The S&P 500 Myth

Mukul Pal
mukul@alphablock.org

July 21, 2022

Abstract

The nearly USD 50 trillion passive investment management industry has grown over the last 50 years primarily based on the claim that 9 out of 10 Asset managers can't beat the market, net of fees over a 5-year rolling period. This claim assumes that the S&P 500 has a superior methodology that is unassailable. Modern financial theory and the industry practitioners never questioned the respective claim, conveniently assuming it to be correct. This led to extreme fee pressure on the nearly USD 100 trillion active managers who found themselves academically defenseless to explain, why they could not generate alpha consistently.

In this paper, the author explains the chequered history of Indexing mathematics and how the S&P 500 methodology of market capitalization weighting came to be, and how, unlike popular performance claims, the method is based on ease of calculation, speculative assumptions, arbitrary methods, scientific and statistical inconsistencies, lack of true purpose, axiomatic conflicts, unwarranted complexity, a disproportionate focus on weights, naive statistical approaches and upward biases.

In its current form, the S&P 500 method is erroneous, has survivorship bias, amplifies market risk, does not offer the diversification, and has been instrumental in propagating a design error across the world of investments, offering sub-optimal solutions to global investors and asset owners, creating an unintended misrepresentation that has harmed and continues to harm the global capital markets.

The author illustrates his case using Polya's urn, a mathematical thought experiment to highlight the design gaps in the incumbent indexing method, why Active performance and S&P market capitalized benchmarks passive performance are two incomparable approaches, and how logic, mathematics, and physics can help the industry build better indices and assist active asset managers to generate alpha and move to a pay-for-alpha fee model.

Content

Archimedes of Syracuse and Rice Vaughan [275 BC – 1675 AD]

Indexing started much before Archimedes of Syracuse in 275 BC when he had the eureka moment, feeling the buoyancy in the bathtub [or the pool]. If it was not for the indexed values of Gold, Silver, and other metals, he had no way of resolving the puzzle of the impure crown. Indexers have always tried to solve the problems of their time. Rice Vaughan in 1675 in "Coins and Coinage" documented the general price rise that accompanied the superfluous lifestyles of King Charles II, when he decreed to suspend payments for funding the Dutch War. Vaughan concluded that the prices had risen six or eight times their level of a century earlier. The idea of studying price changes, and debasement of currencies, using the accessible data and variables would later become a recurring theme for indexers.

William Fleetwood and Adam Smith [1707 - 1776]

While Vaughan was concerned about price changes, William Fleetwood, Bishop of St Asaph concerned himself with the plight of a correspondent who would lose the fellowship of an Oxford college if he had outside income above £5 [a 1440 college statute]. In "Chronicon Preciosum" Fleetwood established a price equivalent in 1706 for the £5 set in the fifteenth century. The result is the earliest treatise on index numbers, where the conception of purchasing power and the measures of changing purchasing power of money are first treated in a modern manner. The book was the first major historical survey of prices, wages, and income and was considered a masterwork on Index-numbers.

Adam Smith used some of Fleetwood's data in "The Wealth of Nations" [1776] but did not develop or adopt the idea of comparing purchasing power at different dates. Admiration for Fleetwood's work and efforts came in the nineteenth century.

Nicolas Dutot and Gian Rinaldo Carli [1738-1764]

Nicolas Dutot was an important figure for both economic history and economic thought before Adam Smith when the word "economist" didn't exist. He was the first to use an unweighted index of prices, now called the "Dutot price index", a price index defined as the ratio of the unweighted arithmetic average of the prices in the current period to the unweighted arithmetic of the prices in the base period. Though elementary, the work influenced the next generation of indexers.

The ensuing debate attracted the attention of a brilliant Italian aristocrat Gian Rinaldo Carli who made extensive investigations into the nature and causes of the rise in prices and created the earliest recorded use of index numbers. An interest in the frequent debasement of Italian coinage stimulated Carli to investigate the monetary question. He proposed that the rise in prices was a consequence of the debasement in coinage rather than an increase in the supply of bullion and attempted to demonstrate that the purchasing power of a unit of bullion had declined slightly while prices had more than tripled since the 15th century. These phenomena, coupled with the declining bullion content of Italian monetary units, indicated to Carli that debasement of the coinage was primarily responsible for rises in the price level.

His indexes were unweighted, computed an arithmetic mean of prices per unity quantity to obtain the individual commodity indexes, and simply averaged the results to arrive at the commodity index. The Carli Index has an upward bias and as we will see later, the respective bias continues to become an entrenched problem in the indexing method.

Joseph Lowe and George Evelyn [1774 - 1820]

Lowe advocated indexation as applied to bonds, but also wage contracts and land rents. His advocacy of a "tabular standard" has been recognized as a technical advance in monetary analysis. Lowe used the terms "standard of reference" or "table of reference". The idea itself, and the term "tabular standard" are attributed to Evelyn. The first to use weighted index numbers, Lowe's work continued the chain of thinking

in Indexing, influencing the next generation of indexers. Most CPIs [Consumer Price Indices] and employment cost indices from Statistics Canada, the U.S. Bureau of Labor Statistics, and many other national statistics offices are Lowe indices.

John and Henry Poor [1849-1862]

Railroad history in the United States is nearly as old as the country itself, dating back to the mid-1820s. Although railways reduced the industry's dependence on rivers to transport timber to the mills, their initial importance was in carrying lumber from the mill to market.

Lumber to railroads to S&P 500 methodology might seem like a strange connection but that's exactly what happened. By investing money in Maine's growing timber industry, the Poor brothers made a fortune. John Poor became a minor railway magnate in association with the European and North American Railway and was heavily involved in the building of the Maine rail network. In 1849, John purchased the American Railroad Journal and Henry became the manager and editor.

In 1860, Henry Poor published History of Railroads and Canals in the United States, an attempt to compile comprehensive information about the financial and operational state of U.S. railroad companies. He later established H.V. and H.W. Poor Co. with his son, Henry William, and published annually updated versions of his book. The process of maintaining data for the essential sector would later become a key ingredient when Poor's publishing and Standard Statistics Bureau merged in 1941 to extend the Index component list to 490 components.

Ernst Louis Étienne Laspeyres and Herman Paasche [1851-1871]

Ernst Louis Étienne Laspeyres is the most prominent figure in the indexing study as his method drives today's benchmarks and is at the heart of modern finance that assumes the benchmark integrity without questioning or challenging its construction. Laspeyres was the representative of *Kathedersozialismus* [The historical school of economics]. The historical school held that history was the key source of knowledge about human actions and economic matters, and hence not generalizable over space and time. The school rejected the universal validity of economic theorems. They saw economics as resulting from careful empirical and historical analysis instead of from logic and mathematics. The school also preferred reality, historical, political, and social, as well as economic, to mathematical thinking.

Ernst Louis Étienne Laspeyres and Herman Paasche formulated the Paasche and Laspeyres price indices. The only difference between the two methods was that the former used current period quantities [period n], whereas the latter used base period quantities [period 0] quantities. Both these methods created an upper and lower bound for prices.

If one already has price and quantity data for the base period, then calculating the Laspeyres index for a new period requires only new price data. In contrast, calculating many other indices [e.g., the Paasche index] for a new period requires both new price data and new quantity data for each new period. Collecting only new price data is often easier than collecting both new price data and new quantity data, so calculating the Laspeyres index for a new period tends to require less time and effort than calculating other indices for a new period. In practice, price indices regularly compiled and released by national statistical agencies are of the Laspeyres type, due to the above-mentioned difficulties in obtaining the current period.

This Laspeyresian ease of the nineteenth century will perpetuate further for two centuries into today's indices and eventually into every other investment basket that is sold as an ETF [Mutual funds]. The method will embed itself as an unquestionable assumption of modern financial models, be it the CAPM, Fama, and French, or any other factor models that the industry takes as given.

3

Axe and Houghton [1854]

The 1880s were the most fertile times for capital market indexing. There was a gold rush to compete for popularity and business. This was also the time when Science started to become mainstream, with the first journal of Science published, the formation of Munich Re, the incorporation of the Canadian Pacific Railway, the liquification of Oxygen, etc.

Mr. Emerson Axe and Ruth Houghton began the Axe-Houghton index in 1854 and till 1869, the index consisted only of railroad stocks. It appeared in "The Annalist" on Friday, March 14, 1930. At that time the Index included 33 industrials. "The Annalist" added 5 rails and 5 utilities, bringing the total stocks carried up to 43. After observing the unusual decline from the high of 1929 to the low of 1932, Axe decided that the system of weighting, and perhaps some of the stocks, had been poorly selected because none of the other popularly used averages had declined to anything like the 1895 levels, which had been the case with the Axe-Houghton. For his work, Mr. Axe recomputed the series back to December 1925 with only 30 industrials as it was more convenient. The Annalist, however, continued to publish the original Axe-Houghton average of 33 industrials. The Index was based on monthly high-low prices. The various stocks in the original index were weighted in proportion to their market importance which was gauged partly by their relative size and importance.

Computation of these adjusted weights of each stock was the most important principle underlying the construction of these averages. Each stock was weighted in inverse proportion to its normal or characteristic width in the fluctuation or normal price range. The method recognized the fact that volatile stocks invariably dominate the fluctuations of a straight average so some of the really important stocks with narrow price ranges have very little influence on the movement of the composite. The arbitrary nature of the work eventually lead to more superior indexes crowding out the Axe-Houghton Index.

Francis Ysidro Edgeworth and Alfred Marshall [1887 - 1925]

As the research progressed, the practitioners started to influence the economists, who saw their work becoming mainstream. And questions about the statistical validity, speculative assumptions, and conflict between the consensus and the desired output came to the forefront. Index number came to be seen as an entity with a peculiar character, differing in some essential attributes from ordinary operations.

In 1890 Marshall followed the 1925 work of Edgeworth led to the Marshall-Edgeworth Index which tried to overcome the problems of under and overstatement by the Laspeyres and Paasche indexes by using the arithmetic means of the quantities. All these indices provide some overall measurement of relative prices between periods or locations. This index uses the arithmetic average of the current and base period quantities for weighting.

Irving Fisher and Correa Moylan Walsh [1901-1902]

Irving Fisher was perhaps the first celebrity economist who experienced the ups and downs of fame. He was mathematically inclined from an early age and brought new rigor to Indexing. Fisher's 1922 book, "The Making of Index Numbers", was considered the fountainhead of almost all the best later work on index numbers dedicated to both Walsh and Edgeworth.

"And if we view economics as a science, in which a department is the science of money, it is intolerable to think that its most important technical term, of a concept which plays the same role as that of energy or force in physics, there still exists ambiguity which leads different minds to mean different things by the same word"

Walsh 1903

4

Walsh's, 1901 "The Measurement of General Exchange-Value", was deemed a game-changer monograph by most of his contemporaries interested in index numbers. The book discusses a statistical instrument at the service of his monetary theory. The same is true for Fisher, whose interest in index numbers arose from the problem of measuring the purchasing power of money. It is not a coincidence, therefore, that both Fisher and Walsh, as they strived for scientific purity, saw the measurement of variations in the exchange value, or purchasing power of money, as the unique purpose of index numbers. The subject saw more contributions from, Frederick Macaulay [1910], John Maynard Keynes [1930], Arthur Cecil Pigou [1932], Leo Törnqvist [1936], Walter Erwin Diewert [1972].

Indexing Methods and Tests

The subject of Indexing is clustered around indexing methods. The fixed basket approach, the statistical approach, the test or axiomatic approach, the stochastic approach, and the economic approach. Strangely all the work on indexing showed convergence to the Laspeyres method.

The Lowe indices were modified Laspeyres index. In the stochastic approach, the reciprocals of the price ratios turned out to be the Laspeyres price index. None of the other index number formulae satisfied the somewhat stringent definition of consistency like the Laspeyres Index. The unobservable Indices are bounded from above by the observable Laspeyres price index. The unobservable Indices were bounded from below by the observable Paasche price index, but it gathered less attention since the Paasche index lacked calculation ease, unlike its Laspeyresian counterpart.

The Harmonic index is always equal to or less than Jevon's index, which is always equal to or less than the Carli index. The inequalities don't tell us by how much the Carli index will exceed the Jevon's index and by how much the Jevon's index will exceed the Harmonic index. There are systematic divergent trends in prices. The Dutot index can change dramatically as the units of measurement are changed. The backward Carli index turns out to be equal to the harmonic index, and so the divergence and convergence continue. The more one looks into the indexing complex the more one sees Laspeyres everywhere as if the indexing world was oscillating around Laspeyres.

Chain Indexing and Laspeyres

Various Indexing tests were established to test these methods. The tests of identity, dimensionality, commensurability, time reversal test, factor reversal test, and transitivity. The body of work was consciously resolving a big conflict. The long-term standardized value had to be regularly updated with new inputs and new weights. The current relevance was as important as maintaining the long-term historical value of a constant base Laspeyresian standard.

Convenience has a high probability of trumping integrity. The Chain index was conceived as a measure of the cumulated effect of successive steps [from 0 to 1, 1 to 2, . . .]. What mattered was not the type of base or weight, but explicitly taking into account the intermediate periods and not just the two endpoints. Adopting a Chain index was a shortcut for the indexing industry to achieve many objectives, including the Laspeyresian calculation ease, the longer-term historical standardized value, and the ability to update to achieve current relevance. Today investment management runs on the combination of chain indexing and Laspeyres. While MSCI Equity Indices are calculated using the Laspeyres concept of a weighted arithmetic average together with the concept of chain linking, many countries have moved from the Laspeyres index to a chain index.

Challenges, Shortcomings and Failure

Taking two different ideas together [Chain and Laspeyres] and forcing them into cohesion does not decrease the conflict but intensifies the potential error. During the conflict, it is hard for a process to

5

choose between criteria. What should come first, what should come second and what should be sacrificed.

Both Laspeyres and Chain Indices have challenges and shortcomings. Under certain conditions, the Laspeyres and Paasche index become the upper and lower bound of an interval encompassing most calculations. Whenever prices and quantities change in opposite directions Laspeyres gives a higher weight to rising prices and vice versa. This upward bias is what we see in all market capitalized weighted benchmarks today. The reasons given for the alleged advantages chain indices enjoy over direct indices are not conclusive. There are undeniable shortcomings of chain indices such as difficulties in aggregating and deflation, path dependence, and inexplicability of axiomatic reasoning.

"There is no conceptual basis from some ingratiating rhetoric in terms of flexibility, adaptability, up-to-datedness, relevance, or so. "I realized that there is not much going on beyond the fixation with, not to say obsession, up-to-datedness of weights, which largely goes unquestioned. Most of the indices known to me are some kind of weighted average, but I never encountered an average in statistics where the concern about the weights dwarfs all other considerations in such an amazing manner as in the case of chain apologetics. In my view index theory should not be reduced to a search for the most up-to-date weights but should strive for a meaningful interpretation."

Peter von der Lippe, Chain Indices, 2001

According to Lippe, a lot of attention was given to the idea of pure comparison at the expense of the representativity of the basket to which the index refers. Wanting an index that takes the most recent changes in the consumption pattern into account would make a chain index preferable. But this implied an impure comparison in the sense that what is supposed to be a measure of change in prices is to some extent also reflecting a change in quantities and structure. A reasonable compromise is not a chain index nor unchanged weights of a Laspeyres index but a Laspeyres index in which weights are reviewed and readjusted in intervals of say five years or so.

The theory of index numbers is dominated by complicated mathematical considerations, concerning index functions apparently in growing sophistication and complexity, such as index theory becomes less and less accessible to "ordinary" economists and statisticians. Moreover, the two groups, the index theoreticians, and official price statisticians, appear to diverge rather than converge. For statisticians, the variety of opinions axioms and approaches in the index theory is not easy to understand.

Chain logic as the theoretical background of chaining is inconclusive or even contradictory. Chain indices could be viewed as describing the particular pattern of a time series, rather than comparing two situations as indices in general do and should do. The method emphasizes "flexibility" at the expense of comparability and meaningful interpretation. The Laspeyresian ease when combined with the chain links became an amplifying error that took the indexing methods away from the representation it was supposed to be. And once the practitioners picked up critical mass, the noise of markets muffled the scientific initiatives of the industry. The error now was ready to propagate inside the stock market indices.

1854 Axe and Houghton
1897 Dow Jones Average
1911 New York Times Average
1925 Herald Tribune Stock Averages
1925 New York Stock Exchange Averages
1926 Standard Statistics Stock Index
1935 Associated Press Average
1936 The Annalist Average
1957 Standard & Poor 500

Charles Dow and William Hamilton [1897 - 1929]

Unlike the Axe-Houghton Index, the Dow Jones Index thrived for more than a century till the company was acquired by the Standard and Poor Indices in 2011. The history of the Dow Jones Average is critical to understanding the S&P myth because Charles Dow was an indexer, a forecaster, and hence traversing both the active and passive investment worlds.

In November 1882, Dow, Jones & Company set up their headquarters in the basement of a candy store, a ramshackle building next door to the entrance of the stock exchange. In November 1883, the company started putting out an afternoon two-page summary of the day's financial news called the "Customers' Afternoon Letter" which soon achieved a circulation of over 1,000 subscribers and was considered an important news source for investors.

The first issue of The Wall Street Journal appeared on July 8, 1889. It cost two cents per issue or five dollars for a one-year subscription. Its objective was to give fully and fairly the daily news attending the fluctuations in prices of stocks, bonds, and some classes of commodities. The stock price average was created on July 3, 1884, by Charles Dow as part of the letter. At its inception, it consisted of 11 companies—9 railroads and 2 non-rail companies, Pacific Mail Steamship and Western Union Telegraph. On September 23, 1889, the "20 Active Stock" index was introduced. It included 18 railroad and 2 non-rail stocks. The wildly speculative market meant investors needed information about the stock activity. Dow took this opportunity to devise the Dow Jones Industrial Average (DJIA) in 1896. By tracking the closing stock prices of twelve companies, adding up their stock prices, and dividing by twelve. Dow came up with his average. The first such average appeared in the Wall Street Journal on May 26, 1896.

According to George Bishop, "Dow did not use a weighted mean or make adjustments of any other nature. There is no evidence that Dow looked upon the averages as containing anything more than an indication of the statistical nature of the trend of the stock market as a whole. It was for this reason that Dow found it desirable to include active stocks [i.e., those stocks with a large volume of transactions] in the averages."

The Dow Jones averages remained averages, or arithmetic means [A Laspeyresian Characteristic], throughout Dow's lifetime. Therefore, he was not concerned with adjustments for stock splits as is the case with the Dow Jones averages today. Dow did substitute one stock for another in the averages and to this extent attempted to have the averages reflect the trend of the market under changing conditions [Need for current relevance].

The present Dow Jones averages are not averages but are indexes. The industrial average of 30 stocks began on October 1, 1928, and was calculated by adding the closing prices of the 30 stocks divided by 16.67. This divisor was arrived at by adjusting for stock splits and stock dividends. A constant division was used for the first time on September 10, 1928, when the industrial average contained 20 stocks. Before that time an attempt was made to compensate for stock splits by applying a multiplier to the respective individual stocks.

The transportation average was introduced on January 2, 1970, and was devised to replace the railroad average. It includes the stocks of railroads, airlines, and other corporations concerned with the transportation industry. In 1897 the index included only 12 stocks. Later these averages consisted of a 30-stock Industrial average, a 20-stock Rail average, and a 20-stock Utility average. On October 1, 1928, The Dow Jones Industrial average was extended to include 30 stocks, instead of the old 20, and at the same time, a few substitutions were made among the old 20. The purpose of these changes was to make the index more representative of the market, to substitute for inactive and unrepresentative issues, stocks of greater activity, and significance. Not only for the markets, but as indices representing the country's business, and to minimize the possibility of unusual fluctuations in any one stock distorting the averages on any given day.

The Dow Jones utility average was started on December 26, 1929, and was then worked back for the whole year. Until October 1, 1928, the Dow Jones averages were published only in the form of closing data. After this date, high, low, and the last were published. The total sales in each of the groups first appeared on April 29, 1931. From October 5, 1932, the Dow Jones Averages were computed and published for every hour of trading time [open, 11 A.M., 12 Noon, 1 P.M., 2 P.M., and 3 P.M.]. The Dow averaged were widely published and hence became popular.

A 14- stock average was computed for over 10 months in 1885; a 20-stock average was compiled for a part of 1889; a 20-stock average was presented for the years 1890-1896, inclusive, and a 12-share composition was figured for part of 1896. What these various combinations of indexes seem to imply is that from 1885 onward Dow was experimenting, searching, and investigating the precise combinations of averages he was seeking, to depict the underlying trends of the market as a whole. His search was concluded to his satisfaction when the calculation of dual Industrial and Railroad averages was initiated at the beginning of 1897. It was from this point, that Dow's Theory as it is known today began to crystallize and take its place in the field of business and stock market forecasting.

"Those who defend the use of the averages point out that the selection of the stocks in the averages is not a random one and the corporations included by their stock issues represent large businesses. As a result, the averages do not include marginal concerns or special situations. Likewise, the Dow Jones averages are still the gauge of the overall market to a great number of investors and speculators and their long historical record makes it possible to use them for comparison purposes."

George Bishop, The Dow Theory Revisited, 1974

The validity of the Dow Jones averages as a proper measure of overall market movements has been questioned due to the relatively small sample used. Two other criticisms often advanced are the lack of weighting by the size of capitalization of the corporations concerned and the change of a stock's weight in the sample due to major changes in capitalization. Likewise, the substitution of stocks in the averages over time has also received critical attention.

Speaking again of Dow's original hypothesis, at some point after the time when the publication of The Wall Street Journal was initiated in July 1889, Dow became convinced that individual stocks, instead of fluctuating solely based on each company's prospects were importantly influenced by the rising or falling tides in general business activity and the coincident bull or bear markets in stocks, all of which were a part of the same long term up or down cycle.

The basic idea of Dow was that the stock price is affected by various factors interacting at the same time, leading to distinct patterns of stock price movement. One of the most important contributions to stock market thought was his theory of the three movements in the market. Dow stated there were three movements in the market—all going on at the same time. The first was the daily fluctuation; then the short swing ranging from about 10 to 60 days, and finally the main movement covering at least four years in duration. Dow was a firm believer in the periodicity of the business cycle and felt that the first action of a trader should be to ascertain if a bull or bear market was underway. This was reflected, of course, by the direction of the main movement over an extended period.

Dow's contributions to stock market thought by his classification of three movements in the market and the formulation of the law of action and reaction no doubt caused a great deal of discussion among the readers. The law applied to movements in individual stocks as well as the general equity market as reflected by a stock average. The rule that a primary move would have a secondary movement in the opposite direction of at least three-eighths of the primary movement became a religion with market technicians in those days.

Next in line among the expounders of Dow's Theory was William Peter Hamilton, who also served as a reporter under Dow after a career as a newspaperman in England and South Africa. While not all of the editorial forecasts scored bull's-eyes, and at least one missed the mark by a wide margin, most were amazingly accurate and the collection of written material, on the whole, demonstrates the efficacy of Dow's original hypothesis and the reliability of the "barometer" he discovered. Hamilton's most famous forecast was titled "A Turn in the Tide." It appeared in The Wall Street Journal on October 25, 1929, just two months before he died in December 1929.

Alfred C. Cowles [1933]

Despite all the value that practitioners brought to the industry, there was a clear performance metric that was printed every day. Most economists did not have such a challenge as they were more focussed on the method, the bounds, the mathematics, and the Science. However, the gaining popularity of indices made it difficult to ignore the closing prices of indices. In time the focus had shifted to the forecast because the comparison was convenient, and theorizing was hard. This is when Cowles' work got noticed.

"Has the Dow Theory successfully predicted bull and bear markets making it possible for Dow Theorists to profit thereby? Evidence that this is the case has been accumulating for 40 years."

Alfred Cowles III read a paper before a joint meeting of the Econometric Society and the American Statistical Association in Cincinnati, Ohio, on December 31, 1932, the bottom of the great depression. This paper was concerned with the stock market forecasting records of 20 fire insurance companies, 16 financial services, 24 financial publications, and the Dow Theory forecasting record of William Peter Hamilton from December 1903 to December 1929.

"From December 1903 to December 1929, Hamilton, through the application of his forecasts to the stocks comprising the Dow Jones industrial averages, would have earned a return, including dividend and interest income, of 12 percent per annum. In the same period, the stocks comprising the industrial averages showed a return of 15.5 percent per annum. Hamilton, therefore, failed by an appreciable margin to gain as much through his forecasting as he would have made by continuous outright investment in the stocks comprising the industrial averages. He exceeded by a wide margin, however, a supposedly normal investment return of about 5 percent. Applying his forecasts to the stocks comprising the Dow Jones railroad averages, the result is an annual gain of 5.7 percent while the railroad averages themselves show a return of 7.7 percent."

Cowles' words regarding Hamilton's record

In 1944 Cowles continued his study on stock market forecasting and extended the records of 11 of the forecasters who had been the subject of his research published in 1933 in Econometrica entitled "Can Stock Market Forecasters Forecast?" Cowles did not identify the forecasters but stated, "These organizations are well known. Names are omitted here because their publication might precipitate controversy over the interpretation of the records. The wording of many of the forecasts is indefinite, and it would be frequently possible for the forecaster after the event to present a plausible argument in favor of an interpretation other than the one made by a reader."

Cowles found that the forecasting agency with the most successful record for the period from 1928 to 1943 could be extended back to 1903. He explained, "While three individuals were for different periods responsible for the forecasts through these 40 years, the general principles followed by them all were similar, and the succeeding forecasters were avowed disciples of their predecessors. It, therefore, seems justifiable to treat the combined record as a continuous one for the 40 years in question." [There is a 70-year documented history of the Dow Theory].

9

Cowles' work created the foundation for a generation of researchers who believed that the S&P 500 was the only benchmark and could not be beaten. His work influenced a generation of thinkers, including John Bogle who lived by the invincibility of the index fund and the futility of active management. Vanguard led the Index fund revolution, becoming a multi-trillion-dollar passive investment manager. In the center of all of this was the S&P500 myth.

Polya's Urn

George Pólya was a Hungarian Mathematician who came up with the 'Polya urn' thought experiment, a type of statistical model used as an idealized mental exercise framework, unifying many treatments. In an urn model, objects of real interest [such as atoms, people, cars, etc.] are represented as colored balls in an urn or other container. In the basic Pólya urn model, the urn contains x white and y black balls. A ball is drawn randomly from the urn and its color is observed. It is then returned to the urn, an additional ball of the same color is added to the urn, and the selection process is repeated. The urn experiment is used to understand the evolution of the urn population and the sequence of colors of the balls drawn out, a way to demonstrate how the rich get the richer or preferential attachment.

A basket of stocks can be visualized as an urn of stocks. The current indexing methodology weights based on market capitalization [size] is a version of the Polya's rich get richer urn [the upward bias]. The more stock gains size, the more the methodology overweighs the stock. In simple terms, the methodology picks up a white stock from the universe [urn] and then replaces it along with another white stock. Invariably the urn has more white stocks. This increases the white stock's probability to be picked up in the next selection, just because there are more white stocks.

But this is not the complete picture. A market capitalization-weighted methodology has a dual influence. While it's creating an intentional rich getting richer bias, it's also naturally reducing the weightage of small size companies [the companies that are not selected], the black balls, which shrink in relative size every time a white ball is selected and replaced with two white balls back in the urn. This means for every rich that gets richer, the poor is getting poorer.

The methodology intentionally creates a rich-get-richer bias for big-sized stocks, which inflate in proportion compared to the total basket. You can visualize it like a magical ball, every time you touch it, it grows in size. And every time you don't touch it, the size reduces. If you touch the same ball twice, it grows significantly, while the balls you have not touched reduce in size. The method is designed to benefit the winners at the expense of the losers. The indexing methodology bets on positivity [winners] and bets against negativity [losers].

S&P 500 Risks

The market capitalization method is a flawed design not just because of its hit and trial indexing history, the ease of convenience but because of the upward bias in Laspeyres Index which when combined with chain indexing creates super-sized components. 6 of the 500 stocks can own up to 30% of market value. Concentration in a few stocks creates major risks for asset owners and investors, there is limited diversification and high idiosyncratic risk, which the broad market is supposed to avoid.

A broad market should never be so concentrated that it can be replicated by a few components. The poor design also leads to long recovery periods, in the case of bear markets. When you concentrate power within a small set of components, the flip side of the stellar blue-chip expression is slow and painful recovery. We have seen it happening time and again across global benchmarks, be it S&P 500 in different periods 1929-1945, the 1970s, 2000-2010s or the Nikkei in 1980s and European STOXX 50, which could never get back above their respective multi-decade peaks.

The slow recovery periods accompany negative consumer sentiment, which propels a vicious cycle of decay, forcing bankruptcy and wealth destruction and increasing survivorship bias. S&P 500's market capitalization methodology creates an upward bias in a small section of the broad market. This is a poor representation of a market that operates with many biases, taking turns, and moving in and out of favor. Singularly focussing on a bias is a double edge sword, which does more harm than good.

Active Managers

Calling an Index construction, the way it is, a Polya urn or a rich get the richer process is important because it resolves the myth of S&P 500 method superiority and the Active manager's incompetence as has been explained and disseminated since the beginning of the S&P 500 index as the incumbent method.

Active manager's performance cannot be compared with the Polya's urn Index because the Index is an ex-post construction, it can see the winner [loser] and then decide to overweigh [underweight]. Active managers have no such liberty to look inside the urn, they rely on ex-ante capabilities, which come with more risk and hence they can't keep up with a construction that is designed to amplify the winners and minimize the losers. The indexes rich get richer approach replenishes itself through a mechanism that allows the poor to get poorer. Such a mechanism will always be better in the long term than an idiosyncratic approach that has no systematic way of replenishing itself.

The active manager works with asymmetric information and is always at a disadvantage with the Index method, which operates like a Polya's urn. Active managers are forced to resort to tactical and timing methods to decipher a market that is like a closed urn, opaque, complex, and unpredictable in the longer term. Active is an ex-ante discretionary sub-selection approach that works with a small subset of components, uses discretionary methods for exit and entry, and hence disappoints against a systematic ex-post mechanism that has access to today's information and the complete list of components to choose from. Polya's urn will always beat a discretionary manager in the long run because he/she has access to information, and a systematic method to enter, exit, and replenish.

Informational Context vs. Informational Content

There are many effective ways to consistently beat the S&P 500 performance. Most ways involve taking the same 500 stocks and weighting them differently but still maintaining similar volatility, similar turnover, low tracking error, and high information ratio. Modern finance considers this an impossible feat. And if one of these effective ways can be connected to a certain starting point [base] and still allow for refreshing the universe, we will have a novel portfolio construction that can solve the indexing conundrum.

Modern finance has not focussed on Indexing mathematics, assumed the market [benchmark] as given, and is hobbled by its conflicts regarding efficiency and inefficiency of information. The binary nature of information drives most financial models and creates challenges for active managers. Information is assumed to be either relevant or irrelevant [redundant], it cannot be both at the same time, a fact Kenneth Boulding highlighted back in 1966. Singularly focussing on informational content creates a disadvantage for active managers who select and deselect based on information relevancy and irrelevance respectively, while information drifts and shifts from relevant to irrelevant states, creating unpredictability.

There is a poor and low quantification of information with market trends. The multi states of information are why modern finance can't figure out the efficient – inefficient nature of markets. 9 out of 10 managers' underperforming the markets is the victory of Polya's informational context over information content. There is no information asymmetry for Passive managers who rely on Polya's mechanism, which is contextual and content agnostic. Polya's experiment is agnostic to what you put into the urn. Even though size weighting may indicate content, size is just another factor like a few hundred other factors that can lend themselves well to Polya's urn mechanism.

In the longer term, systematic context [Today's Passive] will always beat subjective interpretation of content [Today's Active]. To build alpha-generating solutions and a new and better S&P 500 index, we need to revisit Polya's urn. The S&P 500 overweighs the rich – get – richer context while underweighting the poor. The effectiveness of the thought experiment is its mechanistic simplicity, with a lot of unexplored temporal capabilities. The market is a complex entity that operates in many states. Along with the rich getting richer and poor getting poorer states, there is poor getting richer, rich getting poorer, rich staying rich [not getting richer], or poor staying poor [not getting poorer]. Even though Polya did not see his urn as a mechanism of Physics, the urn carries many probabilistic states at the same time. A new improved indexing method will need to understand the constraints of the current S&P 500 and build on Polya's urn multi-state nature to come up with new weighting methods that use other informational states including rich get the richer state to build a new basket of 500 stocks.

The more active managers will focus on the context the lesser the chances of their underperformance. Because none of the informational states including the rich get richer state is secular. Like everything else, all the states oscillate, they wear out and drag into long recoveries. Looking at the urn for a mechanism will reduce active managers' disadvantages to select and deselect, allow them to tactically play between the various states, and not only generate alpha, but bring explainability to their investing process. This will give them a better chance to operate in the pay-for-alpha investing world.

Bibliography

R. Vaughan, "A Discourse of Coin and Coinage", 1675
W. Fleetwood, "Chronicon Preciosum", 1707
F. Y. Edgeworth, "A Defence of Index-Numbers", The Economic Journal, 1896
C,. M. Walsh, "The Measurement of General Exchange-Value", Macmillan, 1901
S. A. Nelson, "The Abc Of Stock Speculation",1903
I. Fisher, "The Best Form of Index Number", American Statistical Association Quarterly, 1921
I. Fisher, "The Making of Index Numbers: A Study of Their Varieties, Tests, and Reliability", Houghton Mifflin Company, 1922
W. P. Hamilton, "The Stock Market Barometer", 1922
A. Cowles 3rd, "Can Stock Market Forecasters Forecast?, Econometrica, 1933
H. M. Gartley, "Profits in the Stock Market", 1935
P. Greiner, "Encyclopedia of Stock Market Techniques", 1963
W. A. Chance, "A Note on the Origins of Index Numbers", The Review of Economics and Statistics, 1966
G. Bishop, "The Dow Theory Revisited", Reason, 1974
W. E. Diewert, "The early history of price index research", NBER, 1998
H. D. Taylor, L. Taylor George Pólya: master of discovery 1887–1985. Dale Seymour Publications, 1993
P. v. d. Lippe, Chain Indices: A Study in Price Index Theory, Statistisches, Bundesamt, 2001