
Avg no. of Indices	22.2	14.3	20.2	22
Statistics BMI				
BMI Mean %	15.4	15.2	17.1	15.5
BMI SD %	8.9	5.8	9.3	10.7
Min BMI	0	0	0	0
Max BMI	98.7	57.4	98.7	70.1
Avg no. of Indices	9	7.5	9	8.3

The table reports the summary statistics for IXI and BMI measures. The first set of numbers in statistics IXI (active share adjusted) refers to the main measure of indexing used in this paper, which adjusts the contribution of each active fund to IXI measure by their active share. Non-adjusted statistics for IXI is constructed in a similar manner to BMI by assuming that all benchmarked active funds are fully closet indexer. The average number of Indices is at stock level. summary statistics for BMI measure is extracted from Table 1 of **?**.

Table A.2: High IXI minus Low IXI sorted portfolios performance Minus Megacaps

	Dependent variable:							
	HILI							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	0.343***	0.342***	0.382***	0.374***	0.290***	0.248**	0.369***	0.284*
	(0.104)	(0.109)	(0.110)	(0.100)	(0.103)	(0.099)	(0.101)	(0.149)
MKT-Rf		0.002	0.040	0.040*	0.067***	0.085***	0.046*	0.069*
		(0.029)	(0.029)	(0.024)	(0.025)	(0.025)	(0.026)	(0.036)
SMB			−0.186***	−0.190***	−0.130**	−0.137***	−0.192***	−0.105
			(0.071)	(0.056)	(0.051)	(0.050)	(0.055)	(0.071)
HML				0.095*	0.040	−0.012	0.099*	−0.053
				(0.054)	(0.049)	(0.058)	(0.053)	(0.047)
RMW					0.152***	0.160***		
					(0.049)	(0.044)		
CMA						0.138*		
						(0.080)		
MOM							0.012	0.049
							(0.035)	(0.031)
GMB								0.064
								(0.075)
Observations	253	253	253	253	253	253	253	144
R^2	0.000	0.00003	0.103	0.134	0.168	0.183	0.135	0.089

Note: *p<0.1; **p<0.05; ***p<0.01

The performance of the high IXI minus low IXI (HILI) portfolio, sorted on a monthly basis, is assessed using data spanning from 2000 to 2021. The market cap is winsorized to remove the top 2.5 percentile at every month. The outcome variable in this context is the HILI - the difference in returns between the high and low IXI portfolios. The excess market return is represented by Mkt-RF. SMB and HML correspond to size-sorted and value-sorted factors as proposed by Fama and French (1993). Moreover, profitability and investment factors are denoted by RMW and CMA as per Fama and French (2015). MOM is the momentum factor of Carhart (1997). GMB is the green minus brown factor of Pástor et al. (2022). Notably, all returns are expressed in percentages and at a monthly frequency, with robust standard errors enclosed in parentheses.

Figures

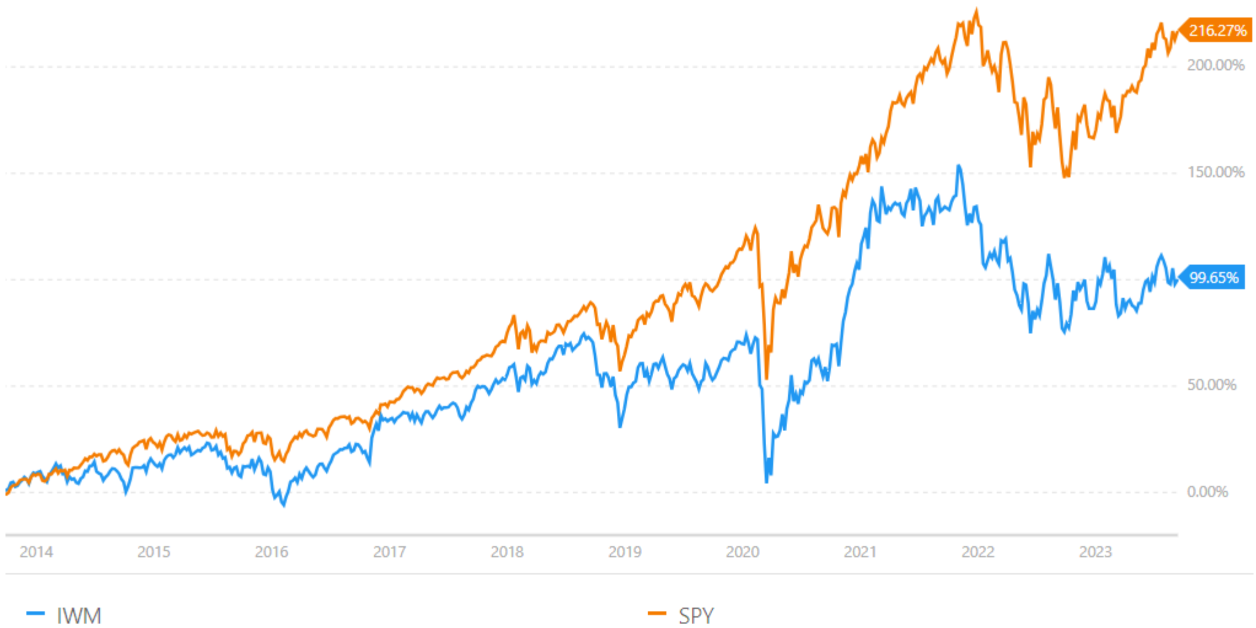

Figure A1: Cumulative return of Large cap and small Cap ETFs)

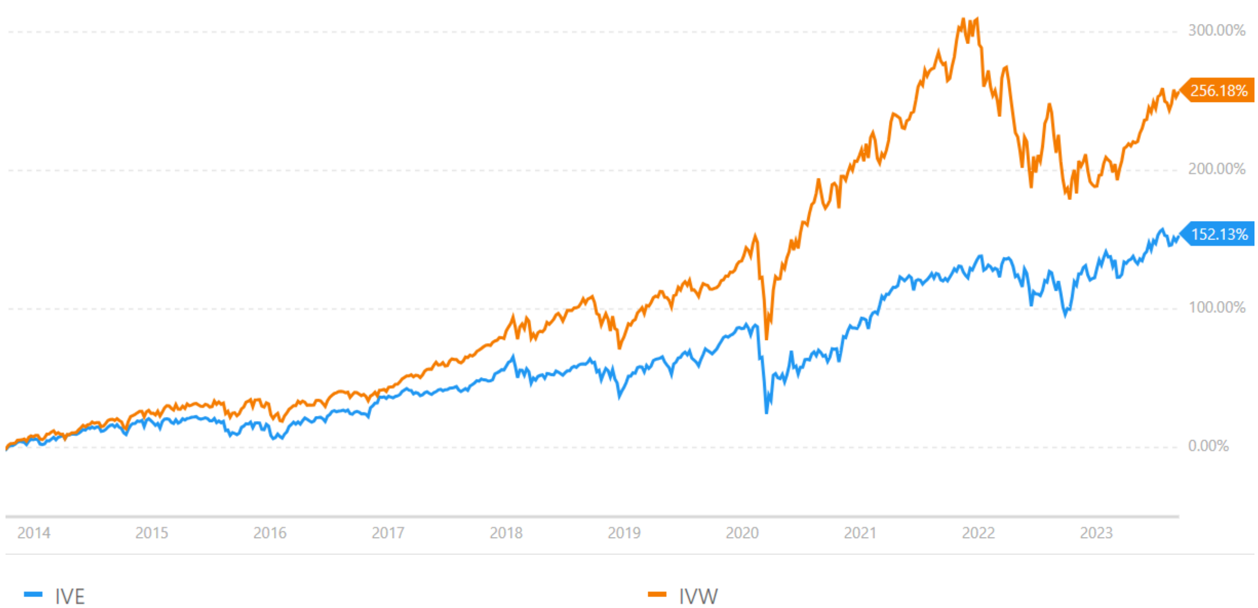

Figure A2: Cumulative return of growth and value ETFs)