A Tale of Two Index Funds: Full Replication vs. Representative Sampling*

Travis Dyer
Brigham Young University
travis.dyer@byu.edu

Nicholas Guest
Cornell University
nguest@cornell.edu

July 2022

Abstract
We examine the two approaches used by equity index funds to track their benchmark index. The first, full replication, mimics the index with exactness. The second, representative sampling, holds a subset of the index. We find that samplers trade 3-4 times more, have 30-50% higher expenses and fees, and earn 50-70 basis points lower annual returns, which is substantial given index funds' mandate to limit tracking error to a few basis points. Samplers' underperformance is not purely driven by higher expenses and transaction costs, but also poor stock picking. Overall, our analyses suggest representative sampling is detrimental to index investors.

Keywords: passive investing, index funds, full replication, representative sampling, fund performance, returns, portfolio turnover, expense ratio, management fees, stock picking
JEL Codes: D02, D14, G11, G23, G51, G53, J32

*We thank Sanjeev Bhojraj, Brian Boyer, Will Cong, Adrian Corum, Peter Cziraki, S.P. Kothari, David Ng, Jeff Sommer, Ian Wright, and Elisha Yu as well as seminar participants at Cornell University for helpful comments and suggestions. We also appreciate the research assistance of Michael Urbinato, Isaac Willyerd, and Elisha Yu and the financial support of Brigham Young University and Cornell University.

1. Introduction

Index funds' regulatory filings reveal relatively little about their investment operations, which are usually discussed in a mere few sentences or paragraphs. Moreover, what the funds do disclose can be misleading or opaque (e.g., deHaan, Song, Xie, and Zhu, 2021). Thus, we were surprised to find that they almost universally (99%) divulge which of two indexing approaches they use: full replication or representative sampling. Unlike full replicators, who attempt to passively mimic an index with exactness, representative samplers select a subset of securities from the index using indicators other than index weights to match index returns. Many of their indicators, such as dividend and earnings yields, are commonly used in active strategies. We find that representative samplers have higher turnover, higher expenses, and lower returns. These results suggest there is substantial heterogeneity in index fund approaches that can have economically large wealth implications for investors.

Passive investing approaches have grown in popularity in recent decades relative to active investing. As the argument goes, it is difficult, if not impossible, to outperform the market, consistent with a large body of research that finds that the average mutual fund's returns are equal to or lower than market returns (e.g., Fama and French, 2010). In addition, as its name suggests, passive investing is relatively easy and transparent, resulting in lower costs for a variety of reasons, including less trading and limited agency concerns. Thus, passive investors seek to match the market while minimizing costs. These investors often gravitate towards index funds, which exemplify passive investing by offering ownership in a large subset of the market with supposedly limited rebalancing and stock selection. As a result, index returns are usually close to the market return, and their fees are typically quite low, especially relative to other, more active mutual funds.

In the purest form of passive equity investing, an investor's portfolio includes each stock in the market in exact proportion to its weight in the market (i.e., the total stock market index). However, for several reasons, including that it is impractical for most investors to hold several thousand stocks, funds typically attempt to replicate only a subset of the market, known as an index. They do so using one of two methods.

First, owning each stock in proportion to the underlying index is known as full replication. This strategy is challenging for many reasons, including that it typically requires adjustments to all (i.e., tens, hundreds, or thousands) of the portfolio's positions each time an index adds or removes a stock.[1] Many of the required adjustments are small and pertain to relatively illiquid stocks, which creates the potential for large trading costs that reduce the benefits of replication. To mitigate these costs, some funds update their portfolio weights only periodically (e.g., monthly) instead of immediately adjusting positions each time an index is reconstituted. In addition, they may not update each position, especially those with a smaller initial index weight or a smaller change in index weight. These practices can reduce trading costs, leading to higher returns, but can also create the potential for substantial tracking errors that violate the fund's mandate to follow the index (Blume and Edelen, 2003).

The second approach, called representative sampling, selects only a subset of index components for inclusion in the investor's portfolio, but retains the goal of matching index returns. Of course, sampling creates the potential for even greater tracking errors and thus strays farther from the passive ideal. However, because the strategy requires holding fewer stocks, it may reduce

[1] The relative costs and benefits laid out in the next few paragraphs are frequently discussed by fund providers and finance pundits. For fund provider examples, see Vanguard (2021), Blackrock (2022), Fidelity (2022), and CIBC (2020). For finance pundit examples, see Investopedia (2020), justETF (2022), ETF Database (2016), ETF.com (2020), and Hedge Fund Research (2022). However, we have not found any large sample evidence to back up their arguments. Our intent with this paper is to address this dearth of empirical evidence.

trading costs, which would enhance returns. For example, because they do not hold the entire index, samplers might be able to avoid the most illiquid stocks or avoid trading following many instances of index reconstitution.

Some sampling funds select stocks based on size, such as by holding only the largest stocks in the index. However, many funds select on a variety of non-size metrics, such as dividends, earnings, and ESG ratings. In this regard, they are similar (at least in form, if not in substance) to traditional active investors who attempt to identify and exploit indicators of high expected returns. As a result, while replication provides exposure to a subset of the market captured by the index, sampling may *also* provide exposure to factors reflected in the sampling variables.[2] Since these variables often change absent changes to the index constituents, they might spur trading that would not otherwise occur if index weights were the only factor. As a result, trading costs could be higher and returns could be lower. Also, the additional effort inherent in sampling might lead to higher operating costs and management fees. Finally, the sampling manager's skill and luck (or lack thereof) at identifying reliable indicators of expected returns could help or hurt fund performance.

In summary, sampling has the potential to affect fund costs and returns through at least two channels. First, because the approach entails holding fewer positions, transaction costs could be lower, resulting in higher returns. Second, because the approach is based on factors other than index inclusion and size, it might entail more trading and effort that could either enhance or detract from cost and return performance. Ultimately, the relative strength of these two channels is an empirical question that we seek to answer in this study.

[2] Unfortunately, the total factor exposures are less clear for samplers because many of them do not disclose which sampling variables they use (see iShares example in Appendix A). Furthermore, even the samplers that do disclose their sampling variables do not disclose the extent to which they rely on each variable, or whether the weight they place on each variable is constant over time (see Fidelity example in Appendix A). While one may wish to circumvent this issue by inferring factor exposures from end-of-period holdings, of course these are not necessarily the holdings (and by extension, the exposures) that prevailed throughout the period or will prevail going forward.

We first identify which method a fund uses by manually examining the strategy descriptions in funds' prospectus filings. Appendix A provides several examples of these descriptions. Consistent with there being significant costs of replication and/or incentives for index fund managers to actively invest, we find that the sampling approach is common. In particular, among the U.S. equity index funds we study from 2010 to 2020, 37% are samplers. Another 11%, which we label "hybrids," claim that they typically use replication but may implement sampling under certain conditions (e.g., periods of market illiquidity). As noted previously, practically all index funds provide this information, suggesting that fund managers believe their replication vs. sampling choice has a significant bearing on their ability to meet the fund's investment objectives.

We also use the prospectuses to manually identify the index that each fund follows so that we can include index (or, in other words, 'benchmark' or 'target portfolio') fixed effects in our regression model.[3] This helps us rule out a variety of concerns, including the possibility that our results arise because it is easier to replicate some indices than others, perhaps because of differences in the number of stocks included or the liquidity of the constituent stocks.[4]

We show that sampling funds have higher turnover than replicating funds. This suggests that the active component of sampling, or the selection of stocks using variables other than index weights, more than offsets any reduction in trading arising from holding fewer positions. We also find that sampling funds have higher expense ratios and management fees, consistent with the costs of active selection more than outweighing the benefits of holding fewer positions, and with fund managers seeking compensation from investors for their efforts to actively invest.

[3] Throughout the paper we refer to an 'index' as the basket of securities that one or more mutual funds may attempt to track. Indices are also frequently referred to as funds' benchmarks or targets. In contrast, we use the term 'index *fund'* (emphasis on inclusion of the word fund) to refer to a mutual fund or ETF that tracks a pre-defined index.

[4] In other words, the tradeoff between tracking error and trading costs/returns might vary from index to index. These cross-index differences in the ease of replication could explain our results if the propensity of the replication and sampling approaches also varies across indices.

However, our examination of fund returns suggests these higher expenses and fees are not warranted because the sampling fund managers do not appear to be skilled at active investing. In particular, sampling funds' returns are lower than replicating funds. This is true of both net and gross returns, suggesting that higher expenses are not entirely to blame. In fact, only about 25% of the return difference appears to be explained by higher expenses, leaving 75% to be explained by other factors such as additional turnover and poor stock picking. To distinguish between turnover and stock picking explanations, we find similar results using measures of portfolio returns that ignore transaction costs, suggesting that both explanations contribute to the underperformance. These findings are potentially surprising given that the prior literature on index funds emphasizes the potential for sampling to reduce trading costs and enhance returns (Blume and Edelen, 2003). We argue that the prior literature overlooks the potential for sampling to lead to additional trading and suboptimal stock picking by index fund managers who are ill-suited for active investing.

We also examine return volatility to understand whether sampling affects funds' riskiness. On the one hand, selecting a subset of the index makes the sampler's portfolio mechanically less diversified, and thus riskier, by definition. On the other hand, the sampling indicator(s) may favor less-risky stocks. Relatedly, holding only the largest stocks in an index may avoid some of the small stock risk premium that has been documented in prior research. The same (or converse, depending on the factor) could be said of sampling that favors high value, profitability, or investment stocks (Fama and French, 2015). Empirically, we observe that the sampling funds have marginally higher return volatility and significantly lower risk-adjusted returns (i.e., return divided by the standard deviation of returns), which suggests the lower returns we document are not the result of taking on less risk.

Several additional analyses support and extend our main results. First, our results hold in subsamples of S&P 500 indexers and other market-cap-based indexers, which helps rule out concerns that our findings are driven by one or a few peculiar indices, by "style" or "sector" funds, or by unobservable cross-index differences. Second, we find that our results are strongest among funds following indices with fewer constituent stocks, and that they entirely disappear for samplers following indices with 1,000 or more stocks. This suggests sampling is not harmful only when it can drastically reduce the number of stocks held in the portfolio. Third, we find that investors' funds increasingly flow to samplers relative to replicators over our sample period, which is puzzling given our cost and return results. While this might be explained by the formidable search frictions facing investors (Hortacsu and Syverson, 2004; Choi, Laibson, and Madrian, 2010), we leave it for future research to more fully resolve this puzzle.

The differences in costs, returns, and flows we document are economically significant. For example, replicators outperform samplers by about 60 basis points (bps) per year on a net return basis. To illustrate the potential wealth effects of this difference, consider a hypothetical investor who makes a one-time index investment of $100K at 35 years old and holds the investment for the next 30 years. Assuming a constant 8% annual return, the investor's holding will be worth about $1,000K at age 65. However, if annual returns are 60 bps lower (i.e., 7.4%), then the value of the investor's holding would only be about $850K at age 65. This $150K, or 15%, difference in portfolio value is approximately equivalent to losing the last two years of returns over the 30-year horizon. Our assumptions in this hypothetical illustration are motivated by the popularity of index funds as a passive vehicle for retirement savings, often through employer-sponsored plans. Financial advisors, the media, and others often encourage workers to "set and forget" their investments for long periods of time, using index funds as a preferred store of value because of

their simplicity, safety, and low costs (e.g., Forbes, 2015; MarketWatch, 2021a, 2021b). In aggregate, the potential wealth implications of investors' choice between samplers and replicators are substantial, e.g., the index funds in our sample collectively manage trillions of dollars.

To our knowledge, this study is the first to compare the investment outcomes of replication vs. sampling index funds, which many still consider to be the closest thing to a purely passive investment. Our findings are especially relevant to the vast literature examining the many types and aspects of active investing. These studies usually exclude index funds based on the assumption that they are passive (e.g., Pastor, Stambaugh, and Taylor, 2020); our evidence suggests this may not be an appropriate design choice in many cases.

Our paper also builds on prior research that studies the extent to which funds' disclosures accurately represent their investment syles and holdings (e.g., Chan, Chen, and Lakonishok, 2002; Wermers, 2012; Chen, Cohen, and Gurun, 2021; deHaan et al., 2021). For example, some funds are accused of being closet indexers, which means that they claim to actively manage assets but end up with a portfolio that looks very similar to their benchmark (Chan et al., 2002; Berk and Green, 2004; Cremers and Petajisto, 2009). Our study suggests the opposite can also be the case, that is, many funds claim to passively manage assets but end up with performance that strays significantly from the benchmark index due to a portfolio that appears to be based on active strategies (see also Akey, Robertson, and Simutin, 2021).

A few other recent papers suggest that some index funds and ETFs are more active than many people realize (Cheng, Massa, and Zhang, 2019; Easley, Michayluk, O'Hara, and Putnins, 2021; Akey et al., 2021). These studies largely use measures of activeness from the academic literature, such as factor exposures (Amihud and Goyenko, 2013), active weights (Doshi, Elkamhi, and Simutin, 2015), and the combination of portfolio turnover and liquidity (Pastor et al., 2020).

In contrast, our paper examines funds' replication vs. sampling choice, a potential dimension of activeness that is reported by practically every index fund but is heretofore neglected by the academic literature.

Most importantly, our findings should be useful to fund managers trying to decide how to track an index, to plan sponsors selecting investment options for an organization's employees, and to the ultimate investors trying to evaluate their index fund managers.[5] The disparate approaches and outcomes of replication vs. sampling have been surprising to financial economists (including both academics and practitioners) with whom we have shared our results thus far. To us, this suggests that most mom-and-pop investors, and even many finance professionals, are likely similarly unaware of the distinctions.

2. Sample and Classification

2.1 Index Fund Data

We collect data on index fund performance from 2010 through 2020 using the CRSP Survivorship-Bias-Free Mutual Fund Database (CRSP MFDB). The beginning of our sample coincides with the availability of 497K prospectus documents that we use to classify funds, which we explain in detail in the next section. We obtain fund holdings from CRSP MFDB instead of from Thomson Reuters Mutual Fund Holdings for multiple reasons. For example, CRSP checks fund prospectuses and contacts fund management to collect voluntarily disclosed holdings more often than Thomson Reuters.[6] Also, Thomson Reuters misses many new U.S. equity mutual fund

[5] While we emphasize the use of index funds by passive investors, many active investors also use index funds as vehicles to get exposure to certain factors (for hedging and other reasons, see Cong and Xu, 2016). Of course, in these cases returns might be secondary to other objectives (e.g., exposure to some smart beta variant). However, we think such investors would still be intrigued to know that, without sacrificing their main objective(s), they can often achieve much higher returns by choosing a replicator over a sampler.

[6] https://wrds-www.wharton.upenn.edu/pages/support/support-articles/crsp/mutual-fund/tfn-mutual-fund-holding-vs-crsp-mutual-fund-holdings/?_ga=2.8750065.514058509.1606764681-1628130839.1581111764

share classes after 2008 (Zhu, 2020), which is potentially important for our study because our sample period begins in 2010.

Consistent with prior literature on index funds, we focus on equity funds due to the vastly different risk-return profiles of other asset classes as well as difficulty identifying the index that the fund is tracking in other asset classes. We keep funds that CRSP flags as tracking an index (CRSP MFDB variable index_fund_flag = 'D'), including ETFs. We require nonmissing data on funds' turnover, expense ratio, returns, total net assets, and fund flows. As explained in the next section, we manually identify the index followed by each fund using their prospectus. We then use the CRSP mutual fund holdings data to approximate the number of stocks in the indices followed by funds in our sample, and we remove any fund following an index for which we can not calculate this measure. We also require that a fund's index be followed by both replicator(s) and sampler(s) (or a hybrid of the two) in our sample in order to ensure variation in our measure of interest once we include benchmark index fixed effects. To be clear, when we refer to a 'benchmark index' or simply to an 'index' without using the word fund, we are referring to the index a fund is tracking, rather than the fund itself. These criteria result in 3,365 fund-year observations.[7]

2.2 Index Fund Classification: Full Replication vs. Representative Sampling

To classify index funds as replicators or samplers, we use Form 497K from SEC's EDGAR database, which is a required filing for all funds beginning in 2010.[8] Form 497K is a summary prospectus filed for individual funds, which is ideal for our purposes because we are interested in individual funds' investment strategies. Funds also file Forms 485APOS and 485BPOS, full (or

[7] Our research design intentionally focuses on *within* benchmark variation, which inherently omits benchmark indices that are tracked by only one fund type. In later analysis (see Section 4.5 and Table 10), we examine the generalizability of our main results to a broader sample of 12,969 fund-year observations that is not subject to such strict selection criteria. Using this broader sample, we reach similar conclusions.

[8] See SEC Rule No. 33-8998, issued in 2009.

"statutory") prospectuses that often pertain to multiple funds (e.g., a fund family). Because there is often substantial variation in strategy within a fund family, using 485APOS and 485BPOS filings would introduce substantial noise into our tests. Another reason to use 497Ks is that we are able to automate the identification and extraction of their Principal Investment Strategies (PIS) section, which is where funds indicate whether they employ full replication or representative sampling.

To determine whether a fund uses full replication or representative sampling, we manually read the PIS in funds' 497Ks.[9] Rather than reading every 497k disclosed by an index fund over time (approximately 26,000 filings), we randomly selected and read one 497k for each fund in order to make this manual task more feasible (approximately 2,300 filings). As a result, a fund's classification is fixed over our sample period. The classification task is fairly straightforward because funds almost always use relatively standardized language to disclose whether they use full replication, representative sampling, or full replication as the default with the option to switch to representative sampling under certain conditions. We refer to the latter group as "hybrid" funds. Appendix A provides several excerpts from the prospectuses of funds from all three groups. Of the funds in our sample, about 52% are replicators, 37% are samplers, and 11% are hybrid. We were able to classify all but a handful (~1%) of funds, which do not discuss replication or sampling and are excluded from the sample.

While the choice to replicate vs. sample appears to be fairly stable over time in general, it is possible that funds could change approaches over time. In untabulated robustness tests (available upon request), we employ a computer-based textual analysis methodology to classify funds and reach similar results and identical inferences as those that are tabulated and discussed in the paper.

[9] We had multiple research assistants complete the initial classification, and then both coauthors manually reviewed and checked the prospectuses and resulting classifications to ensure their accuracy.

Of note, this methodology allows for time-series variation in strategy since the computer classifies each fund-year's 497k filing. The approach is based on keyword identification (similar to Dyer, Guest, and Yu, 2021). However, we find that a keyword approach misclassifies many funds, prohibiting us from distinguishing whether a fund uses replication or sampling. For example, some funds explain that they "use representative sampling instead of an exact replication approach." Other funds explain that they employ "full replication of a benchmark index that is constructed to provide a statistically representative sample of the market." In both of these scenarios, the fund uses the word stems of both "replicate" and "sample" in their description, even though they are clearly one or the other. Thus, to avoid misclassifying or excluding these funds, we rely on our manual classification throughout the main analyses presented in the paper.

2.3 Variable Definitions

Based on the index fund classification described in the prior section, we create the categorical variable *Sampling*, which is set to 1 for samplers, 0.5 for hybrids, and 0 for replicators. This is the key independent variable in our analyses. We control for several additional fund-specific variables of interest. *Fund Size* is the fund's total net assets (TNA), or total assets minus total liabilities, reported in millions (CRSP MFDB variable mtna). *Fund Age* is the inception date of the fund. *Big Three* is an indicator set to one if the fund is offered by Vanguard, BlackRock, or State Street, and zero otherwise. *ETF* is an indicator set to one if CRSP flags the fund as exchange-traded (CRSP MFDB variable et_flag = 'F'), and zero otherwise.

We also examine and control for several aspects of the index that the fund follows. *Domestic* is an indicator set to one if the index is primarily made up of U.S. stocks (e.g., MSCI USA Large Cap 300 Index), and zero if the index is entirely made up of foreign stocks (e.g., S&P Europe 350 Index) or has a global focus (e.g., NASDAQ OMX Global Agriculture Index).

#Constituents is our best approximation of the number of stocks in the index. Rather than manually looking up the number of constituents in each index (which CRSP does not provide), we calculate the average number of stocks held in the portfolios of replicators following the index. This assumes that the replicators are truthful when they say they "generally invest in all stocks included in the index" (see Dreyfus International example in Appendix A).[10]

Additionally, we follow CRSP methodology in creating indicators to capture the three different ways indices segment the market. To be specific, *Market Cap Index* is set to one for funds following an index made up of a segment of the market based primarily on size (e.g., Russell 1000 Index, CRSP US Large Cap Index, and MSCI USA Small Cap Index). *Style Index* is set to one for funds following an index made up of a segment of the market based primarily on a certain style (e.g., JP Morgan US Value Factor Index, MSCI USA Momentum Index, and Pacer US Cash Cows Growth Index). *Sector Index* is set to one for funds following an index made up of a segment of the market based primarily on a certain sector (e.g., Alerian Energy Infrastructure Index, Dow Jones US Consumer Goods Index, and S&P Developed Ex-US BMI Health Care Sector Index). In regressions, we omit *Market Cap Index* to avoid perfect collinearity between these three indicators. Finally, we also include a linear time trend (*Trend*) in our regressions, except when it is subsumed by time fixed effects, to help us understand the extent to which our variables of interest change during our sample period.

The dependent variables in our tests are meant to measure fund costs, performance, and flows. *Turnover* is the minimum of aggregated sales or aggregated purchases of securities, divided by the average 12-month TNA of the fund (CRSP MFDB variable turn_ratio). *Exp Ratio* is the ratio of total investment that shareholders pay for the fund's operating expenses (CRSP MFDB

[10] To the extent they are not truthful and do not hold all stocks in the index, *#Constituents* will be understated.

variable exp_ratio). *Mgmt Fees* and *12b-1 Fees* are the management fees and marketing/distribution fees, respectively, reported in the fund's Statement of Operations (i.e., CRSP MFDB variables mgmt_fee and actual_12b1), expressed as a percentage of TNA. Both are components of *Exp Ratio*. *Net Returns* is the fund's cumulative buy-and-hold return for the year (based on monthly returns, i.e., CRSP MFDB variable mret). *Gross Returns* is the sum of *Net Returns* and *Exp Ratio*. *Return Volatility* is the standard deviation of monthly returns. *Risk Adj Returns* is the ratio of *Net Returns* and *Return Volatility*. *Fund Flows* is the difference between the ratio of ending TNA to beginning TNA and one plus *Net Returns*. We winsorize fund-level continuous variables at the 1% and 99% levels by year to limit the influence of outliers.

2.4 Descriptive Statistics

Panel A of Table 1 presents distributional statistics for the variables used in our analyses. The mean of *Sampling* is 0.42, which indicates that just under half of our sample uses representative sampling either exclusively or occasionally. The average fund manages about $2.8 billion of assets, but the median of $224 million shows that fund size is substantially right skewed. The typical fund in our sample is mature, with a mean and median age of about 11 years and 75 percent of funds having been offered for at least five years. The Big Three provides 28% of the index funds in our sample, 18% of index funds are exchange traded, and 83% focus exclusively on U.S. stocks. An 80% majority of funds follow indices that are primarily based on market capitalization, while 18% follow a style index and only 2% follow a sector index.

Regarding *#Constituents*, the mean (median) index followed by funds in our sample holds 896 (507) stocks. The right skewness arises mainly because of a handful of extremely large indices (especially the FTSE Global All Cap Index as well as the Russell 2000 and 3000). Also, note that

the 25th percentile and median are both 507, which pertains to the third of our sample that is made up of funds that track the S&P 500 index.[11]

As is to be expected for index funds, turnover is relatively low, with a mean (median) of 0.39 (0.14) percent of TNA. This level of turnover is much lower, about half as much to be specific, than in recent studies on more active mutual funds (e.g., mean [median] turnover is 0.70 [0.51] in Dyer, Guest, and Yu, 2021). The average expense ratio is 0.51 percent of TNA, and the average management fee (12b-1 fee) is 0.20 (0.15) percent of TNA. The average fund return is about 12 to 13 percent. Of course, returns are subject to substantial variation, as evidenced by the 13 percent standard deviation of *Net Returns* and *Gross Returns* as well as the 4 percent mean of *Return Volatility*. Finally, consistent with prior research documenting a substantial shift of capital to passive "indexing" approaches in recent decades, average fund flows are positive and almost one percent of TNA. However, the mean is much higher than the median (which is about zero) and the 75th percentile, meaning that a few funds had huge positive flows.

Panel B of Table 1 presents and compares the means of our variables of interest for samplers and replicators. Obviously, many of the means differ significantly across the two groups. Rather than discuss these differences in detail here, we further examine them (or attempt to control for them when necessary) in the regression-based analyses that follow.

2.5 Determinants of Replication vs. Sampling

In Table 2, we regress *Sampling* on other fund and index characteristics to better understand the determinants of funds' choice to use replication vs. sampling. In the first column, we include

[11] As of early 2022, the S&P 500 consisted of 500 companies that issue a total of 505 stocks because a few companies issue multiple classes of shares. In addition, many funds disclose in their prospectus that they may temporarily (and to a limited extent) hold stocks that are not in the index for liquidity and other reasons (e.g., delayed rebalancing following index reconstitution). We think these factors explain why *#Constituents* is slightly above 500 for the S&P 500 index funds in our sample.

fund and index characteristics, but in the second column, the index characteristics are subsumed by Year-Benchmark Index fixed effects. Both columns show that sampling is more likely among older funds and ETFs. The former association is consistent with funds switching to sampling as they gain experience and the trust of their investors.[12] Perhaps as managers become more familiar with the mechanics of index investing, they feel more confident in their ability to take on additional complexity, such as by adding a sampling component that involves collecting and acting on additional data. Alternatively, perhaps managers tire of tracking every single stock in an index over time and gradually switch to sampling. The positive association between sampling and ETFs is consistent with recent research by Cheng et al. (2019) and Easley et al. (2021), suggesting that ETFs are more likely to attempt to develop informational advantages and actively invest than more "passive" index mutual funds.

In the first column, we learn that several of the index characteristics are also associated with the propensity for funds to use sampling. Specifically, the coefficient on *Domestic Index* is significantly negative, suggesting funds focusing on domestic stocks are less likely to use sampling compared to their foreign fund counterparts. This may be because the additional dimensionality facing foreign funds (e.g., whether to focus on stocks in developed vs. emerging markets) increases the usefulness of sampling to whittle down the potential set of holdings. It could also be the case that institutional frictions, such as cross-country differences in regulations and openness to foreign investment, make it more challenging to replicate a foreign index by holding each constituent stock.

[12] Recall that we fixed each fund's classification to facilitate manual data collection, so our time-series "switching" interpretation is based on cross-sectional variation, which is not ideal. Thus, in untabulated analyses we confirmed that the coefficients on *Log(1+Fund Age)* are similar if we instead use the computer-based classification that can vary over time if a fund switches their approach (see Section 2.2 for more detail).

We also find that the log of the number of stocks in the index, *Log(#Constituents)*, is positively associated with sampling, which is consistent with the idea that replication is more difficult for more populated indices. Besides the obvious, that it takes more work to identify, trade, and track more stocks, the indices with more constituents also mechanically tend to include smaller stocks. To illustrate, consider that in early 2022 the average market cap of components of the Dow Jones 30 and S&P 500 were about $370 billion and $76 billion, respectively. Of course, the difference in this illustration would be even more stark if we replaced the S&P 500 with, say, the Russell 3000. Since smaller stocks tend to be more illiquid, replicating an index that includes them can be relatively costly, especially since smaller components of an index are typically added or removed more frequently.

Finally, the coefficients on *Style Index* and *Sector Index* are significantly positive, indicating that funds are more likely to be samplers when they follow an index based on a particular style or sector than when the index is based on market cap (recall that *Market Cap Index* is the omitted category in the regression). This is to be expected, given that funds following a style or sector index have already decided to stray from the passive ideal of holding the entire market to an extent, even ignoring the replication vs. sampling decision. In other words, if they have decided to tie their portfolio to an index that is based on factors other than market cap, then it makes sense that they would also use factors other than market cap to select from the stocks *within* that index. Interestingly, the coefficient on *Trend* is insignificant, suggesting that the popularity of the sampling approach is about the same at the beginning and end of our sample period.

Overall, the fact that many of the fund and index characteristics load in our determinants regressions and that the adjusted R^2 values are 12 to 15 percent, suggest we have made a good start identifying the factors underlying funds' replication vs. sampling decision. These findings also

suggest we should include these determinants as controls in our later regressions, given that many of them are associated with the replication vs. sampling decision and are probably also associated with fund outcomes.

3. Research Design And Main Results

3.1 Research Design

To compare the costs, performance, and flows of index funds that use full replication and representative sampling, we estimate one of the following two regressions:

$$Outcomes_{i,t} = \beta_0 + \beta_1 \times Sampling_{i,t} + \lambda \times Fund\ and\ Index\ Controls_{i,t} + \varepsilon_{i,t} \quad (1)$$

$$Outcomes_{i,t} = \beta_0 + \beta_1 \times Sampling_{i,t} + \lambda \times Fund\ Controls_{i,t} + Fixed\ Effects + \varepsilon_{i,t} \quad (2)$$

where *i* denotes funds; *t* denotes years; *Outcomes* denotes our investment outcomes of interest (including turnover, expenses, returns, and flows); *Sampling* is our categorical variable identifying samplers, replicators, and hybrids; *Fund and Index Controls* and *Fund Controls* denotes the set of control variables defined in Section 2.3; and *Fixed Effects* is a set of indicators set to one for each year-benchmark index, and zero otherwise.

As noted previously, our sample only includes funds if their index is followed by multiple types of funds (e.g., at least one replicator and one sampler) each year in order to ensure variation in our measure of interest once we include year-benchmark index fixed effects. This allows us to use a constant sample in estimating equations (1) and (2), which limits generalizability but avoids the possibility for changes in sample composition across our tests to explain any of our results.[13]

[13] In Section 4.5 and Table 10, we relax this constraint and observe similar results, which suggests the effects of sampling vs. replication are fairly generalizable.

Besides including fixed effects, in all of our tests standard errors are clustered at the fund level to account for the almost certain interdependence among observations of the same fund.

3.2 Turnover and Costs

Given that the primary rationale provided (by funds, pundits, and academics) in support of sampling is to reduce turnover and operating expenses, our investigation begins with these outcomes. In contrast with the reasoning provided in support of sampling, it could also result in higher turnover and operating expenses because it is based on factors (such as dividend yield) that can vary more frequently, or at the very least differently, than index weights. As a result, a sampler might at times adjust their portfolio even in the absence of changes to the index.

In Figure 1, we plot the mean turnover and expense ratios separately for replicators and samplers over our sample period of 2010-2020. The figure shows that samplers have higher turnover and expenses than replicators every year, often by substantial amounts (e.g., 3-4 times more turnover). In order to better understand the extent to which these differences are incremental to other characteristics, whether the results hold within benchmark index, and whether the results are statistically significant, we next report regression estimates of equations (1) and (2).

Specifically, the estimates in Table 3 indicate that sampling funds have significantly higher turnover and operating expenses than replicating funds. In particular, the coefficient on *Sampling* in column one (three) is a significantly positive 0.440 (0.352), which is about the same magnitude as the unconditional mean of turnover in Table 1. This evidence that samplers turnover their portfolio so much more than replicators (i.e., more than double) casts serious doubt on the claim that sampling helps funds limit trading. Similarly, focusing on the more conservative estimates from the expense ratio regression with Year x Benchmark Index fixed effects (column four), we

find that expense ratios are about 0.14 percent higher for samplers than replicators.[14] This difference is economically substantial at about 28 percent of the unconditional sample mean. Together, these results suggest that sampling entails more trading, which generates additional operating costs that managers pass on to the fund's investors.

For the most part, the coefficients on the control variables in Table 3 seem reasonable. For example, larger funds have lower turnover, which is consistent with the principle of inertia, and older funds have more turnover, consistent with their greater likelihood of being samplers. Similarly, funds tracking style indices have higher turnover, which likely arises because their index is updated more frequently as the underlying style variables (e.g., the PE ratio) evolve. However, since we do not have strong theoretically motivated priors or predictions about the control variables, we caution against putting too much weight on these conjectures.

3.3 Return Performance

In this section, we examine whether the different investment operations of replicating and sampling funds affect the returns they provide to their investors. We first examine the returns reported by the fund on both a gross and net basis, in order to understand the extent to which our findings are driven by expenses. We also use return volatility to investigate whether risk differences can explain our results. Finally, we calculate buy-and-hold portfolio returns to help us distinguish between transaction costs and stock picking explanations.

3.3.1 Fund Returns

[14] In Section 4.4 and Table 9 we show that this expense ratio result is partly driven by management fees. These fees are one of the major components of a fund's expense ratio, so much so that the two are sometimes (incorrectly) used interchangeably. We expect any findings regarding turnover and operating expenses to be similar for management fees, which is the compensation managers demand from investors for their efforts to oversee the fund's trading and other operations. In contrast, we do not find similar results for 12b-1 (i.e., marketing and distribution) fees, which arise from a support function rather than the trading strategy and operations the fund chooses to employ.

In this subsection, we examine differences in funds' return performance. Doing so helps us partially understand whether, and the extent to which, the turnover and expense findings from the prior section depress fund returns. Of course, fund returns are not merely a function of trading and other operating costs; they also depend on the portfolio performance during the holding period. To attempt to separate these factors as much as possible, we consider returns both before (i.e., *Gross Returns*) and after fund expenses (i.e., *Net Returns*). We also examine return volatility and volatility-adjusted returns to understand whether risk differences explain any of the return differences we document.

We first plot mean net returns separately for replicators and samplers over time in Figure 2. While the samplers' net returns are consistently below replicators' returns in the top panel, this is difficult to see because market volatility from year to year requires us to include a broad range of values on the y-axis (i.e., -10% to 40%). Thus, the bottom panel of the figure presents abnormal net returns, in which we subtract the full sample mean of net returns each year. This adjustment makes it much easier to see that samplers underperform replicators on average, often by 1-3%, in all but two of the 11 years in our sample period.

The regression results in Table 4 provide consistent evidence, i.e., that both net and gross returns are lower for sampling funds. Specifically, the coefficient on *Sampling* in the *Net Returns* regression is significantly negative, with a magnitude of about 60-70 basis points (bps), depending on whether we include index-level controls (column one) or Year x Benchmark Index fixed effects (column five). This 60 bps underperformance of samplers is about 5 percent of the average fund's return. While 60 bps (or 5 percent) is economically significant for any mutual fund, it is especially substantial for index funds, whose mandate typically requires them to keep tracking error low, such as within 10-20 bps, and who often voluntarily maintain even lower tracking errors (Frino

and Gallagher, 2001; Blume and Edelen, 2003). This evidence suggests that sampling is detrimental to funds' performance and, relatedly, their ability to achieve their investment objectives.

Overall, the evidence from the *Net Return* regressions is consistent with the higher expenses of samplers (see Table 3) translating into lower return performance. However, the *Sampling* coefficient is also negative and significant in the *Gross Return* regressions. Negative gross returns could occur for various reasons. For one, the samplers' additional turnover likely results in more payments to liquidity providers in the form of bid-ask spreads, short-term return reversals (Nagel, 2012), etc. It could also be the case that the index fund managers are not well suited to sampling, which in many ways appears to amount to poor stock picking. We attempt to further distinguish between these two explanations (i.e., turnover vs. stock picking) in the next subsection. For now, if we compare the coefficients on *Sampling* in the net and gross returns regressions of columns five and six, it appears that about 25 percent of the samplers' return underperformance is explained by the expense ratio (i.e., 1 - [0.449/0.589] = 0.24). This leaves 75% to be explained by higher turnover, poor stock picking, or other factors.

Return performance could also be lower among samplers if their strategy results in a less risky portfolio. Thus, we next consider return volatility and a simple measure of risk-adjusted returns. In contrast with the idea that sampling leads to less risk, we actually find some evidence that samplers' returns are more volatile, if anything. That is, the coefficient on *Sampling* is positive but not significant in column three, and positive and significant at the 10 percent level in column 7. We then divide net returns by return volatility, creating a measure which we refer to as *Risk Adj Returns*, and find similar results in columns four and eight. That is, risk-adjusted returns are negatively associated with sampling, consistent with samplers underperforming even after we take

risk into account.[15] Figure 3 shows similar results graphically. In particular, the top panel shows similar return volatility for replicators and samplers (albeit slightly higher for samplers in earlier years), and the bottom two panels show that samplers' risk-adjusted returns are consistently lower.

With respect to the coefficients on the control variables, the most consistent finding is that larger funds outperform smaller funds, consistent with economies of scale in index investing. Also, note the extremely high R^2 of 0.97-0.99 in regressions with Year x Benchmark Index fixed effects. These R^2 values are due to funds matching index returns to a great extent. Of course, this is the whole point of indexing, and is therefore not surprising. The fact that R^2 is less than one means there is some deviation; however, R^2 values nearing one together with our *Sampling* coefficients suggest that sampling explains much of the limited deviation that exists.

3.3.2 Portfolio Returns: Turnover vs. Stock Picking Explanations

In order to distinguish between the transaction costs and stock picking explanations for our results, we examine *portfolio* return performance in this subsection. To be precise, instead of using the overall returns reported by the fund (i.e., the net and gross fund returns) as in our prior analyses, we calculate portfolio returns assuming that the fund maintained the stock positions in its portfolio throughout the period. Because this assumption becomes less and less likely to hold as we examine increasingly long periods, we present results for portfolio returns over the prior 1, 3, or 12 months. Of course, this measure mimics the many studies in finance that investigate the return performance of a portfolio that is constructed after sorting on some factor or anomaly variable. These simple portfolio sorts are notorious for ignoring transaction costs, which in this case we use to our

[15] The coefficient on *Sampling* is not significant in column four. However, this is not because samplers take on less risk, as evidenced by the positive (albeit insignificant) coefficient in column three. We believe it is because of low power; note the low R^2 in column four, which is likely because the ratio of returns and return volatility is a very noisy measure.

advantage.[16] That is, they are based on buy-and-hold returns over the holding period and are calculated using the opening and closing stock price, which do not capture the brokerage fees, commissions, bid-ask spreads, and market impact of fund-specific transactions. In contrast, *fund* returns are based on the change in the fund's total net assets that are available for distribution to investors, and therefore reflect all the transaction costs incurred by the fund during the period.

In Table 5, we repeat the analyses from Table 4 after replacing fund returns with *Portfolio Returns (X months)*, which is the sum of the returns of each of the fund's stock holdings over the X months preceding the end of the reporting year multiplied by the stock's portfolio weight at the end of the year (i.e., the ratio of the dollar value of the stock holding and the fund's TNA). Because returns are unavailable for the stocks held by a few of the funds, mainly those with a foreign mandate/focus, the sample in this table consists of 3,021 of the 3,365 fund-years in the full sample.[17] Panel A shows that the correlations between fund net returns and portfolio returns are high but significantly less than one. There are a few explanations for this, including that portfolio returns ignore transaction costs and assume that the fund did not trade at all during the period. In addition, fund returns include the return on any cash holdings, derivatives, and share lending programs, while portfolio returns are based entirely on the capital appreciation and dividends attributable to the stock portion of the fund's portfolio. Note that *Net Returns* is most directly comparable to *Portfolio Returns (12 months)* because they both measure annual returns.

In Panel B of Table 5, we regress portfolio returns on the fund's classification as a replicator, sampler, or hybrid, as well as several other fund and benchmark index characteristics

[16] Many studies discuss the issue of accounting for transaction costs in general, such as Fama and French (2008), Pontiff (1996), and Pontiff (2006). For examples in the context of a specific anomaly, see the many studies on the post-earnings-announcement drift, including Bernard and Thomas (1989), Bernard and Thomas (1990), and Ng, Rusticus, and Verdi (2008).

[17] For the same reason, we omit the *Domestic Index* indicator from the regressions in Table 5 to avoid collinearity.

(see equations (1) and (2)). The coefficient on *Sampling* is negative and significant across the specifications, consistent with the performance regressions of Table 4. Since these returns exclude transaction costs, this evidence suggests that transaction costs do not fully explain the underperformance we document. Instead, it appears that poor stock picking also contributes to sampling underperformance. In terms of economic magnitude, the coefficient on *Sampling* in the *Portfolio Returns (12 months)* regression of column six is about 75 percent of the analogous coefficient in the *Net Returns* regression in column 5 of Table 4 (i.e., 0.440/0.589 = 0.75). This suggests that the majority of the underperformance we document is attributable to poor stock picking. The visual evidence in Figure 4 using *Portfolio Returns (12 months)* is consistent with the regression-based findings, in that samplers' portfolio holdings consistently underperform replicators' holdings over time. This difference can be seen in both panels whether or not we subtract the mean annual return. In a few years, the difference is upwards of 4-5%.

Overall, the results in Tables 4 and 5 and Figures 2 through 4 are consistent with samplers underperforming (relative to replicators) because they charge higher expenses and fees, because they incur greater transaction costs, and because they are poor stock pickers.

4. Additional Tests

4.1 S&P 500 and Market-Cap Tracking

Despite our inclusion of Year x Benchmark Index fixed effects and our requirement that a benchmark index be followed by multiple types of funds, one may wonder whether our results are generalizable to a broad set of index funds. Alternatively, they may be driven by niche indices that are followed by relatively few funds, such as the S&P 500 Catholic and Sharia Values Indices, or

the Solactive Guru Index (which tracks and includes the top holdings of large hedge funds).[18] Our results could also be driven by style or sector funds, whose index choice suggests they seek exposure to different "factors" (e.g., based on different investment styles, risk factors, and/or industries) from the outset. Of course, the fact that they have already somewhat strayed from the passive ideal (i.e., the market portfolio) in their index choice would make it less surprising to find them straying from the passive ideal in other ways, such as by sampling.

To address these possibilities, in this section we separately examine funds tracking the S&P 500 index and other Market Cap indices. Panel A of Table 6 shows that our main results generally hold within the S&P 500 subsample. That is, S&P 500 index samplers have significantly higher turnover, albeit not as much as the rest of the sample (i.e., coefficient of 0.277 compared to 0.440 in Table 3). Also, while S&P 500 samplers' expense ratio is not significantly higher, their management fees are (untabulated), suggesting that managers require additional compensation for attempting to discriminate among the S&P 500 firms using sampling. As in the main sample, columns three and four suggest that net and gross returns are lower for S&P 500 samplers compared to replicators. Further, the inference that samplers are relatively poor performers does not change when considering return volatility and risk-adjusted returns.

We observe similar patterns in Panel B for the larger subsample of funds that track any index based primarily on market cap (i.e., where *Market Cap Index* =1). Specifically, market cap index samplers have higher turnover and expenses, as well as lower net, gross, and risk-adjusted returns. Thus, the overall results in Table 6 suggest that our main findings apply to the substantial set of funds that track the S&P 500 or other market-cap-based indices and are not driven by more specialized niche indices or by style and sector funds.

[18] Relatedly, Akey et al. (2021) finds greater turnover and expenses among index funds that follow their own proprietary index, which is as "niche" as a fund can get.

4.2 Number of Index Constituents

As noted throughout the paper, one of the primary potential benefits of sampling lies in reducing the number of stocks that need to be included in an index fund's portfolio. Of course, the more stocks there are in an index, the more stocks there are for samplers to leave out of their portfolio. Relatedly, the more stocks there are in the index, the greater the chances are that the index includes illiquid stocks. Thus, we expect the benefits of sampling to be increasing in the number of index constituents. In this section, we test this prediction using cross-sectional variation in *Log(#Constituents)*.

To be specific, Table 7 shows estimates of regressions that include the interaction of sampling and number of index constituents, *Sampling x Log(#Constituents)*. As in earlier analyses, we present regressions that exclude Year x Benchmark Index fixed effects so that we can evaluate the main effects of the index-level controls (Panel A) as well as regressions that include Year x Benchmark Index fixed effects, which of course subsume the main effects of *Log(#Constituents)* and the other benchmark index-level controls (Panel B).

In the first column of Panel A, the positive and significant coefficient of 1.888 on *Sampling* provides further evidence that samplers turnover their portfolios more than replicators. However, because this regression includes *Sampling x Log(#Constituents)*, the coefficient on *Sampling* is the implied turnover difference between samplers and replicators of an index with zero constituents (i.e., where *Log(#Constituents)* = 0), which is impossible. Thus, we must also consider the negative coefficient on *Sampling x Log(#Constituents)*, which suggests that the additional turnover of samplers gradually disappears as the index becomes larger. Interestingly, the coefficients on *Sampling x Log(#Constituents)* are not significantly different from zero in the expense ratio regression, suggesting that the lower turnover for samplers following larger indices does not

translate into lower expenses for the fund's investors. However, the returns results in columns three through six are consistent with the turnover finding. That is, while we continue to find that returns (net, gross, and risk-adjusted) are lower for samplers, this underperformance gradually decreases and eventually disappears for funds following indices with many constituents.

The results in Panel B are largely consistent, except that the positive (negative) coefficient on *Sampling* (*Sampling x Log(#Constituents)*) is not statistically significant in the turnover regression. However, this could be due to low power, perhaps because of limited within-benchmark index variation in *Sampling x Log(#Constituents)*. Alternatively, we could infer that the argument that sampling can benefit index fund managers by helping them keep down trading and operating costs is weak and not supported by the data.

In terms of economic magnitude, in both Panels A and B the coefficients on *Sampling* are about 7-8 times larger (in absolute value) than the coefficients on *Sampling x Log(#Constituents)*. This suggests that the break-even point at which samplers and replicators costs and returns are the same is somewhere between 1000 and 3000 stocks (i.e., *Log(#Constituents)* = 7 implies *#Constituents* = 1000 and *Log(#Constituents)* = 8 implies *#Constituents* = 3000). This means that, on average, sampling funds following indices with fewer than 1,000 constituents, which includes more than 75 percent of our sample, underperform the replicators following the same index.

Overall, the marginal evidence on turnover and strong evidence on returns in this section is somewhat consistent with the argument that sampling helps reduce costs of replicating indices with many (i.e., thousands) of constituents. However, our findings merely suggest that the samplers' underperformance disappears in large indices, but the argument is that sampling could help funds outperform replication, which we do not find except for perhaps the largest couple of indices (e.g., FTSE Global All Cap Index and Russell 3000). Thus, even if the argument in favor

of sampling reducing costs has some merit, it seems to be overwhelmingly offset by the detrimental aspects of sampling, such as additional trading due to variation in the sampling variables, greater tracking errors, and poor stock picking.

4.3 Fund Flows

Given the substantial differences in costs and returns between samplers and replicators, we wonder whether fund flows also differ between the two groups. There are many possibilities. If investors are aware of the average underperformance of samplers, then they would likely shift capital away from samplers toward replicators or non-index fund assets. However, investors may be attracted to sampling funds because they believe the argument that sampling reduces the costs of holding all or most of the constituents of a large index. Also, they may believe that the variables on which the fund samples (e.g., size, dividends, or ESG) enhance returns or other objectives. This belief would seemingly contradict their decision to choose "passive" index investments and is more consistent with an "active" investment philosophy. Finally, it is possible that many investors are unaware that some index funds use replication and some use sampling, and even more likely that the typical index fund investor is unaware of the cost and performance differences we document.[19] Thus, the fund flows of replicators relative to samplers is an empirical question.

In Table 8, we regress fund flows on *Sampling* and the other explanatory variables that we have studied thus far. We also place a particular emphasis on *Trend* and its interaction with *Sampling* to understand any changes in investors' preferences during our sample period. The negative coefficient on *Sampling* in the first two columns suggests that samplers had lower flows than replicators at the start of our sample period. However, the coefficient on *Sampling x Trend* is

[19] Admittedly, we weren't aware that some index funds use replication and others use sampling until recently, and several readers and seminar participants have told us that they weren't aware either until seeing our paper. Of all investors, we and our academic peers should likely have been the most aware of these distinctions. Thus, it seems safe to assume that most mom-and-pop investors, and even many professional investors, are likewise unaware.

positive, significant, and about one-fourth the magnitude of the coefficient on *Sampling*. This suggests that within four years of the start of our sample period (i.e., by about 2014), flows to samplers and replicators were similar, and that flows to samplers were greater thereafter.

Because of the vast body of research studying the relationship between prior return performance and fund flows (e.g., Ippolito, 1992; Sirri and Tufano, 1998; Del Guercio and Tkac, 2002; Berk and Green, 2004), we also add the fund's prior year net return to the regression and report the results in columns three and four. The coefficient on *Prior Returns* is positive but not significant in both columns, which may at first be surprising. However, recall that one of the main purposes of an index fund is to match index returns, so by design there is not much variation in returns in the first place. As a result, when choosing index investments, investors appear to put more weight on other factors, especially fees (Elton, Gruber, and Busse, 2004; Elton, Gruber, and de Souza, 2019; Sun, 2021). Returning to the primary variable of interest, *Sampling*, its coefficient attenuates to the point of being insignificant; however, the coefficient on *Sampling x Trend* remains positive and significant, which strengthens our inference that samplers have attracted more capital over time relative to replicators.

This finding that sampling funds attracted more of investors' capital over time is surprising considering our evidence that they underperform on both a cost and return basis. However, perhaps our evidence can be explained by the recent concerns regarding the risks of too much 'passive' index investing. For example, Easley et al. (2021) discusses how the growing activeness of ETFs may assuage concerns about ETFs harming price discovery, and our evidence on sampling might assuage similar concerns about the more passive replication approach. Our evidence might also (or alternatively) be explained by the recent trends towards certain types of investing that are captured by the main factors used in the sampling approach, such as high-yield investing (Baker

and Wurgler, 2004; Harris, Hartzmark, and Solomon, 2015) and ESG investing (e.g., Krueger, Sautner, and Starks, 2020; Bolton and Kacperczyk, 2021; Pastor, Stambaugh, and Taylor, 2021).

4.4 Expense Ratio Components

In this section, we further examine major components of the expense ratio to better understand the reasons underlying the differences between samplers and replicators. We focus on management fees and 12b-1 fees, which together make up almost 70 percent of the average fund's expense ratio (see the means in Panel A of Table 1, i.e., [0.20+0.15]/0.51=0.69), but represent very different costs. That is, while the management fee represents the compensation that fund managers require for their overall efforts and performance, the 12b-1 relates specifically to the fund's marketing and distribution expenses.

In Table 9, we repeat the regressions from Table 3 after replacing *Exp Ratio* with either *Mgmt Fees* or *12b-1 Fees*. We find that management fees are higher for samplers, consistent with these managers requiring compensation from investors for their efforts to actively invest. In contrast, 12b-1 fees are not significantly higher for samplers. This suggests that the increasing flows to samplers over our sample period cannot be explained by their managers spending more to advertise the fund to current and potential investors. Overall, the results in this section, combined with our earlier results on turnover and the aggregate expense ratio, suggest that the additional trading and operating complexity inherent in sampling is what drives their higher costs and lower returns.

4.5 No Benchmark Index Controls or Fixed Effects

We believe that our inclusion of benchmark index level variables, especially the Year x Benchmark Index fixed effects, is one of the most important and useful aspects of our research design because they help us rule out a host of concerns about cross-benchmark (i.e., index)

variation. However, requiring such data about the index, including that the fund's index be followed by multiple types of funds (e.g., at least one replicator and one sampler), reduces our sample substantially. In particular, our main sample comprises only about 300 funds per year, out of over one thousand U.S equity index mutual funds and index ETFs that operated during our sample period. Thus, in this section, we repeat our main analyses after relaxing our benchmark index level criteria, which results in a much larger sample of 12,969 fund-years (i.e., more than 1,000 per year). However, because these regressions only include fund controls and year fixed effects and exclude index level controls and Year x Benchmark Index fixed effects, they should be interpreted with greater caution. Table 10 shows the results of this analysis, which confirms that samplers have higher turnover and expenses as well as lower returns in this broader sample. This test provides further evidence that our overall finding that samplers underperform generalizes to a broad swath of the index fund population.

5. Conclusion

Index funds are designed to match the returns of an index; however, there are two different approaches that funds use to attempt to achieve this objective. Full replicators hold index constituents in proportion to their index weights, with few exceptions. In contrast, representative samplers select a subset of stocks from the index using variables such as size, dividends, and P/E ratios. By reducing the number of holdings, the latter approach is meant to reduce trading and other operating costs. However, it introduces additional risks, such as a greater possibility of tracking error, poor stock picking, and additional trading if the sampling variables change absent index reconstitution.

In this study, we compare the performance of the two approaches and find that samplers have higher turnover and expenses while earning worse returns. Furthermore, this result does not

appear to be driven by niche indices, as we find similar results in the subsample of funds following the S&P 500 and other market-cap-based indices. We do find that the differences between samplers and replicators attenuate for funds following indices with many more constituents (i.e., 1,000-3,000 stocks), suggesting there is some merit to the argument that sampling can reduce the costs of replication in certain situations. We also document puzzling evidence that sampling attracts greater fund flows relative to replication late in our sample period, despite the evidence that sampling is detrimental to fund performance.

Overall, our findings suggest that almost half of index funds use a sampling approach that strays from the passive ideal and results in worse performance on average. This evidence should be of interest to index fund managers who are deciding between adopting a full replication or representative sampling approach. The replication vs. sampling dimension should also be on the radar of index fund investors, who are often encouraged to focus on factors such as expenses and fees when selecting index investments (Elton et al., 2004; Elton et al., 2019; Sun, 2021). While such factors are certainly crucial, investors may have little awareness or understanding of the extent to which fund expenses and returns are driven by the mechanics, including replication vs. sampling, underlying their fund managers' efforts to track an index.

References

Akey, P., Robertson, A., Simutin, M., 2021. Closet active management of passive funds. Working paper, University of Toronto.

Amihud, Y., Goyenko, R., 2013. Mutual fund's R^2 as predictor of performance. *Review of Financial Studies* 26, 667-694.

Baker, M., Wurgler, J., 2004. A catering theory of dividends. *Journal of Finance* 59, 1125-1165.

Berk, J., Green, R., 2004. Mutual fund flows and performance in rational markets. *Journal of Political Economy* 112, 1269–1295.

Bernard, V., Thomas, J., 1989. Post-earnings-announcement drift: Delayed price response or risk premium? *Journal of Accounting Research* 27, 1-36.

Bernard, V., Thomas, J., 1990. Evidence that stock prices do not fully reflect the implications of current earnings for future earnings. *Journal of Accounting and Economics* 13, 305-340.

Blackrock, 2022. iShares investigates: Market indexes and index investing. Retrieved from https://www.blackrock.com/au/intermediaries/ishares/market-indexes-and-investing.

Blume, M., Edelen, R., 2003. S&P 500 indexers, delegation costs, and liquidity mechanisms. Working paper, University of Pennsylvania.

Bolton, P., Kacperczyk, M., 2021. Do investors care about carbon risk? *Journal of Financial Economics* 142, 517-549.

Chan, L., Chen, H., Lakonishok, J., 2002. On mutual fund investment styles. *Review of Financial Studies* 15, 1407-1437.

Chen, H., Cohen, L., Gurun, U., 2021. Don't take their word for it: The misclassification of bond mutual funds. *Journal of Finance* 76, 1699-1730.

Cheng, S., Massa, M., Zhang, H., 2019. The unexpected activeness of passive investors: A worldwide analysis of ETFs. *Review of Asset Pricing Studies* 9, 296-355.

Choi, J., Laibson, D., Madrian, B., 2010. Why does the law of one price fail? An experiment on index mutual funds. *Review of Financial Studies* 23, 1405-1432.

CIBC, 2020. Managing an index fund is harder than you may think. Retrieved from https://www.investorsedge.cibc.com/en/learn/managing-an-index-fund.html.

Cong, W., Xu, D., 2016. Rise of factor investing: Asset prices, informational efficiency, and security design. Working paper, Cornell University and University of Florida.

Cremers, M., Petajisto, A., 2009. How active is your fund manager? A new measure that predicts performance. *Review of Financial Studies* 22, 3329–3365.

deHaan, E., Song, Y., Xie, C., Zhu, C., 2021. Obfuscation in mutual funds. *Journal of Accounting and Economics* 72, 101429.

Del Guercio, D., Tkac, P., 2002. The determinants of the flow of funds of managed portfolios: mutual funds versus pension funds. *Journal of Financial and Quantitative Analysis* 37, 523–558.

Doshi, H., Elkamhi, R., Simutin, M., 2015. Managerial activeness and mutual fund performance. *Review of Asset Pricing Studies* 5, 156-184.

Dyer, T., Guest, N., Yu, E., 2021. New accounting standards and the performance of quantitative investors. Working paper, Brigham Young University and Cornell University.

Easley, D., Michayluk, D., O'Hara, M., Putnins, T., 2021. The active world of passive investing. *Review of Finance* 25, 1433-1471.

Elton, E., Gruber, M., Blake, C., 2001. A first look at the accuracy of the CRSP mutual fund database and a comparison of the CRSP and Morningstar mutual fund databases. *Journal of Finance* 56, 2415-2430.

Elton, E., Gruber, M., Busse, J., 2004. Are investors rational? Choices among index funds. *Journal of Finance* 59, 261-288.

Elton, E., Gruber, M., de Souza, A., 2019. Passive mutual funds and ETFs: Performance and comparison. *Journal of Banking and Finance* 106, 265-275.

ETF.com, 2020. How to run an index fund: Full replication vs. optimization. Retrieved from https://www.etf.com/etf-education-center/etf-basics/how-to-run-an-index-fund-full-replication-vs-optimization.

ETF Database, 2016. What's the difference between sampling strategy and full replication ETFs? Retrieved from https://etfdb.com/portfolio-management/difference-between-sampling-strategy-and-full-replication-etfs/.

Fama, E., French, K., 2008. Dissecting anomalies. *Journal of Finance* 63, 1653–1678.

Fama, E., French, K., 2010. Luck versus skill in the cross-section of mutual fund returns. *Journal of Finance* 65, 1915–1947.

Fama, E., French, K., 2015. A five-factor asset pricing model. *Journal of Financial Economics* 116, 1-22.

Fidelity, 2022. Why invest in Fidelity index funds. Retrieved from

https://www.fidelity.com/mutual-funds/fidelity-funds/why-index-funds.

Forbes, 2015. Set it and forget it works. Retrieved from https://www.forbes.com/sites/rickferri/2015/06/11/set-it-and-forget-it-works.

Frino, A., Gallagher, D., 2001. Tracking S&P 500 index funds. *Journal of Portfolio Management* 28, 44-55.

Harris, L., Hartzmark, S., Solomon, D., 2015. Juicing the dividend yield: Mutual funds and the demand for dividends. *Journal of Financial Economics* 116, 433-451.

Hedge Fund Research, 2022. Replication strategies. Retrieved from https://www.hfr.com/replication-strategies.

Hortaçsu, A., Syverson, C., 2004. Product differentiation, search costs, and competition in the mutual fund industry: A case study of S&P 500 index funds. *Quarterly Journal of Economics* 119, 403-456.

Investopedia, 2020. How Vanguard index funds work. Retrieved from https://www.investopedia.com/articles/investing/111715/how-vanguard-index-funds-work.asp.

Ippolito, R., 1992. Consumer reaction to measures of poor quality: Evidence from the mutual fund industry. *Journal of Law and Economics* 35, 45–70.

justETF, 2022. Physical replication of ETFs. Retrieved from https://www.justetf.com/uk/academy/physical-replication-of-etfs.html.

Kacperczyk, M., Sialm, C., Zheng, L., 2008. Unobserved actions of mutual funds. *Review of Financial Studies* 21, 2379-2416.

Krueger, P., Sautner, Z., Starks, L., 2020. The importance of climate risks for institutional investors. *Review of Financial Studies* 33, 1067-1111.

MarketWatch, 2021a. Set it and forget it: Try this easy 'subscription' method of investing. Retrieved from https://www.marketwatch.com/story/set-it-and-forget-it-try-this-easy-subscription-method-of-investing-11611267489.

MarketWatch, 2021b. When to 'set it and forget it' - and when not. Retrieved from https://www.marketwatch.com/story/when-to-set-it-and-forget-it-and-when-not-11627495099.

Nagel, S., 2012. Evaporating liquidity. *Review of Financial Studies* 25, 2005-2039.

Ng, J., Rusticus, T., Verdi, R., 2008. Implications of transaction costs for the post-earnings-announcement drift. *Journal of Accounting Research* 46, 661-696.

Pastor, L., Stambaugh, R., Taylor, L., 2020. Fund tradeoffs. *Journal of Financial Economics* 138, 614-634.

Pastor, L., Stambaugh, R., Taylor, L., 2021. Sustainable investing in equilibrium. *Journal of Financial Economics* 142, 550-571.

Pontiff, J., 1996. Costly arbitrage: Evidence from closed-end funds. *Quarterly Journal of Economics* 111, 1135-1151.

Pontiff, J., 2006. Costly arbitrage and the myth of idiosyncratic risk. *Journal of Accounting and Economics* 42, 35-52.

Sirri, E., Tufano, P., 1998. Costly search and mutual fund flows. *Journal of Finance* 53, 1589–1622.

Sun, Y., 2021. Index fund entry and financial product market competition. *Management Science* 67, 500-523.

Vanguard, 2021. Indexing at Vanguard - quality not commodity. Retrieved from https://global.vanguard.com/portal/site/institutional/nl/en/articles/research-and-commentary/portfolio-construction/indexing-at-vanguard-quality-not-commodity.

Wermers, R., 2012. A matter of style: The causes and consequences of style drift in institutional portfolios. Working paper, University of Maryland.

Zhu, Q., 2020. The missing new funds. *Management Science* 66, 1193-1204.

Appendix A. Excerpts from Index Fund Prospectuses

This appendix provides excerpts from the prospectuses of index funds, in which they disclose whether they use full replication, representative sampling, or a combination of the two (which we refer to as the "hybrid" approach).

Full Replication

Vanguard Large-Cap Index Fund (2018): "The Fund attempts to replicate the target index by investing all, or substantially all, of its assets in the stocks that make up the Index, holding each stock in approximately the same proportion as its weighting in the Index."

Dreyfus International Stock Index Fund (2019): "The fund generally invests in all stocks included in the index. The fund's investments are selected to match the benchmark composition along individual name, country and industry weighting, and other benchmark characteristics. Under these circumstances, the fund maintains approximately the same weighting for each stock as the index does."

Invesco S&P 500 Index Fund (2012): "Invesco Advisers, Inc. (the Adviser), the Fund's investment adviser, passively manages the Fund's assets by investing in stocks in approximately the same proportion as they are represented in the S&P 500 Index. For example, if the common stock of a specific company represents five percent of the S&P 500 Index, the Adviser typically will invest the same percentage of the Fund's assets in that stock."

Representative Sampling

iShares Dow Jones U.S. ETF (2018): "BFA uses a representative sampling indexing strategy to manage the Fund. 'Representative sampling' is an indexing strategy that involves investing in a representative sample of securities that collectively has an investment profile similar to that of an applicable underlying index."

State Street Global Equity ex-U.S. Index Fund (2016): "In seeking to track the performance of the Index, the Fund employs a sampling strategy, which means that the Fund is not required to purchase all of the securities represented in the Index. Instead, the Fund may purchase a subset of the securities in the Index in an effort to hold a portfolio of securities with generally the same risk and return characteristics of the Index. The number of holdings in the Fund will be based on a number of factors, including asset size of the Fund."

Fidelity Small Cap Growth Index Fund (2020): "Using statistical sampling techniques based on such factors as capitalization, industry exposures, dividend yield, price/earnings (P/E) ratio, price/book (P/B) ratio, and earnings growth to attempt to replicate the returns of the Russell 2000 Growth Index using a smaller number of securities."

Hybrid

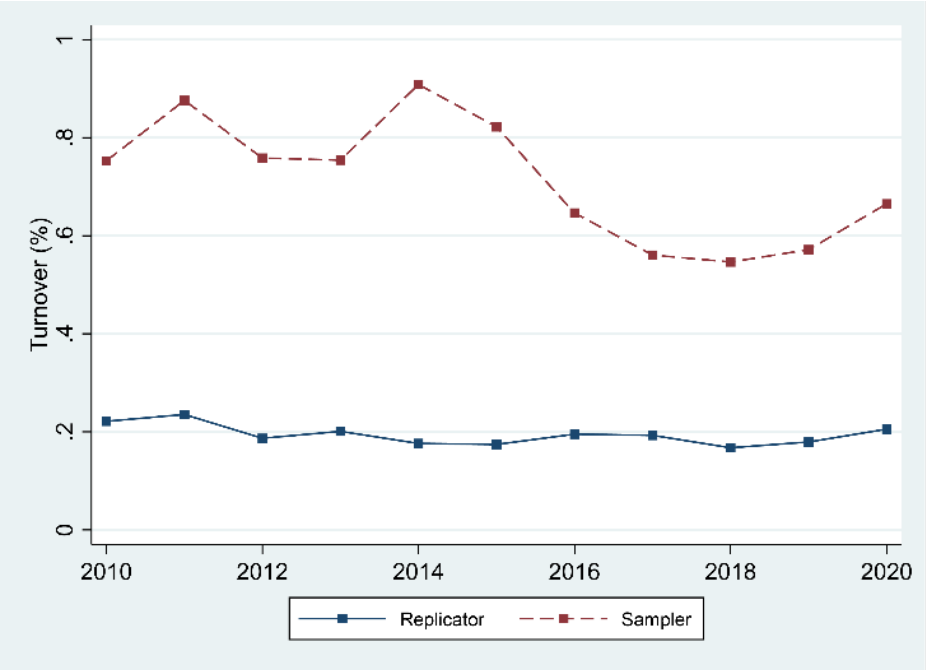
Principal Healthcare Innovators Index ETF (2020): "In seeking its objective, the Fund typically employs a 'full replication' strategy which involves investing in all the securities that make up the Index, in the same approximate proportions as the Index. The Fund can, however, use a 'sampling' methodology to purchase a subset of the securities in the Index in an effort to hold a portfolio of securities with generally the same risk and return characteristics of the Index."

Deutsche S&P 500 Index Fund (2015): "The Portfolio uses an optimization strategy, buying the largest stocks in the index in approximately the same proportion they represent in the index, then investing in a statistically selected sample of the smaller securities found in the index."

BlackRock Russell 1000 Index Fund (2013): "The Fund will invest in the common stocks represented in the Russell 1000 in roughly the same proportions as their weightings in the Russell 1000… At times, the Fund may not invest in all of the common stocks in the Russell 1000, or in the same weightings as in the Russell 1000. At those times, the Fund chooses investments so that the market capitalizations, industry weightings and other fundamental characteristics of the stocks chosen are similar to the Russell 1000 as a whole."

Figure 1. Fund Turnover and Expenses

This figure shows how the mean of turnover and expense ratio vary over our sample period for replicator and sampler index funds. Variables are defined in Table 1. The sample consists of 3,365 U.S.-based equity index fund-years during the period 2010 through 2020.



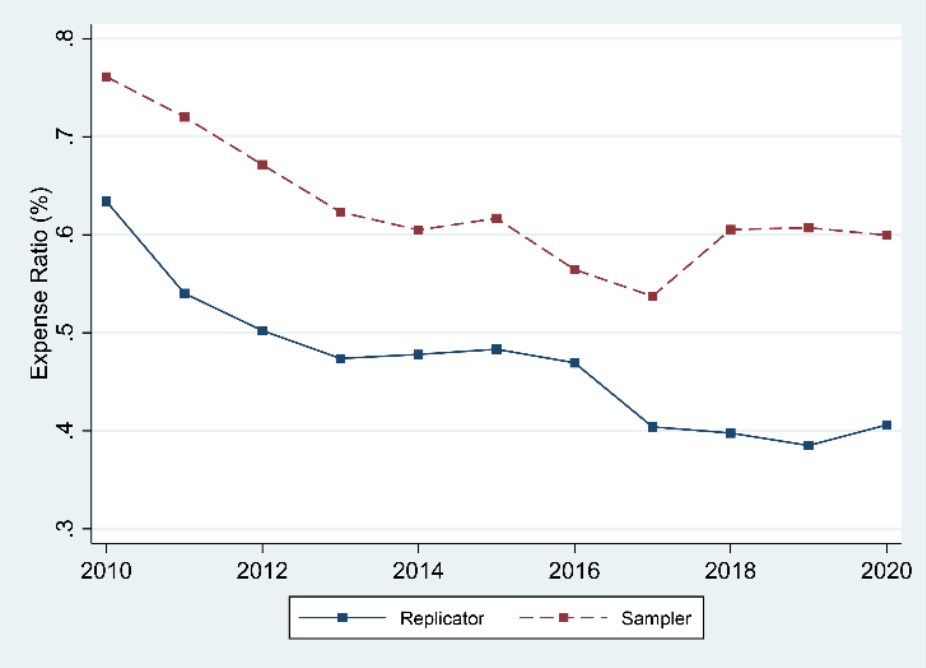
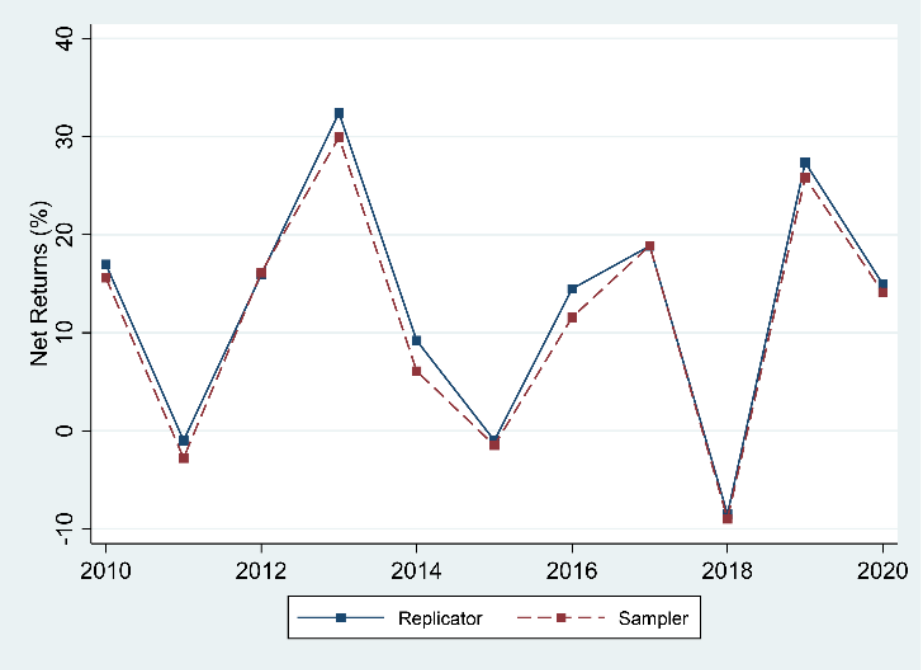


Figure 2. Fund Return Performance

This figure shows how the mean of net returns and abnormal net returns vary over our sample period for replicator and sampler index funds. Abnormal net returns is the difference between net returns and the full sample mean of net returns for the year. Other variables are defined in Table 1. The sample consists of 3,365 U.S.-based equity index fund-years during the period 2010 through 2020.



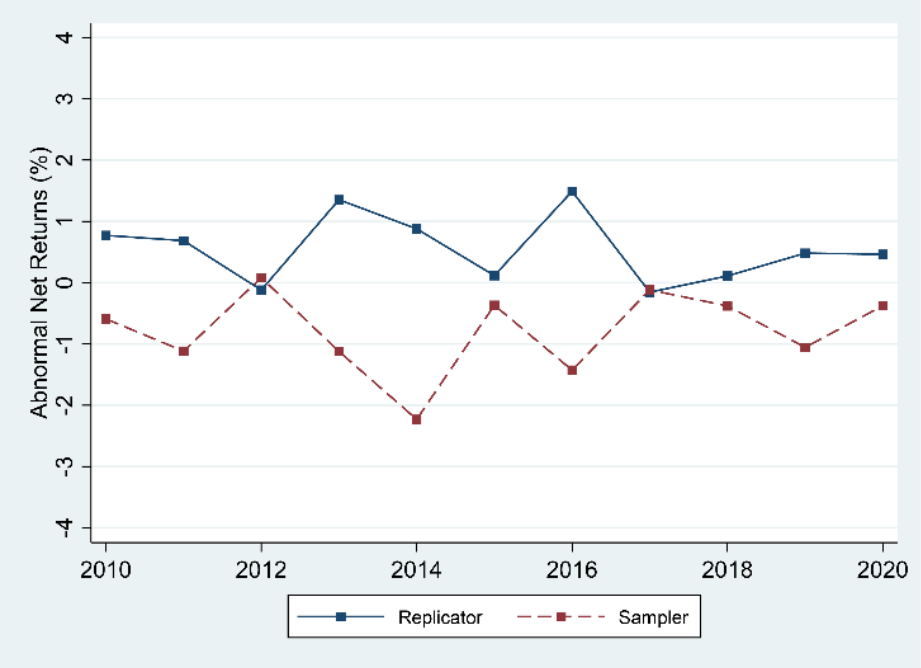


Figure 3. Return Volatility and Risk-Adjusted Returns

This figure shows how the mean of return volatility and risk-adjusted returns vary over our sample period for replicator and sampler index funds. Risk-adjusted returns are net returns divided by return volatility. Abnormal risk-adjusted returns is the difference between risk-adjusted returns and the full sample mean of risk-adjusted returns for the year. Other variables are defined in Table 1. The sample consists of 3,365 U.S.-based equity index fund-years during the period 2010 through 2020.

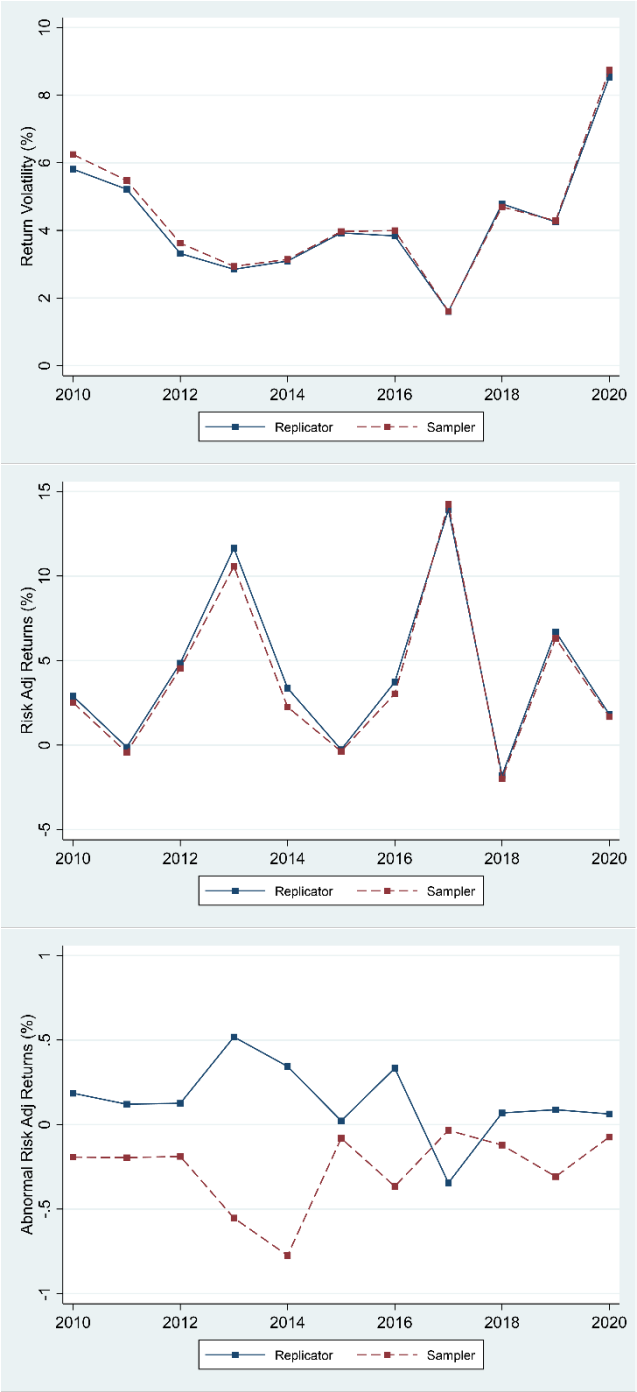
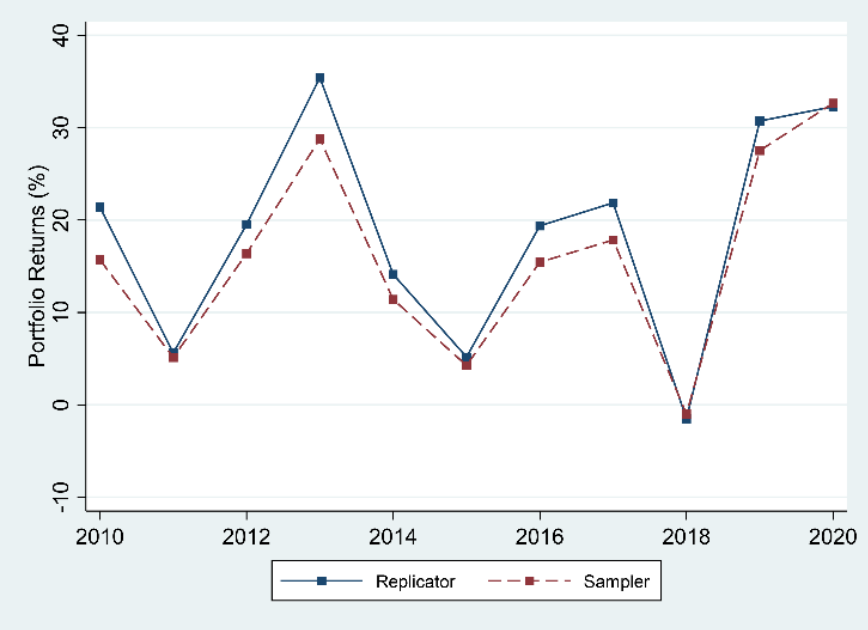


Figure 4. Portfolio Return Performance

This figure shows how mean portfolio returns and abnormal portfolio returns vary over our sample period for replicator and sampler index funds. Our calculations of portfolio returns assume that the fund maintained the stock positions in its portfolio for the prior 12 months. Specifically, *Portfolio Returns* is the sum of the returns of each stock held by the fund over the 12 months preceding the end of the reporting year multiplied by the portfolio weight (i.e., the ratio of the dollar value of the holding and the fund's TNA). Abnormal portfolio returns is the difference between portfolio returns and the full sample mean of portfolio returns for the year. Other variables are defined in Table 1. Because returns are unavailable for the stocks held by a few of the funds, mainly those with a foreign mandate/focus, the sample in this figure consists of 3,021 of the 3,365 total U.S.-based equity index fund-years during the period 2010 through 2020.



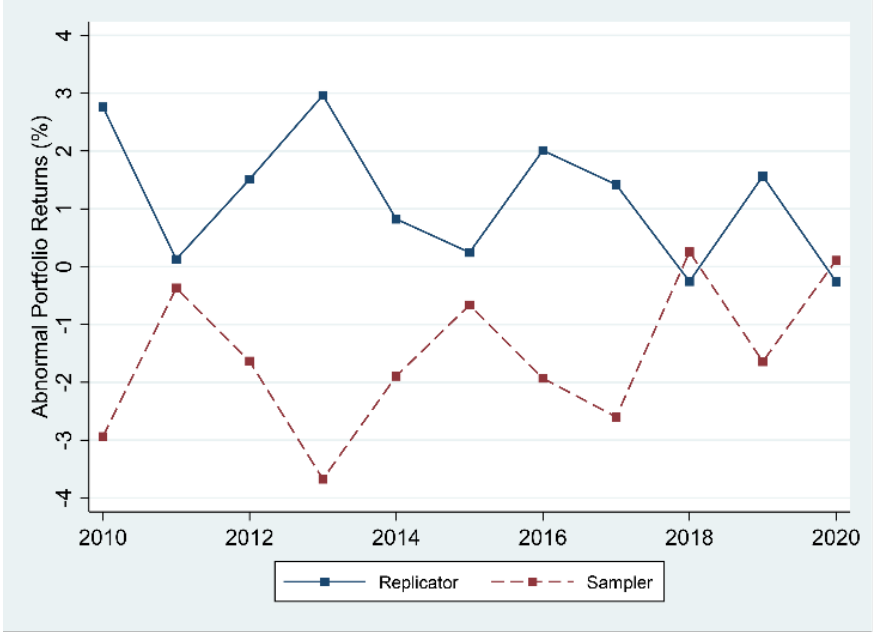


Table 1. Descriptive Statistics

Panel A (Panel B) presents distributional statistics (means) of the full sample (samplers vs. replicators) for the variables used in our analyses. The sample consists of 3,365 U.S.-based equity index fund-years during the period 2010 through 2020. Based on the index fund classification described in Section 3, *Sampling* is set to 1 for samplers, 0.5 for hybrids, and 0 for replicators. *Fund Size* is the fund's total net assets (TNA), or total assets minus total liabilities, reported in millions (CRSP MFDB variable mtna). *Fund Age* is the inception date of the oldest fund class. *Big Three* is an indicator set to one if the fund is offered by Vanguard, BlackRock, or State Street, and zero otherwise. *ETF* is an indicator set to one if CRSP flags the fund as exchange-traded (CRSP MFDB variable et_flag = 'F'), and zero otherwise. *Domestic* is an indicator set to one if the index is primarily made up of U.S. stocks, and zero if the index is entirely made up of foreign stocks or has a global focus. *#Constituents* is an approximation of the number of stocks in the benchmark index; specifically, we calculate the average number of stocks held in the portfolios of replicators following the benchmark index. *Market Cap Index* is set to one for funds following an index made up of a segment of the market based primarily on size. *Style Index* is set to one for funds following an index made up of a segment of the market based primarily on a certain style. *Sector Index* is set to one for funds following an index made up of a segment of the market based primarily on a certain sector. *Turnover* is the minimum of aggregated sales or aggregated purchases of securities, divided by the average 12-month TNA of the fund (CRSP MFDB variable turn_ratio). *Exp Ratio* is the ratio of TNA that shareholders pay for the fund's operating expenses (CRSP MFDB variable exp_ratio). *Mgmt Fees* is the management fee reported in the fund's Statement of Operations (CRSP MFDB variable mgmt_fee) and is a component of *Exp Ratio*. *12b-1 Fees* is the marketing and distribution fee reported in the fund's Statement of Operations (CRSP MFDB variable actual_12b1) and is a component of *Exp Ratio*. *Net Returns* is the fund's cumulative buy-and-hold return for the year (based on monthly returns, i.e., CRSP MFDB variable mret). *Gross Returns* is the sum of *Net Returns* and *Exp Ratio*. *Return Volatility* is the standard deviation of monthly returns. *Risk Adj Returns* is the ratio of *Net Returns* and *Return Volatility*. *Fund Flows* is the difference between the ratio of ending TNA to beginning TNA and one plus *Net Returns*. We winsorize fund-level continuous variables at the 1% and 99% levels by year to limit the influence of outliers. ***, **, and * indicate significance at the 1, 5, and 10 percent two-tailed level, respectively.

Panel A: Full Sample

Variable	Mean	Std Dev	25th	Median	75th	N
Sampling	0.42	0.46	0	0	1	3,365
Replicator	0.52	0.50	0	1	1	3,365
Sampler	0.37	0.48	0	0	1	3,365
Hybrid	0.11	0.32	0	0	0	3,365
Fund Size	2,760	8,339	32	224	1,085	3,365
Fund Age	11.11	6.86	5	11	16	3,365
Big Three	0.28	0.45	0	0	1	3,365
ETF	0.18	0.38	0	0	0	3,365
Domestic Index	0.83	0.38	1	1	1	3,365
#Constituents	896	896	507	507	929	3,365
Market Cap Index	0.80	0.40	1	1	1	3,365
Style Index	0.18	0.38	0	0	0	3,365
Sector Index	0.02	0.15	0	0	0	3,365
Turnover	0.39	1.00	0.05	0.14	0.26	3,365
Exp Ratio	0.51	0.51	0.16	0.31	0.70	3,365
Mgmt Fees	0.20	0.26	0.06	0.16	0.25	3,365
12b-1 Fees	0.15	0.28	0.00	0.00	0.25	3,365
Net Returns	12.25	13.44	1.02	13.47	21.53	3,365
Gross Returns	12.76	13.42	1.42	13.71	21.80	3,365
Return Volatility	4.21	2.01	3.00	3.84	5.00	3,365
Risk Adj Returns	4.36	5.62	0.25	3.75	5.98	3,365
Fund Flows	0.85	4.92	-0.14	0.01	0.28	3,365

Panel B: Means for Samplers vs. Replicators

Variable	Sampler	Replicator	Difference
Fund Size	3,366	2,751	615*
Fund Age	11.83	11.23	0.60**
Big Three	0.33	0.24	0.09***
ETF	0.26	0.13	0.13***
Domestic Index	0.75	0.90	-0.14***
#Constituents	1,109	739	370***
Market Cap Index	0.73	0.85	-0.12***
Style Index	0.25	0.13	0.12***
Sector Index	0.02	0.02	0.00
Turnover	0.70	0.19	0.51***
Exp Ratio	0.62	0.46	0.16***
Mgmt Fees	0.27	0.17	0.11***
12b-1 Fees	0.16	0.15	0.02
Net Returns	11.39	12.83	-1.44***
Gross Returns	12.01	13.29	-1.27**
Return Volatility	4.35	4.14	0.22***
Risk Adj Returns	4.01	4.51	-0.50**
Fund Flows	0.71	0.86	-0.16

Table 2. Determinants of Replication vs. Sampling

This table presents estimates of regressions of the fund's classification as a replicator, sampler, or hybrid on several other fund and benchmark index characteristics. The sample consists of 3,365 U.S.-based equity index fund-years during the period 2010 through 2020. To be clear, when we refer to a 'benchmark index' or simply to an 'index' without using the word fund, we are referring to the index a fund is tracking, rather than the fund itself. Variables are defined in Table 1. t-statistics are reported in parentheses below coefficients and are based on standard errors that are clustered by fund. ***, **, and * indicate significance at the 1, 5, and 10 percent two-tailed level, respectively.

	(1)	(2)
Variable	***Sampling***	***Sampling***
Log(Fund Size)	-0.013	-0.008
	(-1.40)	(-0.84)
Log(1+Fund Age)	0.112***	0.129***
	(4.20)	(4.30)
Big Three	-0.043	-0.084
	(-0.80)	(-1.36)
ETF	0.146*	0.203**
	(1.92)	(2.33)
Domestic Index	-0.188***	
	(-2.89)	
Log(#Constituents)	0.153***	
	(3.99)	
Style Index	0.265***	
	(3.62)	
Sector Index	0.401***	
	(3.27)	
Trend	-0.004	
	(-1.16)	
Observations	3,365	3,365
Adjusted R-sq.	0.121	0.153
Year x Benchmark Index FE	No	Yes

Table 3. Fund Turnover and Expenses

This table presents estimates of regressions of fund turnover and expenses on the fund's classification as a replicator, sampler, or hybrid, as well as several other fund and benchmark index characteristics. The sample consists of 3,365 U.S.-based equity index fund-years during the period 2010 through 2020. To be clear, when we refer to a 'benchmark index' or simply to an 'index' without using the word fund, we are referring to the index a fund is tracking, rather than the fund itself. Variables are defined in Table 1. t-statistics are reported in parentheses below coefficients and are based on standard errors that are clustered by fund. ***, **, and * indicate significance at the 1, 5, and 10 percent two-tailed level, respectively.

	(1)	(2)	(3)	(4)
Variable	***Turnover***	***Exp Ratio***	***Turnover***	***Exp Ratio***
Sampling	0.440***	0.168***	0.352***	0.140***
	(4.39)	(3.80)	(3.65)	(2.99)
Log(Fund Size)	-0.100***	-0.118***	-0.065***	-0.119***
	(-5.11)	(-14.76)	(-4.37)	(-15.29)
Log(1+Fund Age)	0.146***	0.207***	0.122***	0.205***
	(2.82)	(8.64)	(2.87)	(8.53)
Big Three	-0.176**	-0.223***	-0.268***	-0.219***
	(-2.48)	(-6.19)	(-3.36)	(-5.51)
ETF	-0.400***	-0.013	-0.635***	-0.013
	(-3.02)	(-0.28)	(-3.66)	(-0.23)
Domestic Index	0.140**	-0.029		
	(1.98)	(-0.71)		
Log(#Constituents)	-0.026	-0.084***		
	(-0.46)	(-2.63)		
Style Index	1.242***	0.134**		
	(4.89)	(1.98)		
Sector Index	0.113	-0.174		
	(0.70)	(-1.63)		
Trend	-0.006	-0.009***		
	(-1.07)	(-3.17)		
Observations	3,365	3,365	3,365	3,365
Adjusted R-sq.	0.348	0.524	0.460	0.550
Year x Benchmark Index FE	No	No	Yes	Yes

Table 4. Fund Return Performance

This table presents estimates of regressions of fund returns and return volatility on the fund's classification as a replicator, sampler, or hybrid, as well as several other fund and benchmark index characteristics. The sample consists of 3,365 U.S.-based equity index fund-years during the period 2010 through 2020. To be clear, when we refer to a 'benchmark index' or simply to an 'index' without using the word fund, we are referring to the index a fund is tracking, rather than the fund itself. Variables are defined in Table 1. t-statistics are reported in parentheses below coefficients and are based on standard errors that are clustered by fund. ***, **, and * indicate significance at the 1, 5, and 10 percent two-tailed level, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variable	***Net Returns***	***Gross Returns***	***Return Volatility***	***Risk Adj Returns***	***Net Returns***	***Gross Returns***	***Return Volatility***	***Risk Adj Returns***
Sampling	-0.723***	-0.554**	0.091	-0.184	-0.589***	-0.449***	0.030*	-0.165***
	(-2.68)	(-2.11)	(1.17)	(-1.41)	(-3.70)	(-2.98)	(1.79)	(-2.81)
Log(Fund Size)	0.277***	0.159***	-0.095***	0.205***	0.205***	0.086***	-0.002	0.051***
	(5.13)	(3.06)	(-6.43)	(7.94)	(8.89)	(4.09)	(-0.79)	(7.80)
Log(1+Fund Age)	-0.259	-0.052	0.166***	-0.263**	-0.113	0.092	-0.021	0.009
	(-1.12)	(-0.23)	(3.19)	(-2.39)	(-0.79)	(0.64)	(-1.50)	(0.17)
Big Three	-0.085	-0.308	-0.228*	0.179	0.484**	0.265	-0.007	0.159**
	(-0.21)	(-0.76)	(-1.91)	(1.02)	(2.07)	(1.15)	(-0.57)	(2.03)
ETF	-0.841	-0.854	0.710***	-0.896***	-0.584	-0.597	0.003	-0.169
	(-1.48)	(-1.51)	(4.63)	(-4.22)	(-1.19)	(-1.22)	(0.17)	(-1.04)
Domestic Index	6.236***	6.207***	0.170	1.190***				
	(19.63)	(19.64)	(1.40)	(7.47)				
Log(#Constituents)	0.920***	0.836***	0.132	0.062				
	(4.18)	(3.92)	(1.50)	(0.53)				
Style Index	0.776	0.910*	0.257**	-0.209				
	(1.44)	(1.72)	(2.33)	(-0.98)				
Sector Index	-3.037**	-3.211**	0.668	-1.725***				
	(-2.14)	(-2.32)	(1.53)	(-4.29)				
Trend	0.119***	0.110***	0.197***	-0.008				
	(2.81)	(2.61)	(15.99)	(-0.46)				
Observations	3,365	3,365	3,365	3,365	3,365	3,365	3,365	3,365
Adjusted R-sq.	0.032	0.030	0.102	0.014	0.971	0.972	0.990	0.986
Year x Benchmark Index FE	No	No	No	No	Yes	Yes	Yes	Yes

Table 5. Portfolio Return Performance

Panel A presents correlations between fund (net) returns and portfolio returns. Panel B presents estimates of regressions of portfolio returns on the fund's classification as a replicator, sampler, or hybrid, as well as several other fund and benchmark index characteristics. Our calculations of portfolio returns assume that the fund maintained the stock positions in its portfolio for the prior 1, 3, or 12 months. Specifically, *Portfolio Returns (X months)* is the sum of the returns of each stock held by the fund over the X months preceding the end of the reporting year multiplied by the portfolio weight (i.e., the ratio of the dollar value of the holding and the fund's TNA). Because returns are unavailable for the stocks held by a few of the funds, mainly those with a foreign mandate/focus, the sample in this table consists of 3,021 of the 3,365 total U.S.-based equity index fund-years during the period 2010 through 2020. To be clear, when we refer to a 'benchmark index' or simply to an 'index' without using the word fund, we are referring to the index a fund is tracking, rather than the fund itself. Variables are defined in Table 1. t-statistics are reported in parentheses below coefficients and are based on standard errors that are clustered by fund. ***, **, and * indicate significance at the 1, 5, and 10 percent two-tailed level, respectively.

Panel A: Returns Correlations

	Net Returns	*Portfolio Returns (1 month)*	*Portfolio Returns (3 months)*	*Portfolio Returns (12 months)*
Net Returns				
Portfolio Returns (1 month)	0.62			
Portfolio Returns (3 months)	0.46	0.86		
Portfolio Returns (12 months)	0.78	0.64	0.63	

Panel B: Portfolio Returns Regressions

Variable	(1) *Portfolio Returns (1 month)*	(2) *Portfolio Returns (3 months)*	(3) *Portfolio Returns (12 months)*	(4) *Portfolio Returns (1 month)*	(5) *Portfolio Returns (3 months)*	(6) *Portfolio Returns (12 months)*
Sampling	-0.169*	-1.275***	-2.901**	-0.050**	-0.155*	-0.440***
	(-1.77)	(-3.15)	(-2.55)	(-2.44)	(-1.91)	(-2.88)
Log(Fund Size)	-0.050**	-0.201**	-0.081	0.005	0.023**	0.059**
	(-2.35)	(-2.42)	(-0.37)	(1.57)	(2.25)	(2.54)
Log(1+Fund Age)	0.378***	1.284***	1.840***	0.032***	0.074**	0.077
	(5.32)	(5.45)	(3.10)	(2.74)	(2.21)	(1.26)
Big Three	-0.344**	-0.635	-0.254	0.009	0.045	-0.073
	(-2.42)	(-1.24)	(-0.18)	(0.62)	(1.17)	(-0.66)
ETF	0.345**	1.860***	1.358	-0.006	0.020	0.120
	(2.23)	(3.17)	(0.77)	(-0.31)	(0.39)	(0.63)
Log(#Constituents)	0.226**	0.166	-0.720			
	(2.27)	(0.50)	(-0.88)			
Style Index	0.653***	2.377***	3.777**			
	(4.50)	(4.57)	(2.12)			
Sector Index	0.638*	-2.519***	-7.355***			
	(1.81)	(-3.05)	(-3.86)			
Trend	-0.145***	-0.034	0.796***			
	(-9.17)	(-0.80)	(9.34)			
Observations	3,021	3,021	3,021	3,021	3,021	3,021
Adjusted R-sq.	0.014	0.023	0.049	0.992	0.990	0.988
Year x Benchmark Index FE	No	No	No	Yes	Yes	Yes

Table 6. S&P 500 and Market Cap Index Funds

Panel A (Panel B) presents estimates of the main regressions from Tables 3 and 4 after splitting the sample based on whether a fund tracks the S&P 500 Index (tracks an index made up of a segment of the market based primarily on size, i.e., where *Market Cap Index* = 1). The total sample includes 3,365 U.S.-based equity index fund-years during the period 2010 through 2020. The Panel A subsample consists of the 1,114 fund-years that track the S&P 500 Index, and the Panel B subsample consists of the 2,695 fund-years that track a Market Cap Index. To be clear, when we refer to a 'benchmark index' or simply to an 'index' without using the word fund, we are referring to the index a fund is tracking, rather than the fund itself. Variables are defined in Table 1. t-statistics are reported in parentheses below coefficients and are based on standard errors that are clustered by fund. ***, **, and * indicate significance at the 1, 5, and 10 percent two-tailed level, respectively.

Panel A: S&P 500 Index Subsample

	(1)	(2)	(3)	(4)	(5)	(6)
Variable	***Turnover***	***Exp Ratio***	***Net Returns***	***Gross Returns***	***Return Volatility***	***Risk Adj Returns***
Sampling	0.277***	0.092	-0.754***	-0.662***	0.053	-0.236*
	(2.67)	(0.98)	(-2.88)	(-2.68)	(1.19)	(-1.93)
Log(Fund Size)	-0.010	-0.113***	0.214***	0.101***	0.005	0.052***
	(-1.15)	(-9.22)	(5.75)	(2.89)	(1.18)	(3.64)
Log(1+Fund Age)	-0.041*	0.086**	-0.173	-0.087	-0.037	0.106
	(-1.83)	(2.36)	(-0.75)	(-0.37)	(-1.18)	(1.02)
Big Three	-0.044	-0.064	0.686	0.623	-0.030	0.319
	(-1.08)	(-0.96)	(1.23)	(1.11)	(-1.08)	(1.62)
ETF	0.019	0.030	-4.037	-4.007	-0.181*	-1.165
	(0.09)	(0.33)	(-1.17)	(-1.14)	(-1.69)	(-1.04)
Observations	1,114	1,114	1,114	1,114	1,114	1,114
Adjusted R-sq.	0.162	0.424	0.945	0.946	0.977	0.974
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Panel B: Market Cap Index Subsample

Variable	(1) *Turnover*	(2) *Exp Ratio*	(3) *Net Returns*	(4) *Gross Returns*	(5) *Return Volatility*	(6) *Risk Adj Returns*
Sampling	0.202***	0.090*	-0.470***	-0.379***	0.024	-0.156**
	(3.10)	(1.77)	(-3.05)	(-2.63)	(1.27)	(-2.46)
Log(Fund Size)	-0.026***	-0.114***	0.182***	0.068***	0.000	0.049***
	(-3.43)	(-14.94)	(8.86)	(3.62)	(0.09)	(7.14)
Log(1+Fund Age)	-0.015	0.167***	-0.144	0.023	-0.013	0.017
	(-0.90)	(7.34)	(-1.01)	(0.16)	(-0.89)	(0.29)
Big Three	-0.021	-0.164***	0.491*	0.327	-0.006	0.167*
	(-0.73)	(-4.17)	(1.72)	(1.14)	(-0.44)	(1.72)
ETF	-0.016	0.153***	-1.024	-0.871	-0.017	-0.295
	(-0.34)	(3.24)	(-1.38)	(-1.17)	(-0.72)	(-1.20)
Observations	2,695	2,695	2,695	2,695	2,695	2,695
Adjusted R-sq.	0.169	0.489	0.979	0.980	0.991	0.987
Year x Benchmark Index FE	Yes	Yes	Yes	Yes	Yes	Yes

Table 7. Cross-Sectional Test - Number of Index Constituents

This table presents estimates of the main regressions from Tables 3 and 4 after including an interaction between sampling and number of constituents, *Sampling x Log(#Constituents)*. Panel A does not include Year x Benchmark Index fixed effects so that we can evaluate the main effects of the index-level controls, and Panel B includes Year x Benchmark Index fixed effects. The sample consists of 3,365 U.S.-based equity index fund-years during the period 2010 through 2020. To be clear, when we refer to a 'benchmark index' or simply to an 'index' without using the word fund, we are referring to the index a fund is tracking, rather than the fund itself. Variables are defined in Table 1. t-statistics are reported in parentheses below coefficients and are based on standard errors that are clustered by fund. ***, **, and * indicate significance at the 1, 5, and 10 percent two-tailed level, respectively.

Panel A: Index Level Controls

	(1)	(2)	(3)	(4)	(5)	(6)
Variable	***Turnover***	***Exp Ratio***	***Net Returns***	***Gross Returns***	***Return Volatility***	***Risk Adj Returns***
Sampling	1.888**	0.570	-7.081**	-6.511**	0.283	-2.483**
	(2.35)	(1.54)	(-2.36)	(-2.23)	(0.29)	(-2.02)
Sampling x Log(#Constituents)	-0.222*	-0.062	0.975**	0.913**	-0.029	0.352*
	(-1.93)	(-1.12)	(2.18)	(2.11)	(-0.20)	(1.87)
Log(#Constituents)	0.101	-0.048	0.364	0.315	0.149	-0.139
	(1.52)	(-1.42)	(1.16)	(1.04)	(1.35)	(-1.02)
Log(Fund Size)	-0.098***	-0.118***	0.266***	0.149***	-0.095***	0.201***
	(-5.02)	(-14.45)	(4.94)	(2.88)	(-6.46)	(7.76)
Log(1+Fund Age)	0.141***	0.206***	-0.239	-0.033	0.165***	-0.256**
	(2.75)	(8.58)	(-1.03)	(-0.15)	(3.14)	(-2.33)
Big Three	-0.175**	-0.223***	-0.088	-0.311	-0.228*	0.178
	(-2.53)	(-6.27)	(-0.22)	(-0.78)	(-1.91)	(1.01)
ETF	-0.408***	-0.016	-0.806	-0.822	0.709***	-0.884***
	(-3.11)	(-0.33)	(-1.44)	(-1.48)	(4.64)	(-4.13)
Domestic Index	0.148**	-0.027	6.202***	6.175***	0.171	1.178***
	(2.13)	(-0.66)	(19.92)	(19.88)	(1.41)	(7.45)
Style Index	1.216***	0.127*	0.888	1.015*	0.253**	-0.168
	(4.86)	(1.89)	(1.62)	(1.89)	(2.24)	(-0.78)
Sector Index	0.148	-0.164	-3.190**	-3.355**	0.673	-1.780***
	(0.88)	(-1.55)	(-2.25)	(-2.43)	(1.54)	(-4.33)
Trend	-0.006	-0.009***	0.119***	0.110***	0.197***	-0.008
	(-1.08)	(-3.18)	(2.82)	(2.62)	(15.99)	(-0.45)
Observations	3,365	3,365	3,365	3,365	3,365	3,365
Adjusted R-sq.	0.353	0.526	0.032	0.031	0.102	0.014
Year x Benchmark Index FE	No	No	No	No	No	No

Panel B: Year x Benchmark Index Fixed Effects

Variable	(1) *Turnover*	(2) *Exp Ratio*	(3) *Net Returns*	(4) *Gross Returns*	(5) *Return Volatility*	(6) *Risk Adj Returns*
Sampling	1.490*	0.490	-5.326***	-4.836**	0.371**	-0.890***
	(1.67)	(1.34)	(-2.77)	(-2.56)	(1.98)	(-2.70)
Sampling x Log(#Constituents)	-0.174	-0.054	0.726***	0.672**	-0.052*	0.111**
	(-1.34)	(-0.99)	(2.59)	(2.45)	(-1.93)	(2.56)
Log(Fund Size)	-0.064***	-0.118***	0.199***	0.081***	-0.001	0.050***
	(-4.24)	(-15.12)	(9.12)	(4.12)	(-0.63)	(7.81)
Log(1+Fund Age)	0.119***	0.204***	-0.099	0.106	-0.022	0.012
	(2.80)	(8.48)	(-0.69)	(0.73)	(-1.56)	(0.20)
Big Three	-0.267***	-0.219***	0.482**	0.263	-0.007	0.158**
	(-3.43)	(-5.60)	(2.09)	(1.15)	(-0.56)	(2.04)
ETF	-0.638***	-0.014	-0.570	-0.583	0.002	-0.167
	(-3.74)	(-0.25)	(-1.17)	(-1.20)	(0.11)	(-1.03)
Observations	3,365	3,365	3,365	3,365	3,365	3,365
Adjusted R-sq.	0.463	0.550	0.971	0.972	0.990	0.986
Year x Benchmark Index FE	Yes	Yes	Yes	Yes	Yes	Yes

Table 8. Fund Flows

This table presents estimates of regressions of fund flows on the fund's classification as a replicator, sampler, or hybrid as well as several other fund and benchmark index characteristics. In addition to the variables included in the main regressions from Tables 3 and 4, we include an interaction between sampling and the linear time trend (*Trend*), and in columns 3 and 4 we also control for the fund's prior year net return (*Prior Return*). The sample consists of 3,365 U.S.-based equity index fund-years during the period 2010 through 2020. To be clear, when we refer to a 'benchmark index' or simply to an 'index' without using the word fund, we are referring to the index a fund is tracking, rather than the fund itself. Variables are defined in Table 1. t-statistics are reported in parentheses below coefficients and are based on standard errors that are clustered by fund. ***, **, and * indicate significance at the 1, 5, and 10 percent two-tailed level, respectively.

	(1)	(2)	(3)	(4)
Variable	***Fund Flows***	***Fund Flows***	***Fund Flows***	***Fund Flows***
Sampling	-0.698*	-0.714*	-0.076	-0.046
	(-1.90)	(-1.95)	(-0.65)	(-0.35)
Trend	-0.086**		-0.030**	
	(-2.35)		(-2.54)	
Sampling x Trend	0.153***	0.170***	0.046***	0.044***
	(3.19)	(3.66)	(2.99)	(2.72)
Log(Fund Size)	-0.046	-0.040	-0.035**	-0.033**
	(-1.45)	(-1.01)	(-2.16)	(-1.99)
Log(1+Fund Age)	-2.072***	-2.117***	-0.632***	-0.611***
	(-8.71)	(-8.23)	(-6.12)	(-5.28)
Big Three	-0.087	-0.188	0.275	0.225
	(-0.36)	(-0.65)	(1.58)	(1.15)
ETF	-0.460**	-0.199	-0.261*	-0.201
	(-2.04)	(-0.77)	(-1.79)	(-1.24)
Domestic Index	0.402		0.029	
	(1.46)		(0.17)	
Log(#Constituents)	0.238		0.186*	
	(1.54)		(1.86)	
Style Index	-0.208		0.062	
	(-1.03)		(0.55)	
Sector Index	0.003		0.666	
	(0.00)		(1.45)	
Prior Returns			0.143	0.732
			(0.78)	(0.28)
Observations	3,365	3,365	2,929	2,929
Adjusted R-sq.	0.098	0.105	0.069	0.083
Year x Benchmark Index FE	No	Yes	No	Yes

Table 9. Expense Ratio Components

This table presents estimates of regressions of two major components of the expense ratio, management fees and 12b-1 fees, on the fund's classification as a replicator, sampler, or hybrid as well as several other fund and benchmark index characteristics. The sample consists of 3,365 U.S.-based equity index fund-years during the period 2010 through 2020. To be clear, when we refer to a 'benchmark index' or simply to an 'index' without using the word fund, we are referring to the index a fund is tracking, rather than the fund itself. Variables are defined in Table 1. t-statistics are reported in parentheses below coefficients and are based on standard errors that are clustered by fund. ***, **, and * indicate significance at the 1, 5, and 10 percent two-tailed level, respectively.

	(1)	(2)	(3)	(4)
Variable	***Mgmt Fees***	***12b-1 Fees***	***Mgmt Fees***	***12b-1 Fees***
Sampling	0.092***	0.031	0.071***	0.031
	(3.89)	(1.14)	(2.95)	(1.06)
Log(Fund Size)	-0.024***	-0.058***	-0.022***	-0.064***
	(-5.15)	(-10.10)	(-5.62)	(-10.41)
Log(1+Fund Age)	0.068***	0.099***	0.063***	0.095***
	(5.22)	(5.84)	(5.06)	(5.44)
Big Three	-0.164***	-0.007	-0.160***	0.003
	(-6.52)	(-0.36)	(-6.81)	(0.14)
ETF	0.048	-0.018	0.024	0.015
	(1.49)	(-0.72)	(0.71)	(0.49)
Domestic Index	-0.039	0.036		
	(-1.49)	(1.44)		
Log(#Constituents)	-0.055***	-0.028		
	(-3.05)	(-1.56)		
Style Index	0.191***	-0.053		
	(4.36)	(-1.36)		
Sector Index	-0.108	-0.137***		
	(-1.43)	(-2.83)		
Trend	-0.002	-0.006***		
	(-0.97)	(-3.29)		
Observations	3,365	3,365	3,365	3,365
Adjusted R-sq.	0.304	0.322	0.466	0.292
Year x Benchmark Index FE	No	No	Yes	Yes

Table 10. No Benchmark Index Controls or Fixed Effects

This table presents estimates of the main regressions from Tables 3 and 4 after relaxing the sample selection criteria relating to the benchmark index a fund tracks. Specifically, for the analyses in this table, we relax the criteria requiring (1) non-missing data for the benchmark index control variables (e.g., *Market Cap Index* and *#Constituents*) and (2) that the fund's index is followed by both replicators and samplers (which ensures variation in our measure of interest in tests that include year-benchmark index fixed effects). These changes to the selection criteria result in a total sample of 12,969 U.S.-based equity index fund-years during the period 2010 through 2020. To be clear, when we refer to a 'benchmark index' or simply to an 'index' without using the word fund, we are referring to the index a fund is tracking, rather than the fund itself. Variables are defined in Table 1. t-statistics are reported in parentheses below coefficients and are based on standard errors that are clustered by fund. ***, **, and * indicate significance at the 1, 5, and 10 percent two-tailed level, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
Variable	***Turnover***	***Exp Ratio***	***Net Returns***	***Gross Returns***	***Return Volatility***	***Risk Adj Returns***
Sampling	0.299***	0.157***	-1.027***	-0.869***	-0.099	-0.165*
	(4.70)	(7.41)	(-3.34)	(-2.83)	(-1.34)	(-1.85)
Log(Fund Size)	-0.112***	-0.090***	0.423***	0.333***	-0.089***	0.165***
	(-7.70)	(-19.93)	(7.14)	(5.65)	(-6.26)	(9.25)
Log(1+Fund Age)	0.164***	0.188***	0.184	0.372*	0.204***	-0.123**
	(4.36)	(15.76)	(0.88)	(1.78)	(5.34)	(-2.07)
Big Three	-0.316***	-0.234***	0.120	-0.114	-0.043	0.061
	(-10.18)	(-15.96)	(0.39)	(-0.37)	(-0.63)	(0.67)
ETF	-0.069	-0.046**	-1.978***	-2.024***	0.777***	-1.108***
	(-1.41)	(-2.57)	(-7.12)	(-7.31)	(11.74)	(-13.40)
Observations	12,969	12,969	12,969	12,969	12,969	12,969
Adjusted R-sq.	0.143	0.437	0.511	0.511	0.553	0.643
Year FE	Yes	Yes	Yes	Yes	Yes	Yes