JOIM
www.joim.com

Journal Of Investment Management, Vol. 21, No. 4, (2023), pp. 15–31

REIMAGINING INDEX FUNDS

Rob Arnott[a], *Chris Brightman*[b], *Xi Liu*[c] *and Que Nguyen*[d]

"Gold-standard" cap-weighted indices have a buy-high and sell-low dynamic that causes a structural long-term performance drag. Of course, relative to itself, no index can underperform, which is the reason it goes unnoticed. If we use a company's fundamentals to choose stocks—and then cap-weights them—improves the risk-adjusted returns of gold-standard cap-weighted indices. This index, which we call Fundamental-selection Cap-weighted (FS-CW), has outperformed the most popular cap-weighted equity indices around the world over the last 30 years, while reducing risk, and with additional benefits of slightly lower turnover and transaction costs. Live results further support its merits. Building a better index fund that can earn a superior risk-adjusted return versus other cap-weighted indices is not only possible—it is a reality!

1 Introduction

Capitalization-weighted indices have been the indexing gold standard since the launch of the S&P 500 Index in 1957. Because the *market* is cap-weighted, why would we want to weight a market index any other way? In the 1960s, finance theory doubled down on this choice, not once but twice. The efficient market hypothesis posits that preferring one stock over another is a waste of time because all assets are correctly priced and will deliver the same risk-adjusted return. The capital asset pricing model, or CAPM, represents that the market portfolio is unbeatable on a risk-adjusted basis, assuming an array of qualifiers such as no taxes, no trading costs, efficient markets, general agreement across all investors on the expected risk, and reward of each asset.

Seizing on the zeitgeist of those times, the first cap-weighted index funds were introduced in 1972 and 1973 by Dean LeBaron at Batterymarch Financial, Rex Sinquefield at American National Bank, and Bill Fouse at Wells Fargo. Jack Bogle offered the first retail index fund in 1976 with the launch of the Vanguard Group. Since then, index funds have blossomed, a pattern that continues today. In 2019, for the first time, assets under

LLC, 620 Newport Center Drive, Newport Beach, CA 92660 USA.
[a]E-mail: arnott@rallc.com
[b]E-mail: brightman@rallc.com
[c]E-mail: liu@rallc.com
[d]E-mail: nguyen@rallc.com

management in passive US equity mutual funds exceeded AUM in actively managed US equity mutual funds. Few can challenge the notion that indexing and index funds rank with the invention of mutual funds as one of the most important innovations in finance over the last 150 years.

Countless articles have explored a plethora of ways to "beat the market." Few have explored the possibility of a *better-performing* broad-market-*capitalization-weighted index*. With indexing now a core holding in many investors' portfolios, we are pleased to propose such an index.

2 Cap-Weighted Indexing Has Two Achilles' Heels

"Gold-standard" cap-weighted indices have not one, but *two*, Achilles' heels because the approach

(1) inherently overweights overvalued companies and underweights undervalued companies relative to an (unknowable!) fair value-weighted portfolio; and
(2) typically adds high-flying growth stocks after a recent period of outperformance and drops feared and loathed deep-discount value stocks after a recent period of underperformance.

The first weakness is a truism if we define overvalued and undervalued companies based on their (unknowable) future risk-adjusted IRR. Indexing advocates generally accept that this is the case, but sensibly point out that it is a useless argument unless we know—in advance—which stocks are overvalued and which are undervalued.

We have pointed out for years that this counter-argument is not quite true (Arnott *et al.*, 2005; Treynor, 2005) by proposing Market-Valuation-Indifferent (MVI) indexing as solutions. If the weight of a stock in our portfolio is not tied to its price or its market capitalization, then a company's weight in our portfolio may be overweight or underweight *independent of whether the stock is over- or undervalued*. In this case, errors in a stock's portfolio weight and in its price are not linked, so the errors largely cancel. Cap-weighted indexing has a performance drag, as pointed out by Hsu (2006), relative to its opportunity set and is therefore beatable on a risk-adjusted basis, for any market where price equals a random-walk fair value, plus or minus an error term. It disappears only in an implausible world in which market efficiency means the opposite. Fundamental index has not been without its controversies, most centered around this implausible definition of market efficiency.

As Jeremy Siegel observed in a *Wall Street Journal* op-ed in 2006, there is a drag associated with tying the portfolio weight of an asset to its share price, as cap-weighting will do.[1] One of the best discussions of the Fundamental index controversy, exploring views from both sides, can be found in Dempsey (2012) and his assorted references. By adopting a similar model as Hsu (2006), with some updated assumptions, Dempsey disputed a structural drag due to noise on cap-weighted portfolios, but confirmed that noise in price offers potential for improved portfolio performance.

Nevertheless, we are not trying to address the first weakness, in this article, as we retain capitalization weighting in our proposed solution. We focus instead on the second Achilles' heel, namely additions and deletions—security selection—to mitigate this drag on performance.

Cap-weighted indexes are not purely passive. Companies come and go, and some companies may be issuing new stock or buying back stock, so index providers must change the index portfolio, rebalancing from time to time. Also, publishers of commercially available cap-weighted indices typically limit the number of names in an index to improve investability, adding

and deleting stocks from the index over time. Additions are usually growth stocks, trading at premium multiples and with impressive momentum, whereas most deletions—unless occurring because of a merger or acquisition—have the opposite characteristics.[2] Some index additions go on to great success, exceeding all expectations; most do not. Some deletions proceed to stock market oblivion, and others handily recover. We even see flip-flops: stocks added then quickly dropped from the index, and vice versa. Collectively, additions and deletions encompass most of the turnover in a cap-weighted index. Although small in magnitude, some of this turnover can be harmful to our wealth. Buying mainly frothy stocks with strong momentum and selling mainly tumbling stocks that are severely out of favor creates an unhelpful buy-high and sell-low dynamic in a cap-weighted index.

3 Is There a Better Way?

What if we no longer chase soaring winners and abandon tumbling losers, and instead choose stocks based on a more stable metric, namely, the size of the underlying business? Suppose we select stocks based on their economic scale, their relative size in the macroeconomy, rather than on the market's expectation of the company's future success. This strategy will come very close to matching the portfolio held by existing cap-weighted indices. Big businesses are usually large-cap and small businesses are usually small-cap. Thus, if we select stocks based on the economic scale of the underlying business rather than on market-cap, the result is a portfolio with superb liquidity and capacity, fully comparable to popular and commercially available market-cap indices.

We propose the creation of a broad-market capitalization-weighted index by selecting the constituents using fundamental measures of the size of the underlying company, and cap-weighting them. We call this Fundamental-selection Cap-weighted (FS-CW) index. Instead of cap-weighting the largest market-cap stocks, we would be cap-weighting the largest businesses. Additions will be companies that have grown onto the list of the largest businesses, important enough in the macroeconomy to matter, instead of stocks that have soared onto the list of the most popular companies. Deletions will be companies that have diminished in macroeconomic scale, by enough to no longer matter, instead of unloved stocks that have tumbled off the list of the largest market-cap stocks. For illustrative purposes, let's look closely at two versions of the index: FS-CW 500 and FS-CW 1000.

The differences between FS-CW and two popular standard-bearers in indexing, the S&P 500 and Russell 1000, illuminate why selecting constituents by fundamentals has the potential to outperform them. The FS-CW methodology excludes some S&P 500 companies if they are small businesses trading on lofty future growth expectations[3] and retains some companies with a substantial economic footprint whose share prices push them out of the large-cap universe. But won't we embrace growth stocks and drop "value traps" too late, hurting FS-CW relative performance? In an efficient market, the answer is no, because the high growth prospects of the former and the headwinds of the latter *are already fully discounted in the share price.* So, in an efficient market, FS-CW will match the returns of the standard-bearers, give or take a modest random tracking error. In a market that is *not* entirely efficient, these constituent differences bode favorably for fundamental selection. If a stock is mispriced, selection based on market-cap introduces a slightly disproportionate bias that will include some overpriced stocks and exclude some underpriced stocks. As the market seeks to correct the mispricing, this mean-reverting error

should create costly flip-flops in the stocks held by a traditional market-cap-weighted index. FS-CW can often avoid these flip-flops.

4 Indexing and Fama–French "Migration"

In 2007, Eugene Fama and Ken French introduced the term "migration" to the indexing lexicon. We expanded on this work in a 2012 paper entitled "Rebalancing and the Value Effect" (Chaves and Arnott, 2012). Fama and French (2007) primarily focus their attention on the effect of "migration" on value and growth portfolios, and on large- and small-cap portfolios. Value stocks and small-cap stocks see benefit from migration, as stocks trading at neutral book-to-price ratios are kicked out of the value index, to be replaced with new high B/P stocks, boosting the book value of the value portfolio with almost every reconstitution of the index. The opposite mechanism trims the book value of the growth portfolio with almost every reconstitution. *The same effect can be observed in broad-market index funds, which is largely unacknowledged.*

A deeper dive is useful. A growth index consists of stocks trading at high valuation multiples, whether price/book, or price/earnings, or some other measures. For Fama–French value, the metric of choice is (for better or worse) book-to-market.[4] Their growth stocks are low book/market, which is equivalent to the practitioners' high price/book stocks. These stocks are typically richly priced for a reason, whether rapid growth, exciting product, path-breaking innovation, insightful management with a finger on the pulse of the client marketplace, or some other tailwinds. The good news is well-known, is already *in the price*, and so a growth stock will not help our performance unless the company performs even better than already-lofty expectations (or, on a shorter-term basis, unless expectations rise even higher). If expectations for a growth stock falter, it will be "demoted" out of the growth index, and replaced with a new highflier. For the Fama–French growth portfolio, this means a high book/market stock is replaced with a low book/market stock. While there are exceptions to this pattern, the book value of the value portfolio falls with almost every portfolio reconstitution.

Can we quantify the costs of "migration" as it relates to index funds? Yes, we can. In Arnott *et al.* (2023), we find that additions to the S&P trade at an average valuation premium—equally weighting the relative price/book, price/earnings, price/cash flow, price/sales, and price/dividends—of 1.92 times the market multiple, or a 92% premium. The S&P 500 has averaged 4.4% turnover since 1989. The impact on valuation multiples is not as straightforward as the product of these two numbers.[5] But, we find that the S&P 500 gets 2.2% more richly priced per year, on average, just due to the very limited turnover that takes place.

Unless markets are truly efficient, and the premium priced stocks are worth every penny of premium that they cost, this rebalancing from cheap stocks into expensive stocks will hurt us. All we are doing with FS-CW is choosing a rebalancing and reconstitution algorithm that mitigates this pattern, which—empirically and unsurprisingly—improves index performance by a margin that, in the indexing world, is substantial.

4.1 Data

We download S&P 500 and Russell 1000 indices' total returns from Factset and use Compustat and CRSP to simulate the True CW and FS-CW with the indices' descriptions as follows:

- **S&P 500** and **Russell 1000** use proprietary float adjustment in weighting constituent stocks. CRSP does not provide free float data,

so we use the raw cap-weight without the free-float adjustment. The performance difference between the two methods is less than five basis points (bps) annually.[6]

- **True CW 500** selects the 500 largest market-cap US stocks on June 30 of each year and cap-weights them. We call this "True CW (cap-weight)" because the index's construction is utterly formulaic and includes exactly the 500 largest market-cap US-domiciled stocks (i.e., a committee does not choose additions and deletions).
- **Fundamental Selection (FS) CW 500 or 1000** selects the stocks of the 500 (or 1000) largest US companies, based on a blend of fundamental measures of company size, and cap-weights them. The four fundamental measures are—current book value adjusted for intangibles, five-year-trailing-average sales adjusted for the company's equity-to-asset ratio, five-year-trailing-average cash flow plus the company's R&D expenses, and past five-year-trailing-average dividends plus share repurchases. We average the four metrics, each measured as a percentage of all publicly traded US-headquartered companies, for each stock. The 500 or 1000 largest stocks are selected for FS-CW. The portfolio is rebalanced annually in March to capture the large number of updated annual financial statements.

The size, value, and momentum factor returns and risk-free returns are from the Kenneth R French Data Library. Other cap-weighted equity indices data in this paper are from Bloomberg, Factset, MSCI, S&P Global, and FTSE Russell.

4.2 Empirical findings

Markets have periods of comparative tumult and comparative stability. The performance difference (or tracking error) between any two cap-weighted indices with different constituent stocks, and their respective ability to add value, should vary, roughly, in parallel. If markets are perfectly efficient, the choice of market capitalization or fundamentals to select constituents would be irrelevant because market-cap should correctly value a company's future business prospects. If markets are inefficient, pricing errors will exist. Assuming the pricing errors mean revert, then selecting index constituents based on their fundamentals should add value by largely avoiding the buy-high/sell-low dynamics inherent in selecting index constituent stocks by market-cap. During turbulent times, when mean-reverting pricing errors are presumably larger, FS-CW should exhibit higher tracking error against conventional cap-weighted indices and capture greater incremental returns as these outsized pricing errors mean revert. *This is exactly what we see in the historical data.*

5 Historical Performance Over Different Tracking-Error Periods

To test this hypothesis, we examine the trailing 12-month tracking error (TE) of the Russell 1000 and three alternative cap-weighted indices: True CW 500, Fundamental Selection FS-CW 500, and FS-CW 1000, against the S&P 500 from June 1992 through December 2022 in Figure 1. The S&P 500 is our baseline for measuring TE and value-add.

It may come as a surprise to our readers (it did for us!) that the tracking errors of the Russell 1000, True CW 500, and FS-CW 500 relative to the S&P 500 are essentially identical, with each TE rounding to 1.0%. Low tracking-error periods and high tracking-error periods alternate. During turbulent times, the dispersion in stock returns tends to soar. Rising return dispersion magnifies differences in index methodology (e.g., discretionary selection, market impact of announcement,[7] and different reconstitution months) that result in differing

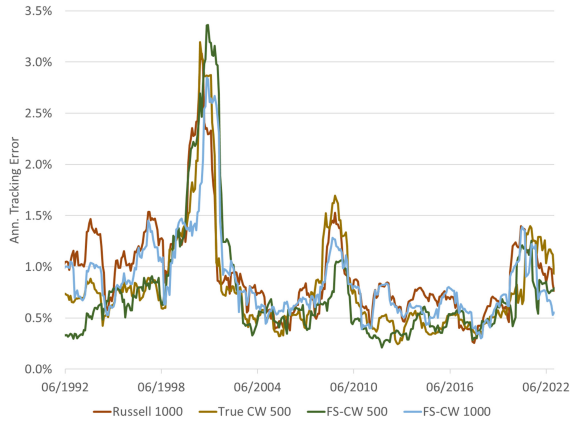


Figure 1 Trailing 12-month tracking error vs. S&P 500 for four US cap-weighted indices, Apr 1992–Dec 2022.

portfolio constituents. Consequently, index performance differences are magnified and tracking error spikes.

Three spans of market turbulence and high dynamic tracking error clearly stand out over our sample period. The most obvious time of turbulence is between September 1999 and March 2002, largely overlapping with the tech bubble and the subsequent tech crash. The global financial crisis (GFC) and COVID crisis were also periods of considerable turbulence, albeit less severe than during the tech bubble and crash. Over our sample period, the highest tracking error versus the S&P 500 for all four indices occurred during the tech bubble and its aftermath. The peak tracking error ranged from 2.8% for FS-CW 1000 to 3.4% for FS-CW 500. Keep in mind that all of these are cap-weighted indices, the only differences among them are the selection of stocks each holds.

During the GFC, the Russell 1000 and True CW 500 had the highest peak tracking errors of 1.5% and 1.7%, respectively. *FS-CW 500, with constituents selected based on the fundamental size of their underlying business, had the lowest peak tracking error during the GFC of just 1.1% versus the S&P 500!* Isn't it amazing that using a stable anchor, such as fundamental size, to select index constituents delivered the lowest tracking error compared to conventional cap-weighted indices during the turbulence of major market events?

We observe moderate tracking error of 1.3–1.4% for each of the four indices against the S&P 500 during the COVID-related turbulence of 2020–2022. The magnitude and timespan of the indices' TE versus the S&P 500 during COVID is smaller and shorter, respectively, compared with the two other turbulence spans we observe during the sample period.

Given the dynamic nature of tracking error, an investor's fear of missing out seems sensible only if, during high TE periods, the benchmark (in this case, the S&P 500) delivers outstanding performance. The empirical evidence, however, suggests otherwise: the benchmark typically does not deliver outstanding performance in high TE periods.

To determine if relative performance differences between the cap-weighted indices are present during high and low TE periods, we divide the full sample (July 1991 through December 2022) into three categories. The categories are based on the average 12-month tracking error of the four indices against the S&P 500: >2% (high), 1–2% (medium), and <1% (low). Transitional spikes or troughs in tracking error of six months or less remain with their adjacent tracking error episodes. We define high, medium, and low TE periods as follows:

High TE period. Average TE 2% or more, September 1999 to March 2002, which was the tech bubble and its aftermath.

Medium TE periods. Average TE from 1% to 2%, three spans: June 1996 to August 1999, October 2007 to January 2010, and September 2019 to October 2021.

Low TE periods. Average TE 1% or lower, four spans: July 1991 to May 1996, April 2002 to September 2007, February 2010 to August 2019, and November 2021 to December 2022.

In Table 1, we show annualized performance characteristics and tracking errors for the Russell 1000, True CW 500, FS-CW 500, and FS-CW 1000 indices against the S&P 500 in the high,

Table 1 Performance of Russell 1000, True CW, FS-CW US 500, and FS-CW US 1000 indices vs. S&P 500 Index during periods of high, medium, and low tracking error, Jul 1991–Dec 2022.

	S&P 500	Russell 1000	True CW 500	FS-CW 500	FS-CW 1000
Full sample (7/1991–12/2022)					
Returns	9.88%	10.00%	9.95%	10.35%	10.40%
Volatility	14.86%	15.05%	14.96%	14.58%	14.74%
Excess returns		0.11%	0.06%	0.46%∗	0.52%∗
Tracking error		1.05%	0.97%	0.97%	1.00%
Information ratio		0.11	0.07	0.48	0.52
CAPM alpha		0.05%	0.03%	0.55%∗∗	0.54%∗∗
CAPM residual volatility		1.04%	0.97%	0.92%	0.99%
High TE period (9/1999–3/2002)					
Returns	−4.11%	−3.26%	−4.84%	−1.92%	−0.95%
Excess returns		0.85%	−0.73%	2.19%	3.16%∗
Tracking error		2.00%	2.15%	2.56%	2.02%
Information ratio		0.42	−0.34	0.85	1.57
CAPM alpha		1.13%	−0.10%	1.34%	2.57%∗
CAPM beta		1.02	1.06∗∗	0.91∗∗	0.93∗∗
CAPM residual volatility		2.00%	1.88%	2.03%	1.68%
Medium TE period (6/1996–8/1999, 10/2007–1/2010, 9/2019–10/2021)					
Returns	12.69%	12.51%	13.33%	13.39%	12.91%
Excess returns		−0.17%	0.64%	0.71%	0.23%
Tracking error		1.20%	1.10%	0.93%	1.14%
Information ratio		−0.14	0.58	0.76	0.20
CAPM alpha		−0.27%	0.59%	0.75%∗	0.22%
CAPM residual volatility		1.16%	1.11%	0.89%	1.15%
Low TE periods (7/1991–5/1996, 4/2002–9/2007, 2/2010–8/2019, 11/2021–12/2022)					
Returns	10.70%	10.81%	10.67%	10.83%	10.96%
Excess returns		0.11%	−0.03%	0.13%	0.26%
Tracking error		0.79%	0.63%	0.52%	0.69%
Information ratio		0.14	−0.05	0.26	0.37
CAPM alpha		0.08%	0.00%	0.18%	0.25%
CAPM residual volatility		0.79%	0.63%	0.52%	0.69%

Notes: All numbers are annualized. Returns are geometric returns. Excess return is calculated as the difference between the index return and the S&P 500. For the CAPM regression, we use the S&P 500 return to proxy the market factor. For the excess return and CAPM alpha, * indicates significance at the 5% level and ** indicates significance at the 1% level. For the high TE period, we provide the CAPM beta and indicate the significance level of the difference from 1.0. For all other periods (data not shown here), CAPM betas for the different market-cap-weighted indices are all close to 1.0. Please refer to Section 4.1 for the definitions of each strategy.

medium, and low TE periods. Over the full sample period, all four cap-weighted indices have nearly identical TE to the S&P 500 of close to 1.0%. The indices' annualized excess returns vary from a low of 6 bps for True CW 500 to a high of 52 bps for FS-CW 1000. FS-CW 500 earned an excess return of 46 bps. The information ratios of the indices range from 0.07 for True CW 500 to roughly 0.50 for both FS-CW indices. The annualized beta-adjusted CAPM alphas for the two FS-CW indices are very similar (with FS-CW 500 at 55 bps and FS-CW 1000 at 54 bps) and significant at the 1% level. Both Russell 1000 and True CW 500 had a CAPM alpha close to zero.

During the highest TE period—the dot-com bubble and its aftermath—the Russell 1000 and both FS-CW indices beat the S&P 500 by a meaningful margin, whereas the performance of the True CW 500 fell well short of the return of the S&P 500. The poor showing of the True CW 500 is a direct consequence of adding dozens of frothy tech names in June 2000 just as they were beginning to crash, and vividly illustrates the pitfalls of selecting stocks based on their most recent market-cap. S&P's Index Committee was more measured in their decision-making and was wary of adding stocks with no history of earning profits.

Despite high tracking errors across all indices, ranging from 2.0% for the Russell 1000 to 2.6% for FS-CW 500, the FS-CW 500 (0.85) and FS-CW 1000 (1.57) indices had much higher information ratios than the Russell 1000 (0.42) and True CW 500 (−0.34) indices. During this highly volatile period, the FS-CW indices had statistically significantly lower betas (0.91 and 0.93 for the FS-CW 500 and 1000 indices, respectively) than the other indices. Many of the pricey high-beta dot-com stocks never made it into the FS-CW indices, because their businesses were not large enough to make the cut, even if their market-cap was substantial. FS-CW 500 and 1000 preserve much of their outperformance even after adjusting for their betas. FS-CW 1000 earned a CAPM alpha of 2.57%—statistically significant at 5%, despite the short span—during this extreme market event.

During the medium TE periods, including the GFC and Covid crash, the four indices had tracking errors that ranged from 0.9% for the FS-CW 500 to 1.2% for the Russell 1000. FS-CW 500 led the way in relative returns, posting an average excess return of 71 bps, followed by True CW 500 at 64 bps. FS-CW 500 also earned a statistically significant CAPM alpha of 0.75%. During the moderate TE spans, the Russell 1000 underperformed the S&P 500 by an average of 17 bps. The underperformance of the cap-weighted indices holding 1000 names, versus those with 500 stocks, is largely due to the underperformance of small caps from 1996 to 1998. Even with this headwind, FS-CW 1000 outperformed the S&P 500 by 23 bps.

The periods of low tracking error (<1%) are predominantly growth-dominated bull markets. The performance differences of the four indices versus the S&P 500 benchmark are small. FS-CW 500 and 1000 add 13 and 26 bps, respectively, of incremental return during these quiescent spans in

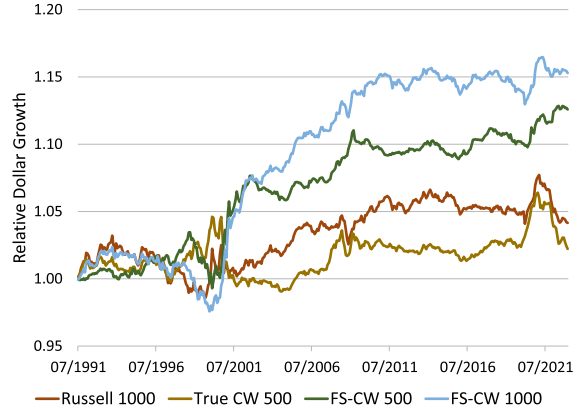


Figure 2 Relative dollar growth of four US cap-weighted indices vs. S&P 500 Index, Jul 1991–Dec 2022.

Table 2 Performance attribution of US cap-weighted indices using Fama–French–Carhart four-factor regression, Jul 1991–Dec 2022.

	S&P 500	Russell 1000	True CW 500	FS-CW 500	FS-CW 1000
Alpha	0.0%	−0.02%	0.07%	0.40%**	0.37%*
Alpha *t*-stat	—	−0.14	0.50	2.83	2.33
Market^{+}	1.00	1.01**	1.01	0.98**	0.99**
Size	0.00	0.06**	0.01	−0.02**	0.03**
Value	0.00	0.00	−0.04**	0.04**	0.04**
Momentum	0.00	0.00	0.01**	0.01*	0.00

Notes: The S&P 500 (minus the **3-month** T-bill return) is used as the market^{+} factor in the Fama–French–Carhart Four-Factor (FFC4) regression, and the *t*-statistic for the market is calculated as the difference from 1.00. Alpha is annualized. * Indicates significance at the 5% level and ** indicates significance at the 1% level. Please refer to the note under the first graph for the definitions of each strategy.

contrast to the excess return of 219 and 316 bps they earned during the most turbulent period. The variation is largely a function of factor loadings: during growth-dominated bull markets, FS-CW's value tilt and low beta detract from return. In this context, keeping up with the S&P 500 and Russell 1000 indices is actually an impressive win for both FS-CW indices. Over the 30 years from July 1991 through December 2022, the FS-CW 500 and 1000 outperformed the S&P 500 by a cumulative 12.6% and 15.3% respectively, measured in terms of accumulated wealth in Figure 2.

6 Impact of Factor Loadings on Index Performance

Sophisticated investors may argue that FS-CW's outperformance arises from a mild value bias because the index methodology is less likely to add stocks at high valuation multiples or drop stocks at low valuation multiples. To explore the style differences among the cap-weighted indices in our analysis, we regress each against a Fama–French–Carhart four-factor model that includes the market (S&P 500 return minus 3-month Treasury bill return), size (small cap versus large cap), value (value versus growth), and momentum factors.

Table 2 illustrates the four indices have similar characteristics and all have a market beta very close to 1.00. The Russell 1000 includes smaller-cap stocks and thus has a noteworthy size tilt. The two FS-CW indices have statistically significant but modest loadings on the value factor, whereas the True CW 500, by mechanistically selecting stocks based on price and market-cap (without a committee to slow down the process), has a growth bias and statistically significant momentum loadings. The buy-high/sell-low nature of the S&P 500 is amplified in the True CW 500. Even after controlling for style differences, both FS-CW indices still have remarkably high alphas with strong statistical significance (40 bps and 37 bps for the 500 and 1000, respectively), while the Russell 1000 and True CW 500 have insignificant alphas.

FS-CW 500's loading on value is small, averaging 0.04, but enough to matter. Stated a different way, the S&P 500 has a modest selection-based growth title by dint of adding highfliers, with a tiny economic footprint, and dropping unloved deep-value stocks, even if their underlying businesses are substantial. Consistent with the tracking error changes over time against S&P 500, the factor loading on value for different cap-weighted

strategies also showed some dynamics over time shown by trailing three-year HML value beta loading in Figure 5. From the Appendix FS-CW strategies will take more bets during turbulent period, e.g., after tech bubble, but even during that time, around 0.1 HML beta is still much smaller comparing to other value strategies, due to FS-CW's constituents are largely overlapping with traditional index funds and cap-weighting nature. FS-CW's value bias is unsurprising, but most readers will be shocked to see that FS-CW 500 has a negative size loading and a positive momentum loading relative to the S&P 500. How does FS-CW have a larger market-cap and better momentum than the S&P 500, when FS-CW's construction method is less likely to add high-fliers or drop deep-value stocks? The answer may surprise our readers.

Because FS-CW typically includes a number of stocks not in the S&P 500 but in the 500 largest market-cap category, the index has a large-cap bias relative to the S&P 500. In addition, the S&P 500 includes about 120 stocks, on average, outside the 500 largest US market-cap stocks. The explanation for FS-CW's positive momentum loading is that although the S&P 500 generally adds stocks with impressive momentum, the momentum fades remarkably quickly over time. These high-momentum stocks have high valuation multiples that often presage negative future momentum. The effect is weak, but strong enough to cause FS-CW 500 to have a slightly higher average momentum loading than the S&P 500.

Next, we examine the cumulative growth of the four indices' Fama–French–Carhart four-factor alphas measured against the S&P 500 in Figure 3. After adjusting for factor exposures, the four indices behave similarly in the first few years. But during the tech bubble build-up, a period of high tracking error, the time series begin to diverge. After a sharp, short-lived loss roughly

Figure 3 Cumulative four-factor alpha for four US cap-weighted indices, Jul 1991–Dec 2022.

Notes: we use S&P 500 for the market factor in FFC4 regression, while Size, value, and momentum factor returns and risk-free returns are from the Kenneth R French Data Library.

coincident with the bursting of the tech bubble, FS-CW 500 launched its strong outperformance trajectory. Contrary to the story told by the four indices' cumulative excess returns, after controlling for style differences, the Russell 1000 and True CW 500 had cumulative (not annualized!) alphas of −0.83% and 2.18%, respectively, compared to FS-CW 500's cumulative alpha of 13.3% over the last 30 years. FS-CW 1000 followed closely behind with a cumulative alpha of 12.0%.

7 Historical Turnover

A popular myth in indexing is that index fund trading is essentially free. In fact, the opposite is true: index fund trading costs are huge. As Arnott *et al.* (2023) demonstrated, the index calculators facilitate an illusion of free trading. They preannounce changes in the index with a specific effective date, allowing index fund managers to trade in a market-on-close order the day the change takes place. In doing so the managers can lock in the price at which a stock is added to or dropped from the index, and thus create an illusion of zero tracking error and zero trading costs. If the

Table 3 Historical average one-way turnover of various US cap-weighted indices, 2013–2022.

Average turnover	S&P 500	Russell 1000	True CW 500	FS-CW 500	FS-CW 1000
2013–2022	3.7%	4.3%	3.6%	2.9%	2.9%

index was changed at the prior closing price, on the day a change in the index is announced, index funds would lag the published index by around 20 bps per year. Therefore, another advantage of using fundamentals to select index constituents is that fundamentals change much more slowly than share prices. The result is lower portfolio turnover, fewer trades, lower costs, and less flip-flops (additions that are quickly dropped, and vice versa).

Over the last 10 years, selecting index constituents by their fundamental footprint has delivered distinctly lower average one-way turnover (2.9%) for the two FS-CW indices in Table 3. In contrast, the Russell 1000 had the highest turnover of 4.3%. The S&P 500 with its discretionary additions and deletions had the second-highest turnover of 3.7%, and the True CW index posted similar turnover of 3.6%. Delinking the choice of index constituents from their recent price movements also benefits the portfolio by lowering the turnover—by more than 20%—while raising the potential for long-term outperformance and slightly reducing overall volatility.

8 Live Fundamental-Selection Cap-Weighted Model Portfolio Performance

Inspired by these research findings, we set up a model portfolio in September 2021 to see how it would perform live as an out-of-sample test.[8] The live results *exceed* the backtest (as is typical in volatile market conditions).

The methodology is similar to FS-CW US 500, but applied globally. Similar to MSCI ACWI, and thus making the comparison more apple-to-apple, we selected nearly 3,000 stocks. Unlike MSCI ACWI, we chose the stocks for the index based on the fundamental economic scale of the underlying businesses rather than on market capitalization; we then cap-weighted them with free-float adjustment. The live model portfolio has exhibited robust outperformance across regions and time. We can carve out specific regions—the US, UK, Japan, developed Europe, and emerging markets—to see how these segments of the global index have fared since inception.

The short, roughly 22-month period FS-CW Global has been live has included a bear market, a meaningful recovery, a stark rebound in value relative to growth and elevated market volatility. A value rebound punishes conventional indices whose additions and deletions are driven by price momentum, and elevated volatility magnifies the opportunities to add (or detract!) performance relative to the conventional cap-weighted indices. Given these tailwinds, FS-CW delivered consistent outperformance across all six regions relative to the appropriate cap-weighted regional benchmark as shown in Table 4. Indeed, four out of six indices already have a live information ratio (IR) above 1.0 and one has an IR above 2.0.

The two best-performing indices were FS-CW Global and FS-CW UK. FS-CW Global outperformed the MSCI ACWI by a 2.3% annualized return with a modest tracking error of 1.1%, resulting in an IR of 2.04. FS-CW UK delivered a 3.3% excess return relative to the FTSE 100

Table 4 FS-CW live performance by region, Sep 17, 2021–Jun 30, 2023.

Annualized performance	Returns	Excess returns	Tracking error	Information ratio	Number of constituents*
Global					
FS-CW Global	0.48%	2.28%	1.12%	2.04	2763
MSCI ACWI	−1.80%	—	—	—	2882
US					
FS-CW US	1.79%	0.52%	0.96%	0.54	482
S&P 500	1.27%	—	—	—	503
Developed Europe					
FS-CW Europe	1.36%	1.55%	1.48%	1.05	402
MSCI Europe	−0.19%	—	—	—	424
UK					
FS-CW UK	6.20%	3.27%	2.00%	1.63	70
FTSE 100	2.94%	—	—	—	100
Japan					
FS-CW Japan	−5.79%	1.10%	0.94%	1.18	371
MSCI Japan	−6.89%	—	—	—	237
Emerging Market (EM)					
FS-CW EM	−6.31%	3.93%	3.34%	1.18	1194
MSCI EM	−10.24%	—	—	—	1373

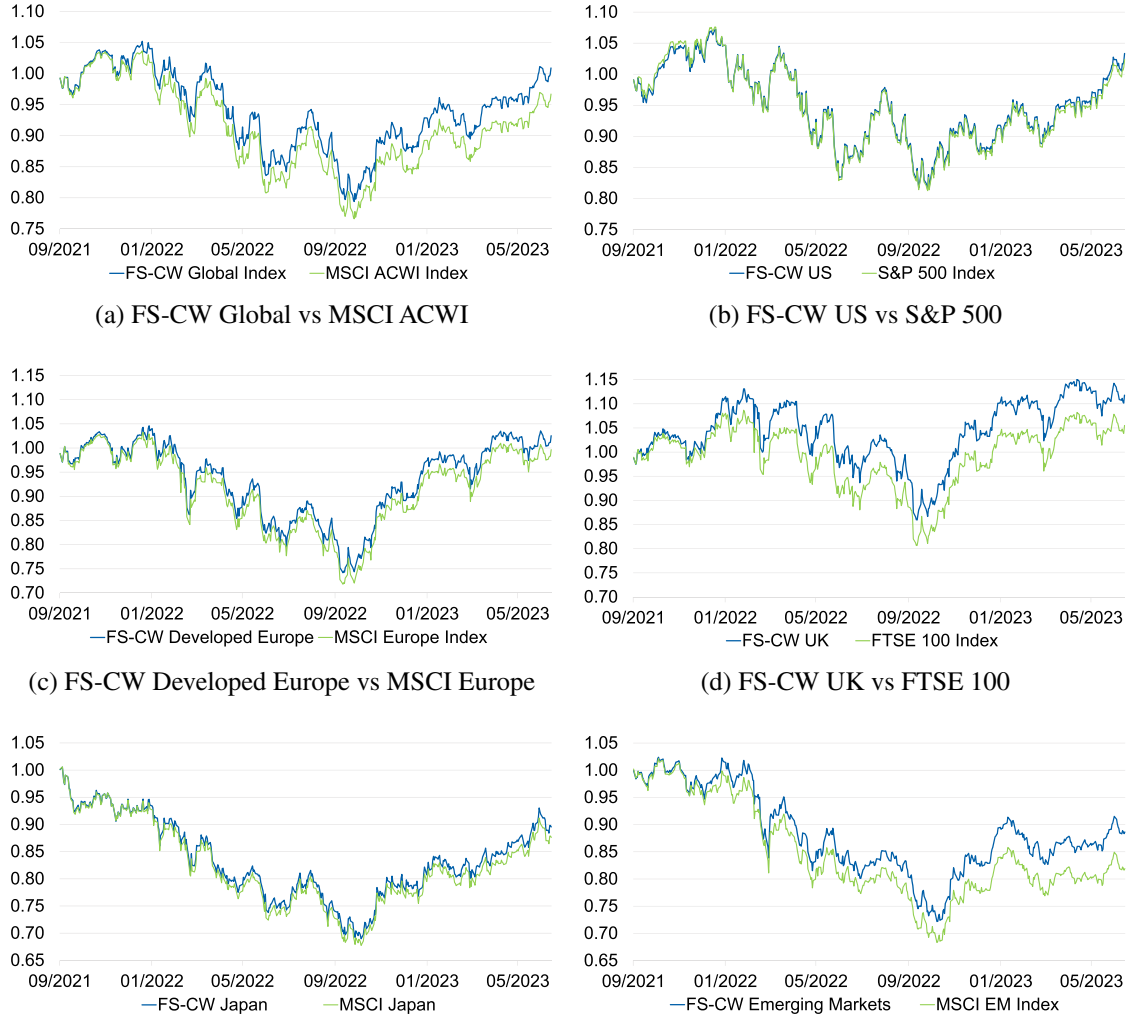
Notes: Calculations use daily returns. * The numbers of constituents are based on February 2023 for illustration purposes. The annualized performances of other regional indices over this period are: FTSE All-World Index, −1.80% (value-add by FS-CW Global of 2.28%); Russell 1000 Index, −0.12% (value-add by FS-CW US of 1.91%); FTSE Europe Index, −0.84% (value-add by FS-CW Europe of 2.20%); FTSE All Share, −0.37% (value-add by FS-CW UK of 6.58%); TOPIX, −7.04% (value-add by FS-CW Japan of 1.25%); and FTSE Emerging Index, −8.87% (value-add by FS-CW EM of 2.56%).

Index. Even with a high tracking error of 2.0%, the resulting IR was 1.63. The excess return in the European region is more modest, producing just under 1.6% incremental performance versus MSCI Europe, but with a tracking error of 1.5% delivering an IR of 1.05. FS-CW Japan shows a comparable IR of 1.18 over MSCI Japan, with 0.9% tracking error and excess return slightly higher than 1%. The results in the United States are the most modest. Even so, with 52 bps over the S&P 500, FS-CW US post a decent IR of 0.54; it bears mention that, against the most popular alternative cap-weighted indexes, the Russell 1000, FS-CW fared much better. Finally, FS-CW EM had the highest tracking error (3.3%) of the six regional indices. With an excess return of 3.9%, its IR is 1.18.

Two cautionary notes. When we launched the Global model portfolio in September 2021, we obviously could not have known what a turbulent time would follow, one that has rewarded value investors. Even a tiny value tilt will deliver alpha over such a span. Therefore, we believe it prudent to frame our expectations from the 30-year historical backtests of the indices and not by this exceptional 21½-month span. We would also counsel against shaping expectations based on

regional differences in returns. FS-CW EM added almost 4% a year over MSCI EM, whereas FS-CW US added only 0.52% a year versus the S&P 500. We have no reason to expect that this short-term result is anything more than two outliers and ascribe far more importance to the finding that the FS-CW approach worked in six out of six regions.

In Figure 4, as a live strategy, we have shown that FS-CW model portfolio has outperformed the traditional indices, globally and in developed and emerging markets, even the least impressive result, in the US market, FS-CW US exceeded the performance of the S&P 500, by a margin almost identical to the backtest. In the first weeks after launch, FS-CW Global mimicked MSCI ACWI's performance closely, but from November 2021, FS-CW persistently outperformed. Through June end 2023, since its launch in September 2021, FS-CW beat the MSCI ACWI by a cumulative 4.2%. *This graph shows what an information ratio of 2.0 looks like!*



(a) FS-CW Global vs MSCI ACWI

(b) FS-CW US vs S&P 500

(c) FS-CW Developed Europe vs MSCI Europe

(d) FS-CW UK vs FTSE 100

(e) FS-CW Japan vs MSCI Japan

(f) FS-CW EM vs MSCI EM

Figure 4 Cumulative dollar growth of FS-CW and traditional indices, Sep 2021–Jun 2023.

9 The Bizarre Obsession with Tracking Error

The more-stable selection algorithm that FS-CW uses creates "tracking error" relative to standard cap-weighted indices. Because passive investors typically are far more interested in minimizing tracking error than in seeking to add value, FS-CW's tracking error may seem to be a serious flaw. This argument makes little sense when well-established and popular broad-market indices exhibit tracking error of 100 bps a year or more *relative to one another*.

Using a tracking error lens in index selection can be quite arbitrary. Consider this thought experiment. Should we presume that the Russell 1000 is flawed because it has 105 bps of tracking error relative to the S&P 500? Or should we prefer the Russell 1000 because the S&P 500 has a disturbing 105 bp TE relative to the Russell 1000? *The question never comes up* because both are widely accepted cap-weighted indices of large US stocks. Ergo, no investor should view tracking error of 100 bps between one cap-weighted index and another with alarm.

We calculated the tracking error of the S&P 500 relative to the Russell 1000, the True CW 500, the FS-CW 500, and the FS-CW 1000 in Table 5. Recall that the True CW 500 chooses the 500 largest market-cap US stocks and cap-weights them. Many investors think that the S&P 500 is exactly that index! But the reality is that the S&P 500 and the True CW 500 have, on average, only 380 stocks in common. This means that the S&P 500 typically includes 120 smaller-cap stocks that fail to make it into the top 500 by market-cap *and excludes 120 large-cap stocks in the top 500 by market-cap*. Put another way, True CW 500 holds 120 stocks larger than the 120 smallest in the S&P 500. That constituent difference will create performance difference and thus cause additional tracking error between the two. Arnott *et al.* (2023) showed that in the year after the reconstitution, discretionary deletions from S&P 500 will on average beat S&P additions by 22%. Please refer to Arnott *et al.* (2023) for more detailed discussion about the difference in performance, turnover, factor loadings and alpha of S&P 500 and True CW 500, along with other market cap-weighted index variants.

The tracking error of FS-CW 500 relative to the S&P 500 is marginally *lower* than the tracking error between the Russell 1000 and the S&P 500. In fact, the tracking error between any of the cap-weighted indices we analyze is nearly identical: all TEs fall in the range of 0.8–1.5% a year. Should investors concern themselves about the tracking error of any cap-weighted index to the S&P 500? The differences between them result from changes in their selection algorithm, which lightly alters the 120 smaller-cap names they hold,

Table 5 Tracking error between US cap-weighted indices, Jul 1991–Dec 2022.

	S&P 500	Russell 1000	True CW 500	FS-CW 500	FS-CW 1000
Geometric ann. return	9.88%	10.00%	9.95%	10.35%	10.40%
			Tracking error		
S&P 500		1.05%	0.97%	0.97%	1.00%
Russell 1000	1.05%		0.85%	1.52%	1.05%
True CW 500	0.97%	0.85%		1.45%	1.30%
FS-CW 500	0.97%	1.52%	1.45%		0.80%
FS-CW 1000	1.00%	1.05%	1.30%	0.80%	

typically encompassing well under 10% of the index.

Do we advocate terminating allocations to the S&P 500 for this reason? Hardly. *We advocate tolerating similar tracking error, if it leads to a better (performing) index! We submit FS-CW is exactly that—a better-performing index.*

10 Conclusion: Superior Cap-Weighted Index = Superior Cap-Weighted Index Fund

By construction, a cap-weighted index puts more of an investor's money into overpriced stocks and less into underpriced stocks, but—as indexers will happily point out—How to know which is which. That said, why should we hasten that process by mostly adding stocks based on newly elevated market-cap, when they are priced at "peak froth," and mostly dropping stocks just after their market-cap has cratered, priced at "peak fear"? We propose a better way to create a cap-weighted index. Using FS-CW, which bases additions and deletions on a company's fundamental measures and thus de-links index constituents from the stock's recent price movement, we can create a superior cap-weighted index fund.

With this simple expedient, FS-CW US 500 earns 46 bps of annualized excess return (with less risk!) versus the S&P 500 in a 30-year historical simulation.[9] The live results of the FS-CW model portfolio, since launched in September 2021, have exhibited a stronger outperformance. An additional benefit of this index is that by anchoring index stock selection with fundamentals, we can lower portfolio turnover and potentially markedly reduce trading costs.

Investors can benefit most from the Fundamental-Selection Cap-Weighted index where and when equity markets are less efficient and thus offer more mispricing opportunities. FS-CW's live portfolio performance, in markets around the world, supports our findings that the index can provide greater outperformance during market turbulence and in higher-volatility markets. After adjusting for relative risk—with FS-CW offering slightly lower turnover in most markets—the result is a superior cap-weighted index, improved by largely eliminating the buy-high/sell-low dynamics inherent in the rebalancing process for most commercially available indexing products.

Appendix

To examine the factor loadings over time for the cap-weighted indices in our test, we run a three-year rolling regression using Fama–French–Carhart four-factor model with S&P 500 served as market factor.

Figure 5 illustrated how much value loads the four cap-weighted indices took over time, comparing to S&P 500. There are time-varying loadings in value factor for all four indices. In general, the two FS-CW indices demonstrated slightly higher value bets compared to Russell 1000 and the True CW 500, but overall, the loadings on value factor

Figure 5 Trailing three-year HML value loadings using four-factor model for the US cap-weighted indices against S&P 500, Jun 1994–Dec 2022.

Notes: we use S&P 500 for the market factor in FFC4 regression, while Size, value, and momentum factor returns and risk-free returns are from the Kenneth R French Data Library.

is not high, even for FS-CW with the highest loading on HML approximately 0.1 post tech bubble. This small value tilt is much smaller than that of traditional value indices.

Endnotes

1 In Siegel's (2006) words, "Current attempts to explain the hidden risks in value stocks remind me of the astronomers in the 16th century who attempted to save the earth-centered Ptolemaic view of the universe. They were forced to add complicated 'epicycles' to the orbits of the planets to rationalize their movements in the evening sky; the model collapsed when Copernicus showed that a simple sun-centered solar system was an easier explanation. As with Copernicus, there is now a new paradigm for understanding how markets work that can explain why small stocks and value stocks outperform capitalization-weighted indexes."

"This new paradigm claims that the prices of securities are not always the best estimate of the true underlying value of the firm. It argues that prices can be influenced by speculators and momentum traders, as well as by insiders and institutions that often buy and sell stocks for reasons unrelated to fundamental value, such as for diversification, liquidity and taxes. In other words, prices of securities are subject to temporary shocks that I call 'noise' that obscures their true value. These temporary shocks may last for days or for years, and their unpredictability makes it difficult to design a trading strategy that consistently produces superior returns. To distinguish this paradigm from the reigning efficient market hypothesis, I call it the 'noisy market hypothesis."'

2 We ignore nondiscretionary deletions resulting from corporate actions, which are most of the deletions.

3 An illustrative example is Tesla. In April 2020 Tesla made its way into FS-CW, nine-and-a-half years after its addition to the Russell 1000 and eight months *before* its addition to the S&P 500.

4 Academe prefers to look at aggregates for a company rather than per-share data, hence book-to-market (total company book value, relative to a company's total market value, its market cap, or B/M), while practitioners prefer per-share data, hence price-to-book (share price relative to book value per share, or P/B). Academe also prefers the inverse of the ratio favored by practitioners, book-to-market, in which a high ratio is value and a low ratio is growth. The resulting Fama–French nomenclature "HML," meaning High Minus Low, the performance of the value portfolio (high B/M) minus the performance of the growth portfolio (low B/M). These are essentially identical concepts, merely expressed differently.

5 The valuation multiples for a portfolio are not the weighted average of the individual stocks' valuation multiples. Suppose we have two stocks, equally weighted, one trading at 100 times earnings and the other at 10×. The portfolio does not have a 55 × P/E ratio. It has $5.50 in earnings for a $100 portfolio, for a P/E ratio of 18×. This means that our index fund has 95.6% in stocks with a relative valuation multiple of 1 × (plus a little bit, because the deletions were cheap), and 4.4% in stocks with a relative valuation multiple of 1.92×. Even ignoring the impact of deletions (which are much smaller than the additions), this gives us a relative valuation multiple for the resulting portfolio of 102.2%.

6 To quantify the performance difference in weighting with and without the free-float adjustment, we reconstruct the True CW 500 using data from Datastream since 2005, the year free-float-adjusted weights became popular and widely adopted by the industry. The annual return difference between raw cap-weighted and free-float-adjusted cap-weighted for True CW 500 is 3 bps.

7 The buy-high/sell-low dynamic of traditional cap-weighted benchmarks suggests they suffer from the price impact of constituent changes after the changes are announced as well as from any subsequent mean reversion in price of the additions and deletions (Arnott *et al.*, 2023).

8 The live index is called the Research Affiliates Cap-Weighted Global index, or RACWI Global, with a Bloomberg ticker of RACWIGLT for the total return index. There are not yet any assets tracking this portfolio. The regional indexes are extracted from this index. For instance, the US index consists of all 482 US-domicile companies in the RACWI Global index.

9 We have always been very wary of backtests. A backtest used to tweak, enhance, or improve the backtest results should have zero credibility. But, as Arnott *et al.* (2019) suggested, a backtest of an extraordinarily simple idea, with no tweaking of results, has far more credibility.

Acknowledgments

The authors would like to thank Kay Jaitly, Amie Ko, Teo Njegovan, and Thomas Verghese for their help, feedback, and edits.

References

Arnott, R. D., Hsu, J., and Moore, P. (2005). "Fundamental Indexation," *Financial Analysts Journal* **61**(2), 83–99.

Arnott, R. D., Harvey, C. R., and Markowitz, H. (2019). "A Backtesting Protocol in the Era of Machine Learning," *Journal of Financial Data Science* **1**(1), 64–74.

Arnott, R. D., Brightman, C., Kalesnik V., and Wu, L. (2023). "Earning Alpha by Avoiding the Index Rebalancing Crowd," *Financial Analysts Journal* **79**(2), 76–97.

Chaves, D. B. and Arnott, R. D. (2012). "Rebalancing and the Value Effect," *The Journal of Portfolio Management* **38**(4), 59–74.

Dempsey, M. (2012). "The Controversy in Fundamental Indexation: Why Both Sides of the Argument are (Mostly) Correct," *Journal of Investment Management* **10**(4), 54–63.

Fama, E. F. and French, K. R. (2007) "Migration," *Financial Analysts Journal* **63**(3), 48–58.

Hsu, J. (2006). "Cap-weighted Portfolios are Sub-optimal Portfolios," *Journal of Investment Management* **4**(3), 1–10.

Siegel, J. J. (2006). "The 'Noisy Market' Hypothesis," *Wall Street Journal*, https://www.wsj.com/articles/SB115025119289879729.

Treynor, J. L. (2005). "Why Market-Valuation-Indifferent Indexing Works," *Financial Analysts Journal* **61**(5), 65–69.

Keywords: Cap-weighting; fundamental selection; tracking error; outperformance; indexation.