Direct Indexing and the Limits of ETF Tax Efficiency*

Thomas Ernst† Marco Sammon‡

August 27, 2026

Abstract

We examine the after-tax tradeoff between direct indexing and index-tracking ETFs. ETFs largely eliminate capital gains through in-kind redemptions, but they cannot harvest losses to offset against other gains. Direct indexing can realize losses but is exposed to taxable events from mergers and buyouts. Using historical simulations, we show that direct indexing often outperforms ETFs for tax-sensitive investors, particularly over shorter investment horizons and when contributions are recurring rather than lump sum. The advantage, however, can depend on the period the strategy is implemented: ETFs dominate in periods with heavy merger activity and lower dividend payouts, particularly the 1970s and 1980s.

Keywords: Direct indexing; ETFs; tax-loss harvesting; transaction costs; tracking error.

*For helpful conversations and suggestions we are thankful to: Mike Faulkender, Mike Gallmeyer, Pete Kyle, Mark Loewenstein, Rich Matthews, Anna Pavlova, and Chester Spatt as well as seminar participants at the University of Maryland.

†ternst@umd.edu

‡mcsammon@gmail.com

I. Introduction

A taxable investor who wants index exposure can hold shares of an index ETF or own the underlying securities directly through a direct indexing strategy. The ETF structure has a large tax advantage: through in-kind creations and redemptions, an ETF can almost entirely avoid creating capital gains for shareholders. But ETF ownership also bundles the underlying securities into fund shares. The shareholder, therefore, cannot harvest losses on individual stocks, and a sale of ETF shares realizes the investor's embedded gain or loss in the *fund* itself. In contrast, a direct-index investor can harvest losses on individual stocks and can meet cash needs by selling high-cost-basis lots. A direct index investor, however, cannot use the ETF creation/redemption mechanism to remove low-basis positions. Recent declines in transaction costs and the removal of constraints on fractional share trading have made direct indexing more feasible with small amounts of capital. The question is therefore not whether ETFs are tax efficient [Poterba and Shoven, 2002, Moussawi et al., 2025]. The question is whether the ETF's ability to avoid stock-level gains is more valuable than the direct-index investor's ability to harvest stock-level losses and to choose which lots to sell.

This paper explores the tradeoff between direct indexing and an ETF for index exposure. The tradeoff is nuanced, and hinges on two critical factors: the extent to which losses can be used, and the timing of any investment. Losses can be carried forward to offset future portfolio gains, but they can also be used to offset personal income and to offset outside-the-portfolio capital gains. We being our analysis by comparing an ETF and its direct indexing analog using modern transaction costs and tax rates against historical realized returns. Specifically, we construct a generic value-weighted index of the 500 largest companies in any year, and we compare the after-tax returns of a direct indexing strategy to a hypothetical ETF on this portfolio.

Under the most strict assumption that the losses can only be used to offset future gains on the portfolio itself, an ETF will inevitably dominate the direct indexing strategy. As the index rises over time, there are larger embedded gains and fewer losses to harvest, a process typically called *portfolio ossification*. While prior realized losses can be carried forward, they are carried forward in nominal terms. And, eventually, forced capital gains from mergers and acquisitions (M&A) will use up all these losses. These gains will leave the direct index investor with a net tax bill, making them worse off than if they had invested in an ETF instead. There is considerable variation in how long this process takes: an investment initiated in the 1920s has net losses more than 50 years later, while an investment started in the 1970s has a net gain in under 20 years. While portfolios formed after 1996

have yet to see net gains, all portfolios prior to this date do. For direct indexing to compete with an ETF at long horizons, an investor must have some use for losses other than carrying them forward to offset portfolio gains.

Using losses to offset income has two major advantages. First, personal income taxes are almost always higher than capital gains taxes. Second, using a capital loss to offset income produces a tax credit which can be immediately invested. Because this tax windfall can be invested in the portfolio, it can grow over time, unlike the value of a loss which is carried forward in nominal terms. The ability to offset personal income is capped, however, so the benefits of offsetting personal income decline with portfolio size. In the time-series, a one-time lump sum investment of $230,627 ensures that in at least 75% of years, an investor will accumulate sufficient losses to maximize a $3,000 annual personal income offset for at least 15 years. For an investor with a small investment, the benefit from the personal income tax almost always dominates an investment in an ETF. For larger investments, the cap on personal income tax binds, and even if the personal income tax credits are re-invested at the portfolio rate of return, the taxes due on M&A activity will eventually dominate. This occurs for a portfolio of as small as $277,000 in 1978 or as large as $2.3 million in 1990. For investments above this quantity, an investor would be better off in an ETF and avoiding capital gains, assuming no other uses for losses.

Losses can also be used to offset an outside-the-portfolio gain e.g., the sale of a private business or a house. This provides an immediate tax credit which, if invested, can grow to exceed the value of any future M&A tax burden. The earlier in the portfolio lifetime this occurs, the more likely it is that the investment has time to grow to exceed future taxes. For investments in the 1930s, 40s, 60s, or 90s, if an investor can use all of their accumulated losses to offset an outside-the-portfolio gain in roughly the first two decades of the portfolio lifetime, the direct index strategy will deliver a superior return compared to an ETF. For investments in the 1950s and 70s to early 80s, there are only a few select years during which an outside-the-portfolio gain is sufficiently beneficial so that direct indexing would dominate an ETF.

An additional advantage of direct indexing is customization. While ETF holdings are decided by the fund sponsor, an investor has full control over their holdings in a direct indexing strategy. One of the specific advantages of customization arises in how an investor can meet an outside the portfolio cash need. A direct-index investor who needs to liquidate a fraction of their portfolio can chose to sell only lots with the smallest taxable gains, while an ETF investor liquidating the same fraction of their ETF shares would effectively sell all lots proportionately. For a 30% liquidation of a

portfolio, the tax savings by selectively selling lots amount to between 1.5 and 2% of the remaining value of the portfolio, with the advantage slightly increasing with portfolio age. This tax savings comes at the cost of deviation from the benchmark index. In addition to tracking error, we find that the median portfolio slightly under-performs post-liquidation, with the liquidated portfolio having a cumulative return of 0.5% less than the never-liquidated portfolio over the subsequent 30 years. Our generic portfolio has a loading on the market portfolio of almost one, but loads somewhat negatively (-.14) on SMB and has a small positive loading (0.02) on momentum. In our samples, the liquidated portfolios have a slightly lower loading on the market risk premium after liquidation, consistent with lower future performance.

While much of our analysis focuses on the costs and benefits of direct indexing for our generic top-N-largest stock index, investors in practice can choose among thousands of indexes tracked by ETFs. We also, therefore, examine the costs and benefits of direct indexing for this broad set of real-world portfolios. Because the ETF structure is able to avoid many types of taxable events, we do not expect direct indexing to generate a large advantage just from being able to offset future gains. Instead, an important benefit of direct indexing may come from flexibility in how the index is implemented, including trading less frequently, avoiding rebalancing rules known to predict low returns, and avoiding crowded trades [Sammon and Shim, 2026b, Murray and Sammon, 2026, Arnott et al., 2023, Wang et al., 2023]. We find that, even when harvested losses can only offset future gains, direct indexing outperforms the ETF by a median 22 to 45 basis points per year over three- to ten-year horizons. Allowing up to $3,000 annually of remaining unused losses to offset ordinary income increases this advantage to roughly 68 to 79 basis points per year. Perhaps surprisingly, the direct-indexing advantage is largest for higher-turnover ETFs. Although churn in these portfolios likely generates more taxable events, it also creates more opportunities for a direct-indexing investor to avoid costly or poorly timed trades.

II. Related literature

Poterba and Shoven [2002] first notes the tax advantage ETFs relative to mutual funds in their ability to avoid distributing capital gains taxes from M&A activity. Moussawi et al. [2025] estimate that this efficiency has increased long-term ETF investors' "after-tax returns by 1.05% per year relative to mutual funds". Both papers compare an ETF to a mutual fund; our main contribution is to compare an ETF to a direct-index strategy. This comparison is a recent question: relevant only

with recent dramatic declines in transaction costs.[1]

Surprisingly, we find that most ETFs are too young to deliver any advantage over direct indexing. Accumulated losses would cover any gains from M&A on a value-weighted portfolio starting at any point from 1996 onwards, and only a handful of ETFs were formed before this date. While accumulated losses (carried forward in nominal terms) are eventually used up, this process typically takes more than 30 or 40 years. So, for an investor with a relatively shorter horizon, direct indexing performs at least as well as an ETF.

Our work connects to substantial prior work on the optimal investment strategy under taxation. Constantinides [1983] shows it is optimal to realize losses and hold gains, but ignores the impact of wash sale rules. Dammon et al. [2001] show that capital gains taxes give investors a tradeoff between mean-variance optimization and minimizing taxes, and that optimal equity holdings increase in investor age due to a combination of embedded gains and the advantage of stepped-up basis at death. Fischer and Gallmeyer [2016] extend this analysis to multiple assets, and show that the optimal balance between diversification and tax considerations can save an investor as much as 6.4% of initial wealth.[2]

III. Data

This section describes the data used in the empirical tests. We first describe how we identify index ETFs and link their returns, fees, taxable distributions and holdings. We then describe the data cleaning process used to identify the stocks eligible to be included in our direct-indexing simulation.

A. ETF data

To construct our ETF dataset, we combine CRSP mutual fund data with Thompson S12 holdings data. CRSP provides fund returns, total net assets (TNA), expense ratios, portfolio turnover, and distributions. These data allow us to measure ETF returns before and after fund fees and as well as pre- and post-tax. Thompson S12 provides the securities held by each fund at the end of each quarter, which we use to calculate end-of-quarter portfolio weights. While a investor engaged in direct indexing does not pay an ETF expense ratio, they may still face an index licensing fee if they

[1] See e.g., Angel et al. [2011] and Adams et al. [2024].

[2] The tax tradeoff also depends on the exact tax rule used (DeMiguel and Uppal [2005]), the use of tax-deferred accounts (Dammon et al. [2004]), the limited use of losses (Marekwica [2012] and Ehling et al. [2018]), and whether the losses are short or long term (Dai et al. [2015]).

wish to to track a particular index. We obtain licensing fee estimates for a subset of indexes from An et al. [2023].

We identify ETFs from the CRSP mutual fund summary file using the ETF flag. Following Appel et al. [2016], we classify index ETFs using the CRSP index fund flag and by searching for keywords in fund names which typically indicate that the fund is tracking an index. We then map each ETF to CRSP stock data using historical CUSIP and trading symbol. Finally, we use the WRDS MF links database to merge CRSP portfolio identifiers to WFICNs and ultimately to S12 fund numbers (fundnos).

Our ETF sample is restricted to funds that are (1) identified as passive using the procedure above and (2) can be linked between the CRSP mutual fund data and the S12 holdings data. We drop leveraged ETFs, which we identify using the Lipper class and Lipper objective fields. For funds with multiple share classes, we aggregate returns, TNA, expense ratios, turnover and distributions to the WFICN-quarter level, weighting each share class by its fraction of lagged TNA at the portfolio level. Fund flows are measured following [Chevalier and Ellison, 1997, Lou, 2012, Barber et al., 2016] as the change in assets under management (AUM) not explained by returns: $flow_{f,t} = AUM_t - AUM_{t-1} * (1 + r_{f,t})$. As a robustness check, for ETFs we can also use changes in shares outstanding as an alternative way to identify flows through net creations and redemptions.

The ETFs' holdings are obtained from Thompson S12. For each fund-quarter, we compute portfolio weights as the value of that position divided by the total value of reported positions for that ETF that quarter.[3] A direct indexing strategy that observes ETF holdings at quarter-end (e.g., through the daily basket disclosures) could use those weights in real time as the target portfolio when rebalancing.

We also compute holdings-implied ETF returns. For each ETF-quarter, we merge the ETF's end-of-quarter holdings with stock returns from CRSP over the following quarter and compute the return of that portfolio. This is the return an investor would earn if the ETF did not rebalance at all throughout the quarter, and will serve as an additional benchmark relative to our tax loss harvesting strategy.

For fund-quarter-level tax burdens, we use data from the CRSP mutual fund database on distributions. This is important for comparing ETFs with mutual funds tracking the same indexes, as well as with direct indexing. Mutual funds typically create taxable events for shareholders when

[3]Importantly, this is portfolio weights relative to *reported* positions, not relative to TNA. We do this because we want the weights to add up to one, especially when we will be using direct indexing to replicate these' ETFs strategies.

they sell securities either to rebalance or meet redemptions. Direct indexing creates taxable events e.g., when stocks are bought out at a premium relative to their purchase price. ETFs, on the other hand, can use the creation and redemption mechanism to minimize these kinds of taxable events. For example, Vanguard's Total Market Index Fund (Ticker: VTI) has not had a taxable event for shareholders (other than dividend payments) in over 10 years, despite significant rebalancing.

To calculate tax burdens, we classify distributions into ordinary or short-term taxable distributions, long-term or qualified taxable distributions and tax-exempt distributions. We assume that ordinary and short-term distributions are taxed at 37%, while long-term and qualified-dividend distributions are taxed at 20%. We aggregate taxable events from the year-share class level to the year-WFICN level again weighting each share class by beginning of year TNA. And finally, we map the resulting estimated tax burden back to S12 fundnos using the MF links database.

Finally, we utilize two different measures of turnover. First, we calculate turnover of ETF shares in the *secondary market*. We compute this measure by identifying our sample of ETFs in the CRSP stock database and calculating turnover each day as dollar volume divided by market capitalization. To reduce the influence of outliers, we drop the first 8 quarters after an ETF first appears, and then take the median of this quantity at the ETF level. The second turnover measure we use is *portfolio* turnover of the fund's holdings. This is the measure of turnover funds report to the SEC and is directly available in the CRSP mutual fund data [Pástor et al., 2017].

B. Security-level data

The security-level data come from CRSP. We filter for ordinary common shares (share codes 10 and 11) traded on major exchanges (exchange codes 1 through 3). For our direct indexing strategy, we use data on these stocks from 1927-2023.

When running our direct indexing strategy, we work with split-adjusted prices and shares outstanding. While this is standard, we would like to highlight that carefully accounting for splits can matter a lot because the same position may be held across multiple corporate actions before it is sold or liquidated for other reasons (e.g., because the stock delists). To account for possible issues with split adjustment factors in CRSP, we drop stock-quarter observations with inconsistent split-adjustment factors for prices and shares outstanding. Specifically, we exclude observations for which the ratio of the cumulative factor to adjust shares outstanding (cfacshr) to the cumulative factor to adjust prices (cfacpr) is below 0.8 or above 1.2.[4] Importantly, this filter uses only information

[4]For almost all stocks, especially in post-1980 data, this ratio is exactly 1. While this filter drops a small percentage

available at the time our direct indexing portfolios rebalance and therefore could be applied in real time.

We also remove days in CRSP where there are fewer than 50 total stocks trading. These are unusual dates, e.g., the market closure around Hurricane Sandy, where a small number of securities still have data in CRSP even though the market was closed. If these dates happen to fall around the beginning or the end of a quarter, they can create significant issues for the tax loss harvesting strategy's rebalancing behavior. As with the observations where cfacshr and cfacpr don't align, this problem is very small as a percentage of total days, but it can matter when the simulation tracks every individual tax lot and needs to assign trades to specific dates.

When calculating the performance of our tax loss harvesting strategy, we use both total returns (ret) and ex-dividend returns (retx). The difference between the two identifies the dividend yield. Dividends are relevant because both ETF investors and direct-index investors pay taxes on taxable dividends.

CRSP delisting data are used to identify an additional source of (possibly forced) taxable events. When CRSP reports a cash delisting payment, we treat the payment as cash proceeds to the direct-index investor on the delisting pay date (dlpdt). If the delisting amount and/or return is missing, to be conservative, we set the delisting return to -100% and assume the investor receives zero. Share-based mergers are treated separately from cash-based delistings.[5] For share-based delistings, we infer the exchange ratio from the acquirer's price on or just before the merger event date. These data allow us to distinguish cash acquisitions, which create a *forced* taxable event, from stock-for-stock transactions, which carry the tax basis into the new security.

IV. Implementing Tax Loss Harvesting

This section describes how we implement the tax-loss harvesting strategy. We first define the two benchmark portfolios: a strategy that holds the top N eligible firms by market capitalization and a strategy that replicates real-world ETFs' end-of-quarter portfolio weights. We then describe the tax-lot accounting procedure. Next, we explain how we compute post-tax returns for these strategies.

of the data, these observations can generate unusually large taxable gains, especially when a stock has had prior splits and later delists.

[5]Specifically, we identify share-based-mergers as cases with CRSP delisting code 231, an identified acquirer in CRSP matched on acquiring permno or acquiring permco, and a valid acquirer price within 10 trading days before the event. For simplicity, we treat all other forms of delisting, including combination cash and share delistings as cash delistings. We do this because for these other forms of delisting, it is difficult to compute the expected exchange ratio from CRSP data alone.

Finally, we describe the parameters we vary across our counterfactual exercises.

A. Tax-loss harvesting strategies

We run the tax-loss harvesting strategy on two sets of target portfolios. The first target portfolio is esentially a value-weighted index of the largest N eligible stocks at each rebalancing date. The second target portfolio uses the actual end-of-quarter portfolio weights of real-world ETFs. In both cases, our tax-loss harvesting strategy is aiming to track the target portfolio in a tax-efficient way i.e., it does not necessarily hold the target portfolio exactly at every point in time.

The largest-N-stock benchmark is meant to be an implementable analog to a large-cap index like the S&P 500 or Russell 1000. The baseline uses the largest 500 eligible stocks, while we also consider a version which targets the 1000 largest stocks. Within each period, the target weights are proportional to each stock's market capitalization.[6] At the end of the quarter, we calculate the market capitalization weights and use them as the target portfolio for the first day of the next quarter. This avoids using information that was not observable when the portfolio was formed. Because this benchmark can be constructed from public data, we assume the baseline version of our strategy does not need to pay an index licensing fee.

The ETF-replicating strategy begins in 2011. While ETFs existed prior to this, we are limited by the availability of transaction cost data, which is necessary to provide a realistic comparison between direct indexing and holding the ETF. For each fund, the replication starts in the first year after 2010 that contains four matched quarters of returns and S12 holdings.

For each strategy, we track every share that would need to be bought and sold. For each purchase, we record the number of shares, the purchase date, and the purchase price. If the investor buys the same stock in multiple quarters, each purchase creates a separate tax lot. Tax-loss harvesting applies to the specific lot with a harvestable loss i.e., we do not require harvesting of the investor's full position in that stock.

At the end of each quarter, the strategy rebalances toward the target portfolio weights. To do this, we compare the current portfolio weights with target portfolio weights and use available cash to buy underweight securities. The word *available* is important, as we do not have the strategy sell overweight securities to eliminate tracking error, as doing so would often create the type of taxable

[6] Unlike popular indexes, such as the S&P 500 or Russell 1000, we do not apply float adjustments to each firm's shares outstanding. While we could add that feature to our tax loss harvesting strategy, doing so would require relying on data from a specific provider, as methodologies for calculating float vary across firms [Sammon and Shim, 2026a]. Furthermore, for this benchmark, we aim to maintain an index without paying licensing fees to an index provider like Russell, MSCI, CRSP or S&P, so we assume their float adjustment data is not freely available.

events we are trying to avoid. If a security was sold at the end of the previous period, we assume it is affected by the wash-sale rule and therefore cannot be bought in the next period. We therefore compute the weight gaps only among securities that are eligible to be bought (i.e., stocks to which the wash-sale rule doesn't apply), renormalize across those eligible securities, and allocate available cash toward the resulting underweight positions.

Residual cash can remain after rebalancing. This can occur because of fractional-share constraints or minimum dollar position constraints. We assume residual cash is invested in the S&P 500 through the State Street SPDR S&P 500 ETF Trust (Ticker: SPY). If there is any remaining residual cash, we assume it is in a money market account that earns the risk-free rate.

Cash enters the account through the initial investment, recurring contributions, dividends, tax-loss harvesting sales, tax refunds, and forced cash delistings. In our simulations, we assume that tax refunds are cash windfalls created by using realized losses to offset either ordinary income (up to $3,000 per yer) or capital gains realized outside the portfolio (e.g., selling a house or a business). Cash leaves the account through stock purchases, taxes on dividends, annual tax payments, as well as fees and transaction costs (when included).

For each target portfolio, we also construct an ETF counterfactual. The ETF counterfactual differs from direct indexing in that the ETF investor owns fund shares and receives the ETF return. When we are replicating real-world ETFs, we compare our tax loss harvesting strategy to the post-fee fund returns. Importantly, unlike the direct index investor, the ETF investor rarely has taxable distributions other than dividends. ETF creation and redemption mechanisms can be used to remove low-basis, high-embedded-gain positions from the fund without a taxable sale [Poterba and Shoven, 2002].

B. Measuring embedded gains and losses

At each portfolio rebalancing date, we compare each lot's market value with its tax basis. A lot is harvested if it satisfies the minimum holding-period rule (which in the baseline strategy is one quarter), and has a loss of at least 2%. When a lot satisfies the harvesting rule, we sell that lot and reinvest the proceeds according to the rebalancing rule described in the previous subsection. We separately account for long-term and short-term realized gains and losses, as their tax implications are different. In the quarterly simulations, gains and losses are short-term if the holding period is four quarters or less and long-term otherwise.

The wash-sale rule prevents the immediate repurchase of a harvested stock. As robustness, we

also run a version of our tax loss harvesting strategy without the wash-sale restriction. This matters because wash-sale restrictions can change the portfolio's factor exposure. For example, preventing immediate repurchase of recent losers (which get harvested) can tilt the portfolio away from losers and toward recent winners, leading the portfolio to positively load on a momentum factor. Turning off the wash-sale rule helps separate the value of harvesting losses from the portfolio tilt effect of selling recent losers and holding on to recent winners.

We assume that taxes affect the strategy when the taxable investor would plausibly face the cash flow associated with that tax event. Dividends are taxed when they are paid.[7] The after-tax dividend cash is then available for reinvestment. Realized gains and losses from selling stocks and cash delistings accumulate through the calendar year. At year-end, short term losses are first offset against short term gains, then any remaining short term losses are combined with long term losses and offset against long term gains. Finally, remaining losses are used to offset income up to the $3,000 limit. Any unused losses carry forward indefinitely. We assume the investor receives any tax refund (e.g., from offsetting their income) after tax day. The tax benefit is therefore paid into the portfolio at the end of the second quarter of the following year.

If the year-end tax bill is positive i.e., the investor *owes* taxes, the same timing convention requires a cash tax payment. A large forced taxable event, such as a cash acquisition at a substantial gain, can sometimes create a tax bill that exceeds the available cash in the account. While rare, as these gains are typically offset by harvested losses and dividends, a significant buyout could result in a tax liability that exceeds *that quarter's* cash proceeds, especially if the proceeds from the taxable event have already been reinvested by the time the tax bill is due. In that case, we assume the investor contributes outside cash into the account to pay the tax bill, rather than forcing them to sell down other positions with embedded gains to meet the obligation. We do, however, count this cash infusion as a drag on the portfolio's return.

In the baseline simulation, we assume ordinary income and short-term gains are taxed at 35%. Long-term gains and dividends are taxed at 20%. In the exercise where we assume there is an outside-the-portfolio gain, any unused losses are converted at the long-term capital gains rate and the resulting tax benefit is invested back into the portfolio after tax day. In the exercise where we assume the investor has a cashflow shock and needs to partially liquidate the portfolio, the tax bill applies short-term and long-term rates to the realized gains produced by the selected lots.

[7] For simplicity and to prevent cash drag, we assume all dividends are paid on the last day of the quarter. Essentially, this means we assume that before quarter end, the dividends are effectivity re-invested in the stock that paid them. We also assume that all dividends are qualified cash dividends and therefore are taxed at 20%.

The main tax object of interest is the cumulative realized tax position after within-portfolio offsets and any ordinary-income offset are deducted. A negative value implies the investor has unused losses. A positive value means forced realizations exceed harvested losses and the investor must put cash into the portfolio to pay the tax bill. This object is central to the ETF versus direct-indexing comparison. ETF ownership avoids many constituent-level taxable realizations, but it does not allow the investor harvest constituent-level losses. Further, the ETF does not allow for an efficient sale of positions with small embedded gains, instead effectively selling all positions pro-rata.

C. Measuring returns, tracking error, and fees

The return calculation is analogous to an NAV return calculation for a mutual fund. We start with beginning of period account value. We add the return on securities actually held, the return on cash, and any tax refund paid into the account. We subtract dividend taxes, fees, transaction costs, and tax payments (when they apply).

The benchmark return depends on the exercise. For the largest-stock benchmark, the benchmark is the value-weighted return on the top 500 or top 1000 eligible stocks. For ETF-holdings simulations, the benchmark is the holdings-implied ETF return computed from the ETF's own S12 holdings. Tracking error is the difference between the direct-index return and the relevant benchmark return. In the exercises where we assume a part of the portfolio needs to be liquidated, the post-liquidation direct indexing portfolio intentionally deviates from benchmark weights. In other words, when the external event arrives, we meet the required cash outflow in a tax efficient way by selling lots with the largest embedded losses or smallest gains. As a result, the portfolio will deviate from its target allocation until new cash infusions or rebalancing from additional loss-harvesting occurs. However, we continue to measure tracking error relative to the original benchmark.

D. Counterfactual portfolio experiments

The counterfactual exercises isolate the economic tradeoff between ETF ownership and direct indexing. The largest-N-stock exercises begin in 1927. The ETF-holdings comparisons begin in 2011 when the trading-cost data are available.

The first counterfactual compares a direct-index investor with an ETF investor holding the same benchmark. The direct indexing investor owns the constituent stocks directly. The ETF investor owns fund shares. The ETF structure largely prevents the ETF investor from having any taxable events when the fund "trades" the underlying securities or when a constituent is acquired. This is

important because even in the case of cash-delistings, ETFs fund can use in-kind redemptions to remove low-basis positions before they are bought out. The direct-index investor can harvest losses on individual securities, but must realize gains when a constituent is acquired for cash.

The second counterfactual varies the timing and size of investment. We simulate one-time investments and recurring contributions. Recurring contributions benefit the direct indexing investor as they continually create new high-basis lots. These new lots are more likely to generate harvestable losses after modest price declines, and therefore can prevent ossification.[8]

The third counterfactual varies how losses can be used. In the baseline case, losses can only offset future gains generated inside the portfolio and ordinary income. We also run a version of our tax loss harvesting strategy which allows losses in a chosen year to offset an outside-the-portfolio capital gain. This exercise is designed to capture the benefit of direct indexing to an investor who realizes a large taxable gain outside the portfolio, for example from selling a private business or a house. Remaining losses after the ordinary-income offset are converted into a tax benefit at the long-term capital gains rate and invested back into the portfolio.

The fourth counterfactual studies partial liquidations. This is meant to capture the possibility that an investor has an outside the portfolio event (e.g., the purchase of a house) that requires raising cash. A direct-index investor can choose which security-level tax lots to sell. We compare this tax-aware liquidation with a pro-rata liquidation of the investor's existing lots, which is similar to what they would experience if they had to sell (appreciated) shares of an ETF. Our tax-aware strategy first sells lots with short-term losses, long-term losses second, long-term gains third, and short-term gains last. Within each category, lots with lower embedded gains are sold first. In this exercise, we sell enough lots to raise the specified fraction of portfolio value and then compute the tax bill. The difference between this tax bill and the pro-rata tax bill is the tax value of selective liquidation.

The fifth counterfactual measures the tracking-error cost of selective liquidation. A pro-rata liquidation tends to preserve benchmark weights more closely, but it can create a larger tax bill. A tax efficient liquidation lowers the tax bill by selling high-basis lots, but it leaves the remaining portfolio farther from benchmark weights. After a liquidation shock, we continue the simulation and compare future cumulative returns, tracking error, and factor exposure to the benchmark portfolio.

[8]Ossification refers to the phenomenon where a direct indexing portfolio has only positions with large, unrealized capital gains. Because direct indexing relies on selling losers to harvest the tax benefits, long-run increases in the value of the market can reduce the pool of stocks that could reasonably generate losses over time, rendering tax loss harvesting less effective.

This exercise separates the immediate tax benefit of selling high-basis lots from the subsequent cost of holding a portfolio that deviates from the benchmark.

We also run a battery of robustness checks. We repeat the largest-stock benchmark with 500 and 1000 stocks. We compare quarterly and monthly harvesting. The monthly exercise uses monthly data and monthly rebalancing, with the same tax-lot logic as the quarterly exercise. We also compare lot-level tax accounting to a dollar-cost-average basis rule. Under the average-basis rule, we compute the value-weighted embedded return across all lots of the same stock and sell only when the average position is sufficiently below basis. This rule collapses individual-lot accounting into one average basis per stock, as in much of the prior literature.

E. Transaction Costs

Transaction costs have fallen significantly in recent years, following the long-term downward trend documented by Angel et al. [2011]. Recently, this decline has been driven by the elimination of fixed commissions (which significantly reduced broker revenue, as noted by Adams et al. [2024]) alongside the removal of fractional share constraints.

We use bid-ask-spread-based transaction costs only in the ETF-holdings comparisons beginning in 2011. For a trade in month i, we use the average effective spread for that stock in month $i-1$. We measure effective spreads using the WRDS Intraday Indicators suite, which is constructed from millisecond TAQ data. The underlying measure is a stock-day effective spread that is volume-weighted across trades within the stock-day. We average this measure to the stock-month level and use the *lagged* value to ensure our strategy does not pay the transaction costs associated with real-world index rebalancing dates.[9] The spread assumption is conservative because we're assuming that you have to cross the whole spread every time you trade, which is not necessarily true in practice.

We implement round trip trading costs in our simulations as follows: First, when we calculate the target dollars to allocate to each stock, we reduce those dollars by the lagged effective spread before converting dollars into shares. When a position is sold, the proceeds are reduced by the lagged effective spread before the cash is available to be reinvested. We do not apply a fixed commission to each trade. This is designed to reflect the current zero-commission trading environment.

[9]By nature, the rebalancing of our tax-loss harvesting portfolios will be correlated with the rebalancing of major value-weighted indexes. Importantly, however, most indices, such as the S&P 1500 family, rebalance on specific dates like the third Friday of the last month of the quarter rather than at quarter-end (which is when our strategy rebalances). By trading at quarter-end, we avoid elevated trading costs associated with actual index rebalancing periods. Therefore, these monthly averages are meant to capture the generic costs of trading these kinds of stocks, rather than the costs associated with trading exactly on index rebalancing days.

The largest-N-stock exercises do not account for trading costs. Comparable stock-level spread data are not available going back to 1927. These exercises therefore focus on the tax mechanics of direct indexing and ETF ownership, not on historical trading implementation costs. The strategy designed to replicate real-world ETFs' holdings include the round-trip trading costs described above.

We do not estimate price impact or the timing cost of predictable benchmark trading. Sammon and Shim [2026b] estimate that direct transaction costs for large value-weighted index funds are 5-10 basis points per year, which are higher than our round trip trading cost estimates from bid-ask spreads, but not by much. Sammon and Shim [2026b] also estimate a performance drag of about 50 basis points from mechanical index rebalancing trades in response to issuance, buybacks and IPOs. Specifically, a value-weighted index must rebalance when firms issue or repurchase shares, or when new firms IPO. This implies that a value-weighted index fund fund tends to buy firms that issue equity and sell firms that repurchase equity, which historically has predicted low returns [Pontiff and Woodgate, 2008]. If an ETF and a direct indexing strategy both follow a value-weighted benchmark, both will face the same mechanical need to trade in response to issuance and buybacks.

Table I: Effective spread summary statistics. We report the distribution of average monthly volume-weighted effective spreads (basis points) by category, from SEC Rule 605 reports for the eight largest wholesalers. Percentiles are calculated as dollar-weighted percentiles: that is, the volume traded in all symbol-months below the median observation is 50% of total volume. Panel A reports the averages for stock-months based on index membership; Panel B reports averages for stock-months based on stock terciles of effective spread.

Panel A: Market Segments

Category	**Mean**	**SD**	**5%**	**25%**	**Median**	**75%**	**95%**
All Stocks	2.20	1.47	1.05	1.30	1.69	2.76	4.41
SP 500 Stocks	1.02	0.74	0.39	0.63	0.80	1.29	2.03
Russell 2000 Stocks	7.28	4.77	2.20	4.50	6.09	9.60	12.89
All ETFs	1.02	0.68	0.34	0.74	0.95	1.08	1.81

Panel B: Terciles

Category	**Mean**	**SD**	**5%**	**25%**	**Median**	**75%**	**95%**
Tercile 1	1.58	0.98	0.69	1.05	1.23	2.05	2.81
Tercile 2	8.99	5.07	4.17	5.84	7.41	11.44	15.53
Tercile 3	21.05	12.05	8.47	11.69	19.85	28.17	38.68

V. Generic Value Weighted Index

A. Tax Position

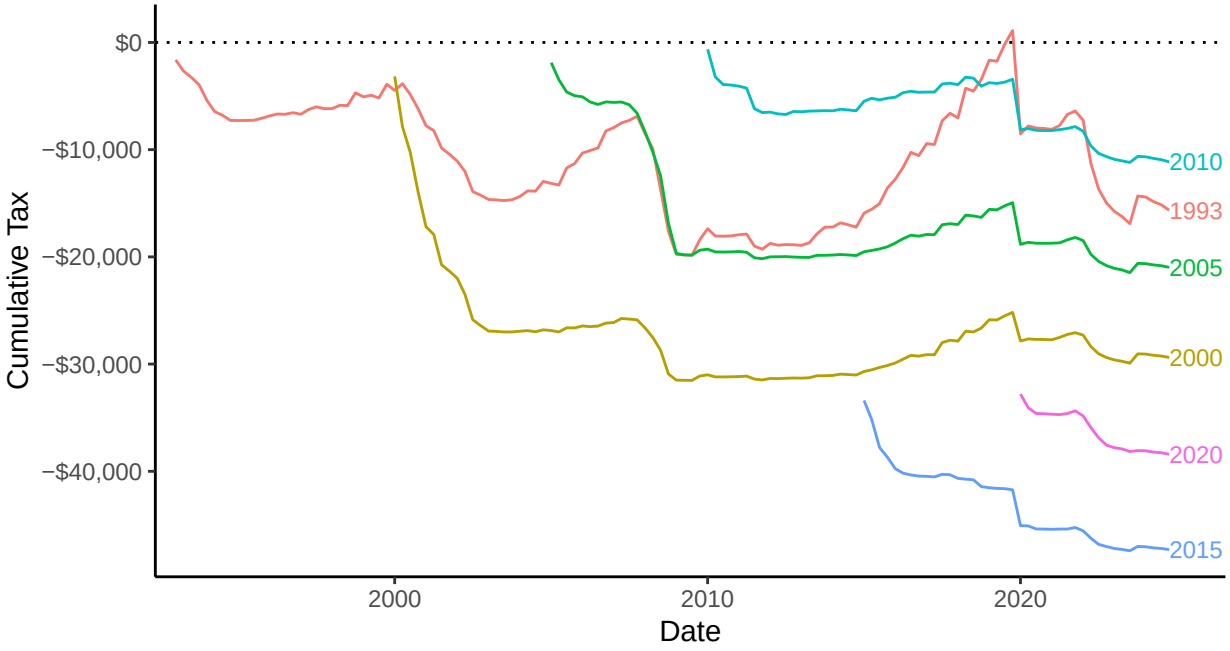
Figure 1 tracks the cumulative harvested losses (or forced gains) on a one-time investment in a value-weighted index of the top 500 stocks. Panel A tracks a portfolio which begins with a single investment of $50,000, with each line plotting the tax position for an investment with a specific start year. The 1993 start year (plotted in red) has a small net tax gain around 2020, though returns to a net loss position in the following years ending with a net loss of $15,000 in 2024. For Panel B, we take the specific start years, and calculate an average across portfolios from each decade. Portfolios formed in the 1920s, 30s, and 40s have substantial net tax losses beyond 40 years after initial investment. Portfolios formed in the 1970s and 1980s have net gains between 20 and 30 years. Portfolios formed in the 50s and 60s have net gains around 30 years, while the average for the 1990s is slightly below zero after 30 years. Figure A2 in Appendix B presents a longer tax history for additional years.

Taxable losses are only valuable to the extent that they can be used. For an investor with no other use than offsetting forced portfolio gains from merger and acquisition activity, Figure 1 Panel B represents how long it takes for an ETF to be superior to a direct-indexing strategy; once the net tax position is positive, the ETF ability to avoid capital gain distributions altogether is superior to the ability of the direct index strategy to use losses to offset gains. This eventually occurs for every initial investment placed between 1927 and 1995 in our sample, taking an average of 26 years, though there is considerable variation in the time series, which is plotted in Figure A3 in Appendix B.

For a buy-and-hold investor, the ability of a direct-index strategy to outperform an ETF entirely depends on the extent to which an investor can utilize the capital losses generated. Figure 2 plots two example scenarios. Panel A presents a worst-case scenario, where the only use of losses it to carry forward to offset forced gains, with direct indexing under-performing an ETF by up to 5% in cumulative return, while Panel B presents the case under the assumption that all losses can be immediately and fully deducted against outside-the-portfolio gains, with direct indexing over-performing an ETF in some decades by as much as 10%. There are three particularly beneficial uses of losses from direct indexing. Personal income (Section V.B) offers one possible use for losses, while capital gains from outside the portfolio (Section V.C) offers another use. For an investor who will sell a large portion of their portfolio (Section V.D), the direct index portfolio also allows selectively selling stocks with lower tax burdens.

Figure 1: Harvested Loss By Vintage. We track net realized tax losses or gains for a one-time investment. Panel A presents the losses and gains for investments in select years between 1993 and 2023. Panel B measures the mean for each decade.

(a) Tax Loss Harvest - Select Years



(b) Tax Loss - Decade Averages

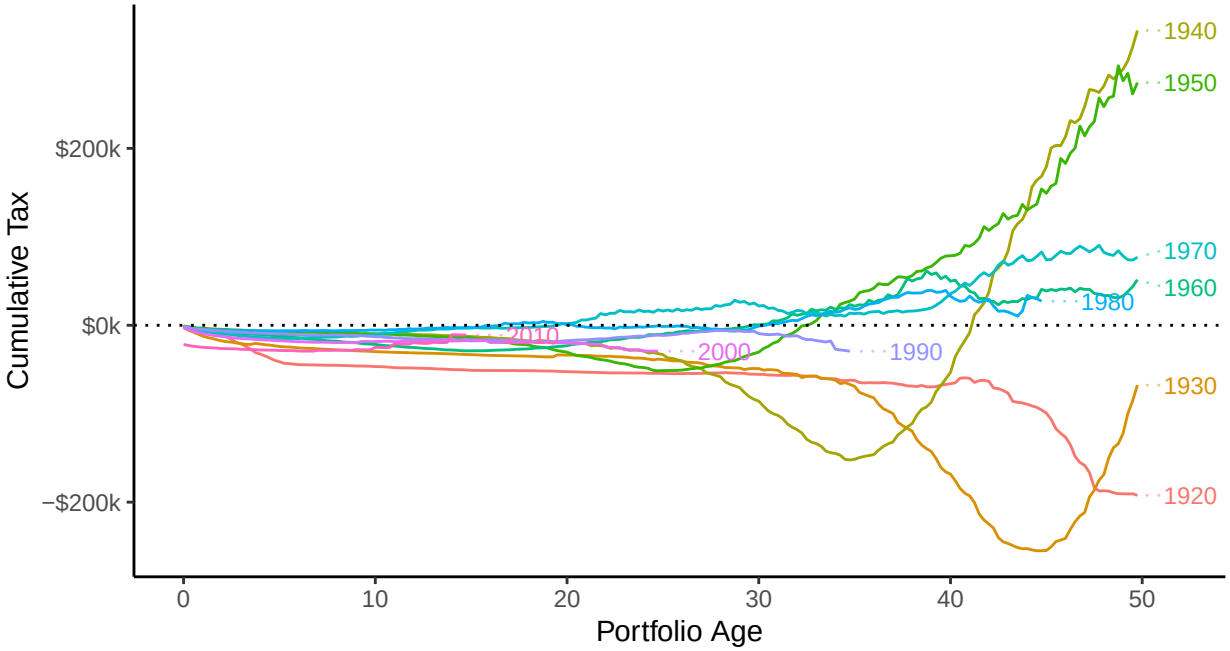
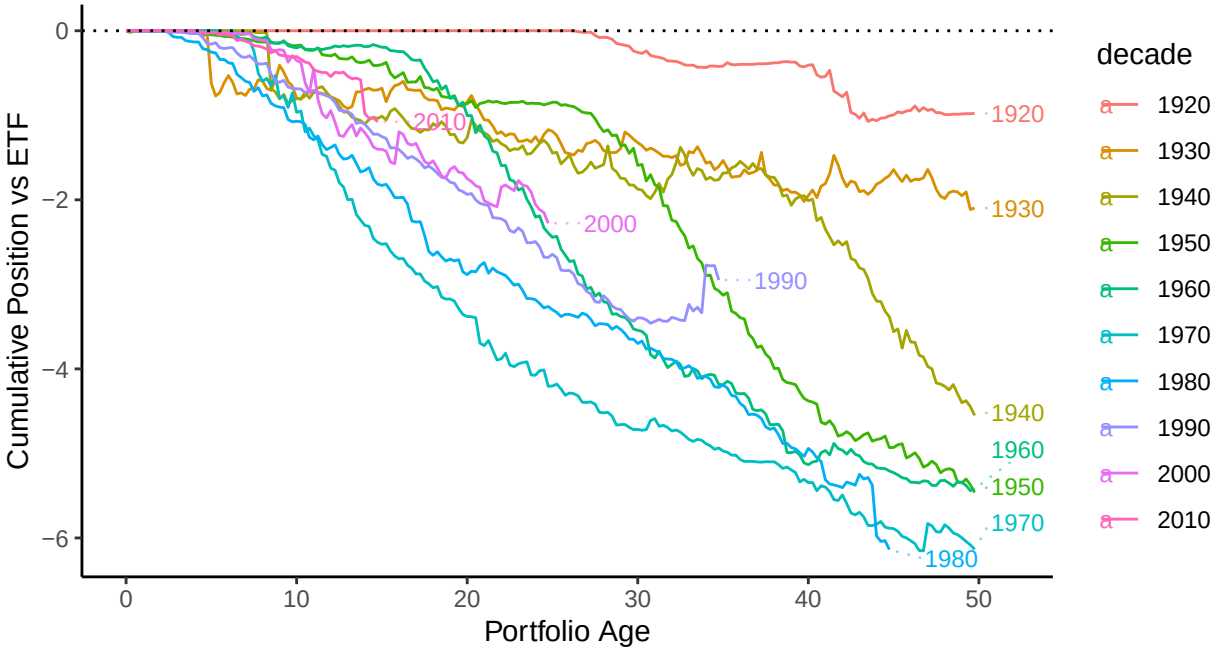
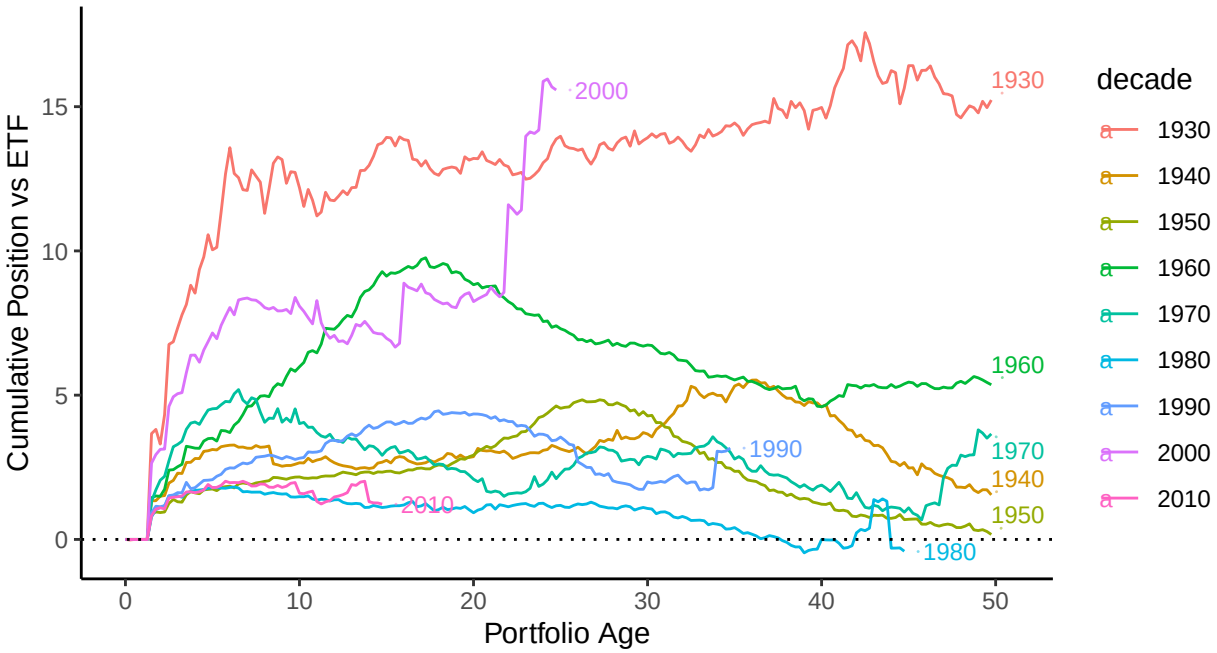


Figure 2: Loss Upper and Lower Bounds. Under the assumption that the only use of losses is to offset future gains, direct indexing will eventually underperform relative to an ETF (Panel A). Under the assumption that all losses can be immediately and fully used to offset a long-term gain, direct indexing will typically outperform an ETF (Panel B).

(a) Tax Impact on Returns: No Use of Losses



(b) Tax Impact on Returns: Full and Immediate Use of Losses



B. Person Income Deductions

Capital losses can be used to deduct up to $3,000 per year from personal income tax. This benefit offers the potential for a direct index strategy to outperform an ETF, particularly as personal income taxes are usually higher than capital gains taxes. [10]

Table II presents summary statistics by portfolio age of the average tax position for a value-weighted portfolio. After 15 years, the median tax position for a $50,000 investment is $1,160 per year. For an investor to have an average tax loss of $3,000 per year for the first 15 years, they would therefore have to invest $129,333 in a one-time lump-sum investment.[11] This one-time lump-sum investment would perform 5 to 10% better than an alternative investment in an ETF, thanks to the income offset from personal income; exact details are provided in Appendix B.

Larger investments eventually leave the direct index investor with excess losses which cannot be used to deduct against income, and would only be of use to the investor in offsetting a capital gain. These unused capital losses can be carried over to future years. For every year prior to 1996 in our sample, forced gains on the portfolio from mergers and acquisitions eventually consume more than the value of carried forward losses. For an investor with no use for capital losses, the ETF will eventually dominate the direct indexing strategy.

Figure 3 plots the maximum value portfolio for which the income offset (assuming it is invested at the portfolio rate of return) will always exceed the value of any future capital gains taxes.[12] For the 1950 vintage, the cutoff-value is $700,418. Investing any amount up to this value would mean that the value of personal income tax rebates (assuming an personal income tax rate of 35%) would, if invested at the portfolio rate of return, exceed the value of any future capital gains taxes on the portfolio. Investing any amount more than this means an investor would eventually mean the capital gains on the portfolio will at some point exceed the value of any invested income tax returns.

The cap of $3,000 limits how much ordinary income can be offset by capital losses. Most vintage

[10] For 2026, a US federal capital gains rate of 15% applies to investors with a personal income between $48,350 and $533,400, and 20% above that. For comparison, federal personal income tax is 22% for income above $48,476, 35% for income above $250,526, and 37% for income above $626,351.

[11] For a loss of $2,274 on a $50,000 investment, the calculation $50,000 \cdot \frac{3000}{1159} = 129,333$ produces the investment needed. For the investor to have a $3,000 loss per year at year 15 for at least 75% of portfolio-year observations, they would have to invest a one-time lump sum of $230,627. At the 95% of portfolio-year observations, the net tax position is positive rather than a loss. If the investor were to make a quarterly recurring investment, rather than a fixed one-time investment, to have a $3,000 loss per year at year 15 would require a quarterly investment of either $2,833 per quarter for the median historical tax position, and an investment of $4,095, or $10,570 per quarter at the 75%, or 95% of portfolio-year observations, respectively.

[12] To calculate the value of personal income deductions, we assume the investor faces a marginal personal income tax of 30% and a capital gains tax rate of 15%.

years have a maximum possible investment of roughly $1 million.[13] Below this value, direct indexing benefits the investor, while above, an investor would be better off in an ETF if there are no other uses for the tax losses generated by direct indexing.

Table II: Net Tax Per Year By Portfolio Age. For each possible starting year, we calculate the average tax position per year, and present summary statistics based on portfolio age. Panel A presents the annualized losses for a one-time investment of $50,000 while Panel B presents the annualized losses for a recurring quarterly investment of $5,000.

Panel A: Tax Position with $50,000 One-time Investment

Portfolio Age	Mean	SD	5%	25%	50%	75%	95%
5	-3,341.5	2,744.7	-8,552.5	-4,316.1	-2,532.6	-1,371.1	-835.9
10	-1,739.4	1,349.1	-4,437.3	-2,547.0	-1,514.2	-809.0	273.0
15	-1,168.6	964.4	-2,816.1	-1,820.1	-1,159.8	-650.4	356.8
20	-948.9	877.8	-2,337.5	-1,474.1	-1,059.6	-623.4	408.8
30	-796.0	1,359.4	-3,448.0	-1,754.6	-493.2	62.0	1,167.4
50	2,272.7	4,733.0	-4,612.3	249.1	1,758.9	4,919.0	10,626.9

Panel B: Tax Position with recurring $5,000 Quarterly Investment

Portfolio Age	Mean	SD	5%	25%	50%	75%	95%
5	-5,735.5	5,679.9	-20,328.2	-6,350.8	-3,405.2	-2,531.2	-1,609.9
10	-5,467.6	3,573.3	-13,756.5	-6,693.9	-4,455.3	-2,998.7	-1,475.6
15	-5,293.7	2,828.1	-10,284.4	-7,010.9	-4,692.1	-3,663.5	-1,419.1
20	-5,364.8	2,779.4	-9,976.7	-7,036.2	-5,292.5	-4,242.6	-616.8
30	-5,941.7	5,119.7	-15,027.8	-8,609.7	-5,954.6	-2,999.6	1,464.5
50	1,194.6	21,509.2	-44,732.5	-3,259.7	3,232.4	13,589.0	34,071.4

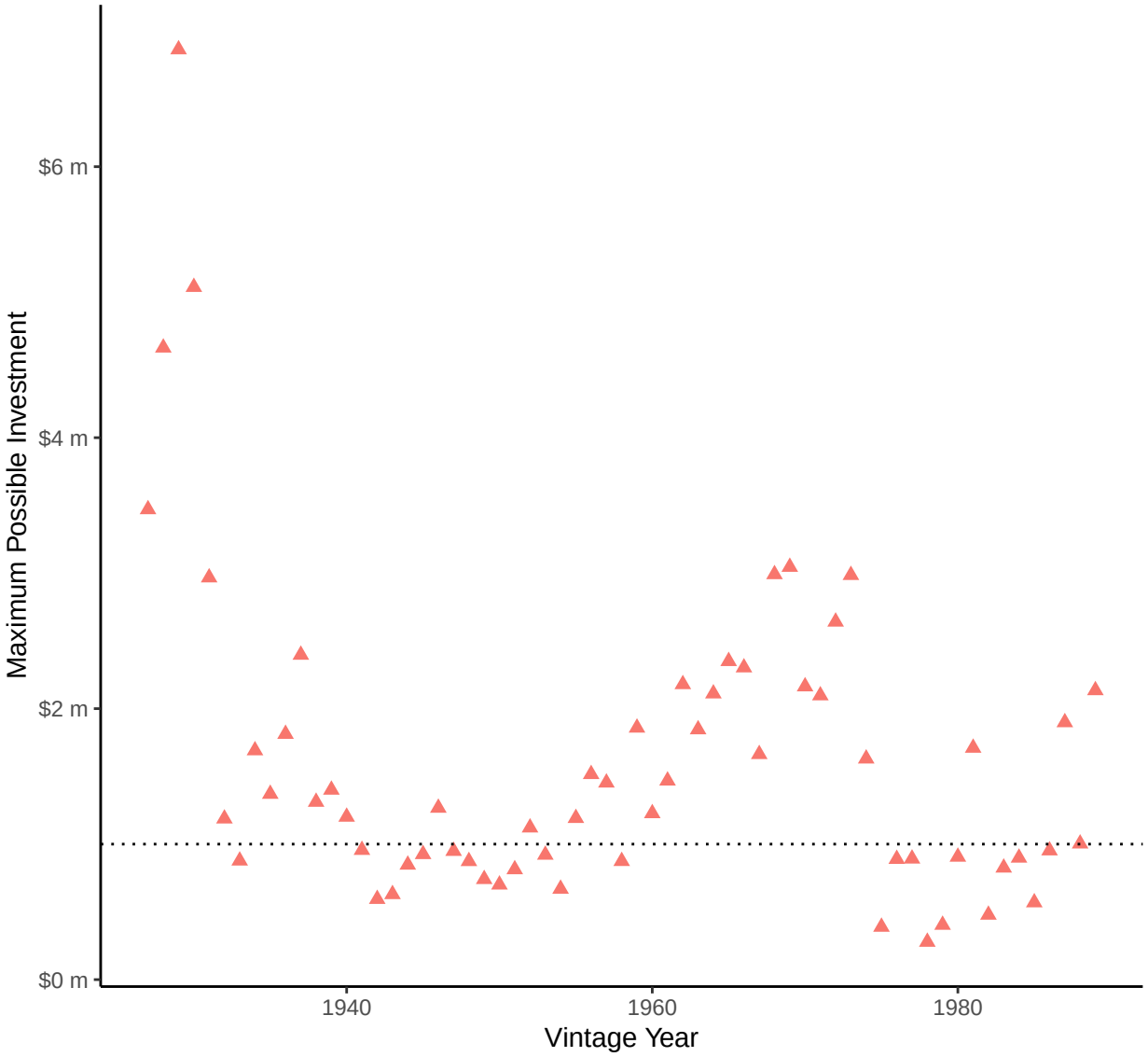
C. Capital Income Deductions

It can take decades before forced gains are substantial. As Figure 1 Panel A shows, the average portfolio formed in the 1980s takes 26 years before the net tax position is positive, while the average portfolio formed in the 1930s takes 50 years.[14] Unused losses carry forward, but are carried forward in nominal terms with no appreciation. If the losses are only used to offset future forced gains, an ETF (which avoids forced gains altogether) will always dominate direct indexing for this reason. If a direct index investor has capital income from outside the portfolio, however, this allows an investor to make immediate use of the losses; the reduction in taxes on taxable income can be invested, and this investment may grow to exceed the value of taxes on any future forced gains.

[13] Table A2 in Appendix B provides exact numbers.

[14] Table A1 and Figure A3 in Appendix B provide exact details.

Figure 3: Income Offset Value. For each vintage year, we calculate the maximum portfolio size (red triangle) such that the income offset (capped at $3,000 per year) exceeds the value of any future taxes on the portfolio, under the assumption that the income offset (with a personal income tax rate of 35%) is invested at the portfolio rate of return. The dotted black line highlights a reference value of $1 million.



For a portfolio formed in each vintage year X, we consider in year Y that the total harvested losses could be used to offset an outside-the-portfolio capital gain. We then consider whether the value of this tax offset, were it to be re-invested in the portfolio, would grow at a rate high enough to cover any future capital gains taxes. Figure 4 plots the results for each portfolio vintage. Areas in blue represent areas where the losses would cover any future gains.

As an example, for an investment in 1927, the substantial losses mean that as little as 2% of the harvested losses would need to be used to offset a gain such that the tax refund, if invested in the direct index portfolio, would exceed any future taxes.[15] For a one-time investment in the 30s, 40s, 60s, or 90s, there is typically a two-decade period where the losses, if they could be used to offset income from outside the portfolio and the subsequent saved taxes invested in the portfolio, would offset any future gains. Under these conditions, a direct index strategy would be superior to an ETF. During the 50s and 70s to early 80s, there are far fewer years during which the losses would be sufficient relative to future gains.

A one-time investment is more sensitive to a specific year in which losses could be used. For a recurring investment, plotted in Figure 4 Panel B, there are typically many more years during which the losses the portfolio generates would be sufficient, if utilized, to offset any future gains. For the most recent years, e.g. from the late 90s onward, there are not enough years to truly evaluate direct indexing, as the available losses still exceed forced gains.

D. Portfolio Sales

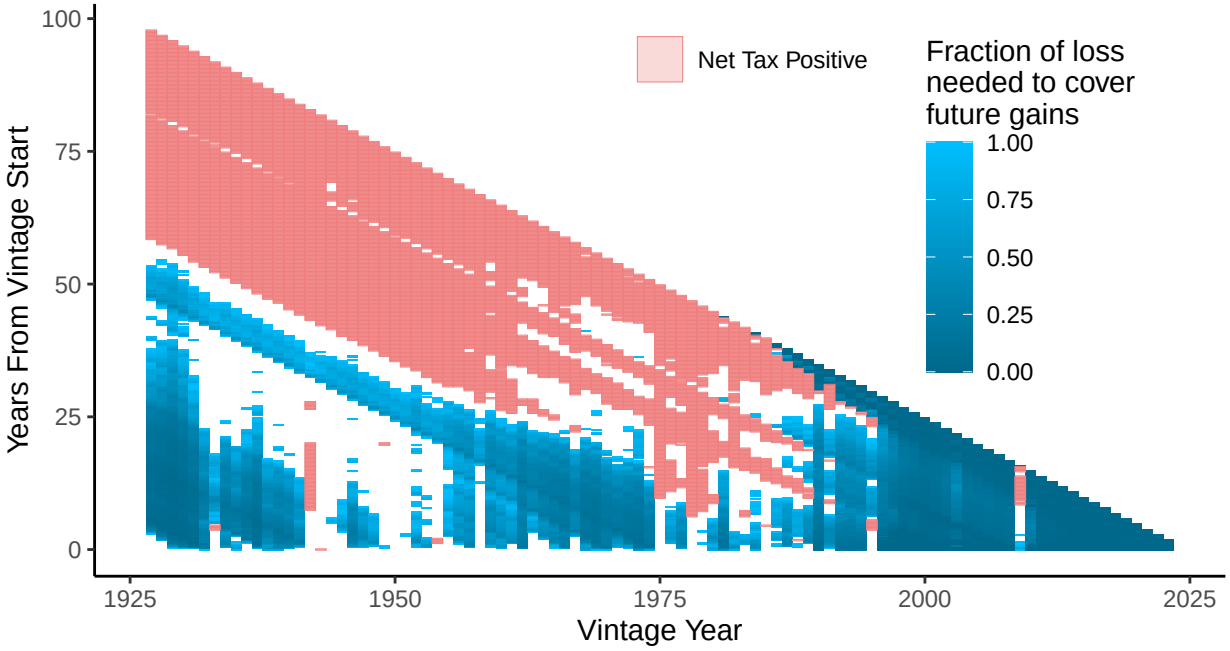
For an investor seeking to sell a large part of their portfolio, a direct index investor can select the individual lots of stock to sell which minimizes capital gains, while an ETF investor can only sell individual lots of ETF shares. This creates another powerful advantage for a direct index investor, as the investor can select only lots with the highest tax basis when selling.

We consider a investor with a preference for liquidating a portion of their portfolio, and compare the tax burden under direct indexing against the tax burden of an ETF investor. Table III presents summary statistics on the savings, measured as basis points of the portfolio assets. For a portfolio which is 20 to 30 years old, an investor liquidating 30% of their portfolio will on average save 191 basis points of (post-tax) portfolio AUM by liquidating highest-basis lots compared to pro-rata

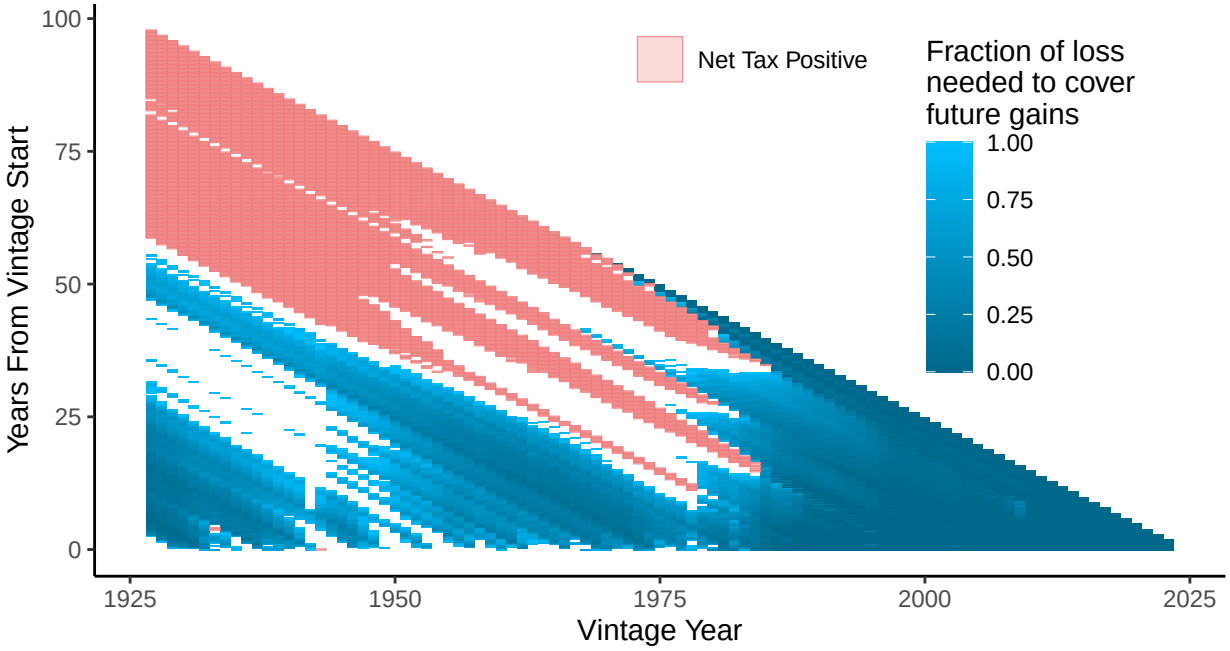
[15] A $50,000 investment in 1927 would be worth $17,700 in 1932, with harvested losses of $43,900. Using just 10% of these losses would reduce tax by $878 (assuming a 20% capital gains tax), and would always be sufficient to cover any net gains in the future. In 1989, the cumulative tax position peaks as a percentage of portfolio value at 6.8% of the portfolio value ($717,000), while in 2019 the tax position peaks in dollar terms at 7.5 million (4.26%).

Figure 4: Harvested Loss Coverage Ratio. For each vintage, we calculate whether the harvested losses, if used to offset gains from outside the portfolio in year Y, would generate a sufficiently large tax credit such that this tax credit, if invested at the portfolio rate of return, would be sufficient to cover any future capital gains. Light blue inidcates that 100% of the harvested losses would need to be invested to cover future capital gains; dark blue that close to 0% of the harvested losses would be needed. White indicates periods where harvested losses would not cover future gains. Red indicates areas with positive net realized gains.

(a) One Time Investment



(b) Recurring Investment



selling. The tax bill on the highest-basis lots will be 30% lower than the bill for pro-rata selling (Table A4 in Appendix C).

Table III: Tax Under Liquidation. We consider a liquidation of 10% (Panel A) or 30% (Panel B) of the portfolio. Under liquidation, we calculate the numerical difference in tax between the tax bill from selling the highest basis lots and the tax bill from selling all lots pro-rata. We measure this tax difference as basis points of total portfolio value.

Panel A: 10% Liquidity Shock

Portfolio Age	Mean	SD	5%	25%	50%	75%	95%
0–5	52	25	2	40	54	65	86
5–10	73	15	51	65	73	82	95
10–15	82	11	64	75	84	90	99
15–20	89	10	73	83	90	97	103
20–30	96	10	80	91	98	103	111
30+	102	9	84	96	103	107	115

Panel B: 30% Liquidity Shock

Portfolio Age	Mean	SD	5%	25%	50%	75%	95%
0–5	131	73	8	93	131	164	223
5–10	166	39	108	145	167	186	215
10–15	178	28	131	160	181	196	221
15–20	183	26	135	169	185	198	226
20–30	191	31	132	174	194	211	237
30+	193	35	127	172	195	217	243

While selectively selling lots lowers the overall tax bill for a partial liquidation of the portfolio, it alters the future returns of the portfolio. Table III plots the returns and tracking error over time. Across each vintage start year, we consider a liquidation of 30% of the portfolio in each possible subsequent year. Compared to a portfolio with no liquidation event, the median portfolio is 2 percentage points larger after 5 years, and 3 percentage points larger after 30 years. Tracking error by quarter is modest, with the median standard deviation of quarterly tracking error at .70 percentage points after five years, and falling over longer horizons. Following the shock, the loading on the market portfolio is lower .04 lower for the median portfolio.

For a 50% liquidation (Table A5 in Appendix C), mean and median performance turns negative, i.e. an investor would be better off selling an equal portion of their portfolio rather than selling the highest-basis lots. Changes in factor loadings following liquidation explain this surprising result. To avoid this negative performance, an investor could adopt an alternative weighting scheme for how dividend cash is used for subsequent re-balancing. As an example, a portfolio approach focused on

tracking the characteristics approach of Daniel et al. [1997] is implemented in Table A6).

Table IV: Performance and Tracking Error. We consider a liquidation of 30% of the portfolio, selling the highest-basis lots. Panel A measures the post-liquidation cumulative return difference between the post-liquidiation portfolio and the never-liquidated portfolio at specific years, in percentage points. Panel B measures tracking error via the standard deviation of quarterly return difference between the post-liquidation portfolio and the never-liquidiated portfolio over specific year ranges, in percentage points. Panel C measures the change in Fama-French factor loadings following liquidation.

Panel A: Cumulative Relative Performance

Years After Shock	**Mean**	**SD**	**5%**	**25%**	**50%**	**75%**	**95%**
5 yrs	1.75	3.10	-3.84	-0.16	2.12	3.57	5.26
10 yrs	-0.61	4.18	-9.47	-2.73	1.02	2.26	4.06
20 yrs	0.33	4.21	-6.50	-1.86	1.31	3.42	4.40
30 yrs	2.30	1.69	0.05	1.01	2.99	3.51	3.78

Panel B: Tracking Error (SD)

Years After Shock	**Mean**	**SD**	**5%**	**25%**	**50%**	**75%**	**95%**
5 yrs	0.70	0.25	0.39	0.51	0.68	0.80	1.17
10 yrs	0.64	0.23	0.36	0.40	0.65	0.77	0.97
20 yrs	0.51	0.23	0.27	0.31	0.50	0.67	0.84
30 yrs	0.29	0.06	0.23	0.25	0.28	0.30	0.37

Panel C: Factor Loadings

Years After Shock	**Mean**	**SD**	**5%**	**25%**	**50%**	**75%**	**95%**
Mkt-RF	-0.04	0.07	-0.13	-0.09	-0.02	0.00	0.05
SMB	0.06	0.21	-0.18	-0.02	0.02	0.08	0.38
HML	-0.02	0.14	-0.25	-0.03	0.02	0.04	0.13

E. Factor Loadings

Our base portfolio is a value-weighted index. Table V shows the estimated factor loadings of the excess returns on the index on the Fama-French 3 factors plus momentum. There is a large loading on the market factor (close to 1, reflecting the value weights of the index). Loading on SMB is large and negative, consistent with the value-weighted index. Finally, momentum has a positive loading, consistent with the portfolio selling assets which have gone down, but it is very small, as the assets are repurchased as soon as the wash sale rule allows.

Appendix E examines the factor sensitivity of tax losses harvested.

Table V: Factor Loadings. We regress portfolio returns on the Fama-French 3-Factor Portfolio plus momentum. We include a fixed effect for each vintage, starting investent, and quarterly investment. Standard errors are clustered by year-quarter and vintage.

Dependent Variable:	Excess Return		
Model:	(1)	(2)	(3)
Variables			
Constant	-1.96×10^{-5}		
	(0.0005)		
Market (MKT-RF)	0.9713***	0.9673***	0.9712***
	(0.0079)	(0.0077)	(0.0079)
SMB	-0.1354***	-0.1402***	-0.1346***
	(0.0108)	(0.0104)	(0.0107)
HML	0.0152	0.0117	0.0155
	(0.0134)	(0.0133)	(0.0133)
Momentum	0.0177**		0.0183**
	(0.0079)		(0.0079)
Fixed-effects			
Vintage		Yes	Yes
Initial investment		Yes	Yes
Quarterly investment		Yes	Yes
Fit statistics			
Observations	99,390	99,390	99,390
R^2	0.98869	0.98850	0.98875
Within R^2		0.98847	0.98873

Clustered (Year-quarter & Vintage) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

F. Long-term Secular Trends

There is considerable time-series variation in the tax performance of a direct indexing strategy. As Figure 1 Panel B highlights, some one-time investments will realize net gains in less than 25 years, while others have unused losses carried forward for over 50 years. Direct indexing cannot avoid realizing gains when mergers and acquisitions occur, and this is a significant driver of time-series variation in the tax position of a direct indexing strategy.

Merger and acquisition activity is cyclical. Figure 6 plots the value-weighted share of firms in the CRSP database which are bought each year. There is a substantial spike in M&A activity in the early 1980s, another spike in the late 1990s, and a broadly elevated level of activity throughout the 2000s. The large spike in M&A activity in the 80s creates a sizable tax burden for any direct index strategies which would have been avoided by an ETF. As Figure 4 highlights, some one-time investments in the late 70s have very few years, if any, in which the losses harvested under direct investing would, even if fully realized as a tax credit from outside-the-portfolio gains and invested, grow to cover future taxes. For these years, an ETF is unequivocally better for a buy-and-hold investor than a direct index strategy.

The late 1990s wave of M&A activity differs from the 1980s wave in composition: while the 1980s buyouts were primarily cash purchases, the 1990s M&A activity was primarily share-based purchases. Crucially, these stock-for-stock transactions do not force realization of capital gains.[16] The direct index strategy can avoid these gains just as the ETF can avoid these gains. At comparable ages, direct index investments started in the 1970s have a much worse tax position than direct index investments started in the 1990s, as the 1970s investments face the 1980s M&A wave of cash purchases, while the 1990s investments face the late 1990s M&A wave of share purchases.

A second, lasting change in the performance of direct indexing is the change in dividend policy of firms. Figure Figure 5 plots the dividend yield over the last 100 years. There is a modest decline in dividends in the 1950s, and a more substantial decline between 1980 and 2000, as firms have switched to buybacks or reinvestment rather than paying dividends. Both ETF investors and direct index investors must pay taxes on all dividends; ETFs do not have a way to circumvent taxable dividends and must distribute them to investors.[17] For a direct indexing strategy, paid out dividends

[16]Some additional qualifications must be met. For example, the target firm must gain a sizable (often 40%) interest in the acquiring firm via the stock from the purchase.

[17]Roundhill Investments offers a " No Dividend Target ETF" which seeks to avoid dividends by switching between two S&P 500 ETFs, selling one in the cum dividend period and buying the other in the ex-dividend period. To our knowledge, this strategy has yet to be adopted by any larger ETFs.

are invested with no embedded gain, and thus can generate harvestable losses from even small downturns in stock prices. This is a benefit from dividends for a direct index strategy which an ETF strategy does not obtain. While buy-and-hold investors in both strategies may prefer to avoid dividend payouts and the associated capital gains, it is a comparative advantage to direct indexing relative to an ETF. Consequently, the decline in the dividend yield over the last century has reduced the potential attractiveness of direct indexing relative to an ETF.

Figure 5: Dividend Yield. We plot the value-weighted average dividend yield from 1927 to 2026, as an annualized percentage of market capitalization. Data obtained from Shiller [2026].

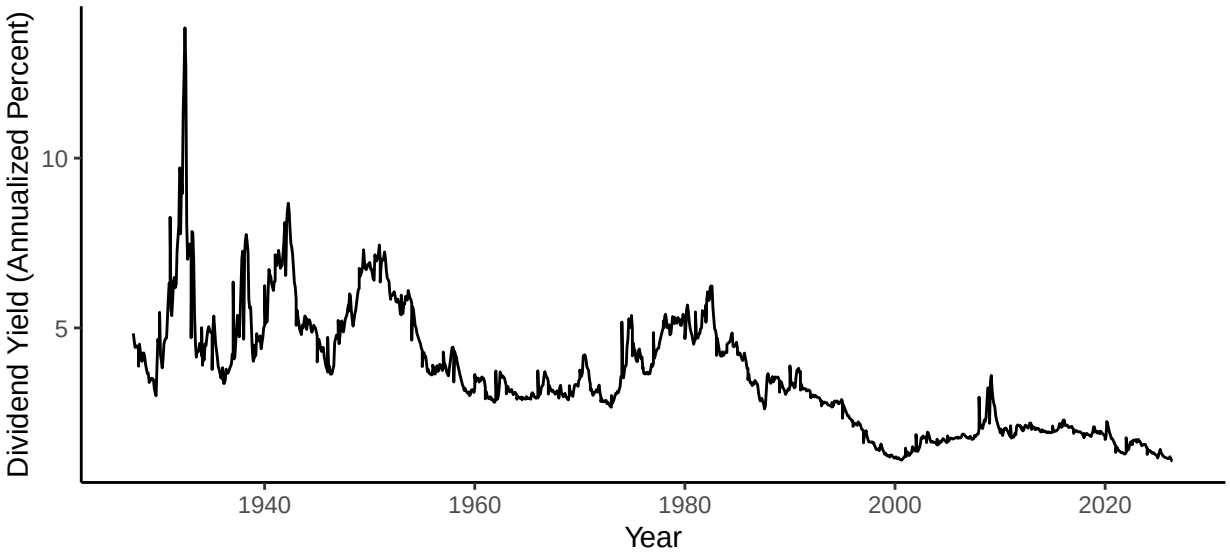
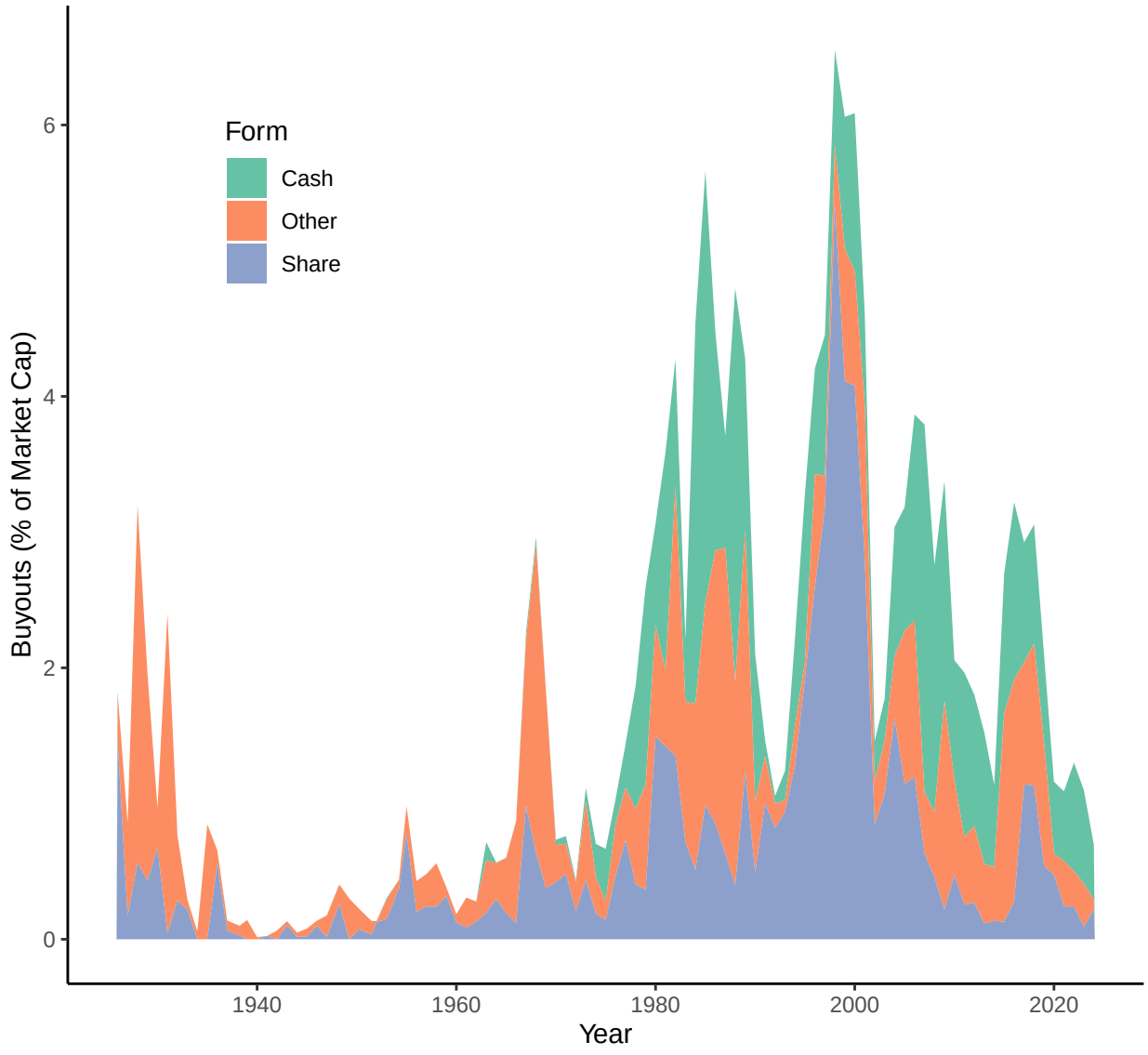


Figure 6: Merger and Aquisition Shares. We plot the value share of firms in CRSP which are bought each year. Cash buyouts are plotted in green, share buyouts in grey, and other buyouts in orange.



VI. Direct Indexing versus Real-World ETFs

Having examined the costs and benefits of direct indexing relative to an ETF for our generic index of the top N largest stocks, we now turn to a broad set of indexes tracked by real-world ETFs. We find that, for the average index, the after-tax ETF and its direct-indexing analogue produce fairly similar outcomes when harvested losses can only offset gains inside the portfolio. Direct indexing becomes relatively more valuable when remaining losses can also offset ordinary income and the total value of the portfolio is relatively small. Perhaps surprisingly, higher turnover indexes tend to benefit more from being implemented through direct indexing, but their direct-indexing analogues also exhibit greater tracking error.

A. ETF Sample and Replication Strategy

In this section, we focus on the portfolios of index ETFs, i.e., ETFs that are mandated to track some kind of index. These range from broad capitalization-weighted indexes (e.g., the S&P 500) to narrower indexes that e.g., track the 100 most profitable stocks in the Russell 1000 weighted by free-cash-flow yield (the Pacer US Cash Cows 100). We would like to highlight that the direct-indexing strategy is designed to replicate the *index* itself, not the ETF. Because, however, the direct-indexing strategy does not license the index directly, we use the ETF's reported holdings as a proxy for the underlying index portfolio that both the ETF and the direct-indexing strategy are trying to track. To simplify the exposition, therefore, we refer to the direct-indexing strategy as replicating the ETF for the rest of the section.

To select the sample of ETFs to replicate, we begin with all ETFs in the CRSP Mutual Fund database that can be matched to CRSP stock data. We then link these funds to Thomson S12 holdings data and exclude active, leveraged, and inverse ETFs, as well as ETFs for which less than 67% of reported holdings can be matched to CRSP securities. This last filter is designed to remove e.g., balanced funds that hold a substantial share of their portfolios in assets other than common stocks. Our final sample contains 652 ETFs, or 25.8% of the 2,528 ETFs that can be matched between the CRSP Mutual Fund database and CRSP stock data. Because our sample includes many of the largest ETFs, our AUM coverage is substantially higher at 53.3%. Appendix Table A10 reports our sample's coverage year-by-year.

For each ETF, the corresponding direct-indexing strategy targets the fund's reported end-of-quarter holdings. Available cash comes from quarterly contributions, after-tax dividends, tax-loss-

harvesting sales, cash delistings, and tax refunds. At each quarter end, we use this cash to move the portfolio toward the ETF's reported weights. Unlike the top-N-largest stock results in Section V, the results in this section account for expected transaction costs based on lagged stock-level bid-ask spreads.[18] We assume that the baseline investor starts with a $50,000 investment and contributes $5,000 per quarter.

When evaluating direct indexing, we consider two possible tax treatments. In the first, harvested losses can offset only current and future gains inside the portfolio. In the second, harvested losses first offset portfolio gains, and up to $3,000 of any remaining annual losses can also offset ordinary income.

When comparing the performance of our direct indexing portfolios to the ETFs themselves, we consider three related benchmarks. The first is the (post-tax) return on the ETF itself. To calculate this, we begin with the ETF's reported return in the CRSP Mutual Fund database (which is already net of the fund's expense ratio) and subtract estimated taxes on dividends and capital gains. Importantly, the direct indexing strategy is assumed not to pay a management or index licensing fee, both of which can be economically large for some ETFs [An et al., 2023].[19]

The second benchmark is a naive direct-indexing strategy that mechanically trades toward the ETF's reported end-of-quarter weights. It does no voluntary tax-loss harvesting, completley sells positions deleted from the index (regardless of whether it is tax efficient to do so or not or not), and pays the same dividend taxes, transaction costs, and annual tax bill on capital gains as the tax-aware strategy. This benchmark is designed to understand how much of the benefit of direct indexing comes from tax-aware trading rather than simply owning the underlying securities directly at each quarter-end.

The third benchmark is the return on the end-of-quarter holdings *without* deducting the taxes or transaction costs generated by rebalancing toward this benchmark. While this is not an investable after-tax strategy, it still provides a useful benchmark. Specifically, it can be compared with the naive direct-index strategy to show how much after-tax performance is lost due to the taxes and transaction costs required to closely track the reported quarter-end ETF holdings.

Finally, we compare our direct indexing strategy with our hypothetical index which holds the

[18] Ideally, we would apply the same transaction-cost adjustment to the portfolios in Section V. Comparable stock-level trading-cost measures, however, are not available over the full historical sample. Our estimates rely on effective-spread measures constructed from millisecond TAQ data, which limits how far back we can estimate these costs.

[19] Many ETFs disclose portfolio holdings daily, so using higher-frequency holdings could in principle allow the direct-indexing strategy to track the ETF more closely. We instead use quarter-end holdings to keep the exercise comparable with the generic top-N-largest stock index in Section V.

500 largest eligible stocks in CRSP. Many of the ETFs in our sample, particularly when measured by fund *count* rather than share of ETF *AUM*, follow narrower indexes with substantially higher turnover than broad value-weighted indexes such as the S&P 500. The top-500 index, therefore provides a useful benchmark for how quickly higher-turnover index strategies generate harvestable losses, forced capital-gains realizations, and opportunities to use those losses. Because not many ETFs have a long history, we start all of the analysis based on available data on or after 2011.

B. Performance Relative to the ETF

Table VI reports three comparisons: the tax-aware direct-indexing strategy relative to the ETF, the tax-aware strategy relative to the naive strategy that mechanically trades toward the ETF's end-of-quarter weights, and the naive strategy relative to the ETF. We begin the comparison at the three-year horizon so that the tax treatment of the ETF and direct-index portfolios is more apples-to-apples.[20]

We find that, even when harvested losses can offset only current and future gains within the portfolio, the median annualized advantage of direct indexing relative to the corresponding ETF is 21.5 basis points per year at three years, 35.3 basis points at five years, and 45.2 basis points at ten years. Allowing up to $3,000 of remaining annual losses to offset ordinary income increases the median advantage to 76.7, 79.0, and 68.1 basis points per year at three, five, and ten years. This shows that the direct-indexing advantage we estimate is not driven entirely by the ordinary-income offset. Many direct-indexing providers emphasize this benefit, but because the $3,000 per-year limit is fixed, the value of this offset is mostly restricted to smaller accounts, particularly once direct-indexing management fees are taken into account.[21]

Some of this advantage is due to the fact that the ETF's return is already net of its expense ratio, while the direct-index strategy is not assumed to pay a management or index licensing fee. Appendix Table A12 shows that when we examine ETF returns *gross* of fees the direct-index advantage decreases by roughly 25 to 30 basis points per year.

A second reason why the direct-indexing strategy may outperform the ETF is by avoiding costly

[20]We assume that shareholder taxes on ETF distributions are deducted as the distributions occur. For direct indexing, taxes on dividends are paid as the dividends are received, while the annual tax payment or refund associated with realized capital gains and losses affects account wealth in the second quarter of the following year. One- and two-year comparisons can therefore be affected by our assumptions about the timing of tax settlements rather than differences in underlying after-tax performance.

[21]For example, at a 37% marginal ordinary-income tax rate, the maximum annual tax benefit from the $3,000 offset is $1,110. At a 9 basis point annual direct-indexing fee, fees would equal this benefit once the account reaches roughly $1.23 million ($1,110/0.0009).

mechanical index rebalancing trades. Index rebalancing often concentrates trading at predictable dates, when trading can be particularly expensive, and there is evidence that delaying some of these trades can generate alpha [Arnott et al., 2023]. More generally, index rules can induce excessive rebalancing or trades that are poorly timed from a return perspective. For example, buying after share issuance predicts low subsequent returns [Pontiff and Woodgate, 2008], yet value-weighted indexes mechanically increase their holdings after issuance [Sammon and Shim, 2026b]. Indexes may also need to scale down existing positions to make room for newly added stocks. For example, CRSP's Fast-Track inclusion rule can require buying newly public firms only days after the IPO, when subsequent returns are relatively poor [Murray and Sammon, 2026]. Similar issues arise for narrower indexes that sort firms on characteristics. A mid-cap index, for example, may be forced to sell a stock after strong performance causes it to graduate into the large-cap universe. A direct-index investor can instead retain the position, potentially avoiding both a taxable realization and a poorly timed sale.

Index rules can also generate inefficient rebalancing on the intensive margin by requiring large changes in the weights of stocks that remain in the index. For example, an ETF that weights stocks in proportion to dividend yield may mechanically increase its weight in a stock after a sharp price decline, even if the high dividend yield reflects deteriorating fundamentals and the dividend has not yet been cut. A direct-index strategy that does not fully rebalance toward the new target weight may avoid putting additional capital into these poorly performing stocks.

The naive benchmark which trades directly toward the ETF's reported end-of-quarter weights allows us to isolate the value of tax-aware trading from the other benefits of direct indexing. Because the naive strategy only trades at the end of the quarter, it can avoid some of the crowded or poorly timed index rebalancing trades of real-world ETFs. The key difference is that it does not harvest losses or avoid selling positions with embedded gains for tax reasons. Perhaps unsurprisingly, the tax-aware strategy outperforms the naive strategy by a median 16 basis points per year at three years, 60 basis points at five years, and 101 basis points at ten years without the personal-income offset. Allowing the $3,000 ordinary-income offset increases these differences to 72, 102, and 121 basis points. This larger gap is also unsurprising because the naive strategy does not deliberately generate losses that can be used against ordinary income.

The third comparison in Table VI is between the naive strategy and the ETF itself. This comparison highlights a different tradeoff between the ETF wrapper and direct indexing: the ETF can avoid taxable events when it has to sell appreciated positions, while the naive direct-indexing

strategy can avoid some crowded or poorly timed index trades by rebalancing only toward end-of-quarter weights. At three years, the naive strategy modestly outperforms the ETF, but at five- and ten-year horizons it underperforms the ETF by 22 and 50 basis points per year at the median.

Table VI: After-Tax Performance across Alternative Implementations

		Tax-aware DI vs. ETF		Tax-aware DI vs. naive		Naive vs. ETF	
Horizon	N	Percent DI wins	Median (bps/year)	Percent DI wins	Median (bps/year)	Percent naive wins	Median (bps/year)
Panel A: No personal-income offset							
3 years	451	60.5	+21.5	57.4	+16.4	59.9	+13.2
5 years	400	63.0	+35.3	72.8	+59.9	31.2	-21.6
10 years	234	70.5	+45.2	83.3	+100.5	20.5	-49.8
Panel B: Personal-income offset allowed							
3 years	451	75.2	+76.7	73.6	+72.2	61.9	+15.9
5 years	400	75.0	+79.0	82.5	+102.3	32.0	-20.6
10 years	234	80.3	+68.1	88.0	+121.2	20.9	-49.4

Notes: The table compares the tax-aware direct-index strategy with two benchmarks: naive direct indexing that mechanically trades toward reported quarter-end holdings and the actual ETF after estimated shareholder taxes. The naive strategy does no voluntary tax-loss harvesting and sells deleted positions. Both direct-index strategies pay the same dividend taxes, transaction costs, and annual tax settlement. The ETF return is net of the ETF expense ratio, while neither direct-index strategy includes a management or index licensing fee. Each row uses the common sample for which all three wealth paths are available. Median differences are annualized basis points per year and positive values favor the first strategy named in each column heading. Neither vehicle is liquidated at the horizon. We omit the one-year horizon because ETF shareholder distribution taxes are deducted as they occur, while the first annual tax payment or refund generated by direct indexing does not affect the account until the second quarter of the following year.

The return differences between the tax-aware and naive strategies also highlight the tradeoff between tax efficiency and closely tracking a target portfolio. Appendix Table A11 shows that the naive strategy tracks the reconstructed quarter-end holdings almost exactly, with a median annualized return difference of less than one basis point. To achieve this low tracking error, however, requires selling $1/4^{th}$ to $1/3^{th}$ of the portfolio each year and therefore realizing gains of 5% to 6.5% of AUM per year. And, this leads the naive strategy to underperform the tax-aware strategy by 2.8-3.5 percentage points per year.

Table VII provides another way to understand why direct indexing outperforms its ETF analogs. Specifically, it separates the portion of the return difference that we can attribute directly to specific tax channels from a "residual" component that reflects differences in the portfolios held by the two strategies. We find that, without the personal-income offset, direct indexing outperforms by 34, 52, and 44 basis points per year at three, five, and ten year horizons. Avoiding taxes on ETF capital-gain distributions accounts for only about 5 to 6 basis points per year of this difference. The much larger

remaining component is a residual that captures all of the other forces discussed above, including ETF expenses, transaction costs, and portfolio holdings differences due to e.g., wash-sale restrictions and retained deletions.[22]

Table VII: Sources of the Direct-Index Advantage

Horizon	N	Remaining after-tax difference	ETF capital-gain distribution tax	Total without personal-income offset	Personal-income offset effect	Total with personal-income offset
3 years	499	+30.6	+6.1	+33.9	+52.8	+86.8
5 years	445	+48.3	+6.3	+51.7	+41.2	+92.8
10 years	254	+42.7	+5.1	+44.1	+24.6	+68.7

Notes: Entries are mean annualized log wealth differences in basis points, with positive values favoring direct indexing. The remaining after-tax difference captures ETF expenses, transaction costs, cash holdings, taxes on realized direct-index gains, retained deletions, wash-sale restrictions, and differences between the actual ETF return and the return on reconstructed quarter-end holdings. ETF capital-gain-distribution taxes provide an additional benefit to direct indexing because these taxes are paid by ETF shareholders but not by the direct-index portfolio. The total without the personal-income offset also incorporates the small net difference in dividend taxes between the two strategies. The personal-income-offset effect is the incremental value of allowing up to $3,000 of remaining annual losses to offset ordinary income. Neither vehicle is liquidated at the horizon. We omit the one-year horizon because the first annual direct-index tax settlement does not affect account wealth until the second quarter of the following year.

Allowing remaining losses to offset ordinary income adds another 53 basis points per year at three years, 41 basis points at five years, and 25 basis points at ten years. As discussed above, this benefit declines with horizon because the $3,000 annual limit becomes small relative to the size of the account, which grows on average over our sample for most ETFs. More importantly, Table VII shows that the direct-indexing advantage remains economically meaningful even without the direct-income-offset channel. This is consistent with the idea that, simply by trading differently from the benchmark, the direct-index portfolio can end up avoiding some trades that are costly or poorly timed on average.

C. Loss Generation and Forced Realizations

Direct indexing is only valuable if it can actually generate losses to harvest. In this subsection, therefore, we ask how quickly losses are generated when direct indexing is used to replicate the wide variety of indexes tracked by real-world ETFs. To this end, Table VIII reports both the losses generated voluntarily through tax-loss harvesting and the forced gains and losses that are realized

[22]The components do not sum mechanically to the reported total because the total also includes the small net difference in dividend taxes between the ETF and direct-index portfolios. For example, at the three-year horizon, the remaining after-tax difference and ETF capital-gain-distribution tax sum to 36.7 basis points, while the reported total is 33.9 basis points because the net dividend-tax effect is negative.

from e.g., cash acquisitions and other forms of delistings. Voluntary harvesting generates losses equal to 12.9% of average AUM per year at one year, 6.5% at three years, 5.0% at five years, and 3.8% at ten years. At the five-year horizon, the 25th and 75th percentiles are 3.0% and 5.9%.

Forced realizations are much smaller, although forced gains are more important than forced losses from a tax perspective. Forced gains rise from 0.47% of AUM per year at one year to 1.15% at ten years, with cash buyouts accounting for 0.35% and 0.64%, respectively. Harvested losses are therefore more than sufficient on average to offset these forced gains. By contrast, the amount of losses used against ordinary income declines with the horizon, in part because the $3,000 annual limit is fixed while the value of the portfolio grows.

Table VIII: Loss Generation and Forced Realizations

Panel A: Voluntary harvested losses

Horizon	N	Mean	P25	Median	P75
1 year	538	12.90	7.32	12.88	17.02
3 years	512	6.49	3.81	5.22	8.07
5 years	453	4.95	3.00	4.27	5.93
10 years	259	3.79	2.35	3.11	4.34

Panel B: Forced gains and loss use

Horizon	N	Cash-buyout gains	All forced gains	Losses used against forced gains	Ordinary-income loss use
1 year	538	0.35	0.47	0.47	4.71
3 years	512	0.55	0.83	0.80	3.10
5 years	453	0.71	1.16	1.02	2.12
10 years	259	0.64	1.15	1.03	1.09

Notes: Entries are annualized percentages of average direct-index AUM under the personal-income-offset specification. Panel A reports losses voluntarily harvested by the strategy. Panel B reports taxable gains created by cash buyouts and other cash delistings, together with the amount of harvested losses used against those gains or against the $3,000 ordinary-income limit. Forced losses are economically small, averaging 0.047% of AUM per year at one year and 0.024% at ten years.

Figure 7 puts these magnitudes in perspective by comparing the real-world ETF portfolios with the top-500-stock portfolio studied in Section V over the same period that the ETFs traded. The real-world ETF portfolios generate both more harvestable losses and more forced gains. At ten years, voluntary losses average 3.79% of AUM per year for the ETF portfolios, compared with 2.54% for the top-500, while forced gains average 1.15% versus 0.46%. This pattern is consistent with the higher turnover of many real-world indexes. For example, a mid-cap index will naturally have stocks leaving as appreciating firms "graduate" to their corresponding large cap index. More turnover

creates more new tax lots that can subsequently generate losses, but it also creates more occasions on which appreciated positions must be sold.

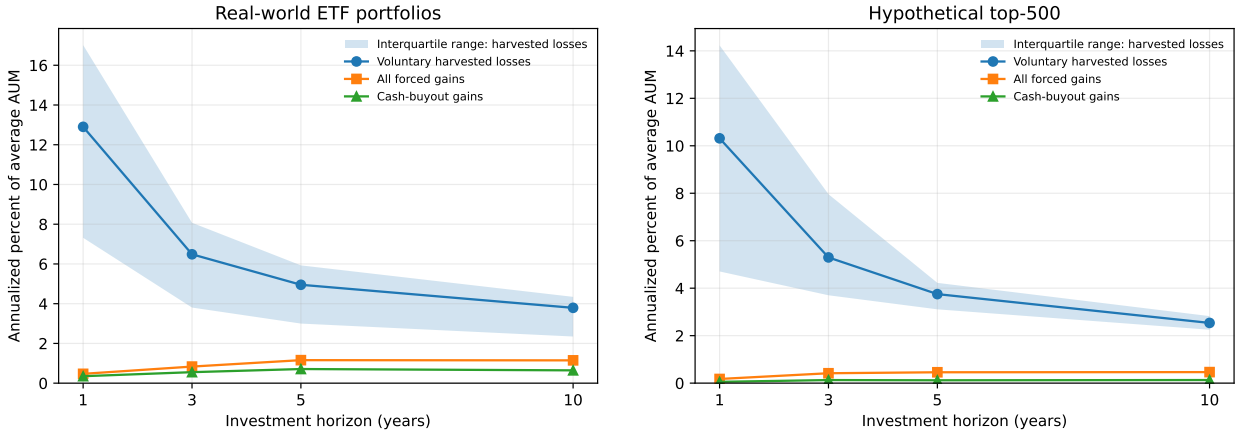


Figure 7: Loss Generation and Forced Realizations. Panel A reports means across real-world ETF replications. Panel B reports the hypothetical value-weighted top-500 strategy using the same start-year support available in the ETF sample at each horizon. The shaded region is the interquartile range for voluntary harvested losses. Both panels use a $50,000 initial investment, $5,000 quarterly contributions, the wash-sale restriction, the personal-income-offset specification, and the modern transaction-cost assumptions. Rates are annualized percentages of average direct-index AUM.

Figure 8 shows how those additional losses are used. The amount applied against ordinary income is almost identical for the real-world ETF portfolios and the top-500-stock portfolio over the same time period. At five years, losses used against ordinary income average 2.12% of AUM per year for the ETF portfolios and 2.13% for the top-500. The main difference is that the real-world ETF portfolios use substantially more harvested losses to offset gains generated inside the portfolio. At ten years, losses used against forced gains equal 1.03% of AUM per year for the ETF portfolios, compared with 0.46% for the top-500.

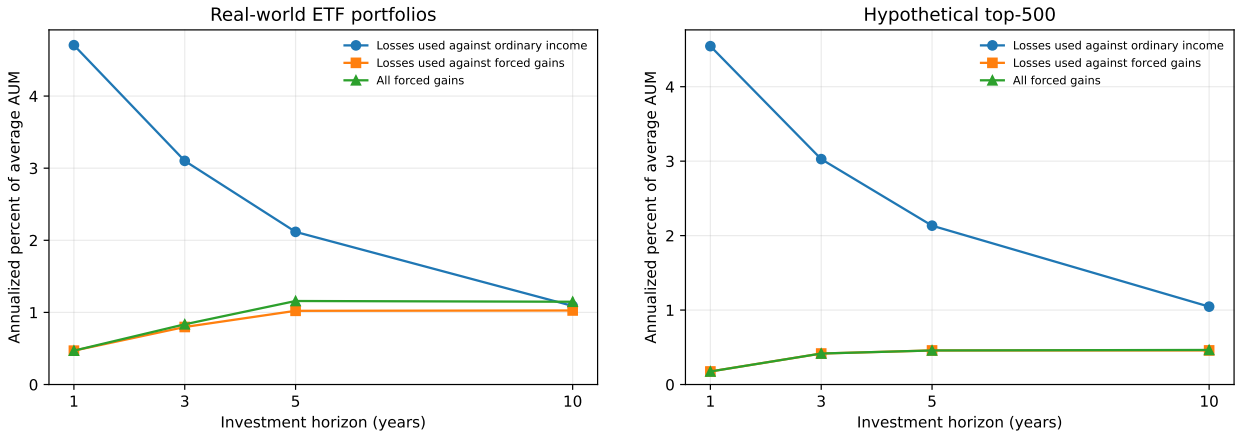


Figure 8: Use of Harvested Losses. Panel A reports means across real-world ETF replications. Panel B reports the hypothetical value-weighted top-500 using the same horizon-specific ETF start years. Both panels use the personal-income-offset specification. Loss use and forced gains are annualized percentages of average direct-index AUM.

D. Which ETF Portfolios Benefit the Most from Direct Indexing?

The results above suggest that the turnover of the underlying index will be particularly important for whether direct indexing adds value relative to the ETF. Higher-turnover indexes create more new tax lots and therefore more opportunities to harvest losses, but they also require more sales of appreciated positions. These forces make the effect of turnover ambiguous ex ante: higher turnover creates more taxable gains, but also more opportunities for the direct-indexing strategy to avoid costly trades (both from an explicity trading cost and implicit market timing cost perspective).

Table IX shows that the second effect is economically important. At five years, the median direct-index advantage relative to the after-tax ETF rises from 52 basis points per year in the low-turnover tercile to 131 basis points in the high-turnover tercile. Direct indexing produces more wealth than the ETF for 72% of low-turnover funds and 80% of high-turnover funds. The fact that the advantage is largest among high-turnover ETFs is particularly striking because these are also the portfolios for which mechanically following the index should create the most taxable realizations.

We find, however, that this advantage comes from the fact that the direct indexing strategy follows the ETF's reported holdings less closely. The median annualized return difference between the direct-index portfolio and the reconstructed quarter-end holdings rises from 2.2% in the low-turnover tercile to 3.7% in the high-turnover tercile. The takeaway from Table IX is that closely following a high-turnover index can be expensive once we account for the taxes and trading costs associated with all of that rebalancing. The tax-aware strategy therefore follows these portfolios less tightly

because avoiding some of those trades is precisely where part of the value comes from.

Another reason the direct-indexing portfolio can drift further from high-turnover ETF weights is that it will not sell appreciated positions simply to generate cash for rebalancing. Cash instead comes from sources such as contributions, dividends, tax-loss-harvesting sales, cash delistings, and tax refunds. We define dividend coverage as after-tax dividend cash divided by the dollars needed to bring underweight positions back toward their target weights. This coverage falls from 2.4% in the low-turnover tercile to 0.8% in the high-turnover tercile. When internally generated cash is limited and selling other positions would be tax inefficient, the strategy cannot fully rebalance toward the target weights, leading to greater deviations from the ETF portfolio.

Table IX: Five-Year Outcomes by ETF Portfolio Turnover

Panel A: After-tax performance

Turnover tercile	N	Median turnover (percent)	Median annualized advantage (basis points per year)	Percent direct indexing wins
Low	149	11.0	+51.6	71.8
Middle	148	35.5	+85.2	73.6
High	148	93.5	+130.9	79.7

Panel B: Tax opportunities and replication

Turnover tercile	Voluntary harvested losses	Cash-buyout gains	Retained-deletion weight	Annualized return difference	Dividend coverage
Low	4.32	0.50	4.3	2.2	2.4
Middle	5.35	0.79	20.4	3.2	1.3
High	5.08	0.75	41.4	3.7	0.8

Notes: ETFs are sorted into terciles using SEC-reported portfolio turnover during the first four simulation quarters. The table uses the personal-income-offset specification and the five-year horizon. The performance benchmark is the after-tax actual ETF. The annualized return difference is the root-mean-squared difference between the pre-tax return on the direct-index holdings and the return on the reconstructed ETF holdings. Dividend coverage is cumulative after-tax dividend cash divided by the cumulative dollars needed to purchase underweight positions. Tax-flow rates, portfolio weights, return differences, and coverage are percentages. These associations are descriptive.

VII. Model

We develop a simple binomial model to explore how direct indexing performs, particularly under policy changes. We have a CRRA investor who lives for six periods ($t = 0, 1, 2, 3, 4, 5$). The investor consumes and invests in each period. At $t = 5$, the investor makes a final choice between consumption and bequest (which is obtained with stepped-up basis) with bequest weight H in the style of Dammon et al. [2001]. Equation 1 gives the investor's utility, with consumption C_t in each period.

$$\text{Utility} = \mathbb{E}_0\left[\sum_{t=0}^{T-1} \beta^t \frac{C_t^{1-\gamma}}{1-\gamma} \; + \; \beta^T \left(\frac{C_T^{1-\gamma}}{1-\gamma} \; + \; H \frac{B^{1-\gamma}}{1-\gamma}\right)\right], \tag{1}$$

A. Assets

There are three identical, independent stocks in the model, along with a risk-free asset. Key model parameters are summarized in Table X. Our default parameters are an initial wealth $W_0 = 100$, labor income $Y_t = 10$ in periods 0,1, and 2. Each stock moves up or down by $e^{\pm\sigma}$. Stocks pay no dividends. The risk-free rate is taxed at τ_I, so the after-tax risk-free rate is $r_f(1-\tau_I)$.

An investor has a choice between buying a direct index (i.e. directly buying or selling each of the individual stocks in whatever ratio they prefer) or buying an ETF. For the direct index strategy, the DI investor must individually decide how to split wealth between consumption and investment in each of the three stocks. The ETF in the model holds an equal weight of each stock; this is the CRRA-optimal holding of the stocks since the stocks are identical and independent. The ETF investor makes a single decision: how to split wealth between consumption and investment in the ETF. At the beginning of each period, the ETF re-balances (tax-free) to always maintain equal weight of stock in each period. For both strategies, the investor's equity holdings are capped at 100% of wealth (no-borrowing), and no shorting is allowed.

Capital gains are taxed on realization at rate τ_C, and we account for each lot individually.[23] The tax basis for each lot is individually tracked, so highest-basis lots are always sold before any lower basis lots. Any time a lot has a capital loss, it can be sold to realize the loss. Losses never expire, but losses are only useful for offsetting gains (either contemporaneous or carried forward to offset future gains).

B. The Investor's Problem

At the start of each period, lots with tax losses are harvested first.[24] The investor decides consumption C_t and investment. For the ETF investor, investment is decided between the risk-free asset and the ETF, while for the direct index investor, investment is decided between the risk-free asset and each of the individual stocks. The investor's Bellman Equation (Equation 2) is solved numerically.

[23] Exact accounting allows us to explicitly track the advantage of DI over an ETF to intermediate states, where the ETF has a net gain but individual stocks have some losses. The number of state variables grows quickly, so we model only three stocks and six periods, trading off more periods for the benefit of a more precise calculation of the exact tax state than rougher approximations of accounting rules would allow.

[24] There are no transaction costs in the model, so it is always weakly optimal to sell a lot for a capital loss, as it can be repurchased with no additional cost.

$$V_t(s_t) \;=\; \max_{c_t,\,\mathbf{a}_t} \; \frac{C_t^{1-\gamma}}{1-\gamma} \;+\; \beta\,\mathbb{E}_t[V_{t+1}(s_{t+1})]\,, \tag{2}$$

The full state vector is $s_t = (W_t, L_t, \{\ell\}_t)$: wealth, the loss carry forward, the full set of both shares, and the basis for each share. We solve the model by backward induction on the full tree, approximating the continuation value at each node on a grid over (W_t, L_t). Investment $\mathbf{a}_t$ is the equity investment weight for the ETF investor, and a set of three weights (which divide holdings between the risk-free rate and the three stocks) for the DI investor. We compare investment strategies by their certainty equivalents at $t = 0$, with $100 \times (CE_{DI}/CE_{ETF} - 1)$ taking positive values whenever DI offers a higher certainty equivalent for the investor.

C. Results

Figure 9 presents summary results on three different scenarios: a low-mean, low-variance scenario, a high-mean, low-variance scenario, and a high-mean, high-variance scenario. Initial equity allocations are similar between direct indexing and the ETF. Whether the certainty equivalent for DI is higher or lower than the ETF is a delicate balance between the tax advantage of DI (which allows the DI investor to use harvested losses to offset some gains in later periods) against the re-balancing benefit of the ETF. In the high-mean, low-variance scenario, both the DI and ETF investor are stuck at the no-borrowing constraint. Increasing risk aversion means that the advantage of DI over the ETF is decreasing in the investor's risk aversion γ. At very high gamma ($\gamma = 5$), the ETF investor has a higher certainty equivalent than the DI investor, as the DI investor must pay costly taxes to re-balance in later periods. In the high-mean, high-variance scenario, the higher variance means that both the ETF and DI investor hold substantial fractions of their wealth in the bond. If the DI investor's holdings become unequally weighted, the DI investor can use some of the bond wealth to invest in an underweight stock. This benefit is increasing in gamma, so the DI investor has a higher certainty equivalent than the ETF investor at high gamma.

While the ETF is always rebalanced to have an equal weight of each stock, the DI investor must sell lots, and incur any relevant taxes, if they wish to re-balance.[25] To avoid taxes, the DI investor will often hold a portfolio which is not equal weight in all stocks. Figure 10a plots the asset deviations from equal weight at the $t = 3$ period. In states with 3 to 8 up moves out of 9, the DI investor can be overweight one stock by nearly 60%. Figure 10b plots consumption over the different

[25] If the DI investor is strictly increasing their equity investment, such as by shifting dollars from the risk-free asset to equity, they can use new purchases to re-balance without a taxable event.

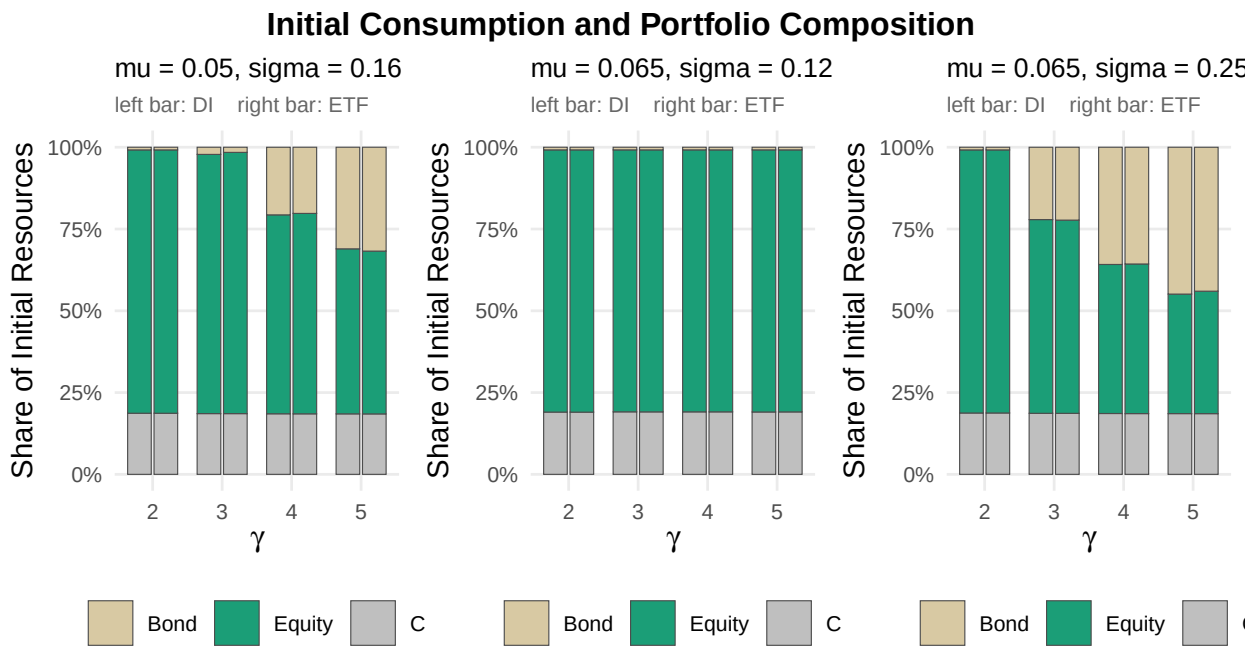
possible states. For fixed number of up moves in the stock, the DI investor has higher consumption volatility. Figure 10c plots these deviations; while the DI investor consumes exactly the same as the ETF investor on average, the DI investor consumes as much as 7% less or 11% more in specific states.

Direct investing allows the investor to harvest losses and potentially use them to offset gains. To gain a better understanding of this ability, we consider an alternative scenario where the investor has an outside-the-portfolio gain of 16 (for context, $W_0 = 100$) at the start of period 3. Figure 11 summarizes the investors position for each of the three scenarios. Both the DI and the ETF investor are able to offset the gain in the lowest states of the world; in the lowest states of the world, the ETF has a harvest-able tax loss since all stocks are down. In the middle scenario (high-mean, low-variance), the DI investor does considerably better than the ETF investor, as the DI investor can offset considerably more of the gain with tax losses in intermediate states of the world. The outside the portfolio gain gives the DI investor a CE which is approximately 0.05% of its baseline ($t = 0$) level compared to the ETF CE gain of 0.02 to 0.04% (from low to high risk aversion). In the high-mean, high-variance scenario, the ETF investor gains more than the DI investor. In this scenario, the DI investor is able to offset a substantial portion of the gain, but doing so uses up most of the DI investor's tax credits. In subsequent periods, the DI investor holds a considerably unbalanced portfolio in some states of the world, as there are no tax losses leftover to use for re-balancing.

Taxes remain constant in the base model. We consider an alternative to the base model where capital gains taxes in period 3 increase from 20% to 30%, and remain at 30% through periods 3, 4, and 5. Figure 12 plots the change in certainty equivalent gap for each of the three regimes. We consider two possible changes: a countercyclical tax increase and an acyclical tax increase. In the countercyclical tax increase, we apply a tax increase to a node (and all child nodes) if the net return on the stock market is negative for that node at the beginning of period 3. The change in certainty equivalent between DI and the ETF is nearly identical, as these are exactly the states of the world where the ETF also has tax losses which have been harvested. Under an acyclical tax increase, where taxes increase for all nodes, DI does comparatively better than the ETF; the CE for the ETF investor falls by approximately 0.1 pp more than the CE for the DI investor. In the high-mean, high-variance state, the result depends on gamma, for the same reason as in the base model that at low gamma, the DI investor faces unbalanced stock holdings due to the prohibitively high tax cost of rebalancing. At higher levels of gamma, the DI investor has higher bond holdings and can use some bond holding proceeds to buy underweight stocks.

Figure 9: Gamma Comparsion. For a sample set of parameter values, we plot the consumption and investment at the initial node in Panel A. The difference in certainty equivalent between direct indexing and an ETF (positive means DI has higher CE, while negative means ETF has higher CE) is plotted in Panel B.

(a) Initial Equity Holdings



(b) Certainty Equivalent Comparison

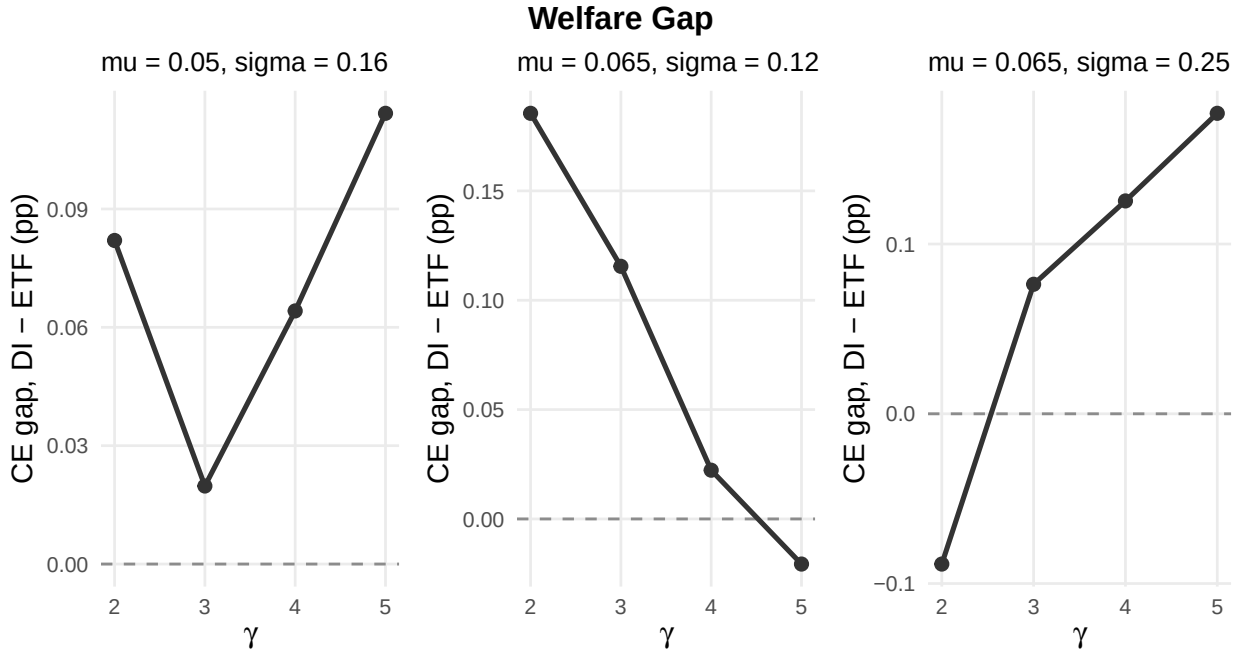
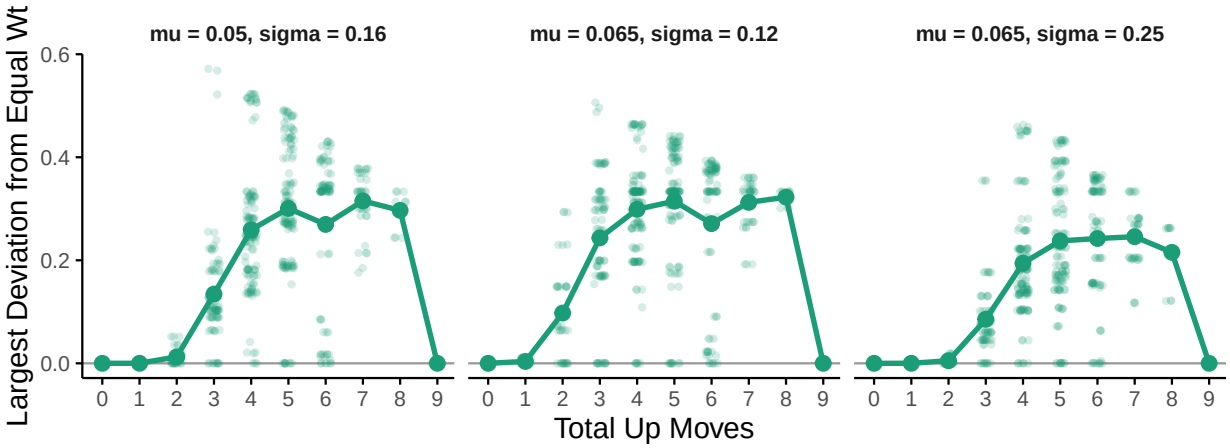
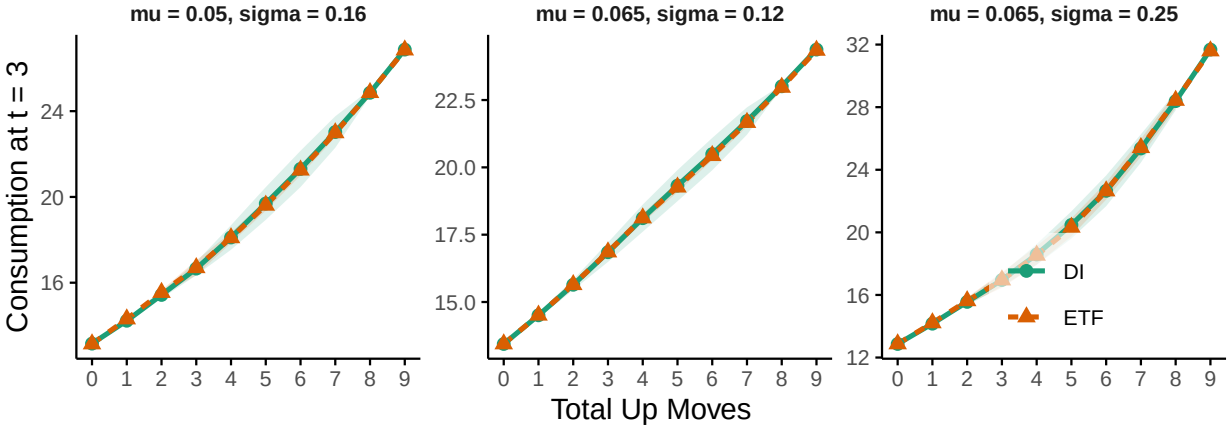


Figure 10: Path Volatility. The DI investor avoids selling deeply embedded gains. Panel A highlights the deviations from equal weight stock holdings at $t = 3$, with the solid line highlighting the average over states. Panel B presents consumption, with the solid line highlighting the average and the ribbons higlighting the range of possible values. Panel C presents the differences in consumption, with the line representing the average of the individual points.

(a) Asset Investments



(b) Consumption



(c) Consumption Differences

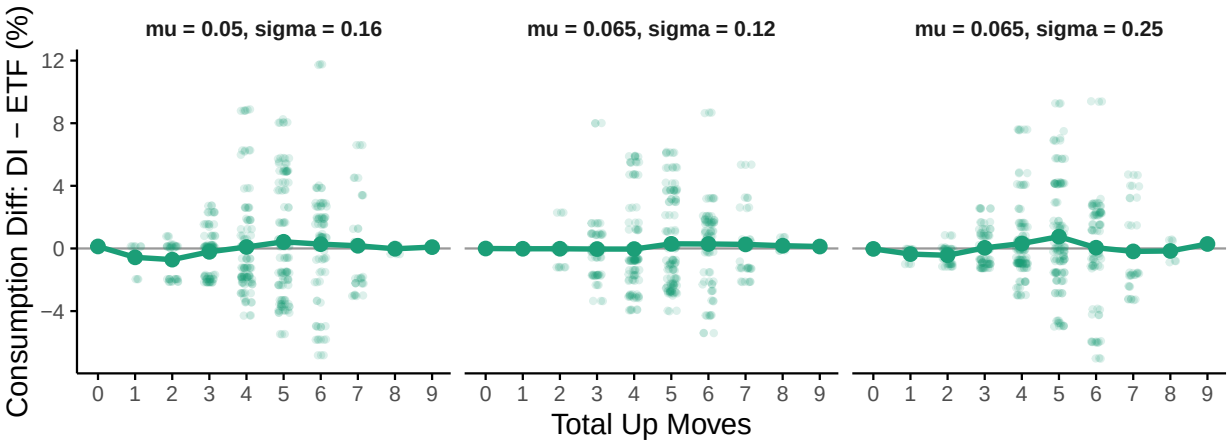
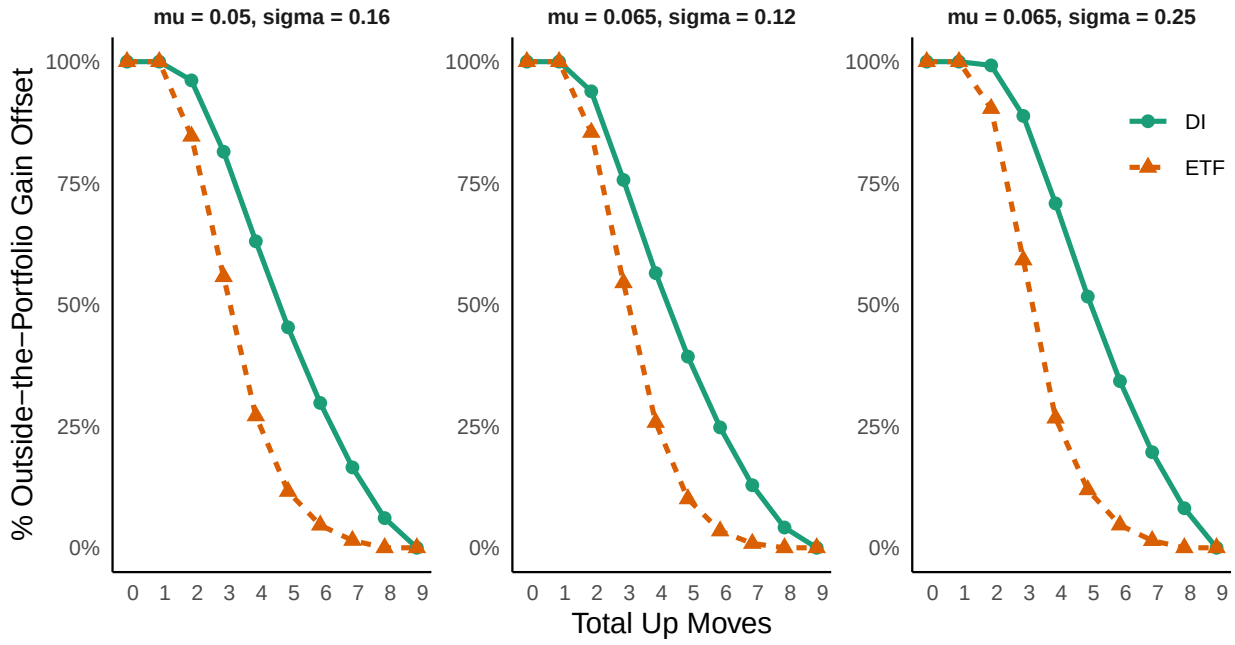


Figure 11: Outside-The-Portfolio Gain. We consider an investor receiving an outside-the-portfolio gain of 16 units of wealth. In Panel A, we plot the mean percentage of the gain which can be offset by harvested tax credits. The x-axis gives the total number of up-moves of the binomial tree; for the states with 0 up moves, the investor has enough unused losses to offset 100% of the gain. In Panel B, we plot the change in certainty equivalent for the ETF and for DI, plotted for each value of risk aversion gamma (x-axis).

(a) Percentage Income Tax-Offset



(b) CE Change From Gain

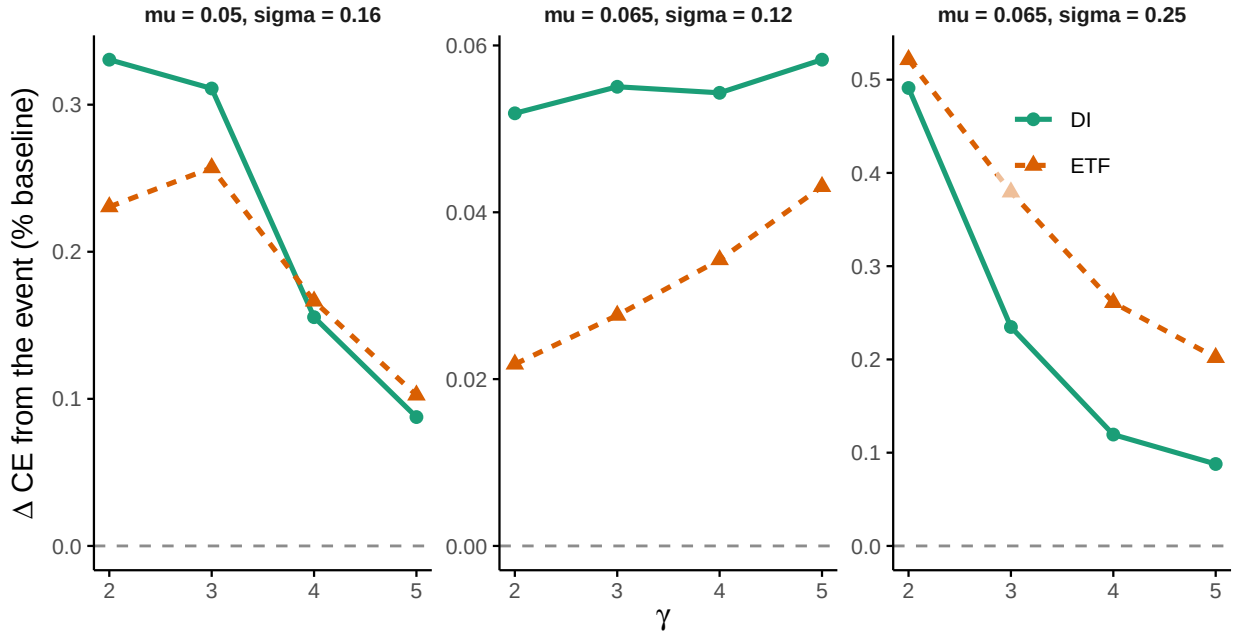


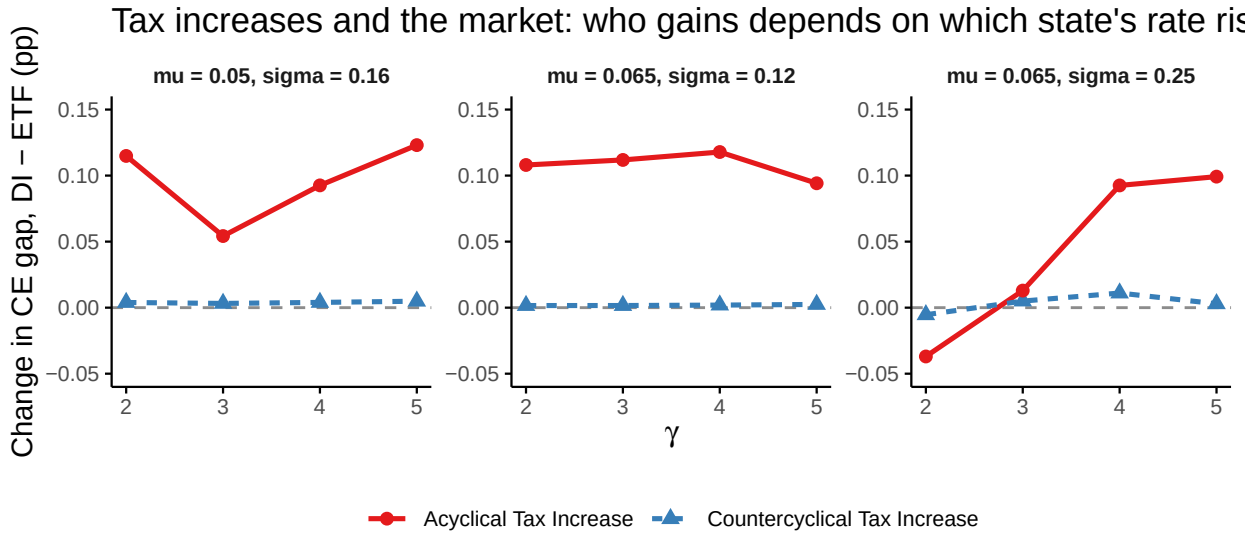
Table X: Model Parameters. We summarize the key parameters of our model. We consider a CRRA investor who lives for 5 periods, and makes a bequest in the terminal period. There are three identical stocks along with a risk-free bond.

Panel A: Common Parameters	
Risk aversion (CRRA)	$\gamma \in \{2, 3, 4, 5\}$
Discount factor	$\beta = 0.96$
Bequest weight	$H = 1$
Horizon	$T = 5$ periods
Initial financial wealth	$W_0 = 100$
Labor income	$Y = 10$ per period, received $t = 0, 1, 2$
Number of stocks	$N = 3$, i.i.d.; ETF holds equal weights, rebalanced internally without tax
Risk-free rate	$r_f = 5.127\%$; interest taxed as ordinary income
Capital gains tax rate	$\tau_C = 20\%$
Ordinary income tax rate	$\tau_I = 35\%$
Lot accounting	highest-in-first-out; losses harvested each period
Loss offsets	against realized gains, with carryforward; no offset against ordinary income
Terminal date	Optimize between taxable sales (funding final consumption) and in-kind bequest with basis step-up

Panel B: Stock price scenarios

	Scenario 1	Scenario 2	Scenario 3
Expected log price growth μ	0.050	0.065	0.065
Volatility σ	0.16	0.12	0.25
Up / down factors e^{σ}, $e^{-\sigma}$	1.174, 0.852	1.127, 0.887	1.284, 0.779
Per-period up probability	0.620	0.749	0.571
After-tax equity premium (pp)	1.7	3.2	3.2

Figure 12: Taxes and Welfare Change. We measure the period-zero difference in certainty equivalents between direct indexing and the ETF from a tax change in periods 3-5. Under the acyclical tax increase (red line), capital gains taxes increase from 20% to 30% in all states in periods 3, 4, and 5. Under the countercyclical tax increase (blue line), capital gains increase from 20% to 30% only in states with an average loss across the three stocks at $t = 3$.



VIII. Conclusion

A taxable investor who wants exposure to an index must chose not only the index, but the method of obtaining the index. Recent declines in transaction costs and the relaxation of fractional share constraints have made direct indexing strategies, where the investor directly owns the entire portfolio, more competitive than ever before. The key advantage of an ETF is the ability to avoid realizing capital gains around M&A activity thanks to in-kind redemption. A direct index strategy cannot avoid these gains, but it can selectively harvest losses on individual securities of the portfolio, as well as meet cash needs by selling high-basis lots.

Our results show that ETFs do not unconditionally dominate direct indexing, and for some investors a direct index can be superior. Under the most restrictive assumption that an investor has no use for losses, an ETF's ability to avoid gains will eventually dominate. This can take as little as ten years in the late 1970s to over 50 years for the early 1920s. If an investor can use losses to reduce their personal income taxes, direct indexing always dominates an ETF, but the $3,000 a year cap means that this only applies to investments under roughly $1 million (depending on the time horizon). Losses can also be used to offset outside-the-portfolio gains, which allows direct indexing to dominate an ETF; this does not work for all years in the time series, and the gains must occur within the first decade or two for the tax credits to be worth more than the ETF's ability to avoid future forced capital gains.

In the historical time series, merger and acquisition activity varies greatly. Cash mergers are costly for direct indexing because they create unavoidable taxable events. The early 1980s therefore produce poor outcomes for many direct-index vintages. Stock-for-stock mergers are different because they can carry basis into the acquirer. The late 1990s share buyout merger wave is therefore less damaging for direct indexing than the 1980s cash buyout wave. Dividend policy also matters. Dividends are taxable for both ETF and direct-index investors, but reinvested dividends create new high-basis lots for the direct-index investor. The decline in dividend yields reduces this source of future harvestable losses.

Direct indexing also changes how investors can raise cash to meet outside the portfolio obligations. An ETF investor who sells fund shares effectively sells a pro-rata claim on the embedded gains and losses in the the underlying stocks. A direct-index investor can instead choose which security-level lots to sell. In our liquidation exercises, this flexibility produces meaningful tax savings. These savings, however, are not free. Selling high-basis lots moves the remaining portfolio away from

benchmark weights and creates tracking error. The measured performance cost is modest in our simulations, but it is a real cost. Direct indexing therefore gives the investor a choice that the ETF does not provide: pay more tax today to stay closer to the benchmark, or accept some tracking error while deferring the tax bill.

Beyond the issue of raising cash, direct indexing also allows an investor to deviate from an index according to their individual preferences. For example, an investor may want exposure to an index of all companies except one specific industry, either for hedging reasons or personal moral reasons. As ETFs lock an investor to a specific index, a personalized portfolio can only be obtained via direct indexing. Our results on the tax tradeoff between an ETF and direct index can also be interpreted as a calculation of the cost of this personalization of the portfolio. The fact that direct indexing in many cases dominates an ETF for tax efficiency suggest that in many cases, the tax cost to investing in a direct index for personal reasons (whether financial or moral) is a low or even negative cost for some investors.

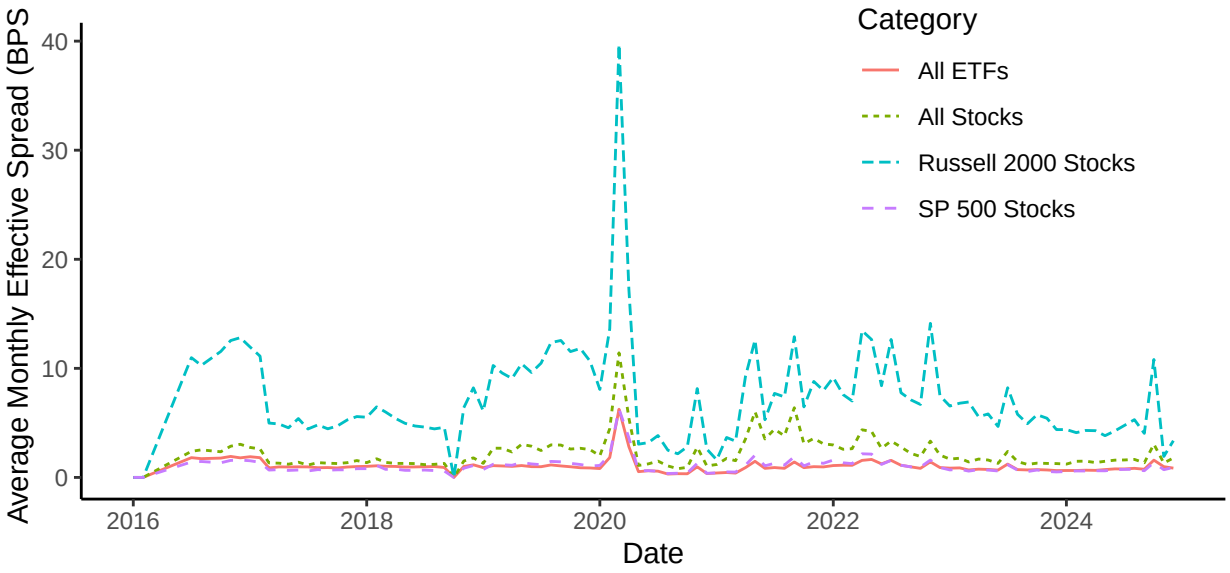
Our results on the costs and benefits of direct indexing for indexes tracked by real-world ETFs provide another angle on the value of this flexibility. Many ETFs track their benchmark indexes extremely closely, with tracking error below five basis points per year. However, there is significant evidence that rigidly tracking these benchmarks may reduce returns [Murray and Sammon, 2026, Sammon and Shim, 2026b, Arnott et al., 2023, Wang et al., 2023]. We find that, across a broad sample of index ETFs, direct indexing outperforms the ETF, even when harvested losses can only offset future gains inside the portfolio. Given the tax efficiency of the ETF wrapper, we argue that much of the remaining advantage comes from allowing the direct-indexing portfolio to deviate from the benchmark. Specifically, direct indexing's flexibility allows it to avoid some costly rebalancing trades. The key tradeoff, however, is that while these deviations can improve after-tax performance, they can only do so by taking on greater tracking error.

Zooming out, one broader lesson from our results is that ETFs and direct indexing each come with their own tax costs and benefits. The ETF wrapper is highly effective at avoiding forced capital-gain realizations inside the portfolio, while direct indexing gives investors greater control over when gains and losses are realized and how tightly they follow the benchmark. Which approach is better, therefore, depends on the investor's ability to use losses, especially if they have gains outside the portfolio to offset, as well as their investment horizon, cash needs, and willingness to accept tracking error. As the cost of holding diversified portfolios directly continues to fall, the relevant question is not whether one structure is universally more tax efficient, but which set of benefits is

more valuable for a particular investor.

Appendix A. Transaction Costs

Figure A1: Average Spreads. We plot average effective spreads, in basis points, calculated as a volume-weighted average across stock-months from SEC 605 reports from the eight largest wholesalers.



Appendix B. Historical Tax Position

Figure A2 presents the net tax position on a portfolio for every possible starting year, in five year increments, of a one-time investment of $50,000 in a value-weighted index of the 500 largest stocks. Figure A3 tracks the number of years until a portfolio first has a net loss, while Table A1 provides a table of summary statistics by decade of the number of years until a portfolio first has a net loss.

Table A2 presents summary statistics on the maximum investment (by decade) for which the personal income tax credit (assuming a personal income tax rate of 35% and a cap of $3,000) would, if invested at the portfolio rate of return, exceed any future capital gains taxes on the portfolio.

Table A3 presents details on the return advantage from the personal income tax deduction relative to a portfolio which does not take the personal income tax deduction.

Figure A2: Cumulative Realized Tax Position. For years in five-year increments, we track the cumulative realized gains or losses of a direct indexing strategy with a one-time $50,000 investment in a value-weighted index of the top 500 stocks in that year. Any sale or delisiting proceeds and dividends are reinvested in the contemporary value-weighted index.The x-axis is the date while the y-axis is the net tax position as a percentage of contemporaneous portfolio value.

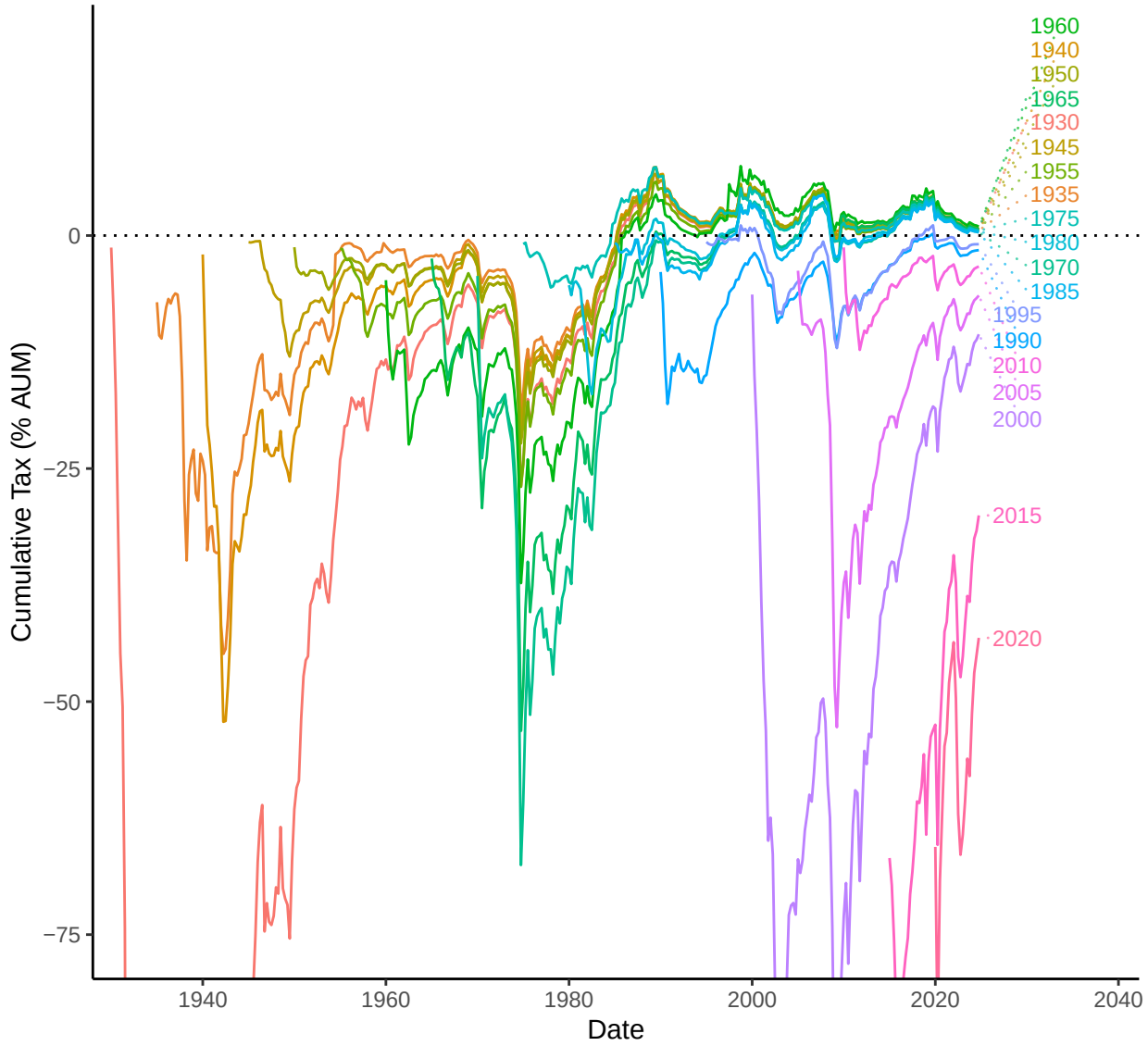


Figure A3: Years Until Net Positive Tax. For a one-time investment in a direct index portfolio, we plot the number of years until the realized net tax position is positive. The x-axis is the year the investment is made, while the y-axis is the number of years until the first positive net tax position. Years with no positive net tax position are marked with a red x.

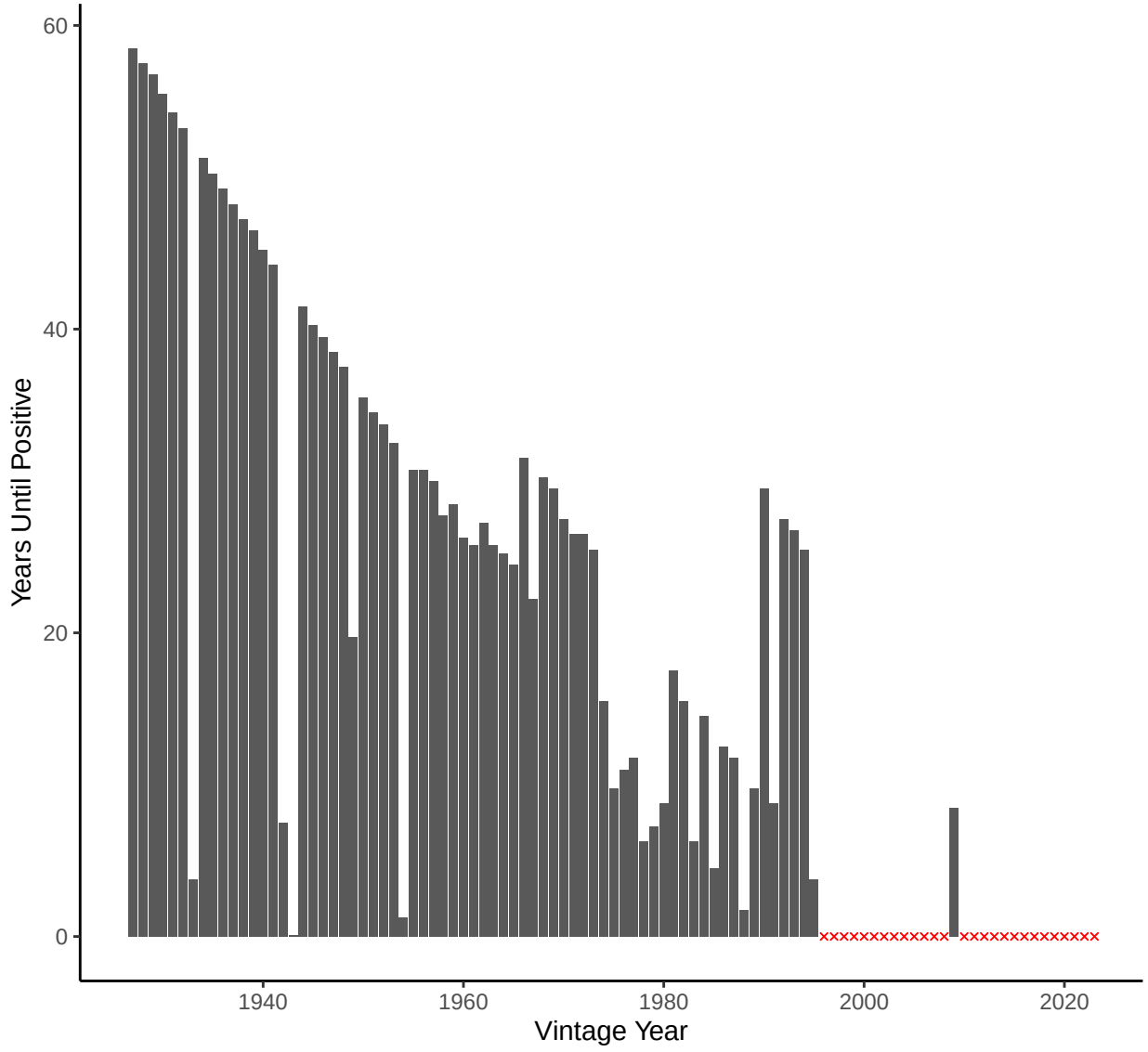


Table A1: Years Until Net Positive Tax Position. For each possible starting year, we calculate the number of years until the net tax position is positive (i.e. gains exceed losses including carry-forward losses). Panel A presents the tax position for a one-time investment of $50,000 while Panel B presents the losses for a recurring quarterly investment of $5,000.

Panel A: Tax Position with $50,000 One-time Investment

decade	Mean	SD	5%	25%	50%	75%	95%
1930	49.7	16.4	24.8	51.8	54.0	57.0	59.0
1940	41.0	12.1	22.3	41.8	44.0	46.2	49.0
1950	38.1	3.8	32.9	35.8	38.0	40.4	43.3
1960	39.7	7.6	32.2	34.2	37.2	44.0	52.2
1970	26.1	12.9	10.8	16.6	24.8	29.5	46.9
1980	26.6	10.3	12.6	18.2	27.2	35.2	38.9
1990	21.2	11.3	0.0	24.0	25.4	27.5	29.3

Panel B: Tax Position with recurring $5,000 Quarterly Investment

decade	Mean	SD	5%	25%	50%	75%	95%
1930	55.0	3.0	51.0	52.8	55.0	57.2	59.0
1940	47.0	3.1	43.0	44.8	47.0	49.2	51.3
1950	45.2	3.0	41.2	43.0	45.2	47.5	49.3
1960	35.2	3.0	31.2	33.0	35.2	37.5	39.3
1970	25.2	3.0	21.2	23.0	25.2	27.5	29.3
1980	23.7	11.5	7.1	17.2	25.0	33.2	35.3
1990	8.9	13.3	0.0	0.0	0.2	21.1	28.8

Table A2: Largest starting portfolio ($) whose personal income-offset tax credits cover all future realized taxes, by vintage decade. N is the number of observations. NoTax is the number of vintages with no capital gains taxes.

Maximum covered starting investment by vintage decade

Decade	N	No tax	Mean ($)	Min ($)	Median ($)	Max ($)
1920s	3	0	4,999,818	3,470,637	4,663,813	6,865,004
1930s	10	0	2,012,589	877,556	1,545,790	5,111,359
1940s	10	0	898,385	594,222	899,376	1,267,234
1950s	10	0	1,111,850	669,448	1,021,058	1,858,576
1960s	10	0	2,118,457	1,227,069	2,145,167	3,045,401
1970s	10	0	1,436,875	277,897	1,261,385	2,985,194
1980s	10	0	1,137,621	477,746	929,119	2,134,724
1990s	10	4	12,612,027	817,292	12,533,492	25,031,540
2000s	10	9	380,068	380,068	380,068	380,068
2010s	10	10				
2020s	4	4				

Table A3: Tax Advantage from Personal Income Offset. We calculate the value of the personal income offset tax advantage by comparing the cumulative returns of a portfolio which takes the personal income tax offset (if there are losses to deduct) at a rate of 35% to a portfolio which does not take the personal income tax offset. Panels A, B, and C present use specific values calculated to deliver an average of $3,000 a year for the first 15 years of the portfolio life in 50% of historical years (Panel A), 75% of historical years (Panel B) or 50% of years from a recurring rather than lump-sum investment (Panel C).

Panel A: IO Return Advantage For $129,815.50 One-time Investment

Age_Group	Mean	SD	5%	25%	50%	75%	95%
0–4	1.0	1.1	0.0	0.0	0.8	1.6	3.0
5–9	3.7	2.1	1.4	2.4	3.2	4.5	7.9
10–14	5.6	3.4	2.1	3.2	4.6	6.8	11.6
15–29	7.4	4.9	3.0	4.3	6.0	8.9	16.4
30–50	8.8	6.6	3.4	5.0	7.2	9.7	27.5

Panel B: IO Return Advantage For $232,732.20 One-time Investment

Age_Group	Mean	SD	5%	25%	50%	75%	95%
0–4	0.6	0.6	0.0	0.0	0.5	0.9	1.7
5–9	2.1	1.1	1.0	1.5	1.8	2.5	4.4
10–14	3.2	1.8	1.5	2.1	2.7	3.8	6.5
15–29	4.3	2.7	1.9	2.6	3.5	5.0	8.9
30–50	5.1	3.5	2.2	3.2	4.2	5.6	14.6

Panel C: IO Return Advantage For $1,303.65 recurring Quarterly Investment

Age_Group	Mean	SD	5%	25%	50%	75%	95%
0–4	2.0	2.8	0.0	0.0	1.2	2.6	7.0
5–9	5.7	3.7	2.1	3.1	4.4	7.5	12.1
10–14	7.1	3.7	3.2	4.3	5.8	9.2	14.2
15–29	8.4	3.3	4.5	5.7	7.5	10.5	14.3
30–50	9.4	3.3	5.7	6.9	8.8	11.3	17.2

Appendix C. Portfolio Liquidations

Table A4 presents summary statistics on tax savings for liquidity shocks by portfolio age. Tax savings are measured as the difference between tax from selling highest-basis lots and tax from selling pro-rata lots, as percentage points of the tax due on the pro-rata lots.

Figure A4 plots the liquidation tax savings (in basis points of portfolio value) by vintage for a portfolio which is 20 years of age. These savings represent the difference between tax from selling highest-basis lots and tax from selling pro-rata lots, measured in basis points of portfolio value.

Table A5 presents the change in performance, tracking error, and factor loadings for a liquidation of 50% of the portfolio. Cumulative return performance on the median portfolio is negative for 10, 20, and 30 years after liquidation due to a change in factor loadings from the tax-efficient liquidation.

Following a tax-efficient liquidation changes the factor loadings, the investment strategy of our default scenario assumes investments are set to minimize deviations from portfolio weights, buying stocks which are the most underweight first. An alternative strategy would be to buy stocks according to Daniel et al. [1997] characteristics. Replacement purchases are matched on a 3×3×3 sort of size, book-to-market, and momentum (27 cells). After the shock sale, new cash goes to cells that are underweight relative to the benchmark's cell weights. Table A6 presents cumulative performance, tracking error, and factor loadings for this strategy.

Table A4: Tax Under Liquidation. We consider a liquidation of 10% (Panel A) or 30% (Panel B) of the portfolio. Under liquidation, we calculate the numerical difference in tax between the tax bill from selling the highest basis lots and the tax bill from selling all lots pro-rata. We measure this tax difference as percentage points of the tax due under the pro-rata liquidation.

Panel A: 10% Liquidity Shock

Portfolio Age	**Mean**	**SD**	**5%**	**25%**	**50%**	**75%**	**95%**
0–5	78	16	54	65	75	93	100
5–10	64	11	48	57	63	71	83
10–15	61	9	48	54	59	66	77
15–20	59	8	47	54	59	65	72
20–30	59	8	46	54	59	65	72
30+	58	8	44	52	57	62	71

Panel B: 30% Liquidity Shock

Portfolio Age	**Mean**	**SD**	**5%**	**25%**	**50%**	**75%**	**95%**
0–5	53	23	29	38	45	59	100
5–10	38	9	26	33	37	42	52
10–15	34	8	24	29	33	38	47
15–20	32	6	21	28	32	36	42
20–30	31	6	19	27	31	35	41
30+	28	6	18	24	28	33	40

Figure A4: Liquidation Savings. For each vintage year, we calculate savings (in basis points of portfolio AUM) of the tax difference between selling highest-tax-basis lots and selling pro-rata. We assume this liquidation shock occurs 20 years after initial investment, and is a liquidation of 10% (red dots) or 30% (blue triangles).

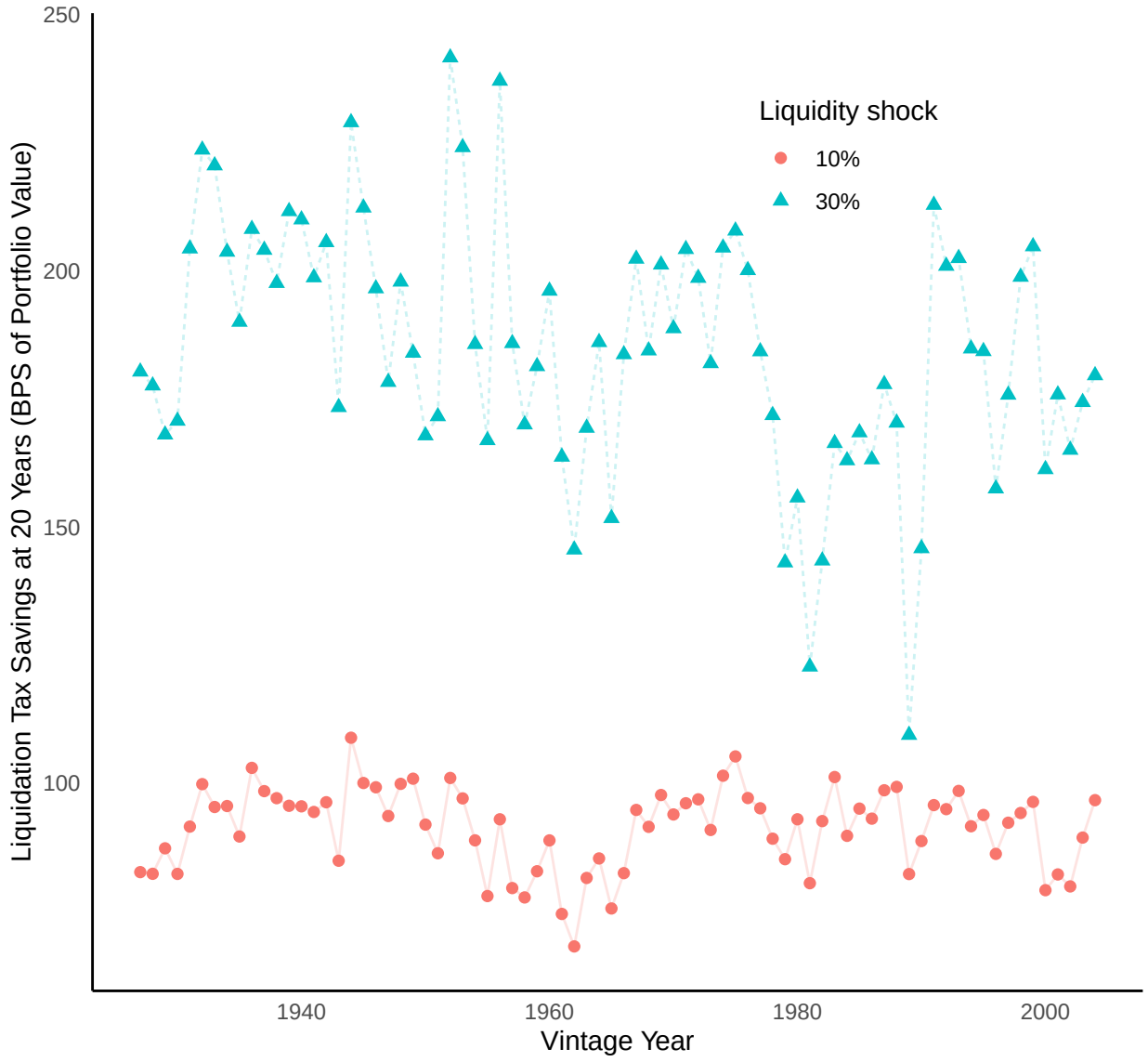


Table A5: Performance and Tracking Error. We consider a liquidation of 50% of the portfolio, selling the highest-basis lots. Panel A measures the post-liquidation cumulative return difference between the post-liquidiation portfolio and the never-liquidated portfolio at specific years, in percentage points. Panel B measures tracking error via the standard deviation of quarterly return difference between the post-liquidation portfolio and the never-liquidiated portfolio over specific year ranges, in percentage points. Panel C measures the change in Fama-French factor loadings following liquidation.

Panel A: Cumulative Relative Performance

Years After Shock	**Mean**	**SD**	**5%**	**25%**	**50%**	**75%**	**95%**
5 yrs	-0.73	5.83	-11.21	-3.74	0.06	3.08	7.00
10 yrs	-1.55	6.79	-14.10	-4.90	-0.73	2.76	7.86
20 yrs	-1.93	7.76	-15.42	-7.06	-1.01	3.69	8.61
30 yrs	-1.42	7.24	-14.74	-5.02	-0.46	2.98	9.58

Panel B: Tracking Error (SD)

Years After Shock	**Mean**	**SD**	**5%**	**25%**	**50%**	**75%**	**95%**
5 yrs	1.22	0.48	0.68	0.90	1.15	1.43	2.05
10 yrs	1.16	0.40	0.65	0.90	1.12	1.35	1.83
20 yrs	1.05	0.32	0.59	0.85	1.03	1.21	1.55
30 yrs	0.89	0.25	0.55	0.73	0.86	1.01	1.32

Panel C: Factor Loadings

Years After Shock	**Mean**	**SD**	**5%**	**25%**	**50%**	**75%**	**95%**
Mkt-RF	-0.06	0.06	-0.14	-0.11	-0.08	-0.03	0.05
SMB	-0.02	0.10	-0.15	-0.06	-0.03	-0.01	0.16
HML	0.05	0.08	-0.06	0.02	0.06	0.09	0.15

Table A6: Performance and Tracking Error using DGTW characterstics. We consider a liquidation of 50% of the portfolio, selling the highest-basis lots. Compared to Table A5, stocks are replaced according to Daniel et al. [1997]. Panel A measures the post-liquidation cumulative return difference between the post-liquidiation portfolio and the never-liquidated portfolio at specific years, in percentage points. Panel B measures tracking error via the standard deviation of quarterly return difference between the post-liquidation portfolio and the never-liquidiated portfolio over specific year ranges, in percentage points. Panel C measures the change in Fama-French factor loadings following liquidation.

Panel A: Cumulative Relative Performance

Years After Shock	Mean	SD	5%	25%	50%	75%	95%
5 yrs	-0.70	7.16	-12.97	-4.56	0.47	4.05	8.74
10 yrs	-2.15	8.44	-17.63	-6.88	-0.72	3.51	9.27
20 yrs	-3.46	9.38	-19.30	-10.32	-1.40	3.95	8.43
30 yrs	-0.67	6.26	-11.33	-4.48	-0.12	3.62	8.54

Panel B: Tracking Error (SD)

Years After Shock	Mean	SD	5%	25%	50%	75%	95%
5 yrs	1.34	0.50	0.76	1.01	1.24	1.52	2.22
10 yrs	1.31	0.43	0.77	1.00	1.24	1.56	2.06
20 yrs	1.37	0.30	0.90	1.23	1.36	1.52	1.81
30 yrs	1.22	0.20	0.89	1.09	1.20	1.35	1.55

Panel C: Factor Loadings

Years After Shock	Mean	SD	5%	25%	50%	75%	95%
Mkt-RF	-0.03	0.09	-0.16	-0.10	-0.03	0.05	0.11
SMB	-0.07	0.06	-0.16	-0.11	-0.07	-0.03	0.04
HML	0.04	0.09	-0.10	-0.01	0.04	0.10	0.18

Appendix D. HIFO and Monthly Analysis

Table A7 presents summary statistics of the comparison between dollar-cost average and HIFO (individual lot tracking) accounting for losses. Figure A5 produces the same plot as Figure A6, but for a recurring investment of $50,000 per quarter rather than an initial one-time investment of $50,000.

Appendix A. Alternative Measures

Our base exercise uses quarterly investments, FIFO tax accounting, a minimum holding period of 1 quarter, and a minimum loss of 2% before selling. We consider several possible variations to this exercise.

We first consider dollar cost average tax (DCA) accounting rather than FIFO accounting. While HIFO is the most tax-efficient method, it requires tracking the tax basis of each individual lot to ensure highest basis lots are sold first. Prior theoretical literature typically models dollar-cost average tracking, as this collapses the number of state variables from a variable for each individual lot basis to a single variable which tracks the average basis for each stock. In principle, the differences may seem slight for a buy-and-hold portfolio: stocks have upward drift, and gains are quickly embedded for all but the largest downturns in the stock price. In practice, however, substantial cash flow is generated from dividends and M&A which is reinvested at a higher basis.

Panel A of Figure A6 compares loss harvesting with HIFO individual basis against loss harvesting with a DCA basis. For long-term losses, the two methods are very similar: the median portfolio (across historical vintages) for a one-time investment generates a loss of around 3.5% of portfolio value, and the difference between HIFO and DCA is less than 0.5%.[26] For short-term losses, the methods diverge considerably: the median portfolio for a one-time investment harvests a peak loss of 10% under DCA and 12.5% under HIFO, and a long-term difference is around 4%. Table A7 and Figure A5 in Appendix D provide additional details. The difference is meaningful for comparing direct indexing to an ETF. Panel B of Figure A6 illustrates that for the median portfolio, direct indexing with HIFO accounting will have a net tax loss for a decade longer than DCA accounting. An ETF can be superior to direct indexing only when long-term gains (which the ETF avoids) exceed losses, so the extension of the period of net losses for direct indexing by 10 years (i.e. a 50% increase from the DCA median of 20 years) is substantial.

[26]HIFO harvests slightly fewer long-term capital losses than direct-indexing, as some long-term loss opportunities which DCA harvests have already been partially harvested as short-term losses under HIFO.

Table A7: HIFO vs Dollar-Cost Averaging. For each vintage-year, we calculate the difference between the losses harvested by HIFO individual lot tracking and the losses harvested by dollar-cost averaging; a positive number is consistent with HIFO harvesting more losses. Panel A presents results for short-term losses, Panel B for long-term losses, and Panel C for the combination of short and long-term losses.

Panel A: Short-Term Loss Advantage (% AUM)

PortfolioAge	Mean	SD	5%	25%	50%	75%	95%
5	-17.4	10.6	-35.6	-25.3	-13.9	-8.9	-5.4
10	-7.8	5.4	-17.5	-11.3	-4.3	-3.1	-2.7
15	-2.9	2.0	-5.7	-5.0	-2.5	-1.2	-0.6
20	-0.1	1.3	-2.1	-1.4	0.1	1.1	1.4
25	0.7	0.6	-0.3	0.3	1.0	1.1	1.6
30	1.7	0.4	0.9	1.6	1.7	1.9	2.2
35	2.0	0.3	1.6	1.7	1.9	2.1	2.5
40	2.6	0.8	1.7	2.0	2.3	2.8	4.1
45	4.1	1.6	2.0	2.8	4.0	5.5	6.6
50	4.3	0.6	3.5	3.9	4.1	4.5	5.5

Panel B: Long-Term Loss Advantage (% AUM)

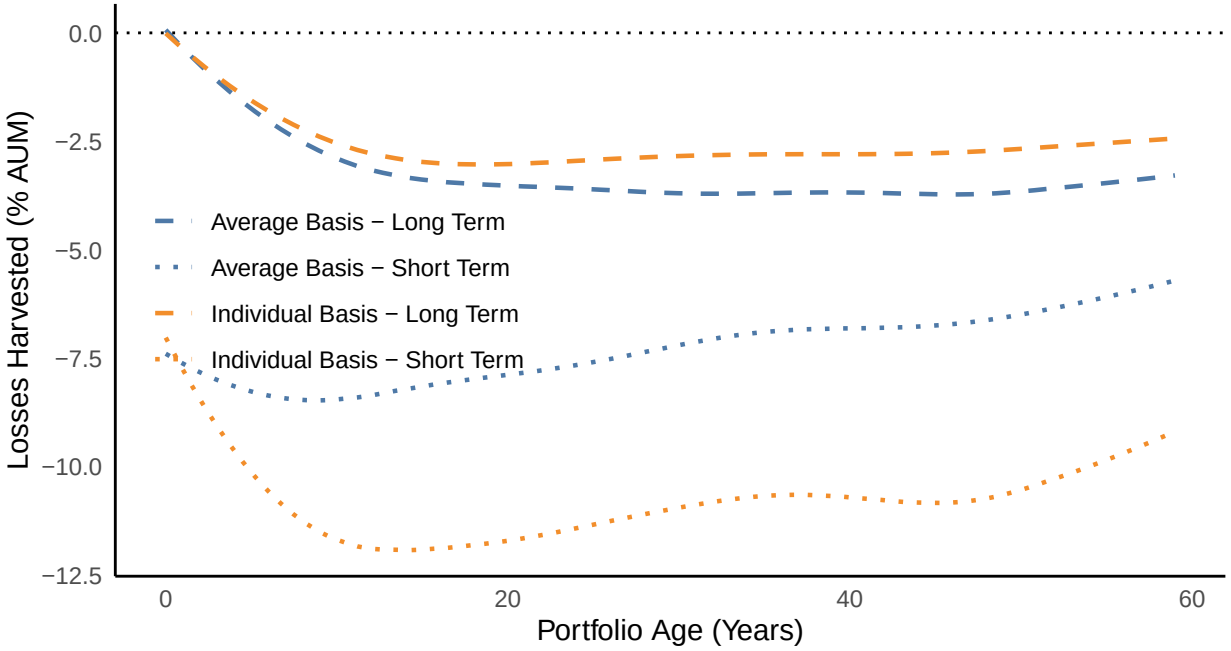
PortfolioAge	Mean	SD	5%	25%	50%	75%	95%
5	16.0	10.1	4.8	8.1	12.2	22.6	34.4
10	8.6	5.9	3.1	3.7	4.7	12.5	19.0
15	3.5	2.5	1.0	1.4	2.3	6.4	6.9
20	2.0	1.5	0.5	0.7	1.1	3.7	4.5
25	0.9	0.7	0.2	0.3	0.4	1.4	2.1
30	0.4	0.4	0.1	0.1	0.2	0.7	0.9
35	0.3	0.3	0.0	0.0	0.2	0.5	0.7
40	0.2	0.1	0.1	0.1	0.2	0.3	0.5
45	-0.6	0.4	-1.2	-0.9	-0.5	-0.2	0.0
50	-0.7	0.2	-1.0	-0.8	-0.6	-0.5	-0.4

Panel C: Total Loss Advantage (% AUM)

PortfolioAge	Mean	SD	5%	25%	50%	75%	95%
5	-1.4	3.9	-7.5	-1.3	-0.8	-0.4	2.9
10	0.9	0.8	-0.1	0.4	0.6	0.9	2.8
15	0.5	0.7	-0.3	0.1	0.2	0.7	1.9
20	1.9	0.5	1.1	1.6	1.9	2.2	2.6
25	1.6	0.3	1.2	1.4	1.5	1.8	2.1
30	2.1	0.4	1.6	1.8	2.0	2.3	2.9
35	2.2	0.5	1.7	1.8	2.2	2.6	3.0
40	2.8	0.9	1.9	2.2	2.5	3.0	4.5
45	3.5	1.2	2.0	2.5	3.4	4.5	5.4
50	3.6	0.5	3.0	3.3	3.6	3.9	4.5

Figure A5: HIFO vs Average Cost Harvesting with Recurring Investment. The gold line plots HIFO harvesting, where the tax basis of each lot is individually tracked while the blue line plots dollar-cost averaging, where the tax basis is calculated as the average of all lots. Panel A compares the losses harvested, both short-term and long-term. Panel B plots the overall tax position (including both gains and losses) in dollars. Relative to Figure A6, the investment is a recurring investment of $50,000 per quarter. The x-axis is portfolio age in years, and each line represents the median across vintages.

(a) Losses Harvested



(b) Total Tax

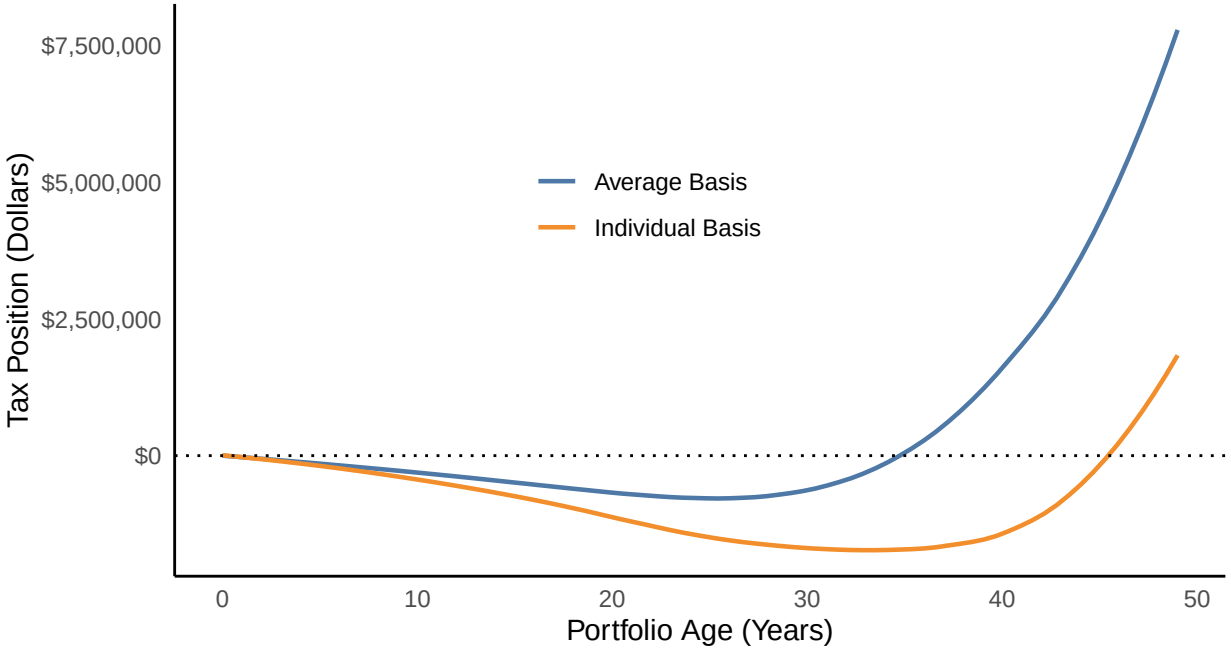
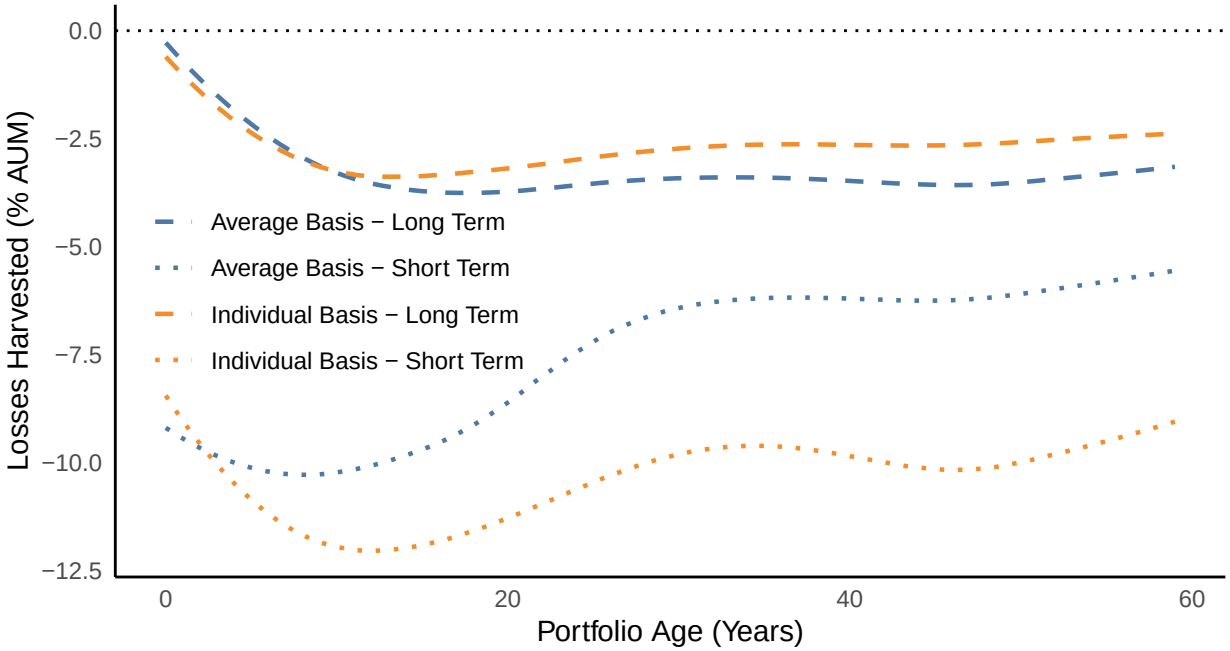


Figure A6: HIFO vs Average Cost Harvesting. The gold line plots HIFO harvesting, where the tax basis of each lot is individually tracked while the blue line plots dollar-cost averaging, where the tax basis is calculated as the average of all lots. Panel A compares the losses harvested, both short-term and long-term. Panel B plots the overall tax position (including both gains and losses) in dollars for a $50,000 initial investment. The x-axis is portfolio age in years, and each line represents the median across vintages.

(a) Losses Harvested



(b) Total Tax

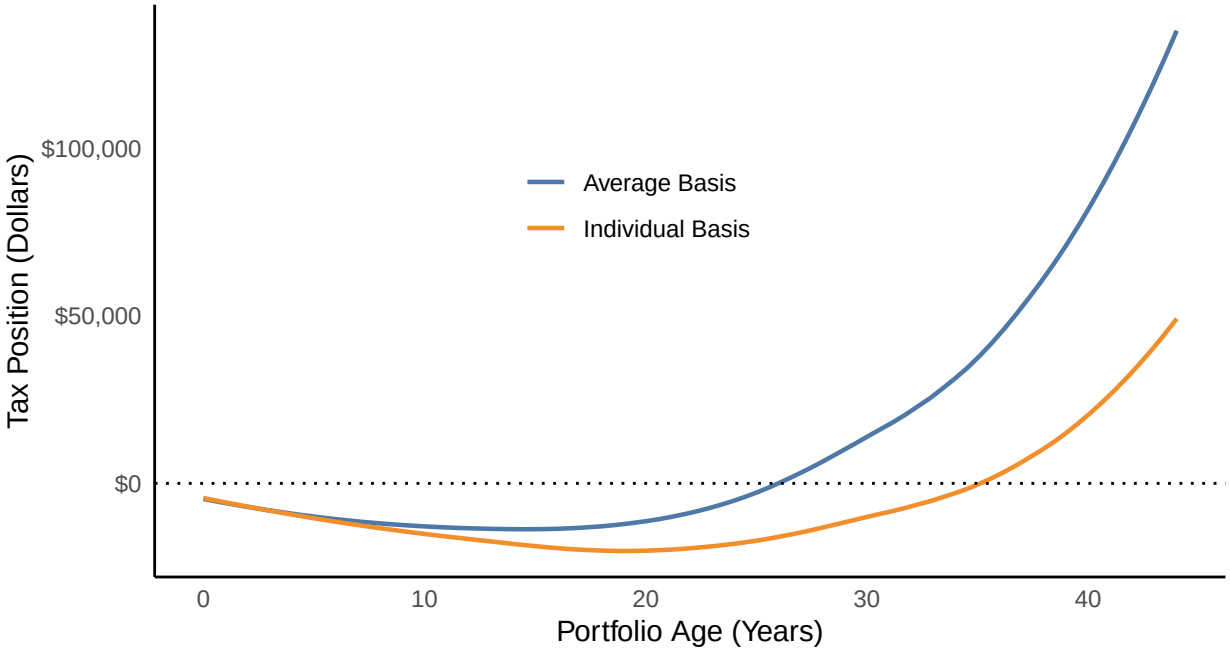
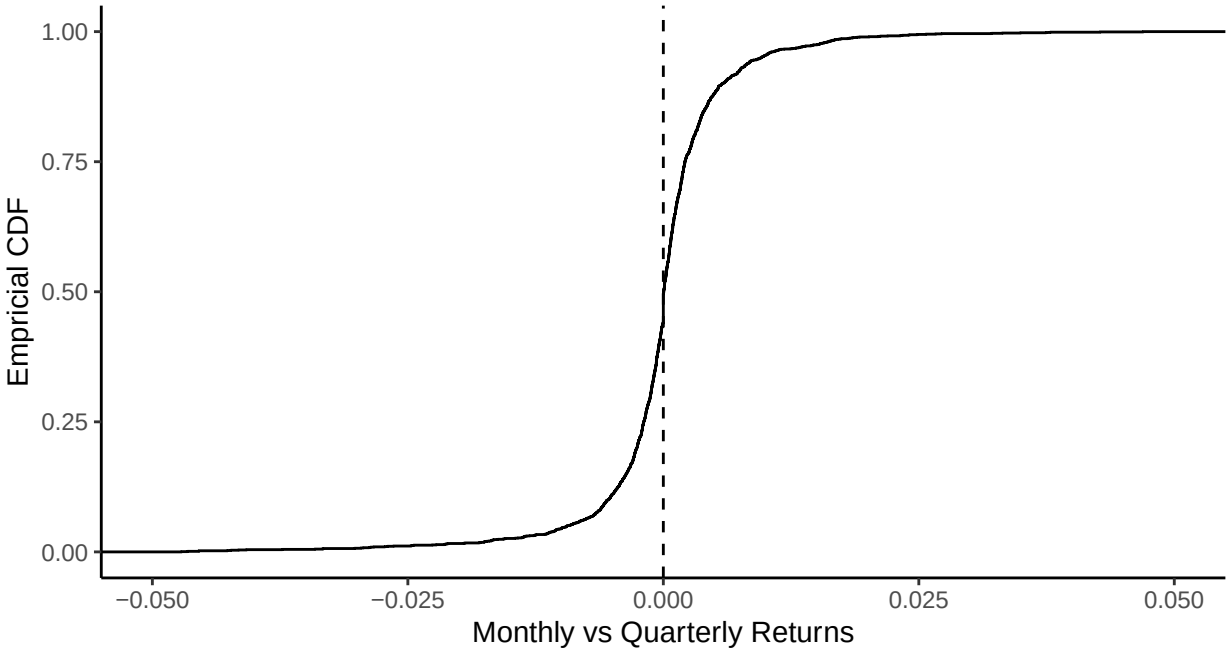
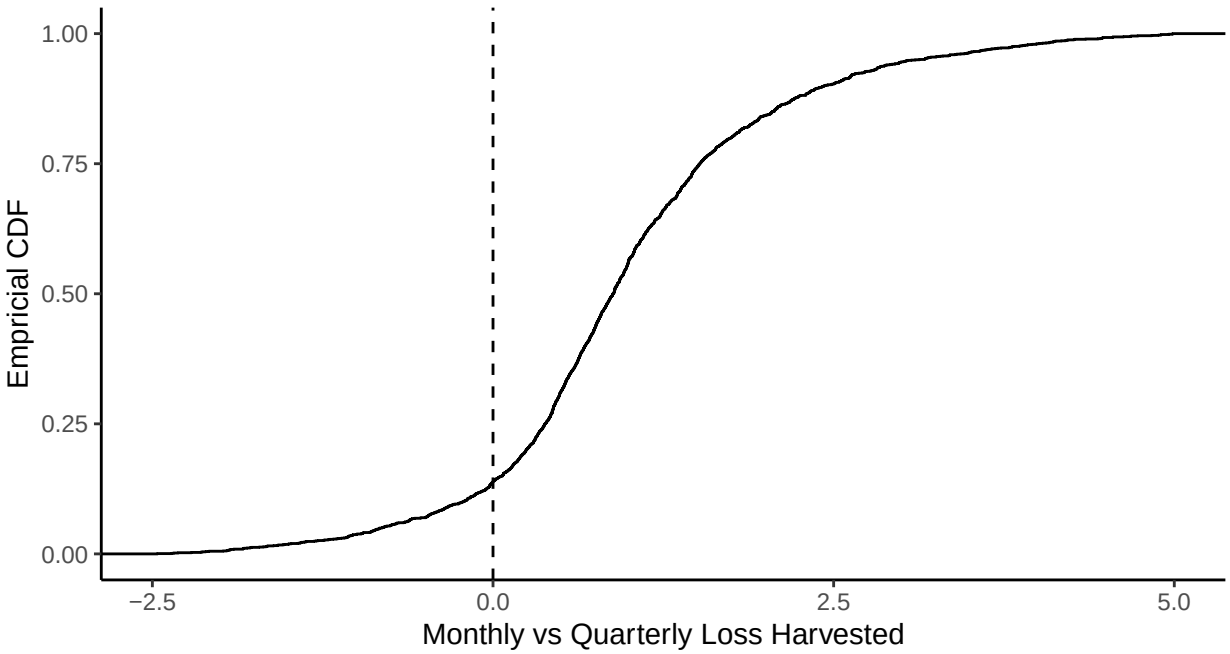


Figure A7: Rebalancing Frequency Comparsion. We compare portfolios which are rebalanced monthly against portfolios which are rebalanced quarterly. Panel A presents the annualized log return spread (a positive number would indicate a better monthly return). Panel B presents difference in tax losses harvested (as a percentage of portfolio value), with a positive number indicating more losses harvested by the monthly portfolio.

(a) Monthly Log Return Advantage Comparison



(b) Monthly Harvest Tax Advantage Comparison



Appendix E. Factor Sensitivity Analysis

There are cyclical patterns in the harvesting of losses (e.g., Figure A2 in Appendix B), with most vintages harvesting considerable losses in 1973, 2001, and 2008. However, the overall correlation with market returns remains low.

Table A8 presents the results of a regression of realized tax gains or losses on the 3 Fama French factors plus momentum. Short-term losses exhibit the highest sensitivity at 0.05, while long-term losses have a sensitivity of 0.01. These coefficients suggest that for every 100 basis point increase in market returns, realized losses increase by 5 and 1 basis points, respectively. This positive sensitivity implies that as the market trends upward, the increased dispersion and broader set of tax lots provide more opportunities to harvest specific losses. While the market factor shows this modest correlation, factors such as size and value do not, likely because broad market movements drive the systematic direction of individual security prices more uniformly than the characteristic-based premiums of style factors. Broadly, the performance of a direct indexing strategy in terms of loss harvesting is not particularly correlated with asset pricing factors.

Table A9 presents factor sensitivity based on portfolio age. For short-term losses, sensitivity to the market is 0.11 for young portfolios (less than 10 years) and a slightly lower 0.3 for older portfolios (40 to 50 years). For long-term losses, sensitivity to the market portfolio is 0.02 for younger portfolios and a slightly lower 0.01 for older portfolios.

Appendix F. Supplementary Results for the Real-World ETF Analysis

This appendix reports the supporting analyses referenced in Section VI. We provide annual sample coverage, define the benchmark objects, report implementation outcomes for the naive quarter-end replication, examine the role of ETF expenses, report additional results for the matched-period hypothetical top-500, and provide sensitivity and replication diagnostics.

Appendix A. Annual Sample Coverage

Appendix Table A10 reports the year-end fund counts and AUM. The annual counts include funds present at each calendar-year end, rather than all 652 funds that appear at some point in the final sample. The 53.3% AUM share reported in the main text is the simple average of the annual shares

Table A8: Factor sensitivty of realized gains and losses. We regress the realization of taxable gains or losses, as a fraction of portfolio value, on the Fama-French 3-Factor Portfolio plus momentum. Column 1 has short-term gains as the dependent variable, column 2 short-term losses, column 3 has long-term gains, and column 4 has long-term losses. We include a fixed effect for each vintage, starting investent, and quarterly investment. Standard errors are clustered by year-quarter and vintage.

Dependent Variables:	ST Gains	ST Losses	LT Gains	LT Losses
Model:	(1)	(2)	(3)	(4)
Variables				
Market (MKT-RF)	0.0002	0.0480***	-0.0008	0.0143***
	(0.0002)	(0.0075)	(0.0036)	(0.0026)
SMB	-0.0002	-0.0108	-0.0052	-0.0008
	(0.0003)	(0.0070)	(0.0037)	(0.0020)
HML	0.0001	0.0052	-0.0033	0.0059*
	(0.0002)	(0.0073)	(0.0043)	(0.0031)
Momentum	0.0002	0.0019	0.0013	-0.0004
	(0.0002)	(0.0047)	(0.0031)	(0.0013)
Fixed-effects				
Vintage	Yes	Yes	Yes	Yes
Initial investment	Yes	Yes	Yes	Yes
Quarterly investment	Yes	Yes	Yes	Yes
Fit statistics				
Observations	99,390	99,390	99,390	99,390
R^2	0.02434	0.27613	0.03386	0.31828
Within R^2	0.00232	0.23116	0.01633	0.30399

Clustered (Year-quarter & Vintage) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table A9: Factor sensitivty by portfolio age. We regress the realization of taxable gains or losses, as a fraction of portfolio value, on the Fama-French 3-Factor Portfolio plus momentum. Relative to tab:factor_sensitivity, Panel A uses portfolios less than 10 years old, while Panel B uses portfolios between 30 and 50 years old. Column 1 has short-term gains as the dependent variable, column 2 short-term losses, column 3 has long-term gains, and column 4 has long-term losses. We include a fixed effect for each vintage, starting investent, and quarterly investment. Standard errors are clustered by year-quarter and vintage.

Dependent Variables:	ST Gains	ST Losses	LT Gains	LT Losses
Model:	(1)	(2)	(3)	(4)
Panel A: Portfolio age < 10 years				
Variables				
Market (MKT-RF)	0.0003	0.1142***	0.0003	0.0266***
	(0.0003)	(0.0149)	(0.0016)	(0.0053)
SMB	-9.74×10^{-5}	-0.0181	-0.0035*	0.0017
	(0.0006)	(0.0181)	(0.0021)	(0.0050)
HML	-0.0002	0.0206	-0.0012	0.0149**
	(0.0004)	(0.0148)	(0.0019)	(0.0072)
mom	0.0005	-0.0007	0.0005	0.0012
	(0.0004)	(0.0104)	(0.0014)	(0.0032)
Fit statistics				
Observations	14,868	14,868	14,868	14,868
R^2	0.06683	0.38578	0.14576	0.32762
Panel B: Portfolio age 40–50 years				
Variables				
Market (MKT-RF)	0.0002*	0.0313***	-0.0010	0.0104***
	(0.0001)	(0.0051)	(0.0038)	(0.0019)
SMB	-0.0003	-0.0037	-0.0063	-0.0006
	(0.0003)	(0.0048)	(0.0041)	(0.0015)
HML	9.91×10^{-6}	0.0037	-0.0032	0.0034
	(0.0001)	(0.0050)	(0.0046)	(0.0021)
mom	1.86×10^{-5}	0.0039	0.0010	-0.0007
	(0.0002)	(0.0031)	(0.0033)	(0.0009)
Fit statistics				
Observations	27,330	27,330	27,330	27,330
R^2	0.05977	0.47452	0.08064	0.47226

Clustered (Year-quarter & Vintage) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

from 2011 through 2023.

Table A10: Annual Coverage of the Real-World ETF Replication Sample

Year	Matched ETFs	Final-sample ETFs	Final share of matched ETFs (percent)	Matched AUM ($ billions)	Final-sample AUM ($ billions)	Final share of matched AUM (percent)
2011	841	260	30.9	908.5	400.1	44.0
2012	884	269	30.4	1155.8	496.8	43.0
2013	948	265	28.0	1528.6	764.1	50.0
2014	1,070	289	27.0	1816.6	958.0	52.7
2015	1,244	321	25.8	1919.4	969.8	50.5
2016	1,367	365	26.7	2307.5	1232.4	53.4
2017	1,428	401	28.1	3120.0	1597.3	51.2
2018	1,552	497	32.0	3027.4	1629.3	53.8
2019	1,642	535	32.6	3925.6	2208.1	56.2
2020	1,711	558	32.6	4769.6	2735.4	57.4
2021	1,901	584	30.7	6327.9	3787.6	59.9
2022	1,944	587	30.2	5553.7	3337.2	60.1
2023	1,821	572	31.4	6908.6	4233.3	61.3

Notes: Fund counts and AUM are measured at each calendar-year end. AUM uses `tna_latest`. The final sample is the set of portfolios passing the paper filters and represented in the direct-index simulation. The simple average of the annual AUM shares is 53.3%.

Appendix B. Naive Quarter-End Replication

Appendix Table A11 reports the implementation differences behind Table VI. The naive strategy stays almost exactly on the reconstructed holdings return, while the tax-aware strategy accepts a larger return difference. The naive strategy achieves this fit by selling roughly 25% to 31% of AUM per year and realizing gains equal to about 5% to 6.5% of AUM per year. These trades also generate transaction costs. The table begins at three years because the first-year comparison is distorted by the timing of the annual tax settlement.

Table A11: Tracking and Implementation under Naive Quarter-End Replication

Horizon	N	Tax-aware return difference (percent)	Naive return difference (percent)	Naive annual sales (percent of AUM)	Naive realized gains (percent of AUM per year)	Naive transaction costs (basis points per year)
3 years	451	2.76	0.001	30.6	5.33	7.7
5 years	400	2.91	0.001	30.3	6.40	6.4
10 years	234	3.49	0.001	25.2	6.48	4.2

Notes: The table uses the common sample from Table VI, the no-personal-income-offset specification, a $50,000 initial investment, and $5,000 quarterly contributions. The return difference is the annualized root-mean-squared difference from the reconstructed quarter-end holdings return, not from the ETF's official index. Naive annual sales and realized gains are annualized percentages of average account value. Transaction costs include the same buy- and sell-side spread assumptions used by the tax-aware strategy.

Appendix C. ETF Expenses and Short-Horizon Performance

Appendix Table A12 examines how much of the short-horizon after-tax direct-index advantage reflects expenses already deducted from ETF returns. We construct an approximate after-tax ETF

return gross of its expense ratio by reversing one-fourth of the annual CRSP expense ratio each quarter while retaining estimated shareholder taxes. This removes the ETF's entire expense ratio, not only its index licensing fee, and should therefore be interpreted as a diagnostic of the expense channel.

Reversing ETF expenses reduces the mean advantage by approximately 24 to 35 basis points per year. The median one-year advantage falls from 29 basis points to approximately zero, while the median remains positive at longer horizons. Appendix Table A13 shows that the one-year result also varies substantially across entry years. The 2015 cohort has the largest positive mean advantage, while the 2020 cohort has the largest negative mean advantage. Removing ETF expenses reduces the advantage in every cohort but does not explain all of the variation across entry years.

Table A12: After-Tax Performance and ETF Expenses

Horizon	N	Mean advantage, after-tax ETF	Mean advantage, ETF gross of expenses	Mean expense-ratio effect	Median advantage, after-tax ETF	Median advantage, ETF gross of expenses
1 year	472	+85.1	+50.0	34.7	+28.7	+0.6
3 years	434	+63.0	+33.0	29.8	+21.3	-0.1
5 years	376	+60.6	+32.9	27.5	+35.9	+12.1
10 years	207	+63.2	+38.7	24.4	+49.1	+21.6

Notes: The table uses the no-personal-income-offset specification and the subset of ETFs with CRSP expense-ratio data. Entries are annualized basis points. The after-tax gross-of-expense benchmark reverses one-fourth of the annual CRSP expense ratio each quarter while retaining estimated shareholder taxes. This removes the ETF's entire expense ratio, not only its index licensing fee. The direct-index simulation does not include an analogous management or licensing fee.

Table A13: One-Year After-Tax Performance and ETF Expenses by Entry Year

Entry year	N	Mean advantage, after-tax ETF	Mean advantage, ETF gross of expenses	Mean expense-ratio effect	Median advantage, after-tax ETF	Median advantage, ETF gross of expenses	ETF return difference vs. reconstructed holdings (percent)
2011	209	+130.2	+96.2	33.4	+20.0	-6.9	1.74
2012	6	-16.8	-51.3	34.6	-45.4	-73.0	1.69
2013	2	+117.3	+91.6	25.4	+117.3	+91.6	0.94
2014	21	+47.8	+23.4	24.3	+67.6	+41.5	0.92
2015	29	+328.2	+265.0	61.2	+357.1	+310.8	3.48
2016	24	+88.0	+57.3	30.5	+29.9	+7.5	1.23
2017	24	+224.4	+190.6	33.1	+141.3	+116.4	1.17
2018	80	-92.9	-124.6	32.0	-72.6	-109.9	1.29
2019	34	+77.1	+41.5	35.3	+52.2	+16.4	2.12
2020	13	-232.3	-269.0	37.9	+24.9	-40.3	2.84
2021	13	+76.0	+34.7	41.5	+34.8	-15.2	3.69
2022	17	+95.6	+64.0	30.9	+58.8	+45.2	1.60

Notes: Advantages and expense-ratio effects are annualized basis points. The final column is the annualized root-mean-squared difference between the ETF's reported return and the return on its reconstructed quarterly holdings. It is not tracking error relative to the ETF's official index. Several entry-year samples are small, so the cross-cohort comparisons are descriptive.

Appendix D. Matched ETF-Era Top-500 Benchmark

Appendix Tables A14 and A15 report additional results for the hypothetical value-weighted top-500 used in the main text. This portfolio provides a simple, broad, relatively low-turnover benchmark of

the type studied elsewhere in the paper. It is not intended to represent the average ETF, many of which are sector, style, or other specialized portfolios.

To hold the sample period fixed, the hypothetical top-500 uses the same post-2011 start-year support available for the ETF sample at each horizon. The comparison therefore uses 12 start years at one year, 11 at three years, 9 at five years, and 4 at ten years. Loss generation is nearly identical across the two personal-income-offset specifications. The specifications can differ slightly because tax refunds are reinvested and can affect the subsequent portfolio path, but these differences are small.

Appendix Table A15 reports forced gains and the use of harvested losses. In both specifications, losses first offset gains generated inside the portfolio. The personal-income-offset specification additionally allows up to $3,000 of remaining annual losses to offset ordinary income. These tables provide the top-500 results underlying Figures 7 and 8 in the main text.

Table A14: Loss Generation in the Matched-Period Hypothetical Top-500

Horizon	Measure	N	Mean	SD	P5	P25	P50	P75	P95
Panel A: No personal-income offset									
1	Voluntary harvested losses	12	10.314	7.978	2.077	4.705	7.456	14.230	30.788
1	Forced losses	12	0.007	0.018	0.000	0.000	0.001	0.003	0.063
3	Voluntary harvested losses	11	5.269	2.323	2.212	3.670	4.330	7.936	9.454
3	Forced losses	11	0.020	0.035	0.000	0.001	0.003	0.036	0.116
5	Voluntary harvested losses	9	3.721	1.053	2.137	3.079	3.888	4.172	5.344
5	Forced losses	9	0.020	0.016	0.004	0.008	0.015	0.031	0.053
10	Voluntary harvested losses	4	2.516	0.380	2.031	2.233	2.553	2.799	2.926
10	Forced losses	4	0.015	0.000	0.015	0.015	0.015	0.016	0.016
Panel B: Personal-income offset allowed									
1	Voluntary harvested losses	12	10.314	7.978	2.077	4.705	7.456	14.230	30.788
1	Forced losses	12	0.007	0.018	0.000	0.000	0.001	0.003	0.063
3	Voluntary harvested losses	11	5.296	2.336	2.217	3.701	4.363	7.960	9.478
3	Forced losses	11	0.020	0.035	0.000	0.001	0.003	0.036	0.115
5	Voluntary harvested losses	9	3.750	1.060	2.160	3.109	3.903	4.224	5.381
5	Forced losses	9	0.020	0.016	0.004	0.008	0.015	0.031	0.053
10	Voluntary harvested losses	4	2.537	0.385	2.047	2.249	2.574	2.824	2.953
10	Forced losses	4	0.015	0.000	0.015	0.015	0.015	0.016	0.016

Notes: Entries are annualized percentages of average direct-index AUM. The hypothetical portfolio holds the value-weighted 500 largest eligible stocks. The initial investment is $50,000, quarterly contributions are $5,000, and the wash-sale restriction is imposed. At each horizon, the hypothetical portfolio uses the same start-year support as the real-world ETF sample. This gives 12 start years at one year, 11 at three years, 9 at five years, and 4 at ten years.

Table A15: Forced Realizations and Loss Use in the Matched-Period Hypothetical Top-500

Horizon	N	Cash-buyout gains	All forced gains	Losses used against forced gains	Ordinary-income loss use
Panel A: No personal-income offset					
1	12	0.055	0.175	0.175	0.000
3	11	0.129	0.418	0.418	0.000
5	9	0.120	0.459	0.459	0.000
10	4	0.130	0.465	0.465	0.000
Panel B: Personal-income offset allowed					
1	12	0.055	0.175	0.175	4.544
3	11	0.129	0.416	0.416	3.027
5	9	0.120	0.457	0.457	2.134
10	4	0.130	0.464	0.458	1.046

Notes: Entries are mean annualized percentages of average direct-index AUM. The hypothetical portfolio, investment amounts, wash-sale rule, and horizon-specific start-year support are the same as in Appendix Table A14. Losses first offset forced gains in both panels. Panel B additionally allows up to $3,000 of remaining annual losses to offset ordinary income.

Appendix E. Start-Year Effects and ETF Characteristics

Table A16 asks whether the cash and replication outcomes in the main text are associated more strongly with when the simulation begins or with the characteristics of the ETF. Start-year fixed effects capture differences across market periods. ETF characteristics include portfolio turnover, distance from market weights, concentration, factor exposure, and whether the fund is a broad or specialized ETF.

At five years, ETF characteristics explain 56.9% of the variation in retained-deletion weight, compared with 4.8% for start-year fixed effects. ETF characteristics also explain more of the variation in dividend coverage and total cash coverage. For the annualized root-mean-squared return difference between the direct-index holdings and the reconstructed ETF holdings, ETF characteristics explain 28.9%, start-year fixed effects explain 11.9%, and the combined specification explains 39.4%. These comparisons are descriptive.

Table A16: Start-Year Effects and ETF Characteristics at Five Years

Outcome	N	Start-year-only R^2	ETF-characteristics-only R^2	Combined R^2	Incremental ETF characteristics	Incremental start year
Retained-deletion weight	444	4.8	56.9	58.2	53.3	1.3
Wash-blocked weight	444	10.9	38.3	48.7	37.9	10.5
Dividend coverage of total purchase need	444	10.3	30.3	34.3	24.1	4.0
Dividend coverage of additions	444	11.2	33.0	37.8	26.6	4.8
Total cash coverage	444	12.1	37.4	41.3	29.2	4.0
Direct-index return difference vs. reconstructed holdings	444	11.9	28.9	39.4	27.6	10.6

Notes: Entries are percentage points of explained variation. Start-year fixed effects capture differences across entry periods. ETF characteristics include turnover, distance from market weights, concentration, factor exposure, and broad-versus-sector/style classification. The return-difference outcome is the annualized root-mean-squared difference between the pre-tax return on the direct-index holdings and the return on the reconstructed ETF holdings. The factor-exposure measures use all available ETF quarters. The decomposition is descriptive, not causal.

Appendix F. Sensitivity to the Personal-Income Offset

The current first-stage output fixes the initial investment at $50,000. Table A17 and Figure A8 therefore vary quarterly contributions while holding the initial investment fixed. Because the ordinary-income offset is capped at $3,000 per year, its value declines as the account becomes larger. At three years, for example, the mean incremental benefit falls from 100 basis points per year with no quarterly contributions to 10 basis points with $50,000 of quarterly contributions.

Both the table and figure measure the direct-index advantage relative to the after-tax ETF. They compare ongoing account wealth and exclude taxes that would arise from liquidating either portfolio at the horizon. Additional first-stage simulations with different initial investment amounts would be required to study initial-investment sensitivity directly.

Table A17: Sensitivity to the Personal-Income Offset and Contributions

Quarterly contribution	Horizon (years)	N	Median without personal-income offset (bps/year)	Median with personal-income offset (bps/year)	Mean incremental offset effect (bps/year)	Percent direct indexing wins without income offset	Percent direct indexing wins with income offset
0	1	526	+69.6	+69.6	+0.0	58.4	58.4
0	3	499	+29.5	+129.1	+99.5	58.9	77.4
0	5	445	+71.7	+143.4	+79.2	63.6	80.2
0	10	254	+94.6	+122.7	+40.9	77.6	83.9
5,000	1	526	+47.8	+47.8	+0.0	57.4	57.4
5,000	3	499	+21.7	+76.4	+52.8	59.9	74.1
5,000	5	445	+38.9	+83.2	+41.2	64.0	75.1
5,000	10	254	+45.5	+68.7	+24.6	70.1	79.9
50,000	1	526	+35.3	+35.3	+0.0	55.1	55.1
50,000	3	499	+23.0	+33.4	+10.2	61.1	64.5
50,000	5	445	+27.3	+33.9	+7.3	62.7	66.5
50,000	10	254	+29.2	+33.9	+4.0	66.9	68.9

Notes: The initial investment is $50,000 in every row because the first-stage output contains no other initial-investment values. All advantages are measured relative to the after-tax ETF. The results compare ongoing account wealth and exclude taxes on embedded gains that would be realized upon terminal liquidation.

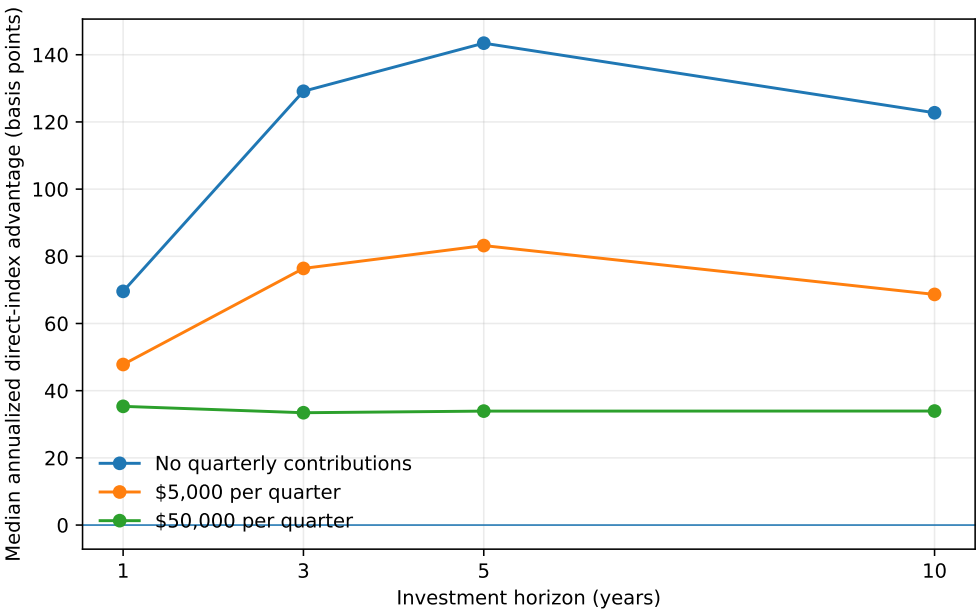


Figure A8: Contribution Sensitivity. The figure reports the median annualized direct-index advantage relative to the after-tax ETF under the personal-income-offset specification. All paths start with $50,000. Neither vehicle is liquidated at the horizon, so the figure excludes taxes on embedded gains that would be realized upon liquidation.

Appendix G. Cash Availability and Replication Frictions

Appendix Table A18 provides the cash and replication details behind the turnover results in Table IX. Dividend coverage is the cumulative after-tax dividend cash divided by the cumulative dollars needed to purchase underweight positions. At the median ETF, dividends cover 1.9% of this purchase need at one year and 1.3% at ten years.

Total cash coverage also includes contributions, sale proceeds, prior cash, and tax payments or refunds. It declines from 149% at one year to 31% at ten years, while the fraction of fully funded portfolios declines from 74% to 2%. As available cash becomes small relative to the purchase need, the strategy retains more stocks that have left the ETF's reported holdings. Median retained-deletion weight reaches 16.8% at five years.

The median annualized root-mean-squared difference between the pre-tax direct-index holdings and the reconstructed ETF holdings rises from 2.6% at one year to 3.5% at ten years. This measures how closely the direct-index portfolio replicates the return on the ETF's reported holdings. It is not tracking error relative to the ETF's official index.

Table A18: Cash Availability and Replication Frictions

Panel A: Cash availability

Horizon	N	Dividend coverage	Dividend coverage of additions	Total cash coverage	Percent fully funded
1 year	538	1.9	2.3	149.4	73.8
3 years	512	1.4	2.9	65.3	17.4
5 years	453	1.3	3.4	48.3	7.5
10 years	259	1.3	4.5	30.8	1.9

Panel B: Replication frictions

Horizon	Average retained-deletion weight	Maximum retained-deletion weight	Wash-blocked weight	Direct-index return difference	ETF return difference
1 year	9.5	15.5	17.8	2.6	1.1
3 years	13.4	23.9	17.2	2.9	1.3
5 years	16.8	29.8	20.0	3.0	1.3
10 years	16.4	31.3	20.3	3.5	1.1

Notes: Entries are medians or percentages. Dividend coverage is cumulative after-tax dividend cash divided by the cumulative dollars needed to purchase underweight positions. Total cash coverage also includes contributions, sale proceeds, prior cash, and tax payments or refunds. The direct-index return difference is the annualized root-mean-squared difference between the pre-tax return on the direct-index holdings and the return on the reconstructed ETF holdings. The ETF return difference is defined analogously using the ETF's reported return. Neither measure is tracking error relative to the ETF's official index.

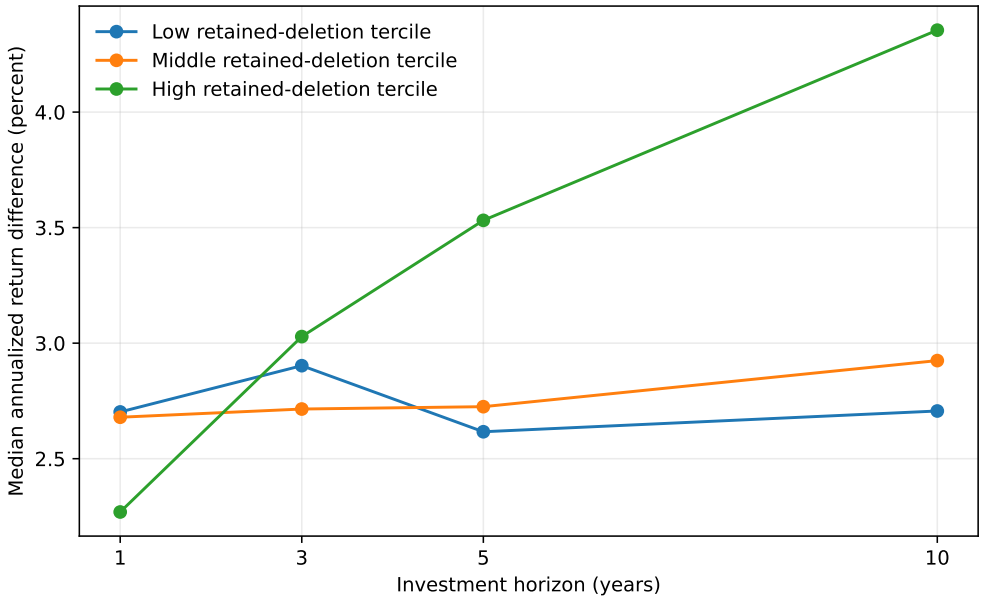


Figure A9: Retained Deletions and Return Differences. ETFs are sorted into terciles based on the average portfolio weight held in stocks that no longer appear in the target ETF holdings. The figure reports the median annualized root-mean-squared return difference between the pre-tax direct-index holdings and the reconstructed ETF holdings. A larger value means that the direct-index portfolio tracks the reconstructed holdings less closely. This measure is not tracking error relative to the ETF's official index.

REFERENCES

Samuel W Adams, Connor Kasten, and Eric K Kelley. How free is free? retail trading costs with zero commissions. *Journal of Banking & Finance*, 165:107226, 2024.

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

James J Angel, Lawrence E Harris, and Chester S Spatt. Equity trading in the 21st century. *The Quarterly Journal of Finance*, 1(01):1–53, 2011.

Ian R Appel, Todd A Gormley, and Donald B Keim. Passive investors, not passive owners. *Journal of Financial Economics*, 121(1):111–141, 2016.

Robert D Arnott, Christopher Brightman, Vitali Kalesnik, and Lillian Wu. Earning alpha by avoiding the index rebalancing crowd. *Financial Analysts Journal*, 79(2):76–97, 2023.

Brad M Barber, Xing Huang, and Terrance Odean. Which factors matter to investors? evidence from mutual fund flows. *The Review of Financial Studies*, 29(10):2600–2642, 2016.

Judith Chevalier and Glenn Ellison. Risk taking by mutual funds as a response to incentives. *Journal of political economy*, 105(6):1167–1200, 1997.

George M Constantinides. Capital market equilibrium with personal tax. *Econometrica: Journal of the Econometric Society*, pages 611–636, 1983.

Min Dai, Hong Liu, Chen Yang, and Yifei Zhong. Optimal tax timing with asymmetric long-term/short-term capital gains tax. *The Review of Financial Studies*, 28(9):2687–2721, 2015.

Robert M Dammon, Chester S Spatt, and Harold H Zhang. Optimal consumption and investment with capital gains taxes. *The Review of Financial Studies*, 14(3):583–616, 2001.

Robert M Dammon, Chester S Spatt, and Harold H Zhang. Optimal asset location and allocation with taxable and tax-deferred investing. *The Journal of Finance*, 59(3):999–1037, 2004.

Kent Daniel, Mark Grinblatt, Sheridan Titman, and Russ Wermers. Measuring mutual fund performance with characteristic-based benchmarks. *The Journal of finance*, 52(3):1035–1058, 1997.

Victor DeMiguel and Raman Uppal. Portfolio investment with the exact tax basis via nonlinear programming. *Management Science*, 51(2):277–290, 2005.

Paul Ehling, Michael Gallmeyer, Sanjay Srivastava, Stathis Tompaidis, and Chunyu Yang. Portfolio tax trading with carryover losses. *Management Science*, 64(9):4156–4176, 2018.

Marcel Fischer and Michael F Gallmeyer. Heuristic portfolio trading rules with capital gain taxes. *Journal of Financial Economics*, 119(3):611–625, 2016.

Dong Lou. A flow-based explanation for return predictability. *The Review of Financial Studies*, 25 (12):3457–3489, 2012.

Marcel Marekwica. Optimal tax-timing and asset allocation when tax rebates on capital losses are limited. *Journal of Banking & Finance*, 36(7):2048–2063, 2012.

Rabih Moussawi, Ke Shen, and Raisa Velthuis. The role of taxes in the rise of ETFs. *The Review of Financial Studies*, 38(10):2988–3039, 2025. doi: 10.1093/rfs/hhaf044.

Chris Murray and Marco Sammon. Primary capital market transactions and index funds. *Review of Asset Pricing Studies*, 16(2):163–202, 2026. doi: 10.1093/rapstu/raaf013.

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

Jeffrey Pontiff and Artemiza Woodgate. Share issuance and cross-sectional returns. *The Journal of Finance*, 63(2):921–945, 2008.

James M Poterba and John B Shoven. Exchange-traded funds: A new investment option for taxable investors. *American Economic Review*, 92(2):422–427, 2002.

Marco Sammon and John J Shim. Index rebalancing and stock market composition: Do indexes time the market? *Journal of Financial Economics*, 177:104229, 2026a. doi: 10.1016/j.jfineco.2025.104229.

Marco Sammon and John J Shim. Index rebalancing and stock market composition: Do indexes time the market? *Journal of Financial Economics*, 177:104229, 2026b. doi: 10.1016/j.jfineco.2025.104229.

Robert J. Shiller. Online data: U.s. stock prices, dividends, earnings, and the cyclically adjusted price-earnings ratio (cape). `https://shillerdata.com/`, 2026. Accessed July 2026.

George Jiaguo Wang, Chelsea Yaqiong Yao, and Adina Yelekenova. Etf rebalancing, hedge fund trades, and capital market. *Available at SSRN 4324054*, 2023.