The making of a better benchmark

Everyone knows what makes a good benchmark—right? Every so often, an organization announces having discovered or built “a better index.” Its purported merits are usually backed by statistics indicating that it has demonstrated higher returns and lower risk than the index it proposes to replace. Users of indexes can be confused by these data and persuaded that they should adopt a “new” index before its benchmarking potential has been fully explored.

Our objective here is to shed some light on this issue and to help investors assess arguments for change. We focus on the important role an index plays in the investment process as a benchmark, and identify the most important characteristics an index must have in order to play this role effectively. We explain some of the important trade-offs providers must address in the construction of better benchmarks—i.e., the inherent trade-offs among construction principles, and between the purity and costs of those principles. We focus on some key moments in index-construction history, providing further support for the importance of good benchmark design. We conclude with a brief discussion of the expansion of the definition of “index,” given that indexes are moving beyond being market benchmarks and into the next generation of methodologies. New methodologies extend the uses of indexes as tools for implementing investment strategies.

The role of an index as a benchmark in the investment process

Indexes are constructed and used for a variety of investment purposes. At the most basic level, a properly designed index functions as a measure of an overall market’s characteristics and performance. An index representative of a market or asset class plays an important role at every step of the investment process: in asset allocation, in implementation and, finally, in performance measurement and evaluation. Well-constructed benchmark indexes embody the risk/return characteristics of the markets and asset classes they represent and inform the forecasting process that creates inputs into the asset-allocation decision. An index that proxies an asset class can be used as a basis for

---

1 The author thanks David Cariño, Mary Fjelstad, Mat Lystra, Mark Paris, and Suzanne Ropeta for valuable comments.

investable products, such as mutual funds or ETFs, to allow for passive investment in the asset class. Investors who choose active investment management can use an index to communicate their goals to the manager, and can compare the actively managed portfolio’s performance to the index’s performance to assess the value of active management. The index represents the performance characteristics of an asset class or market and functions as the performance benchmark for the active portfolio.

**Benchmark construction and principles**

Several important principles guide the construction of benchmark indexes so that they can play these essential roles in the investment process. Benchmark indexes represent “naïve investment” in a given asset class or market. Thus they should be complete; they should be investable and representative; and they should be simple, transparent and reliable.

**The “naïve” alternative**

Benchmark indexes should measure the return and risk an investor with no proprietary or exceptional knowledge of the investment opportunity set could experience when investing in an asset class. In other words, a benchmark index should represent a “naïve” returns series.

This principle most significantly affects index constituent selection and weighting. Capitalization-weighted indexes accept the constituent membership, weights and values that the overall market exhibits. Equal-weighted indexes that accept membership but presume no insight – not even that of the market – into weights or values also meet this criterion.

**Completeness**

A benchmark should cover all of the practical opportunities in an asset class.3 If an index fails to include expected securities, such omission may lead to unintended consequences for investors. An investor will want to know what return could have been achieved had the missing assets been included. An evaluation using an index that fails to include expected securities in the opportunity set may lead to errors in the asset allocation structure of the total portfolio, or, if active managers are used, to charges of unfairness in managers’ evaluation.

**Investability and representativeness**

A benchmark should reflect assets available to all investors and should adjust constituent weights for restricted shares, cross-owned shares or shares that are not available to the public. Index weights should thus reflect the free float or amount of capitalization readily available for purchase. This index characteristic is called investability. Further, a benchmark should be built to minimize the advantage of any class of investors over another. Front-running or other manipulation by large market players should not be possible, or minimized. This fairness requirement is also a component of investability.4

Benchmarks should be capitalization weighted and then float adjusted, when possible. The market price or value of a security in the index represents both market sentiment regarding the security and market opportunity in the proportions of the security that can be purchased by the average investor. A stock’s price is a result of supply and demand, rather than an

---

3 Information, data and investment limitations make investing in some “opportunities” impractical, so “practical opportunities” means those assets one could reasonably be able to purchase or sell without extraordinary effort.

4 This section benefits from the discussion in Laurence B. Siegel, *Benchmarks and Investment Management* (Charlottesville, VA: The Research Foundation of AIMR, 2003).
**index constructor’s estimate of what a stock is worth.** The benchmark should not try to second-guess the value that market participants assign to each stock. The number of shares being traded and the prices market participants are willing to pay for those shares are unambiguous measures of what the stocks or assets are truly worth at a point in time.

Float-adjusted, capitalization-weighted indexes represent investments that all investors may buy into. Any investor who wanted to buy the market portfolio represented by a float-adjusted, cap-weighted index – with every stock at the exact percentage it is in such an index – could do so. Alternatively weighted indexes and equal-weighted indexes do not meet this requirement, as the index constituents cannot be held by all investors at those weights.5

**Simplicity, transparency and reliability**

A benchmark should be constructed in a simple, transparent method that can be replicated by users. The provider should publish clear, unambiguous construction rules. The data should be reasonably available; otherwise, the average investor will be unable to re-create the index and to understand or anticipate changes in it. The index should include only assets for which complete and reliable data can be repeatedly obtained in a timely manner. The users of the index should have confidence that the index membership and valuations are consistent and reliable.

**Principles of benchmark construction: Inescapable trade-offs**

During construction of a benchmark there will inevitably be some trade-offs concerning the recommended core principles described above. A trade-off may involve balancing one principle versus another, or balancing costs – primarily, transaction costs – versus pure implementation of principles. As examples, we highlight two of these trade-offs: completeness vs. investability and representativeness vs. turnover levels.

**Completeness vs. investability:** In some markets there is inherent conflict between completeness (representing all securities in an asset class within the index) and investability. Some of the securities in the broadest ranges of some markets will not be available for trading all of the time, or their prices will be difficult to obtain.

**Representativeness vs. turnover levels:** There is also conflict inherent in maintaining an index’s representativeness while at the same time seeking to maintain its simplicity and cost-efficiency. As successful companies grow, their stock prices rise, and so do their market caps. Other companies are less successful, and their prices and market caps decline. During index reconstitution (conducted to ensure that representativeness is maintained), some securities fall out of benchmark indexes and new securities come in. Whenever this happens, transaction costs may be incurred by actively managed funds benchmarked to the index and by passive funds seeking to replicate the index. The more frequent the reconstitution, the better the index’s representation of the pure market, but the greater the transaction costs in tracking the index.

**Key moments in the history of benchmark construction**

The principles we’ve outlined above are essential for the construction of benchmark indexes that both represent the behavior of markets and help investors evaluate the performance of active managers. Here we discuss some illustrative moments in the history of indexation and provide additional support – focal points – to show how these principles can affect the construction and use of indexes as market measures and as benchmarks.

---

5 See further discussion of this issue below, in focal point # 4.
Focal point #1: The Wilshire 5000, the Russell 3000® and the trade-off between Investability and completeness

The Wilshire 5000 Index was the first broad index of the U.S. stock market, launched in 1974. When index funds were first created to match the index, it turned out that the smallest 2,000 securities (approximately) in the Wilshire 5000 were essentially untradeable. These stocks were not, in practice, investable, because they could not be bought and sold by fund managers in sufficient volume. Investors could not have fully replicated the Wilshire 5000 index without incurring exceptionally high transaction costs or unusual delays in buying and selling the illiquid securities.

Russell took note of the illiquidity of these very small cap stocks – essentially, securities beyond the largest 3000 – and created its broad market index in 1983. The Russell 3000 Index was constructed to capture within the U.S. stock market the broadest investment opportunity set in securities available for timely purchase and sale at reliable and publicly available prices.

Focal point #2: The Yahoo! effect and the investability principle

Modifying constituent weights to reflect shares that are freely available for purchase by the average investor is an important aspect of the “investability principle.” Since their launch in 1983, Russell indexes have always been constructed by use of the free-float method: index constituents are weighted not by total market cap but by float-adjusted market cap, the value of the shares available for purchase. Standard and Poor’s, in contrast, did not apply float-adjusted methodology to its flagship S&P 500 Index until 2004.

The market upheaval that resulted from S&P’s addition of Yahoo! to the S&P 500 Index in 1999 at total capitalization – even though only 10% of Yahoo! stock was available for public purchase – is a textbook example of free-float mismatch. The event occurred on December 7, 1999, the day before Yahoo! Inc. was added to the S&P 500 index. On that one day, Yahoo!’s price rose by $67.25 per share, or 24 percent. Why? Yahoo! was to be included at its full capitalization weight in the S&P 500 even though large numbers of its shareholders were not at liberty to sell. (Most Yahoo! shares were held by employees, venture capitalists and other investors who were restricted from selling.) In actuality, only about 10 percent of the shares were truly available for purchase – in other words, the float did not match the index capitalization. At the time, the S&P 500 was the index associated with the most passively-managed assets in the world. Since so many index funds had to buy the stock in order to track the S&P 500, an obvious mismatch between supply and demand for Yahoo! stock occurred, and a huge price spike was the result. This event is the quintessential example of the importance of float adjustment in index construction.

Focal point #3: The S&P 500, the Russell 2000® and the trade-off between representativeness and pricing pressures at index reconstitution

Index investing does not occur in a vacuum. Investors tracking indexes and active managers benchmarked to indexes must buy and sell the securities that make up the indexes. When constituency is changed to maintain or regain an index’s representativeness, the change can have an impact on asset prices and index performance. The histories of both the S&P 500 Index and the Russell 2000 Index illustrate this point. The

---

8 A similar event happened in Japan’s TOPIX index in the 1980s, when Nippon Telephone and Telegraph spun off DoCoMo, its cell phone subsidiary. DoCoMo was included in TOPIX at full capitalization when in fact NTT retained a substantial portion of the capitalization.
S&P 500 inclusion or exclusion effect is well documented.\(^9\) It has been shown that, on average, when a stock goes into the S&P 500, its price rises, and when a stock is dropped from the S&P 500, its price falls. Among the reasons for this price change is a supply-and-demand effect. Generally, one would expect that when a stock enters a popular index, demand for the stock may increase, possibly due to temporary imbalances in buy and sell orders around the effective date of the inclusion.

These imbalances can arise because investors who wish to replicate the index, such as index funds, may feel compelled to buy the stock on the effective date in order to minimize tracking error. Active managers might anticipate these trades, buying the stock in advance with the expectation to sell it around the date of inclusion, thus supplying liquidity to the indexers. The same sort of well-publicized concern surrounds the Russell 2000 Index rebalancing that occurs every year in June. Russell Indexes has introduced a number of measures to mitigate this effect.\(^10\) Starting in 2004, IPOs have been added to the Russell Indexes on a quarterly basis – i.e., spread out throughout the year instead of waiting for reconstitution day. Further, the Nasdaq closing cross prices have been used at reconstitution for Nasdaq-listed stocks since 2004. Also since 2004, Russell has published provisional indexes; these enable portfolio managers to track indexes that rebalance on dates other than the official reconstitution date, spreading the demand from indexers across different dates surrounding reconstitution.\(^11\) Finally, in 2007, Russell implemented a banding methodology to reduce unnecessary turnover between the Russell 1000\(^8\) and the Russell 2000, whereby stocks near the market cap breakpoints remain in their respective index if they have not materially changed in size. In a study\(^12\) using simulated indexes, banding reduced the average turnover of the Russell 2000 at reconstitution by 5% from 1997 to 2009. Banding significantly reduces the number of securities that change size classifications from year to year. Without banding, the number of stocks that would have moved from the Russell 2000 to the Russell 1000 and then back again the next year averaged about 27% over the simulated period. With banding, that number was reduced to 2.8%.

**Focal point #4: Float-adjusted capitalization-weighted Indexes embody both the market’s definition of an asset class and its valuation of firms**

Traditional capitalization weighting is used in most broad market benchmarks for compelling reasons. Weighting by market value is an objective way of measuring the relative economic importance of index constituents. Valuing available shares at market prices measures the market’s assessment of the relative values of the firms. Hence, market value is, literally, just that – the market’s view of the value of a firm – and it represents a naïve investment approach independent of any propriety information or active management insights.

A cap-weighted index requires less rebalancing than indexes using other types of weighting, thus reducing turnover. If an investor buys a capitalization-weighted index portfolio and


\(^11\) For more information on provisional indexes, see http://www.russell.com/indexes/data/provisional_legacy_returns_US.asp.

there are no changes in the constituency, then the portfolio automatically rebalances each day, right along with the index – no trading is necessary.\textsuperscript{13}

Another compelling reason for capitalization weighting is an index property called macro consistency.\textsuperscript{14} If all investors held cap-weighted index funds, and if there were no active investors, then all shares would be held, with none left over. Laurence Siegel calls capitalization weighting “by far the most important innovation in equity index construction” and “the central organizing principle of good index construction.”\textsuperscript{15} As an objective, practical and theoretically grounded method, capitalization weighting is widely regarded as the standard weighting method for indexes.

The disadvantages of capitalization weighting of benchmarks are not inconsequential, however. Many investment managers, particularly large cap managers, tend not to capitalization-weight their portfolios, for a variety of reasons. The most oft-cited reason is aversion to putting too much money into any one stock: typically, managers want stock name diversification. Unfortunately, when compared against a capitalization-weighted portfolio, this aversion can be viewed as a bet against certain sectors and stocks. Capitalization weighting may weight some sectors of the market more than the average institutional money manager would. Manager deviations from capitalization weighting may cause differences in the average characteristics of the manager’s portfolio, as well as differences in performance. Finally, during periods of “market irrationality” or “market exuberance,” such as the Tech Bubble, herding behavior drives the prices of some securities far above intrinsic values as suggested by Graham and Dodd.\textsuperscript{16} In such times, capitalization weighting takes the cap-weighted investor along for the ride.

\textbf{Focal point \#5: The S&P 500 Index committee sampling construction and the representative principle}

The constituents of the S&P 500 Index are determined by a committee, rather than by a set of consistent rules. The sampling methodology used by the index committee is inconsistent, across several dimensions, with the desirable criteria we have outlined. First, the stock list is a sample and hence suffers from a lack of completeness. When some stocks are selected and others are not, errors in measurement occur in the segment of the equity market under consideration. This can have a material effect on the estimation of the value added by active management and on the accurate representation of that asset class in other stages of the investment process.

Imagine yourself an investment manager holding market weights in both GM and Ford when Ford is outperforming the overall market and GM is not. The user of an index that excludes Ford may believe you are a superior investor when in fact you are not. From a money manager point of view, at one point you may look like a great investor and at another point you may look incompetent – with the variable being not your skills, but the stocks in your portfolio. All else being equal, there are times when differences in portfolio performance are attributable to differences in stock lists and the stock weightings that must occur.

An easy way to understand the implications of not including all of the stocks in an investment universe is to imagine a market in which there are only three stocks: A and B (which are very nearly identical) and C. Let us assume that the total float-adjusted market

\textsuperscript{13}Index adjustments resulting from corporate actions – that is, mergers, spin-offs, and so forth – can result in weight changes for any type of weighting scheme.

\textsuperscript{14}Laurence B. Siegel, \textit{Benchmarks and Investment Management} (Charlottesville, VA: The Research Foundation of AIMR, 2003), p. 5.

\textsuperscript{15}Ibid., pg. 1.

capitalization is $100b. Stock A has a capitalization of $45b, or 45% of the total market capitalization; Stock B also has a capitalization of $45b, and 45% of capitalization as well. Stock C has a capitalization of $10b, or 10% of capitalization. Further, let us assume that our stock index committee decides Stock A and Stock B are indistinguishable, and so it includes Stock A and excludes Stock B. Finally, let us assume that over a measurement period, Stock A and Stock B achieve a return of 5%, and Stock C a return of 20%.

Now we compute the capitalization weights for each of the securities in our sampled index. Stock A now has an 81.8% weight and Stock C has an 18.2% weight. What about index returns? The return for the total market index (all stocks included) is 6.5%, and the return for our index committee–selected index is 7.72%. It is clear that for an investment in the total market, the 6.5% return is a more accurate reflection of what happened to the investment than is the 7.72%.

We conclude that even though Stock A and Stock B are very nearly identical and are weighted by float-adjusted capitalization, the exclusion of any stock changes the weights of all of the stocks included in the index. Therefore, the result is different index returns and index equity characteristics, such as price/book. Furthermore, the weight on included stocks must be larger in the resulting index than in the total market, because it is a larger percentage of the index capitalization. This is inevitable, because when the index maker excludes stocks, the weights of the included stocks must change to compensate for the weights of the excluded stocks. The 45% of capitalization in Stock B is proportionately distributed between Stock A and Stock C, thereby raising the weight of each. The differences in weights and performance of the stocks interact to produce differences in index performance that can be substantial. These kinds of distortions occur whenever stocks are excluded from the opportunity set arbitrarily, when the index is not representative of the entire opportunity set.

**Focal point #6: Role of performance in index worthiness**

As a final point: better performance is not one of the criteria we have defined as a principle of good benchmark design. Benchmark performance is an outcome, not a design objective or an index characteristic. The argument for performance as a key criterion is like saying one pollster has created a better sample of citizens than another pollster because the new sample has a higher average height, implying that increased tallness makes the sample more representative. Claiming that superior performance is the hallmark of superiority in a benchmark confuses the purpose of a benchmark with the purpose of active management. A benchmark seeks to tell us the return a dollar invested passively in the market earns, while an active manager seeks, at minimum, to beat the passive market alternative and, more broadly, to maximize the return on a dollar invested.

**Expansion of the definition of “index”**

Capital markets are ever changing, and the English language is always evolving. The word “index” as it is used in the financial world is no longer confined solely to benchmarks as we have defined them here – namely, as measures of broad markets or asset classes that are suitable for passive investment and as performance benchmarks for active managers. Every now and again, someone discovers that for some past time period, retroactively applying a set of different weights to a list of stocks results in greater hypothetical returns with less risk than those achieved by the cap-weighted market benchmark – and then markets this strategy as an “index.” Robert Haugen did this in the late 1980s with his

---

17 These comments benefit from the insights of Don Ezra.
“Efficient Index,” and some alternative-weighted indexers, such as those marketing equal-weighted or fundamentally weighted indexes, are making similar arguments today.

Additionally, work has been done to extract consistent trading strategies from actively managed portfolios and to build "indexes" that mimic their disciplined sets of trading rules. Some of these have become the bases for investable products such as ETFs. These "indexes" fall into the gray area between classic indexes (benchmarks) and traditional active management. They are not passive, because they utilize active (non–market cap), sometimes proprietary weighting strategies. They are not fully active, either, because they passively track a list of securities like an index does, or they trade securities according to a set of consistent rules to produce an active-strategy risk/return profile or isolate a particular beta. Excess return over the cap-weighted index may or may not be an objective, but the marketing literature for many of these products implies that it is. Other indexes provide specialized, targeted exposures to exotic betas or risk style factors. We believe there is a role for these products in investors’ portfolios, but exactly how they can be most effectively used has yet to be fully explored.

The choice of benchmark has an enormous impact on the quality and nature of the investor experience. Construction issues make a difference, as we have shown, and investors should be familiar with the different choices and trade-offs made by index providers. Even when investors allocate to one or more of the new alternatively weighted or dynamic-strategy indexes, they are still well advised to use the capitalization-weighted total market index – the benchmark – to evaluate how much return enhancement and risk reduction they obtain with the new “index” products, just as they would do for any other active investment vehicle.

